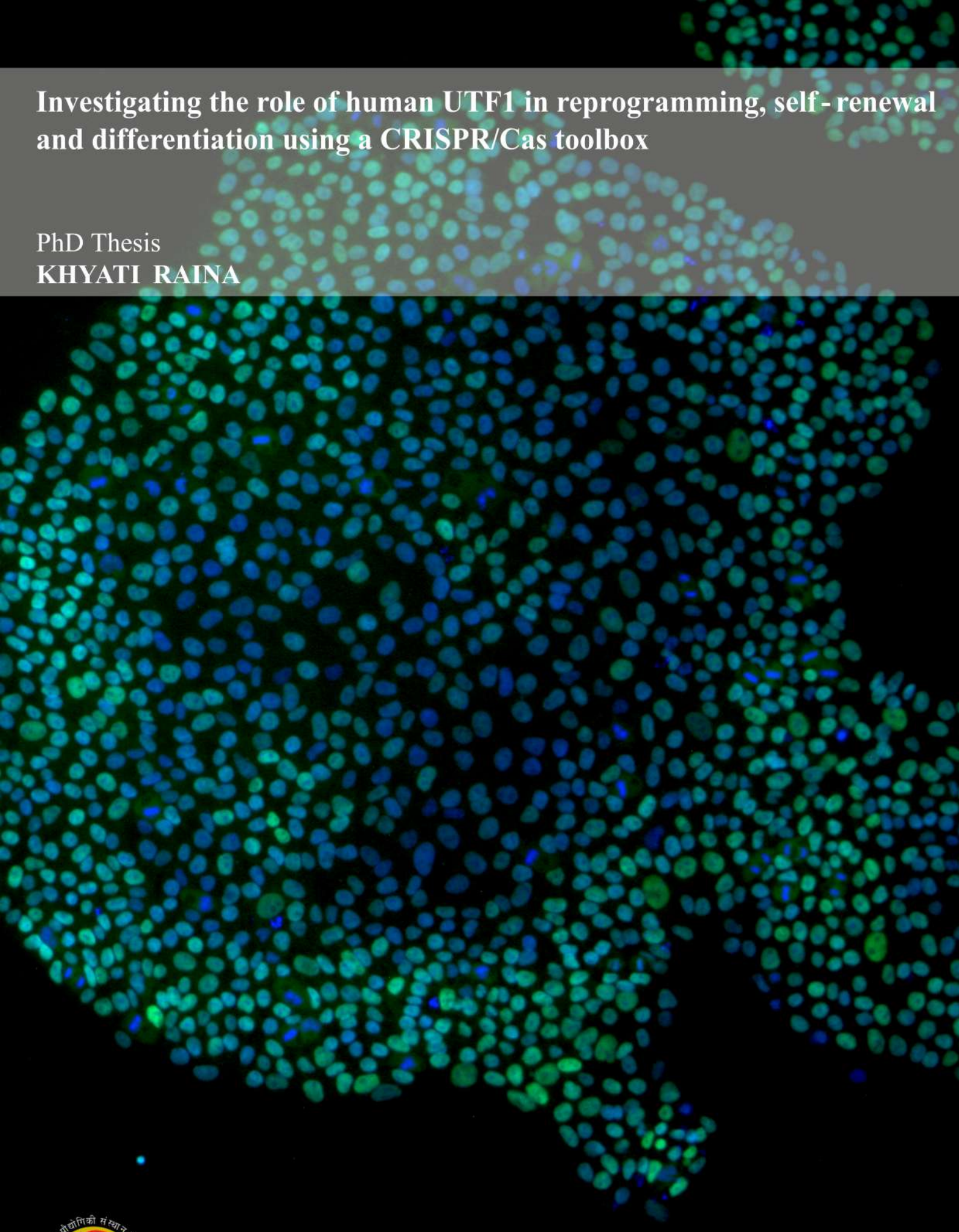
Investigating the role of human UTF1 in reprogramming, self-renewal and differentiation using a CRISPR/Cas toolbox

PhD Thesis

KHYATI RAINA



Indian Institute of Technology Guwahati

Investigating the role of human UTF1 in reprogramming, self-renewal and differentiation using a CRISPR/Cas toolbox

Thesis submitted in partial fulfilment of the requirements
for the award of the degree of

Doctor of Philosophy

by

Khyati Raina
(Roll No. 176106110)

Under the Guidance of

Dr. Rajkumar P. Thummer

Laboratory for Stem Cell Engineering and Regenerative Medicine
Department of Biosciences and Bioengineering
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
Assam - 781039, India
June, 2024

Dedicated to the experiences I would have never expected, and the paths that were redirected. To the family and friends, I found along the way.

DECLARATION

I do hereby declare that the content embodied in this thesis entitled **"Investigating the role of human UTF1 in reprogramming, self-renewal and differentiation using a CRISPR/Cas toolbox"** is the result of investigations carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati for the award of degree Doctor of Philosophy, under the supervision of **Dr. Rajkumar P. Thummer**.

As per the general norms of reporting research findings, due acknowledgments have been made wherever the research findings of other researchers have been cited in this thesis.

.....

Khyati Raina
(176106110)
Department of Biosciences and Bioengineering,
Indian Institute of Technology Guwahati,
Guwahati-781039, Assam, India.

CERTIFICATE

It is certified that the work described in this thesis entitled **"Investigating the role of human UTF1 in reprogramming, self-renewal and differentiation using a CRISPR/Cas toolbox"**, by **Ms. Khyati Raina** (Roll No. 176106110) for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India. This work has not been submitted elsewhere for the award of any degree or diploma.

.....

Dr. Rajkumar P. Thummer
Assistant Professor,
Department of Biosciences and Bioengineering,
Indian Institute of Technology Guwahati,
Guwahati-781039, Assam, India.

ACKNOWLEDGEMENTS

"*The roots of all goodness lie in the soil of appreciation for goodness.*"— Dalai Lama

As I present before you this dissertation, the satisfaction that accompanies it would be incomplete without a mention of people who were part of my long adventurous journey - some in flesh while some in spirits. They have played an integral role in my academic and personal journey, and shaped my scholarly pursuits and self-discovery. I am indebted to my school teachers, namely **Mrs. P. Neerja**, **Mrs. Kalpana Sahu Majumdar**, **Mrs. Sonali Ghosh**, and **Mr. Jyoti P. Nayak**, as their guidance laid the groundwork for my intellectual development. I also owe a debt of gratitude to my under-grad professor, **Dr. Navneet Batra**, whose mentorship has been invaluable. Over time, my academically inclined parents, **Mrs. Suman Raina** and **Prof. P. K. Raina**, have cultivated my thoughts leading to a doctoral study. I have also realized that my educational journey has been influenced by generations past, with the seeds of learning first sown by my grandparents, **Late Smt. Durga Devi**, **Late Shri. G. L. Raina**, **Late Smt. Kamla Devi**, and **Late Shri. R. L. Sharma**. Together, all of them have instilled within me a love for education that continues to serve as a foundation of my academic pursuits.

For making the dream of a doctorate a reality I would foremost, extend my appreciation to **Dr. Rajkumar P. Thummer**, Boss of the Laboratory for Stem Cell Engineering and Regenerative Medicine (SCERM) and my thesis supervisor. His constant support and unwavering belief in my abilities have been the cornerstone of my research endeavors. Our collaborator and my guiding light for this journey is **Prof. R. Velayudhan Shaji**. He generously offered his guidance when I had nothing but enthusiasm and determination to work on my thesis. His mentorship and motivation, even during the most challenging times, has nurtured my scientific temperament which helped me discover my true research potential. He is an inspiration, and I hope to hold on to his teachings in the future. I'm grateful to have known **Dr. Shirisha Nagotu** from the Department of Biosciences and Bioengineering (BSBE), being a part of the SCERM-OBCAL family.

Having discussions with her about her experiences has been incredibly inspiring for me as a female researcher.

I would like to extend my gratitude to the members of my doctoral committee, **Dr. Anil M. Limaye** (Chairperson) and **Prof. Manish Kumar** of the Department of BSBE, and **Prof. Lal Mohan Kundu** of the Department of Chemistry for their valuable insights and expertise. My committee played a crucial role in guiding my efforts towards success by navigating experimental complexities while seeing the broader picture.

I also acknowledge the Ministry of Education, India, for the research fellowship to pursue my PhD and the Department of BSBE for all state-of-the-art facilities to carry out my work. I express my special thanks to **Dr. Krishna P. Bhabak** and his student **Dr. Debojit Bhatacharjee**, Department of Chemistry, IITG for making me part of establishing an interdisciplinary work environment. I would like to extend my sincere thanks to **Dr. B. Sandhya Rani** for helping me understand the world of flow cytometry and fluorescent scanning microscopy at CSCR Vellore. I would also like to extend my thanks to **Mrs. Chitra Premkumar**, **Mrs. Dhavapriya Palani**, **Mrs. Praveena Rajesh**, **Mr. Senthil Kumar G.**, **Ms. Sweety Priyanka**, **Ms. Rakshini Ravichandran**, **Mr. Abdul Muthallib**, and **Mr. Joseph Joel** at CSCR and CMC Vellore for helping me learn lab techniques and creating an enriching environment.

Having worked in three labs (SCERM lab, CSCR Vellore and CMC Vellore) during my thesis work I have numerous seniors and colleagues I want to convey my heartfelt thanks to - **Dr. H. Krishna Kumar**, **Dr. Chandrima Dey**, **Dr. Gloria Narayan**, **Dr. Madhuri Thool**, **Dr. Dinesh Babu**, **Dr. Abhirup Bagchi**, **Dr. Krittika Nandy**, **Dr. Gaurav Joshi**, **Dr. Esther Sathya Bama Benjamin**, **Kirti Modak**, **Sonam Rani**, **Karthik Chambayil**, **Anurag Dutta Chaudhury**, **Debanjan Roy**, **Remi Treasa Eugine**, **Daniel Zechariah Paul**, **Yazhini Sivamani**, and **Betty K. Kumary**. Their insights, perspectives, and camaraderie have enriched my intellectual and personal growth, and I am honoured to have worked with them. Again, I want to acknowledge some of them, specifically **Krishna**, **Debojit** bhaiya, and **Kirti**. I have been profoundly influenced and helped by their scientific persona. I thank them for teaching me the research basics and helping me troubleshoot my experiments. I am genuinely thankful for everything they have done for me. I hope to continue troubling them in the future.

I extend my gratitude to my current lab colleagues, each of whom has contributed uniquely to my academic journey. **Pradeep** and I have engaged in insightful discussions, fostering a bond rooted in shared experiences within the lab. Discussions

with **Atreyee** and **Naveen** have been insightful in sharing valuable knowledge into their respective research areas of cardiomyocytes and β-cells, fostering a mutually beneficial exchange of insights. Special thanks to **Allan** for his help during my final stages in the lab. Gratitude extends to **Harsh**, **Shubham**, and **Asmita** for their collective support and camaraderie. I am also thankful to other past SCERM lab members for their unique and fruitful contributions throughout my PhD journey, such as **Sankalita**, **Poulomi**, **Akriti**, **Tirthankar**, **Ronima**, **Arnab**, and **Sujal**. I am truly fortunate to have had such an exceptional experience to work alongside these people.

"*Memories with the right people always remain priceless.*"

I am blessed to find some great friends on campus who are more like family now. Firstly, I would love to thank - **Dr. Deepshika Jha**, my first friend for all the love, warmth, laughter and wonderful times we shared. I would also like to thank **Dr. Chandrima Dey** for her amazing cooking and soulful music and **Dr. Gloria Narayan** and **Dr. Madhuri Thool** for it was effortless comfort and warmth with both of you. Next, I thank my hostel gang - **Dr. Tanmayee Samantaray**, **Dr. Pratibha Maurya**, **Dr. Priyanka Panigrahi**, **Dr. Tasrin Shahnaz**, **Dr. Tinka Singh**, **Begum Emte Ajom**, and **Roushni Kumari**. I would like to extend thanks to my chai time friends - **Ankita**, **Deepa**, **Shaurya**, **Pavan**, **Sandhya**, and **Satyendu.** Additionally, I thank my **Btwin bicycles** for being my faithful companion in my campus life. Thank you all for making my stay at IITG joyful and more memorable. I express my heartiest gratitude to my dear friend beyond campus - **Pooja** for her love, unending support and consistent pep talks, fuelling my spirits, and making this journey smooth and doable. I am deeply grateful to my best friend **Uma** and her husband, **Kartik**, who supported me from a different continent. Their love and understanding kept my belief in me alive. I extend my heartfelt gratitude to the Bansal family - **Dr. Rajni**, **Dr. Sunil**, thier children **Dr. Nitika**, **Dr. Suprav**, and their granddaughter **Myra**, in Boston. They provided me a home away from home ensuring that my trip to the USA for ISSCR 2023 was a great success.

I can't thank enough my serine green IITG campus, with beautiful landscapes along with facilities like library, hostels, hospital, food court etc., that made my campus life easier and assisted me in focusing on my work. Without hesitation, I can call this institute my second home, which provides a healthy work-study environment.

One person whom I met in the name of CRISPR/Cas9 and somehow, he got "knocked

in" into my life for good is **Kamal**. He has witnessed my journey from the worst days to the best days. He is an incredible science enthusiast which allowed me to share the world of stem cells with him, an avid reader which kept my spirit alive with motivation and captivating science fictions. He has been my supporter and a great advisor in tough times. Now, as my partner, he ensures that my life will be an ever growing one. He has held my hand through thick and thin and made this journey incredible and worth cherishing. I thank him for appreciating and celebrating our time!

"*Families are like branches on a tree. We grow in different directions yet our roots remain as one.*"— Rise Up and Salute the Sun: The Writings of Suzy Kassem

This journey reaches its destination also because of the unconditional love and support of my family, who took care of my every need and accepted me for the person I am. This long journey sometimes became tiring but these gems in my life always stood solid as a rock. My words are not enough to show my gratitude, so I owe them my life - **Mummy**, **Daddy**, **Bhabhi (Dr. Arushi Sharma)**, and **Bhaiya (Dr. Ashutosh Raina)**. My parents have not only patiently awaited the completion of my work but have also actively listened to my tales of cells, their health, and related experiments. I am obliged to my parents for adjusting their schedules to accommodate my work commitments. They have selflessly devoted four months to stay with me during my stay in CMC Vellore. To my brother and sister-in-law, I am thankful for their understanding and guidance during my PhD journey, even as they navigated their own paths in academia. Their love, support, and encouragement have been a source of strength and inspiration. It is their countless personal sacrifices that have propelled me to this pivotal juncture in my life. I also extend my heartfelt gratitude to my extended family for all the love and care.

Lastly, I acknowledge **the ultimate power** for all your blessings and love as it's your will that we walk on.

Thank you to every one of you. Your blessings guide my hand in the art of stem cell generation.

Signing off,

Khyati Raina

"Most experiments don't work the first time. The more ambitious the experiment, the more likely that it will fail. In science, if you are to succeed you need huge passion and determination to overcome countless difficulties. You must pay the same attention to detail as to the big picture."

— Magdalena Zernicka-Goetz & Roger Highfield (**The Dance of Life**)

Abstract

Stem cells are self-renewing cells present at the apex of the lineage hierarchy and, therefore, serve as the founder cells during organismal development. Embryonic Stem Cells (ESCs) are pluripotent cells that can differentiate into all the cell types belonging to the three germ layers: ectoderm, mesoderm, and endoderm. However, human ESCs are not considered ideal for cell therapy applications because of ethical issues and their inability to be used for autologous therapy. Circumventing these limitations, a ground-breaking study was published in 2006, in which pluripotency was induced in terminally differentiated cells (fibroblasts) by the introduction of a cocktail of transcription factors, namely OCT3/4, SOX2, KLF4, and c-MYC (popularly called Yamanaka factors) in mouse fibroblasts to generate induced Pluripotent Stem Cells (iPSCs). Subsequently, the first human iPSCs were reported from fibroblasts using retroviral and lentiviral transduction of reprogramming factors. Since its inception, various reprogramming approaches and combinations of reprogramming factors have been explored to generate iPSCs with higher efficiency and quality. Among these, a pluripotent cell-specific transcription factor Undifferentiated embryonic cell Transcription Factor 1 (UTF1) was believed to be a promising factor for the generation of quantitatively and qualitatively better human iPSCs due to its high expression in pluripotent stem cells. Therefore, our aim was to elucidate the role of human UTF1 in the generation and maintenance of human iPSCs.

Firstly, the thesis delves into the generation and characterization of a human fibroblast-derived iPSC line IITGi-001A by transfection of oriP/EBNA-1 based episomal plasmids expressing OCT3/4, SOX2, KLF4, L-MYC, LIN28 and a p53 shRNA. This iPSC line expressed core pluripotency markers, maintained normal karyotype, and showed trilineage differentiation potential. Further, genomic PCR confirmed the absence of episomal plasmid integration in this iPSC line, which indicated that the cell line generated was indeed integration-free. In addition, DNA fingerprinting of fibroblasts and iPSCs DNA by microsatellite analysis confirmed the genetic identity of this cell line. This iPSC line was free from mycoplasma contamination.

Secondly, the thesis focuses on establishing the importance of human UTF1 in reprogramming by generating a *UTF1* knockout toolbox using the CRISPR/Cas9 technology. Previously, murine studies have successfully generated iPSCs from *UTF1* knockout or UTF1 knockdown fibroblasts with typical pluripotency features, and only one study reported a significant reduction in reprogramming efficiency. However, the essentiality of UTF1 in human iPSC generation remained unexplored to date. Here, the generation of human iPSCs from *UTF1* knockout fibroblasts revealed that the targeted deletion of the human *UTF1* gene exhibited a significant decline in the reprogramming efficiency of human iPSCs. Moreover, the few iPSC clones that did emerge in the absence of UTF1 showed instability upon expansion, hinting at the importance of UTF1 in maintaining pluripotency.

Thirdly, the thesis explores the effect of *UTF1* knockout in iPSCs. Various studies have deleted the mouse *Utf1* gene in vivo and observed a spectrum of outcomes, ranging from developmental arrest to embryonic lethality or developmental delay resulting in death within two days of birth. These outcomes indicated that UTF1 is crucial for proper murine embryonic development at different stages. Here, we established that, the absence of human UTF1 protein resulted in a loss of viability of iPSCs due to the induction of apoptosis.

Lastly, the thesis delves into understanding the effect of reduced UTF1 levels in human iPSCs. To achieve this, a shRNA against UTF1 was expressed in iPSCs. The decline in UTF1 levels was slower than UTF1 levels observed in *UTF1* knockout iPSCs. Also, the effects were not as pronounced as observed upon *UTF1* knockout iPSCs. These iPSCs did not undergo apoptosis but showed spontaneous differentiation of human iPSCs.

In this thesis, we report the generation and characterization of a human iPSC line named IITGi001-A, derived from a human fibroblast cell line, using a non-integrative reprogramming method. Subsequently, we explored the involvement of human UTF1 in both the reprogramming of human fibroblasts to iPSCs and their maintenance. This investigation of the importance of human UTF1 was carried out by loss of function studies. This was achieved by using a CRISPR/Cas9 gene editing tool to generate human *UTF1* knockout fibroblasts and iPSCs, and gene silencing was achieved using a shRNA targeting human UTF1. The thesis collectively indicates that human UTF1 plays a vital role in the generation and maintenance of human iPSCs.

TABLE OF CONTENTS

List of Figures

List of Tables

iPSCs	induced Pluripotent Stem Cells
PGRN	Pluripotency Gene Regulatory Network
ESCs	Embryonic Stem Cells
SSCs	Spermatogonial Stem Cells
ARF	Alternative Reading Frame
ICM	Inner Cell Mass
ECC	Embryonic Carcinoma Cell
UTF1	Undifferentiated Embryonic Cell Transcription Factor 1
CD	Conserved Domain
ExPASy	Expert Protein Analysis System
PGC	Primordial Germ Cells
ATF-2	Activating Transcription Factor-2
PRC2	Polycomb Repressive Complex 2
H2K27me3	Histone 3 Lysine 27 Trimethylation
H3K4me3	Histone 3 Lysine 4 Trimethylation
DMEM	Dulbecco's Modified Eagle Medium
FBS	Fetal Bovine Serum
STRs	Short Tandem Repeats
HEK293T	Human Embryonic Kidney 293 T-antigen
DBSs	Double-Strand Breaks
BFP	Blue Fluorescent Protein
SDS	Sodium Dodecyl Sulfate
PAGE	Polyacrylamide Gel Electrophoresis
gRNA	guide RNA
PGK	Phosphoglycerate Kinase
EGFP	Enhanced Green Fluorescent Protein
PAM	Protospacer Adjacent Motif
dgRNA	dual guide RNA
kDa	kilo Dalton
KO	Knockout
DOX	Doxycycline
LV	Lentiviruses
ZsGreen	Zoanthus Species Green
shSCA	Short Hairpin Scaffold

rtTA	Reverse Tetracycline Transactivator
miR	microRNA
AAVS1	Adeno-associated Virus Integration Site 1
TALEN	Transcription Activator-Like Effector Nucleases
RHA	Right Homology Arm
LHA	Left Homology Arm
CBA	Chicken β-actin
EP	Endogenous Promoter
PPP1R12C	Protein Phosphatase 1 Regulatory Subunit 12C
CopGFP	Copepod Green Fluorescent Protein

CHAPTER 1

INTRODUCTION

An article based on this chapter is published as follows:

Raina K, Dey C, Thool M, Sudhagar S, Thummer RP **(2021)**. An insight into the role of UTF1 in development, stem cells, and cancer. *Stem Cell Reviews and Reports*. DOI: 10.1007/s12015-021-10127-9.

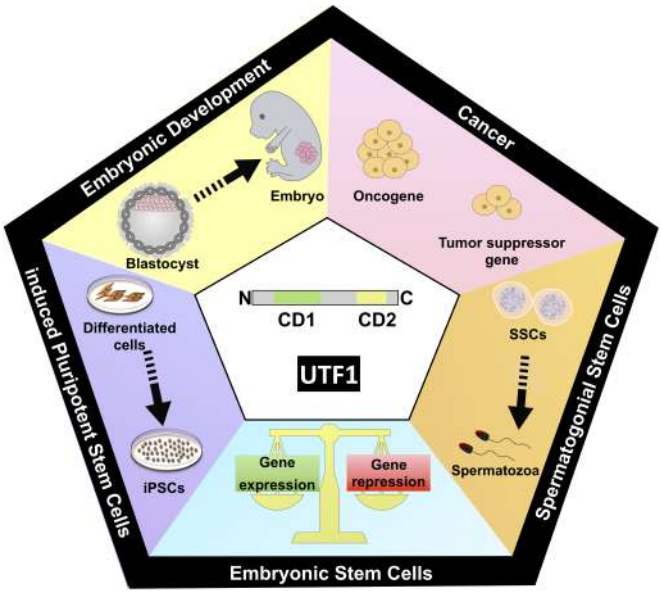


Figure 1.1: Schematic representation of the role of UTF1 in Embryonic Development, iPSCs, Embryonic Stem cells, Spermatogonial Stem Cells, and Cancer.

BRIEF OVERVIEW OF CHAPTER

Chapter 1 offers an introduction about stem cells and the organisms in which these cells have been identified, specifically focusing on induced Pluripotent Stem Cells (iPSCs) and the pluripotency factor UTF1. Commencing with a historical overview of the terminology "stem cells" and their sources, the chapter then proceeds to the emergence of iPSCs, detailing the characteristics of the human iPSC colonies, crucial for their identification and establishment. These cells, possessing embryonic stem cell-like potential yet sidestepping ethical dilemmas, demonstrate pluripotency by differentiating into ectodermal, mesodermal, and endodermal cell lineages. The intricate pluripotency gene regulatory network (PGRN) is regulated by the core pluripotency factors OCT4, SOX2, and NANOG. These three factors cooperatively regulate gene expression to maintain the pluripotent state in human iPSCs.

Pluripotency networks have been of interest since the first human iPSCs were generated. Various other pluripotency-associated genes have been discovered and studied. Based on the critical review of the literature, we choose human UTF1 as our gene of interest. The chapter delves into the UTF1 gene and its protein structure. Highlighting the pivotal role of UTF1, the chapter underscores its significance in embryonic development, Embryonic Stem Cells (ESCs), Spermatogonial stem cells (SSCs), and in cancer and iPSCs. In conclusion, the chapter elucidates the motivation and rationale driving the research within this thesis. The primary objective is to unravel the role of UTF1 in human somatic cell reprogramming and to understand its importance in maintaining the pluripotency network.

1.1 Introduction of the term “Stem Cell”

A German biologist, Ernst Haeckel was the first to use the term “stem cell” in scientific literature from an evolutionary point of view as early as 1868 [48]. In this study, the phylogenetic trees were called “Stammbäume” (German for family trees or “stem trees"). These Stammbäume represented the evolution of organisms by descent and the term “Stammzelle” (German for stem cell) described the unicellular organism from which all multicellular organisms evolved, supporting the Darwinian theory of evolution [48, 49]. In subsequent research, the focus of the term "stem cells" transitioned from evolutionary studies to embryology [50]. This terminology referred to a distinct cell in the embryo capable of giving rise to more specialized cells. This terminology was supported by others too in the late 19^{th} century. Embryologists of this era believed that there were specialized cells (now referred to as the germ cells) that would be distinct from the rest of the body (now referred to as the somatic cells) [147]. Further, Theodor Boveri while working on nematode *Ascaris*, proposed that the stem cells are cells along the germline lineage between the fertilized egg and committed germ cells [14, 15]. Another researcher, Valentin Häcker, stated that the stem cell undergoes asymmetrical cell division, giving rise to mesoderm and germ cells [51]. This work was popularized in the book “The Cell in Development and Inheritance” [155], and stem cell was used to refer to an unspecialized mother cell of the germline [112]. Interestingly, in 1896, Pappenheim used the term “stem cell” to describe a precursor cell capable of giving rise to red and white blood cells [105]. Thereafter, the term “stem cell” gained popularity in the early 1900s in hematology and histology to denote the blood precursor cells in bone marrow [23]. During wartime, the lifesaving potential of bone marrow transplants in irradiated animals was discovered [121]. Eventually, the existence of the hematopoietic stem cells was established [7, 134, 135]. By 1983, Potten compiled a comprehensive collection of reviews, expanding the concept of stem cells beyond hematology to encompass similar cells in epithelial tissues, tumors, and lower organisms [110]. Lajtha’s chapter in this volume marked a pivotal moment, solidifying the modern definition of stem cells [70]. The definition was based on the perceptions of radiation biologists, hematopoiesis specialists, and leukemia specialists that hematopoietic stem cells endure as long as the organism, and divide more slowly than their progeny. Lajtha’s contribution significantly shaped the contemporary understanding of stem cells, while rooted in hematological tradition, it integrates concepts from embryology [121]. These combined influences create the modern

definition of stem cells, which reads as follows:

1. Stem cells reproduce themselves.
2. Stem cells generate progeny destined to differentiate into functional cell types.
3. Stem cells persist for a long time.
4. Stem cell behavior is regulated by the immediate environment (the niche).

The key features of stem cells are illustrated in 1.2. The first two points emphasize their ability to renew themselves and produce an offspring. "Destined to differentiate" means cells divide for a limited time before specializing. The third point distinguishes stem cells by their ability to grow indefinitely in tissue culture or persist throughout an organism's lifespan. The fourth point highlights that stem cells operate within specific microenvironments, controlling their division and differentiation. Even when grown in vitro, stem cells require specialized media mimicking natural conditions. Stem cell behavior is influenced by their intrinsic properties and environmental factors such as animal lifespan, niche characteristics, or culture medium composition. Understanding stem cells necessitates considering both their inherent traits and their surrounding environment [121].

1.2 Do all living beings have stem cells?

As previously described, one of the defining properties of stem cells is their ability to give rise to themselves and a progeny that can differentiate into other cell types. They are limited to multicellular eukaryotic organisms; therefore, stem cells are absent in prokaryotic cells and unicellular eukaryotes. Further, these cells play an essential role in the development, growth, tissue repair, and regeneration of living beings, thereby functioning as a critical component of their biology [30]. The cellular structure, growth, and development of eukaryotic protists and fungi are much simpler, and thereby, the term stem cells is limited to plants and animals [117].

In plants, the term "meristem" is more commonly used to denote stem cell-like functions. These meristematic cells are responsible for the growth and regeneration of plants. Interestingly, these cells are present throughout the plant body and are not restricted to specific regions, which is the case in animals [52]. In animals, "stem cells"

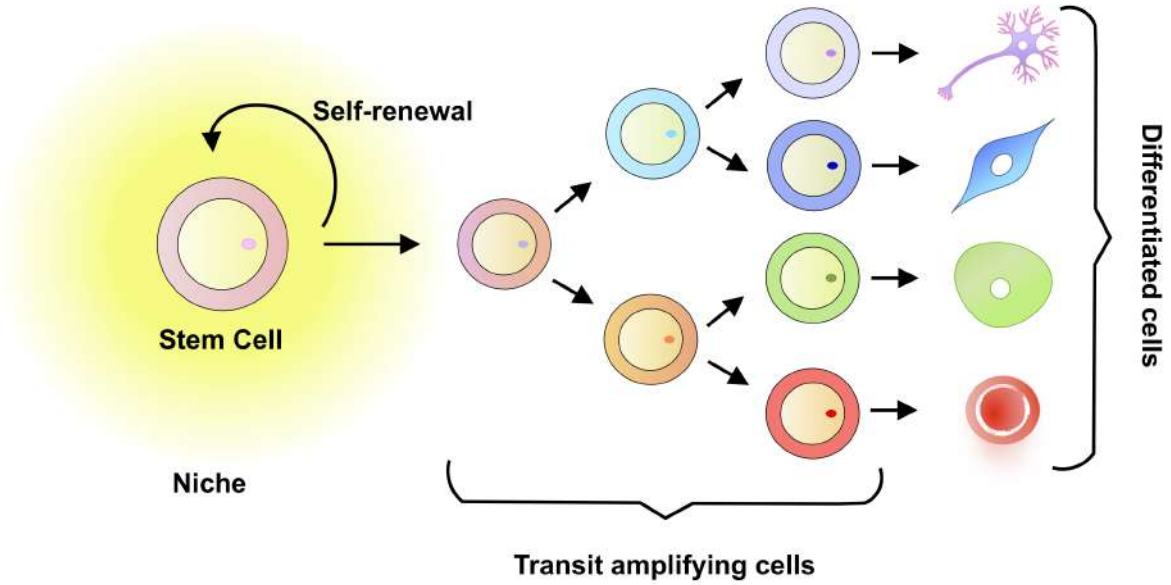


Figure 1.2: An illustration depicting stem cell dynamics. The stem cells reside within a specialized environment, known as a niche (represented by yellow shading), facilitating their self-renewal through divisions. Additionally, they produce differentiated cells through a population of committed transit-amplifying cells. It is important to note that not all stem cell types generate multiple cell types. Adapted from Slack JM. What is a stem cell?. Wiley Interdisciplinary Reviews: Developmental Biology. 2018.

can be observed in simple organisms like sponges and complex organisms like humans. Therefore, stem cells are limited to plants and animals and are not present in all living organisms.

1.2.1 Stem cells in animals

The simplest animals with stem cells are in the phylum **Porifera**, commonly known as sponges. Sponges are simple, aquatic, filter-feeding animals that lack complex tissues and organs. These sponges have specialized cells known as “archaeocytes”, often referred to as amoeboid cells or totipotent cells [28, 35]. These archaeocytes are distributed throughout the sponge’s body and can transform into various cell types as and when needed. Thereby functioning in tissue repair, nutrient transport and reproduction [90, 35]. These archaeocytes possibly demonstrate a primitive form of stem cells, allowing sponges to maintain their simple body plan and carry out essential functions.

The phylum **Cnidaria** hosts one of the extensively studied animals for stem

cells, called Hydra. Its potential immortality and extensive capacity to regenerate and self-renew is due to the presence of three distinct stem cell lineages: ectodermal and endodermal epithelial stem cells and interstitial stem cells [13]. If a Hydra is cut into pieces or injured, the multipotent somatic interstitial stem cells can rapidly differentiate to replace lost or damaged cells and tissues, enabling the organism to recover and regenerate [55]. Despite morphological and functional differences and more than 500 million years of phylogenic separation between Hydra and humans, common signaling pathways are responsible for stem cell maintenance, lineage determination, and differentiation [13].

The phylum **Ctenophora**, commonly known as comb jellies, is not believed to possess stem cells but is thought to provide insights into the early origins of regeneration. They have shown undifferentiated cells, which are not as highly regulated as stem cells, and interestingly, they can dedifferentiate and reorganize the existing cells to replace missing structures. This can also be attributed to the fact that the degree of differentiation is less in Ctenophores in comparison to complex animals [113].

The phylum **Platyhelminthes**, commonly known as flatworms, have evolved a remarkable stem cell system with a single pluripotent adult stem cell type called “neoblast”. This cell can give rise to the entire range of cell types and organs in the planarian body plan [115, 108]. Neoblasts are abundantly present and drive the rapid self-renewal of the entire animal within a matter of weeks [108].

Annelids, such as earthworms and leeches, have been found to possess stem cells that play crucial roles in their remarkable regenerative abilities [3, 91, 166]. These stem cells are involved in forming buds that differentiate into new head or tail segments during anterior and posterior regeneration of missing body parts [166].

Arthropods, including insects, have been a focus of stem cell research, with stem cells identified in both embryonic and adult tissues. However, much of the emphasis on insect stem cell research has been limited to *Drosophila* [20]. **Mollusca**, which includes organisms like snails, clams, and octopuses, have also been found to possess stem cells. Notably, one cell line, the Biomphalaria glabrata embryonic or Bge cell line, has been successfully established in the long history of molluscan tissue culture research [163]. **Echinoderms**, such as sea urchins, have undifferentiated coelomocytes and amebocytes, as well as differentiated phagocytes, which are recruited to damaged areas during regeneration, forming a blastema of undifferentiated, proliferating cells until the missing tissues are restored [32].

Hemichordata have been found to possess stem cells that play a crucial role in their extensive regeneration capabilities. Research has shown that hemichordates deploy functional homologs of critical genes in the mammalian reprogramming factor gene network, such as Oct4, Sox2, Nanog, and Klf4, during regeneration, suggesting the presence of stem cell reprogramming genes in these organisms [58].

In **chordates**, the urochordates have remarkable regenerative abilities. They can asexually propagate to form colonies [136, 141]. In contrast, in **cephalochordates**, the scanning electron microscopy of amphioxus (lancelet) suggests that stem cells may participate in adult notochord growth [54], and in the **vertebrates**, proliferating cells called chordoblasts in the notochord cortex might be stem cells [54].

In summary, stem cells or stem cell-like cells have been identified in most of the subphylum of the kingdom Animalia.

1.2.2 Stem cells in humans

The earliest cell known in human development is the fertilized egg, which is a totipotent cell that can give rise to an entire individual. The first-lineage specification gives rise to trophectoderm and the inner cell mass (ICM); the latter houses the primitive stem cells called the ESCs (typically between days 4-8 after fertilization). These cells can give rise to all three germ layers (ectoderm, mesoderm, and endoderm) [131]. Another pluripotent cell called the embryonic carcinoma cell (ECC) originates from germ line tumors and was extensively used for experimental studies before in vitro ESC culture came into existence, as the latter lacked any tumor identity [123]. As the embryo transitions into a fetus, the stem cells transition from pluripotent to multipotent [12]. With the birth of the baby and its growth into an adult, the role of stem cells shifts from development to repair and maintenance of the somatic cells. Commonly studied adult stem cells are hematopoietic stem cells, mesenchymal stem cells, neural stem cells, epithelial stem cells, and skin stem cells. These multipotent cells are often called somatic stem cells (Mayo Foundation for Medical Education and Research).

With ESC use strongly limited by ethical limitations and adult stem cells limited by their potency, the birth of iPSCs brought a spur in the field of stem cells [64]. Being pluripotent, they could be used to study development. As they could be autologous, the absence of immunosuppressants could become a reality in human tissue transplants. They also brought a new outlook to disease modeling and drug testing.

1.3 Timeline of cellular reprogramming and the birth of iPSCs

The curiosity of understanding how life begins laid the foundation for generating iPSCs about five decades before they actually came into existence (Figure 1.3). Developmental biology contributed to understanding the sequential events that a single cell undergoes, giving birth to offspring. Differentiation, one of the primary cellular events was believed to be unidirectional. Waddington's landscape (1957) depicted this phenomenon where a totipotent cell is depicted as a ball rolling down the hilly landscape into the bottommost inescapable valleys [144]. This doctrine did not believe in the possibility of reprogramming [126]. However, in Cambridge, two researchers, Briggs and Kings, with a contradictory belief, transplanted *Rana pipiens* (Northern leopard frog) nuclei from blastula into enucleated frog eggs, which produced adult frogs. However, the results were challenged because the nuclei were from an undifferentiated state [16]. John Gurdon, in the early 1960s, transplanted tadpole intestinal epithelial cell nucleus (differentiated) in *Xenopus laevis* (African clawed frog) enucleated egg, which gave birth to fertile adults [46]. Extending this success in mammals was a task accomplished by Ian Wilmut in 1997 with the birth of media sensation Dolly sheep using the technique now referred to as somatic cell nuclear transfer [154]. These studies established that nuclei of differentiated cells retain their totipotency and shed light on the cytoplasm's potential to change the chromatin state and, thus, expression, resulting in the altered cellular state [114, 122].

The next breakthrough came in 1981 when two groups, Martin Evans and Matthew Kauffman and Gail R Martin, cultured the first in vivo mouse ESCs [83, 29]. Subsequently, Thomson and group isolated human ESCs from donated cleavage stage embryos in 1998 [131]. Even though a breakthrough, this study could not be used to its full potential because of the ethical concerns involving the destruction of a human embryo.

Circumventing this ethical concern required using human cells which were pluripotent but not acquired from an embryo. The fact that such cells were not present in nature once an embryo developed into an adult led to the development of the first-ever iPSCs derived from differentiated adult cells. The pioneering discovery of these cells was accomplished by generating murine ESC-like cells, now referred to as iPSCs [125].

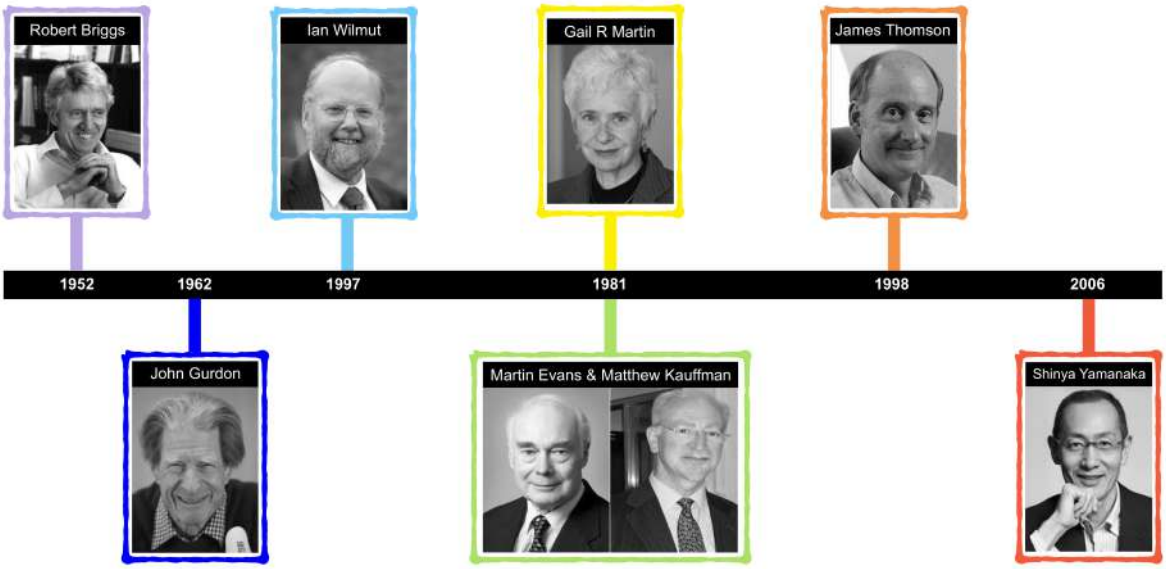


Figure 1.3: Evolution of cellular reprogramming towards the conceptualization and development of iPSCs.

Soon, this pioneering study was performed on human cells by over-expression of the four Yamanaka factors: OSKM (OCT4, SOX2, KLF4, and c-MYC) through retroviral transduction in human fibroblasts [127]. Thomson lab was also able to generate human iPSCs using lentiviral vectors overexpressing a different set of reprogramming factors: OCT4, SOX2, NANOG, and LIN-28 [164].

Since then, numerous groups have studied iPSC generation in other mammals. The studies range from rats to the elephants. Some of the mammals studied to date are listed in Table 1.1.

1.4 Morphology and characteristics of human iPSCs

Human iPSCs have distinct morphological characteristics that distinguish them from differentiated cells. They are as follows:

1. **Colony morphology**: iPSC colonies have a smooth, well-defined border with a compact appearance [92, 84]. The human iPSC colonies have a refractile appearance, giving them the appearance of a shiny boundary [31].

2. **Cell morphology**: Individual iPSCs have a high nucleus-to-cytoplasm ratio and prominent nucleoli [145, 47]. They exhibit an epithelial-like morphology with

Table 1.1: A list of mammals whose iPSCs have been derived or studied.

Mammal	Common Name	Scientific Name	Reference
Monkey	Rhesus Monkey	*Macaca mulatta*	[78]
Rat	Brown Rat	*Rattus norvegicus*	[75]
Pig	Domestic Pig	*Sus scrofa domesticus*	[148]
Dog	Domestic Dog	*Canis lupus familiaris*	[150]
Cat	Domestic Cat	*Felis catus*	[26]
Horse	Domestic Horse	*Equus ferus caballus*	[151]
Cattle	Domestic Cattle	*Bos taurus*	[109]
Rabbit	Domestic Rabbit	*Oryctolagus cuniculus*	[1]
Sheep	Domestic Sheep	*Ovis aries*	[79]
Ferret	Domestic Ferret	*Mustela putorius furo*	[38]
Rhinoceros	Sumatran Rhinoceros	*Dicerorhinus sumatrensis*	[168]
Elephant	African Elephant	*Elephas maximus*	[4]
Gorilla	Gorilla	*Gorilla gorilla*	[42]
Chimpanzee	Chimpanzee	*Pan troglodytes*	[36]
Bonobo	Bonobo	*Pan paniscus*	[160]
Orangutan	Orangutan	*Pongo pygmaeus*	[42]
Gibbon	Lar Gibbon	*Hylobates lar*	[6]
Baboon	Olive Baboon	*Papio anubis*	[124]
Bat	Horseshoe bat Mouse-eared bat	*Rhinolophus ferrumequinum* *Myotis myotis*	[25]
Platypus	Platypus	*Ornithorhynchus anatinus*	[152]

specialized cell-cell junctions [87, 149].

3. **Cytoskeleton**: iPSCs show actin localization under the upper cytoplasmic membrane [81]. The cytoplasm contains ribosomes, glycogen associated with lipid droplets, and elongated mitochondria [43].

4. **Nuclei**: iPSC nuclei are euchromatic with reticulated nucleoli [22]. The expression of endogenous pluripotency markers such as OCT4, SOX2, and NANOG, as well as surface markers like TRA-1-60 and SSEA-4, are commonly used for the characterization of human iPSCs [145, 9].

5. **Differentiation**: Upon differentiation, iPSCs can give rise to cells of all three germ layers: ectoderm, mesoderm, and endoderm [127]. iPSCs lose their distinct colony morphology [84]. Partially reprogrammed iPSCs can be identified by the lack of transgene repression [22].

These morphological and molecular characteristics are used to identify and monitor the pluripotent state of human iPSCs during generation, culture, and differentiation. Automated image analysis and machine learning algorithms have the potential to be used for identifying iPSC colonies based on these features [47].

1.5 The Pluripotency Gene Regulatory Network

The pluripotency gene regulatory network (PGRN) is a complex, interconnected network of genes that cooperatively regulate gene expression to maintain the pluripotent state in human iPSCs [74, 73]. Key features of the PGRN in human iPSCs include:

1. **Core pluripotency factors**: The PGRN is centered around core transcription factors like Oct4, Sox2, and Nanog that are essential for maintaining pluripotency [74, 73]. These factors form an interconnected network that regulates the expression of genes involved in self-renewal and differentiation.

2. **Extrinsic signals**: Specific extrinsic signals and signaling pathways, such as LIF, BMP, FGF, and Wnt, provide important inputs into the PGRN to regulate pluripotency [74].

3. **Heterogeneity**: The PGRN exhibits transcriptional heterogeneity in self-renewing iPSC cultures, with core pluripotency factors showing variable expression levels [74]. This heterogeneity may be part of the regulatory assets of the network.

4. **Pluripotent states**: The PGRN can stabilize iPSCs in distinct pluripotent states, such as the ground state, primed state, and alternative pluripotent states, which have different transcriptional and epigenetic features [74]. These states are interconvertible in vitro.

5. **Post-transcriptional regulation**: In addition to transcriptional regulation, the PGRN is also subjected to post-transcriptional controls, particularly those induced by RNA-binding proteins and alternative splicing [74].

In summary, the PGRN in human iPSCs is a complex, dynamic system that integrates multiple layers of regulation to maintain the pluripotent state and control differentiation potential.

1.6 UTF1 gene

Since the inception of the first iPSCs, various combinations of factors have been investigated for their generation. Among these less explored human pluripotency factors is a single GC-rich gene, no other gene having a similar sequence - the Undifferentiated embryonic cell Transcription Factor 1 *UTF1* (Table 1.2). This gene exists only in placental (eutherian-specific) mammals [99]. It is localized on mouse chromosome 7 at 7F5, and in humans, it is located on chromosome 10 at 10q26 [34, 96]. The 5' region of the gene comprises a TATA-less promoter with four GC boxes (Figure 1.4), typical of a housekeeping gene [96]. Of these four GC boxes, the first GC box serves the most significant function as its mutation showed a 60% decline in promoter activity with lower or no effect in the case of other boxes [96]. Multiple transcription start sites are reported to exist downstream of the promoter, of which 15 have been identified so far [102, 96]. Interestingly, some of these transcription start sites are present after the start codon because a second methionine codon is present downstream of the first start codon (Figure 1.4). As a result, two different mouse UTF1 proteins co-exist in nature, with the longer 339 amino acid (aa) protein being more abundant than the shorter version (297 aa) [96]. However, in humans, only one UTF1 protein of 341 aa is reported [34, 67]. In the 3' region of the gene, an enhancer region was identified with a regulatory element called M1 (Figure 1.4), a conserved octamer sequence essential for NANOG expression [116, 128]. This is soon followed by another regulatory element, namely the Octamer and Sox binding motifs (Figure 1.4) selectively interacting with UTF1 regulators, OCT4, and SOX2 [102, 98], which, when mutated, resulted in a drastic decrease in the expression levels of UTF1 [95]. Therefore, it is a transcriptional target of key pluripotency regulators OCT4 and SOX2 in ESCs.

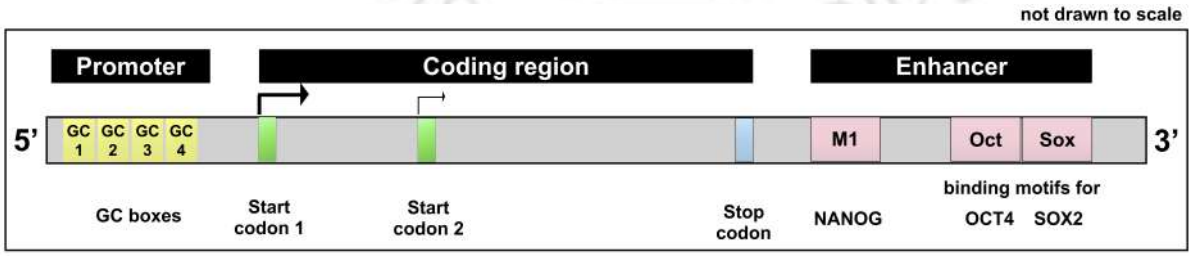


Figure 1.4: The human UTF1 gene. The promoter region of the gene contains four GC boxes. The coding region contains two start codons. The 3'end has an enhancer region with NANOG, OCT, and SOX binding motifs. GC1: GGGCGG; GC2: GGGCGG; GC3: GGGCGG; GC4: GGGCGG; M1: GTCTGGGT; Oct: ACTAGCAT; Sox: AACAATG. Not drawn to scale.

Table 1.2: Physicochemical parameters of human UTF1 sequence.

Parameters	**Values**
Protein-coding DNA Sequence	
Number of nucleotides	1026 bp
Average GC%	78.16%
Nucleotide composition	Adenine (A): 11.99% (123 bases)
	Thymine (T): 9.84% (101 bases)
	Guanine (G): 31.87% (327 bases)
	Cytosine (C): 46.29% (475 bases)
Protein	
Number of amino acids	341
Molecular weight	36.438 kDa
Theorerical pI	10.9
Amino acid composition	Proline (P): 19.4% (66 residues)
	Alanine (A): 12.3% (42 residues)
	Arginine (R): 11.7% (40 residues)
	Leucine (L): 10.9% (37 residues)
	Glycine (G): 7.0% (24 residues)
Total number of negatively charged residues	Aspartic acid+Glutamic acid: 31
Total number of positively charged residues	Arginine+Lysine+Histidine: 50

1.7 UTF1 protein

The sequence of UTF1 protein was found to be proline-rich (5-273 aa), with a majority of basic amino acids like arginine contributing to its high isoelectric point (Table 1.2). Moreover, the sequence of UTF1 protein is unique as it shows no homology to any previously reported protein sequences apart from a Myb/SANT domain and the presence of a leucine zipper motif [34, 139, 67]. Furthermore, two highly conserved regions were identified in UTF1 protein, one at the amino-terminal end and the other at the carboxyl-terminal end, giving rise to their names, i.e., conserved domain (CD)1 and CD2, respectively (Figure 1.5); [34, 139, 67]. A comparison between human and mouse UTF1 protein sequences showed a 63% similarity (Figure 1.6). Further, a comparison between CD1 and CD2 of human and mouse protein sequences showed 84.5% and 85.9% similarity, respectively (Figure 1.6). The determination of the secondary structure of the human UTF1 protein using bioinformatics software showed that it primarily constitutes of random coils and α-helices (Table 1.3). The CD1 is homologous to a Myb/SANT DNA-binding domain [139, 67]. UTF1 protein is localized to the nucleus (excluded from the

nucleoli) and is associated with DNA during all the stages of cell cycle [139, 67, 132]. The deletion of CD1 resulted in the mislocalization of the protein to the cytoplasm [139, 67], indicating that CD1 is responsible for the nuclear localization of the protein. Moreover, studies showed that both human and mouse UTF1 proteins were tightly bound to the chromatin [139, 67]. The earlier mentioned leucine zipper motif is present in the CD2 domain and was reported to contribute to long-term immobilization by tightly binding to the DNA, much like the histones [102, 34, 139, 67, 133]. Functional characterization of two missense single nucleotide polymorphisms (rs11599284 (G73R) and rs4480453 (L275M)) present in each CD, when mutated together, enhanced the mobility of UTF1 protein, indicating a decrease in chromatin association [133]. Whether this change in chromatin association has altered chromatin structure, target gene specificity, or differentiation defects in pluripotent cells remains to be investigated. However, the nuclear localization and transcriptional repressor activity remained unaffected [133].

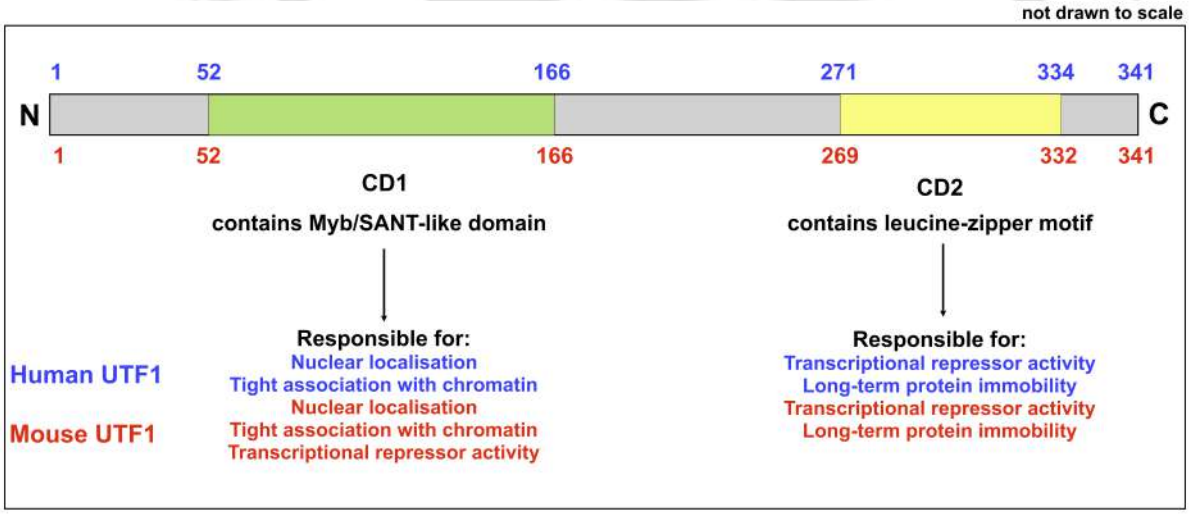


Figure 1.5: The human UTF1 protein. The 341 amino acids long UTF1 protein comprises of a Myb/SANT-like DNA binding domain (in CD1 domain) and a protein-protein interaction leucine zipper motif (in CD2 domain). N: amino-terminal; C: carboxyl-terminal; CD: conserved domain. Not drawn to scale.

It was also reported that UTF1 is likely to undergo phosphorylation at S18, T35, S42, S54, and S245 residues [34, 153, 140]. Furthermore, UTF1 is likely to get dephosphorylated by carboxyl-terminal domain small phosphatases [56] the latter are involved in various functions, including cell cycle regulation and differentiation. It was recently observed that UTF1 is a sumoylation target, and sumoylation modulates the chromatin association of UTF1 [21]. A detailed investigation is required to understand

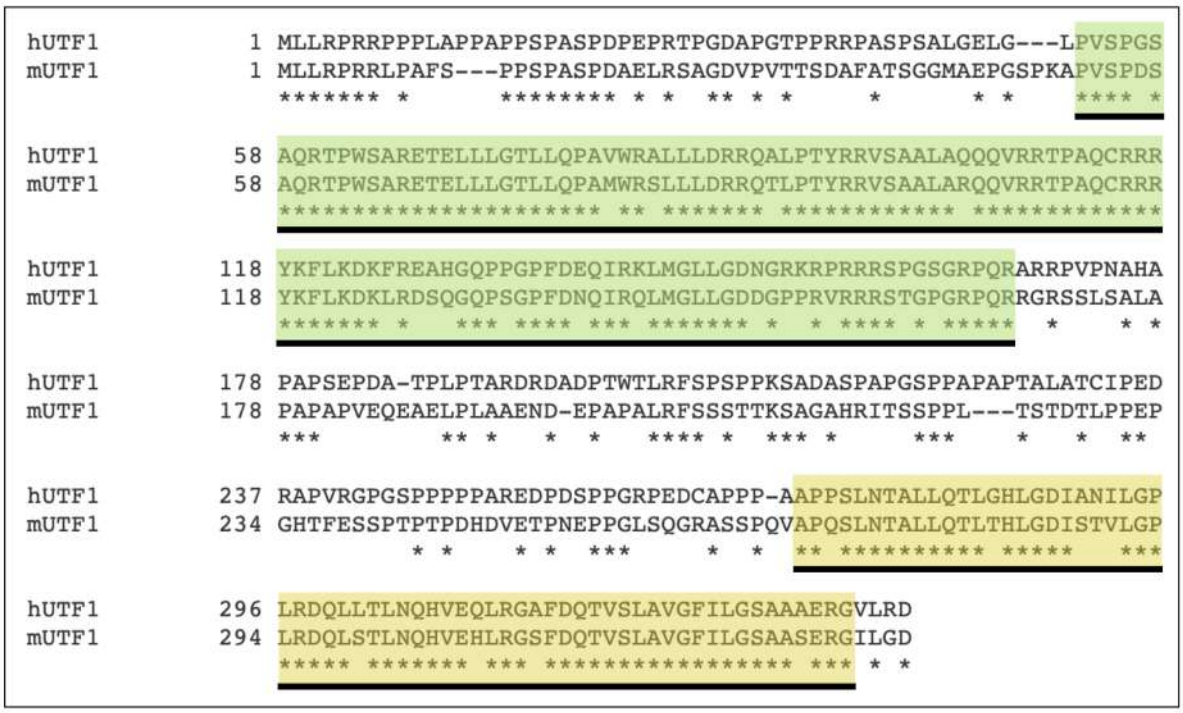


Figure 1.6: Sequence similarity between human and mouse UTF1 proteins. The amino acids conserved between human and mouse proteins are marked with an asterisk (*). The CD1 and CD2 domains are represented by light green and light yellow, respectively. The alignment was generated using the Expert Protein Analysis System (ExPASy) SIM with default parameters

Table 1.3: Predicted secondary structure of human UTF1 protein using different bioinformatics software.

Software	**α-helix**	**β-sheet**	**Random coil**	**Other**
PSIPRED [61]	36.3%	-	63.6%	-
	(124 residues)	-	(217 residues)	-
SOPMA [41]	33.72%	3.52%	59.53%	3.23%
	(115 residues)	(12 residues)	(203 residues)	(11 residues)
GOR4 [39]	32.84%	1.76%	65.40%	-
	(112 residues)	(6 residues)	(223 residues)	-

PSIPRED: PSI-blast based secondary structure PREDiction; SOPMA: Self-Optimized Prediction Method with Alignment; GOR4: Garnier-Osguthorpe-Robson fourth version

the biological significance of the phosphorylation and sumoylation sites in the UTF1 protein.

1.8 Role of UTF1 in embryonic development

UTF1 is expressed during early embryogenesis (Figure 1.7). It exhibited no measurable expression at the early cleavage stages before blastocyst formation. In the early blastocyst stage, UTF1 expression was observed only in ICM, followed by its expression in the epiblast [102, 37]. Further, UTF1 mRNA expression was restricted to the primitive ectoderm (an ICM derivative), and a faint expression was observed in the extra-embryonic ectoderm (a trophectoderm derivative) [102]. Upon the development of the primitive streak, a decrease in UTF1 expression was observed from the posterior to the anterior part of the embryo [102]. Further, through the embryonic development, expression of UTF1 was gradually lost in the neural fold and the chorion, restricting itself to the PGCs in the developing embryos [5, 37, 63, 99, 102, 24]. The highest expression of UTF1 was detected in the epiblast and PGCs during mouse embryonic development (Figure 1.7) [37].

To further understand the requirement of UTF1 during embryonic development, the effect of its genetic deletion was studied in mice. At least one functional *Utf1* allele in both the parental germline was critical to contributing to the complete development of *Utf1* null embryos, suggesting an intergenerational epigenetic inheritance of UTF1 protein or mRNA in parental germ cells [5]. Genetic deletion of *Utf1* in mice either resulted in a developmental arrest [5], or embryonic lethality [63], or developmental delay in midgestation embryos and newborn mice and delay in placental growth, possibly due to placental insufficiency [99]. Notably, the *Utf1* homozygous mutant mice could not survive for more than two days after birth [99]. The ubiquitous inactivation of *Utf1* in mice with a hybrid (C57BL/6×129) genetic background resulted in embryonic lethality at E10.5 [63]. The developmental arrest observed in $Utf1^{-/-}$ embryos, generated from heterozygous *Utf1* parents, was reabsorbed in the uterus and identified by their residual placenta [63]. On the other hand, the $Utf1^{-/-}$ pups born, irrespective of their genetic backgrounds, were smaller in size than the $Utf1^{+/+}$ and $Utf1^{+/-}$ pups, suggesting that the absence of UTF1 resulted in the developmental delay, which could be explained by placental insufficiency [99, 63, 5]. UTF1 expression was observed in the placentas of wild-type mice till the late developmental stage; thus, the placenta of knockout mice was phenotypically smaller in size, resulting in placental insufficiency [99]. Therefore, the deletion of *Utf1* in vivo displayed all possibilities ranging from developmental arrest to embryonic lethality to death within two days of birth (Table 1.4). These variations could

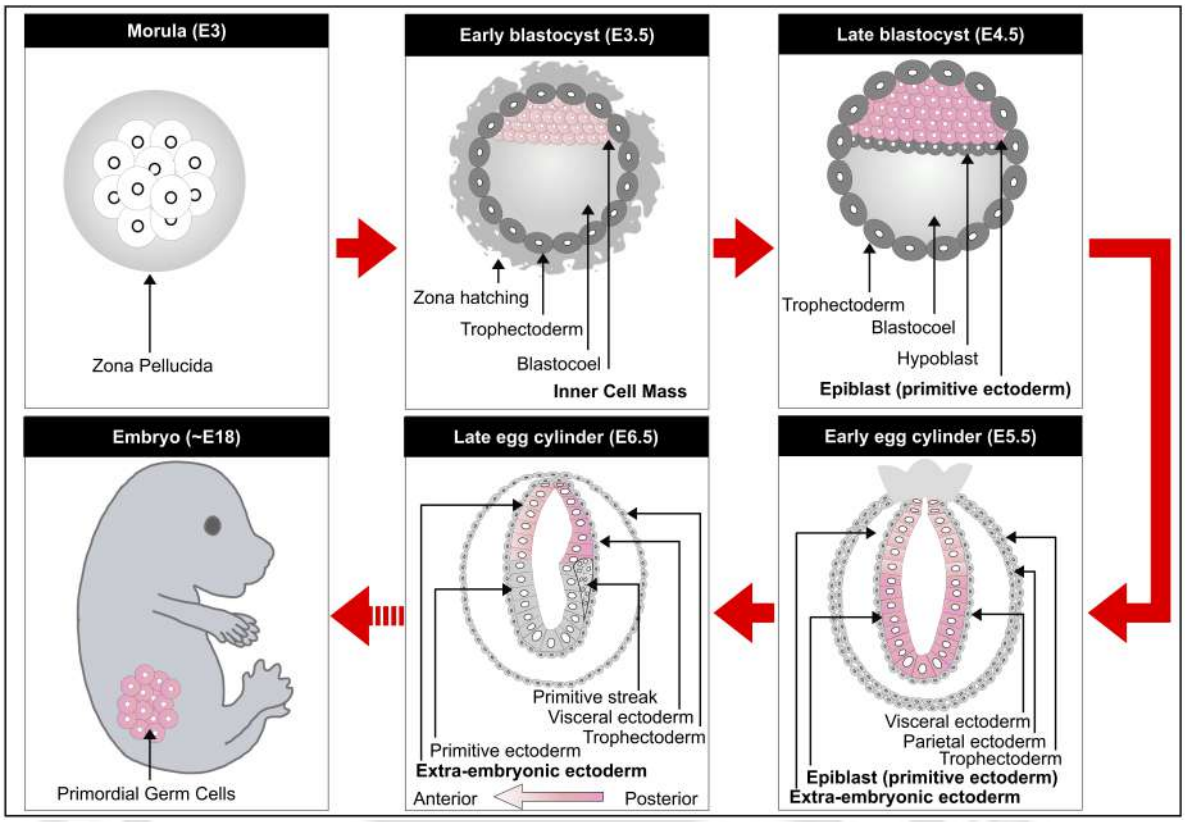


Figure 1.7: Expression of UTF1 during mouse embryogenesis. Post-fertilization, UTF1 is not expressed till the blastocyst stage. A rapid onset occurs during the early blastocyst stage, with enhanced expression in the epiblast of the late blastocyst stage. Moreover, its expression decreases in the extra-embryonic ectoderm and is restricted to the primitive ectoderm. Further, along with the formation of the primitive streak, UTF1 expression becomes restricted to the extra-embryonic ectoderm. It shows a gradual decrease from the posterior to the anterior side of the embryo. Eventually, the expression gets restricted to primordial germ cells (PGCs) only.

be attributed to the variable genetic backgrounds of the mice under experiment (Table 1.4), contributing to UTF1-mediated translational and epigenetic outcomes [99, 5, 63]. Overall, it contributed to proper murine embryonic development.

1.9 Role of UTF1 in ESCs

UTF1 was first identified in undifferentiated mouse and human pluripotent cells (ESCs and ECCs) [102, 34]. This restricted expression, specifically in undifferentiated cells, contributed to its name, “Undifferentiated Embryonic cell Transcription Factor 1” [102]. The core pluripotency regulators, OCT4 and SOX2, are known to control UTF1 expression [128, 5, 97, 98]. Though the expression pattern of UTF1 is similar to OCT4

Table 1.4: Outcome of mouse UTF1 knockout studies.

Study	Parental genetic background	Embryonic lethality	Mendelian ratio (+/+ : +/- : -/-)
Nishimoto et al., 2013 [99]	C57BL/6J	Developmental delay; died within 2 days	Displayed expected Mendelian ratios during gestational stages
	C57BL/6J (25%) and ICR (75%)	Viable and fertile	Displayed expected Mendelian ratios during gestational stages
	(mixed genetic background)	but with growth defects	
Kasowitz et al., 2017 [63]	FVB/N, C57BL/6 and 129	Interbreeding of Utf1 heterozygous mice failed to yield any $Utf1^{-/-}$ viable offspring	Offspring of $Utf1^{+/-}$ x $Utf1^{+/-}$ crosses: 104:146:0
	(mixed genetic background)		
Bao et al., 2017 [5]	C57BL/6J	Expected numbers	1:2:1)
	♀(-/+) × ♂(-/+)	Half the expected numbers; frequent early developmental arrests	2:1 (instead of the expected 1:1)
	♀(-/+) × ♂(-/-)		
	♀(-/-) × ♂(-/-)	Most embryos arrested (<12 dpc)	

C57BL/6J: C57 black 6 Jackson laboratory; FVB: Friend leukemia virus B; ICR: Institute of Cancer Research

and SOX2, these two factors act as self-renewal and pluripotency regulators, whereas UTF1 contributes towards proper differentiation of pluripotent cells and PGCs in the presence of differentiation cues [59, 139, 68, 5]. Upon differentiation, a decrease in expression was observed for UTF1 and other pluripotency factors like OCT4 and SOX2, though the decline in expression was more rapid for UTF1 [102, 34, 95].

Understanding the molecular role of UTF1 in pluripotent cells, Eggen lab silenced UTF1 expression to study its effect in mouse ESCs and ECCs [139]. This study showed that UTF1-depleted pluripotent cells maintained their self-renewal and pluripotency characteristics but failed to differentiate properly, i.e., particularly towards mesodermal and endodermal lineages [139]. The depletion of UTF1 also leads to an increase in the doubling time of ESCs and ECCs [139]. Additionally, slow-growing ESCs showed reduced UTF1 expression [98]. The role of UTF1 in ensuring the rapid proliferation of ESCs could be due to its ability to block Myc-mediated activation of the Arf feedback loop [60]. Although the role of UTF1 in cell proliferation was demonstrated [139, 98], it was dispensable for ESC self-renewal [139, 60]. Further, overexpression of UTF1 in ESCs and ECCs interfered with their differentiation potential [132]. Thus, UTF1 levels can alter the differentiation potential of pluripotent cells. In summary, UTF1 plays a crucial role in the proliferation and differentiation of ESCs but not for maintaining self-renewal and pluripotency.

Functionally, the role of UTF1 as a transcriptional regulator has been controversial in pluripotent cells. Although its exact role as a classical transcription factor has not been determined to date (no consensus DNA binding sequence reported so far), various studies have reported that it can function either as a transcriptional coactivator [102, 34] or as a transcriptional repressor [34, 139, 67, 68]. As a coactivator, UTF1 mediates the physical interaction between the activating transcription factor-2 (ATF-2) and the transcriptional machinery (possibly through basal transcription factor II D), thereby contributing to transcriptional activation in an ATF-2-dependent manner [102, 34]. On the other hand, in the absence of ATF-2, UTF1 showed repressor activity [34]. Also, luciferase reporter-based studies showed it acts as a transcriptional repressor [139, 67]. Both CD1 and CD2 domains are reported to be essential for their transcriptional repressor activity in the case of mouse UTF1 [139]. In contrast, only the CD2 domain is responsible for the same in the case of human UTF1 [67]. Complying with this observation, microarray analysis on mouse UTF1-depleted cells showed an upregulation of 1090 genes (including NANOG, a core pluripotency transcription factor) and a downregulation of only 131

genes [68], suggesting that it primarily functions as a transcriptional repressor.

Mechanistically, UTF1 acts as a critical chromatin constituent, preventing superfluous chromatin decondensation and aberrant gene expression in ESCs to facilitate proper differentiation during exit from pluripotency [68]. It may be crucial for the decrease in transcriptional noise [68], which is vital for reducing the transcription of differentiation-inducing genes [167]. Subsequently, it was reported to regulate the multiprotein enzyme complex Polycomb repressive complex 2 (PRC2), which establishes the repressive epigenetic mark histone 3 lysine 27 trimethylation (H3K27me3) and they compete for the same bivalent genes to control their expression in ESCs [60]. The bivalent state of the chromatin in ESCs is created by both the repressive and active epigenetic marks, H3K27me3 and histone 3 lysine 4 trimethylation (H3K4me3), respectively [8]. This characteristic feature of pluripotent cells keeps the developmentally essential genes in a poised state, which becomes committed upon induction of differentiation [8]. Functional attribution of UTF1 was also observed in mRNA pruning. It involves the tagging of mRNAs transcribed from insufficient silencing of these bivalent genes for cytoplasmic degradation by the mRNA decapping complex [60]. These opposing functions on the bivalent genes, i.e., preventing the establishment of repressive H3K27me3 mark and regulating translation of mRNA from bivalent genes, are believed to maintain the "just right" level of bivalency [60, 71]. Interestingly, the *UTF1* gene itself is also regulated by the H3K27me3 mark and regulates its expression by a feedback mechanism [5]. Furthermore, upon differentiation, loss of UTF1 resulted in a shift of bivalency towards the repressive mark (H3K27me3) on developmental genes (OLIG2, NESTIN, and T) and a decrease in pruning activity of the HOXA1 gene [60]. Possibly, UTF1 is involved in the effective reduction of "transcriptional noise" in ESCs by performing various functions at different levels (Figure 1.8), and upon induction of differentiation, it prepares the cells for proper lineage specification.

1.10 Role of UTF1 in spermatogonial stem cells

During early embryogenesis, germ cells differentiate to form male and female gametes. In mice, UTF1 is implicated in male germ cell development, with studies demonstrating its necessity in this process but not in ovarian development [5, 63]. Failure of spermatogenesis, occurring in the seminiferous tubules of the testes at the onset of puberty can lead to male infertility [57].

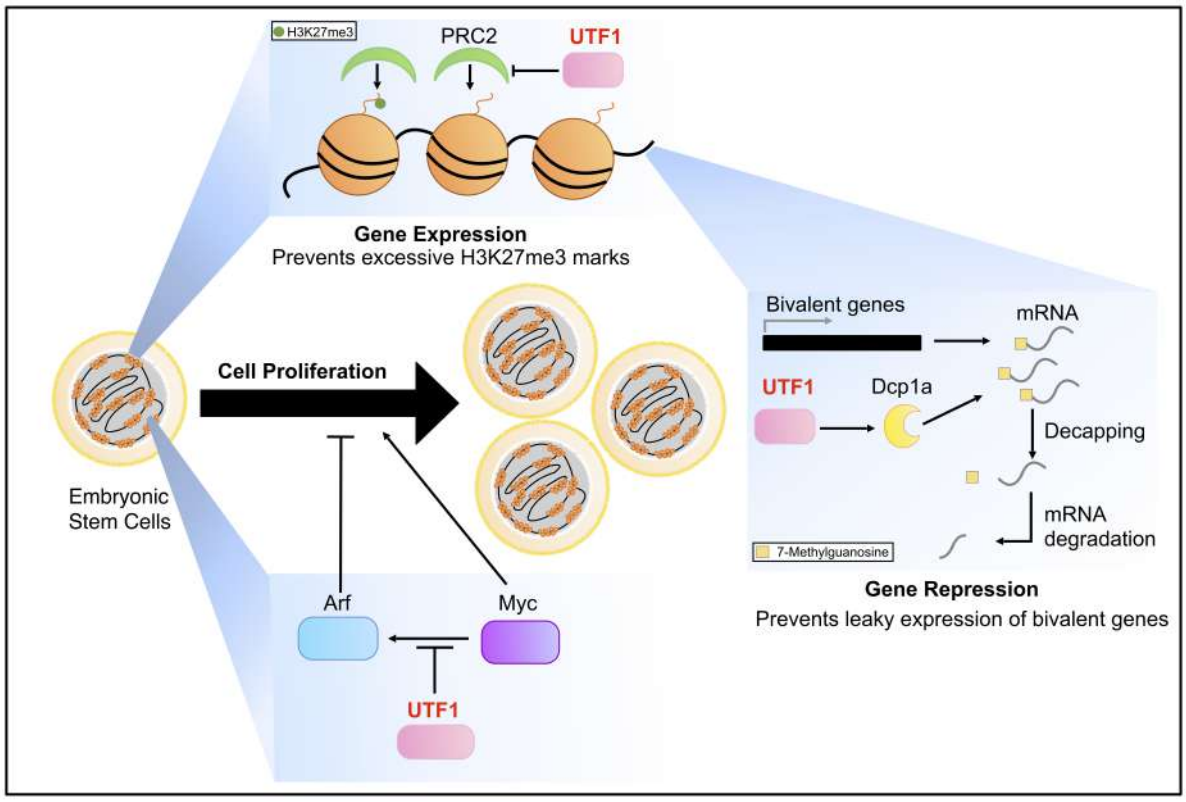


Figure 1.8: UTF1 maintains bivalency and regulates cell proliferation in ESCs. UTF1 regulates the bivalent genes by mRNA pruning and maintaining the epigenetic threshold. It prevents excessive loading of PRC2, thereby preventing excessive establishment of the repressive epigenetic mark H3K27me3. Further, it recruits Dcp1a for decapping mRNA transcripts from insufficiently silenced bivalent genes. These decapped mRNAs are eventually degraded in the cytoplasm by exonucleases, preventing their translation. Additionally, UTF1 ensures rapid proliferation of ESCs by inhibiting the Myc-mediated activation of Arf. PRC2: Polycomb repressive complex 2; Dcp1a: decapping mRNA 1a. Figure prepared from observations and results of [60].

UTF1 emerges as an excellent biomarker for PGCs and spermatogonial stem cells (SSCs) in various studies [138, 129, 119, 94]. Its inactivation resulted in decreased gonocyte numbers at birth and defective spermatogenesis in adult mice [63]. Notably, UTF1 expression is detected in gonocytes, P0 testes, and a subset of spermatogonia but not in spermatocytes or spermatids [63]. Additionally, UTF1-positive spermatogonia exhibit quiescent characteristics, maintaining their ability to differentiate [143, 138].

In contrast, although UTF1 mRNA and protein are present in metaphase II oocytes during female murine ovarian development, no ovarian defects are observed in knockout mice [119, 63]. Furthermore, UTF1 expression is reported in spermatocytic seminomas, suggesting its role in the self-renewal of SSCs and seminomas [69]. UTF1

is crucial for the proper development and functioning of male gonads, contributing to male fertility [5]. Its absence leads to compromised testicular vasculature and inefficient spermatogenesis, indicating potential male infertility [63]. Overall, UTF1 plays a vital role in embryonic gonocyte development and postnatal spermatogenesis.

1.11 Role of UTF1 in cancer

UTF1 expression is documented in various cancer tissues of both germ cell and non-germ cell origin, with notable abundance in ECCs, testicular germ cell neoplasms, and central nervous system germinoma, among others [139, 67, 146, 77, 132, 69, 104]. Elevated UTF1 levels are observed in endometrial and prostate cancer, while reduced levels are noted in colon, renal, and breast cancer compared to normal tissues [89, 161]. Additionally, UTF1 is overexpressed in endometriosis, suggesting potential hormonal regulation of its expression [33].

In cervical cancer, conflicting reports exist regarding UTF1 expression. One study associates its overexpression with hypermethylation of the UTF1 promoter, while another suggests downregulation due to promoter hypermethylation [45, 159]. Its potential as a prognostic marker in brain cancer is notable, with implications for tumor grading and prognosis [85].

Functionally, UTF1 is essential for ECC proliferation, probably similar to its role in ESCs, possibly altering chromatin structure and gene expression profiles to facilitate proliferation [139, 60, 98]. Depletion of UTF1 reduces teratoma tumorigenicity, while its overexpression promotes tumor formation in mice, hinting at an oncogenic function [60, 98].

On the contrary, UTF1 may act as a tumor suppressor by upregulating p27Kip1 and inducing G1/S arrest [159]. Notably, hypermethylated UTF1 is observed in cervical and pharyngeal cancers, but its relationship with gene expression is complex [45, 159, 93].

The variability in UTF1 expression across cancers suggests its potential as a diagnostic biomarker, particularly in endometriosis, cervical cancer, and germ cell tumors [11, 103, 69, 146]. Its DNA methylation status holds promise for diagnostic applications, though conflicting data in cervical cancer necessitates further investigation [45, 103, 159, 93]. Understanding its roles in cancer development offers insights into prognosis and therapy.

1.12 Role of UTF1 in iPSCs

The discovery of iPSCs has unfolded a new research area in stem cell biology due to their wide-scale applications, ranging from understanding developmental biology, disease modeling, drug discovery and toxicity testing, and autologous cell therapy. These cells have a functional and molecular resemblance to ESCs. The similarity in the gene expression profiles of ESCs and iPSCs is directed towards using other pluripotency-associated genes and/or genes involved in early embryonic development for the generation of iPSCs. The first iPSCs derived were by reprogramming mouse fibroblasts using a cocktail of four transcription factors (OCT3/4, SOX2, KLF4, and c-MYC; abbreviated as OSKM; also popularly known as Yamanaka factors) [125]. This reprogramming cocktail was identified from a list of 24 potential factors [125]. Though UTF1 was one of the factors screened, it did not qualify among the four key reprogramming factors in this study. However, the importance of UTF1 in cellular reprogramming was soon highlighted when human foreskin fibroblasts were reprogrammed with UTF1, small interfering RNAs against p53, and the Yamanaka factors, generating iPSCs with enhanced quality and efficiency (>200-fold higher efficiency) (Figure 1.9) [165]. This could be attributed to the improved transition of pre-iPSCs to iPSCs, resulting in enhanced reprogramming efficiency. UTF1 may have aided in establishing a favorable chromatin state or epigenetic profile ideal for switching to pluripotency [165]. Additionally, the removal of oncogenic c-MYC from this six-factor cocktail enhanced reprogramming efficiency by 100-fold compared to the Yamanaka factor cocktail alone. Also, UTF1 (in the presence of OSK or OSKU and a small interfering RNA against p53) resulted in emergence of a majority of iPSC colonies in comparison to other reprogramming cocktails used in this study [165]. In mouse ESCs, MYC shares ~20% of UTF1 target genes [71]. Furthermore, UTF1 was identified to suppress the leaky expression of bivalent genes by tagging their mRNA for cytoplasmic degradation, possibly regulating the expression of the alternative reading frame (ARF) (an inhibitor of ESC proliferation), whereas MYC was known to activate ARF in ESCs. The repression of this MYC-ARF feedback loop in mouse ESCs by UTF1 ensures rapid cell proliferation [60]. Moreover, the UTF1 KO study showed enhanced ARF protein levels and suggested that the proliferation of ESCs was supported by UTF1 [99]. It possibly inhibits the cell cycle inhibitor ARF gene to enhance proliferation during the reprogramming process to facilitate reprogramming. To date, the actual mechanism of how UTF1 stimulates reprogramming in the presence or absence of exogenous MYC is

unknown and requires investigation.

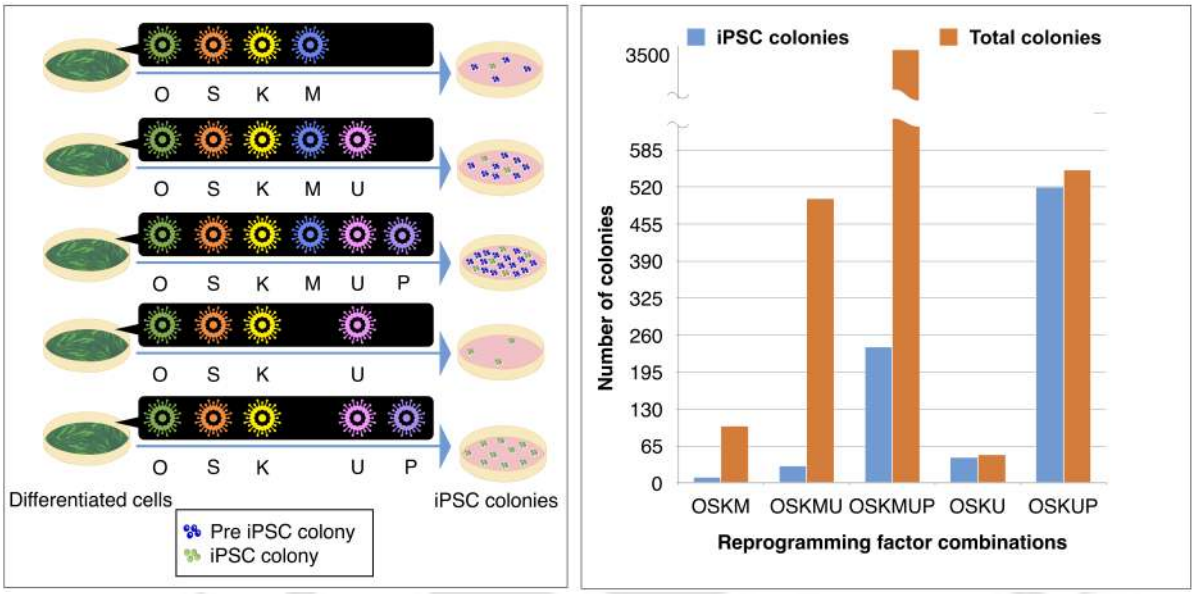


Figure 1.9: Effect of UTF1 on the generation of human iPSCs. (A) The introduction of UTF1 in the reprogramming cocktail along with the Yamanaka factors (OSKM) enhanced the efficiency of iPSC generation. Further, the addition of a p53 siRNA into this cocktail enhanced the efficiency even more. Also, UTF1 could replace oncogenic c-Myc from the reprogramming cocktail. **(B)** The efficiency of colony generation with p53 siRNA (P) and UTF1 (U) was higher in comparison to the same cocktail along with c-Myc. Additionally, the presence of UTF1 in the reprogramming cocktail mostly generated iPSCs with very few transformed colonies, unlike the colonies obtained in the absence of UTF1 in the reprogramming cocktail. O: Oct3/4; S: Sox2; K: Klf4; M: c-Myc; U: Utf1; P: p53 siRNA. Adapted from [165].

The role of UTF1 in cellular reprogramming was evident as one of the crucial genes for the formation of mature iPSCs [162]. Recently, a study reported that UTF1-depleted cells displayed significantly reduced reprogramming efficiency [118]. On the contrary, studies have also shown that UTF1 is dispensable in forming mouse iPSCs [5, 99]. These studies have reported that UTF1-null mouse fibroblasts could be reprogrammed to iPSCs with typical pluripotency characteristics but with very low reprogramming efficiency. Therefore, a detailed investigation is required to examine the importance of UTF1 in iPSC generation. Other studies have reported the importance of UTF1 as a biomarker in iPSC formation. Notably, UTF1 is an early-stage indicator for successful reprogramming [17]. Further, its expression is crucial for identifying clones/cells that will attain full pluripotency [17, 37, 18]. For instance, somatic and partially reprogrammed cells have significant methylation in the enhancer region of UTF1, whereas mature iPSCs lack methylation in the same region [86]. To date, many studies have used UTF1 as one

of the markers to confirm pluripotency [153, 44, 76, 157, 162]. Furthermore, UTF1-based reporters are highly sensitive in identifying mature pluripotent stem cells and, therefore can provide a platform to derive and/or maintain homogeneous pluripotent cells with a high degree of pluripotency [17, 37, 88, 107, 120, 128].

1.13 Study Objectives

Motivated by the current scope and necessities in the field, we have designed the following four objectives in this thesis work.

1. Generation of human induced pluripotent stem cells using a non-integrative approach
2. Generation of *UTF1* knockout human fibroblasts and determining its role in iPSC formation
3. Generation of *UTF1* knockout human iPSCs and determining its role in maintenance of stem cell characteristics
4. Generation of UTF1 knockdown human iPSCs and determining its role in maintenance of stem cell characteristics

CHAPTER 2

GENERATION OF HUMAN INDUCED PLURIPOTENT STEM CELLS USING A NON-INTEGRATIVE APPROACH

An article based on this chapter is published as follows:

Raina K, Joshi G, Modak K, Premkumar C, Priyanka S, Rajesh P, Velayudhan SR, Thummer RP **(2023)**. Generation and characterization of induced pluripotent stem cell line IITGi001-A derived from adult human primary dermal fibroblasts. *Stem Cell Research*. DOI: 10.1016/j.scr.2023.103159.

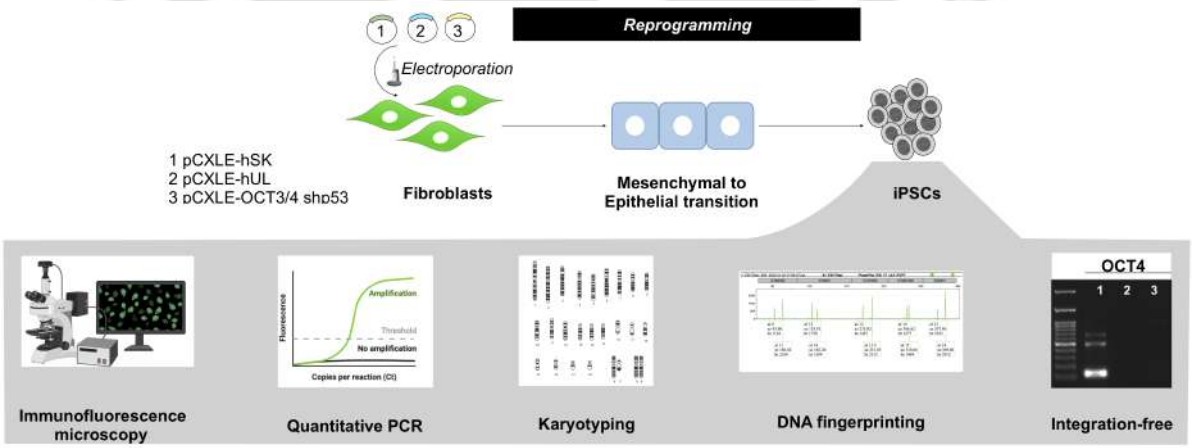


Figure 2.1: Schematic representation of the nucleofection of human fibroblast cells with three episomal plasmids encoding for six reprogramming factors. During the reprogramming process, the mesenchymal phenotype of fibroblasts undergoes a transition to epithelial phenotype, eventually giving rise to iPSCs. Generated iPSCs were further characterized by immunofluorescence microscopy, real-time PCR, and karyotyping.

BRIEF OVERVIEW OF CHAPTER

Adult human primary dermal fibroblast cells were electroporated with a reprogramming cocktail of oriP/EBNA-1-based episomal plasmids expressing OCT3/4, SOX2, KLF4, L-MYC, LIN28, and a p53 shRNA. Initially, these cells were cultured in fibroblast growth medium and then seeded on a matrigel-coated dish with mTeSR^{TM}-E7^{TM} reprogramming medium for the next 15 to 20 days. All the embryonic stem cell-like colonies were picked based on morphology and expanded on a matrigel-coated dish. Out of 14 clones that were picked, one was further characterized to confirm its identity as a pluripotent cell. This clone showed the expression of core pluripotency nuclear (OCT4, SOX2, and NANOG) and surface (SSEA-4 and TRA-1–81) markers using immunofluorescence staining. Additionally, the expression of pluripotency markers OCT4, SOX2, NANOG, and REX1 was confirmed by real-time PCR. Further, in vitro trilineage differentiation into the three germ layers showed the expression of ectodermal (PAX-6, DLK1, and OTX1), mesodermal (Brachyury, HAND1, and TBX6), and endodermal (SOX17, CER1, and FOXA2) markers. The absence of any episomal vector integration in the generated iPSC line was confirmed by genomic PCR. The iPSCs showed normal karyotype (46, XX) after 14 passages, and no mycoplasma contamination was detected. DNA fingerprinting analysis of the fibroblast DNA and iPSC DNA confirmed the genetic identity of the cell line. Altogether, we have generated an integration-free human iPSC line, which can be used for various biomedical applications and basic research.

2.1 Materials and Methods

2.1.1 Generation of iPSCs from human fibroblasts

Adult human primary dermal fibroblast cells [ATCC (PCS-201-012)] were cultured in Dulbecco's modified Eagle medium (DMEM) (MP Biomedicals) supplemented with 10% fetal bovine serum (FBS) (Goibco), 2 mM L-glutamine, 100 U/mL penicillin and 100 μg/mL streptomycin (Thermo Fisher Scientific). A total of 0.5×10^6 fibroblasts at passage 13 were nucleofected with three reprogramming episomal plasmids, at a concentration of 1 μg each, namely pCXLE-hOCT3/4-shp53-F (Addgene ID: 27077), pCXLE-hSK (Addgene ID: 27078) and pCXLE-hUL (Addgene ID: 27080) with Neon Transfection system (Life Technologies) and seeded on a 35 mm dish for the next 6 days. The cells were cultured in fibroblast culture medium comprising of DMEM supplemented with 10% FBS and 1% Penicillin-Streptomycin (Thermo Fisher Scientific) and the medium was changed every 48 hours. The cells were then re-seeded on a matrigel (Corning) coated 35 mm dish with TeSR^{TM}-E7^{TM} medium for Reprogramming (2-Component) (STEMCELL Technologies) until human iPSC-like colonies appeared. The medium was changed every day during reprogramming. Between days 19 and 25, these colonies were picked and mechanically dissociated and seeded on a fresh matrigel-coated plate containing Essential 8 complete medium (E8)^{TM} (Life Technologies) with media changed every day. The colonies were dissociated using 0.5 mM EDTA (Thermo Fisher Scientific), and split in a 1:4 ratio every 3–4 days till 14 passages. The colonies were cryopreserved for the long term using stem cell banker (Amsbio). Fibroblasts and derived iPSCs were maintained in humidified 5% CO_2 incubator (Eppendorf).

2.1.2 Immunofluorescence staining

Immunofluorescence staining was performed as described in a previous study [10]. Briefly, passage 4 cells were fixed using 4% paraformaldehyde in Phosphate-bufferred Saline (Hyclone) for 20 minutes at room temperature. Post-fixation, the cells were incubated in a blocking buffer (1% bovine serum albumin and 5% FBS) for 45 minutes at room temperature. For nuclear markers, 0.2% triton X-100 (Sigma) was also added to the blocking buffer. Treated cells were then incubated with primary antibody overnight at 4 °C followed by incubation with respective secondary antibody for 2 hours at room temperature. Nuclei was stained with 1 μg/mL DAPI (Thermo Fisher Scientific) prepared

in 1X PBS for 5 minutes and imaged using a fluorescence microscope (DS60000, Leica Microsystems) and DAPI filter.

2.1.3 Real-time PCR

For quantitative real-time PCR, RNA was extracted from passage 14 iPSC line, and cDNA was synthesized by reverse transcribing the RNA using Prime Script™ reverse transcription kit (Takara). Real-time PCR was performed relative to pluripotency gene expression of the established iPSC line named TJC8iCas9 [82, 130] (a gift from Prof. R.V. Shaji from Christian Medical College, Vellore, India) and was normalized with GAPDH expression. The real-time PCR was run with GoTaq q-PCR master mix (Promega) using gene-specific primers (Table 2.1) and analyzed using QuantStudio™ 6K Flex (Applied Biosystems).

Table 2.1: Pluripotency gene-specific primers used for PCR.

Target	Forward primers (5'-3')	Reverse primers (5'-3')	Product Size (bp)
OCT4	GTGGAGGAAGCTGACAACAATG	TCTCACTCGGTTCTCGATACTG	105
SOX2	GCACAACTCGGAGATCAG	TAATCCGGGTGCTCCTTC	126
NANOG	GTCAAGAAACAGAAGACCAG	GCCACCTCTTAGATTTCATTC	184
REX1	TTGCATACGCCTGTGTTCC	GGCTCTTGCTGTTATCCAGTC	127
GAPDH	GTCTCCTCTGACTTCAACAGCG	ACCACCCTGTTGCTGTAGCCAA	131

2.1.4 Trilineage differentiation

iPSCs (passage 4 and 15) were differentiated into all three germ layers (ectoderm, mesoderm and endoderm) as per the manufacturer's protocol using STEMdiff™ Trilineage Differentiation Kit (STEMCELL Technologies). Briefly, iPSCs were cultured on matrigel-coated dishes and treated with an ectoderm differentiation medium for 7 days and mesoderm or endoderm differentiation medium for 5 days, with media change every day. For immunofluorescence staining, cells were fixed and incubated overnight with the primary antibodies against PAX-6 (ectoderm), Brachyury (mesoderm), and SOX17 (endoderm) (Table 2.2) followed by a secondary antibody (Goat anti-rabbit IgG Alexa-fluor 488) and then imaged using a fluorescence microscope (DS60000, Leica Microsystems). Real-time PCR was performed as mentioned previously. The expression of trilineage

markers DLK1 and OTX1 (ectoderm), HAND1 and TBX6 (mesoderm), and CER1 and FOXA2 (endoderm) were calculated relative to undifferentiated IITGi001-A and was normalized with GAPDH expression. Primer sequences are mentioned in Table 2.3.

Table 2.2: Antibodies used for immunofluorescence assay.

Antibody	Dilution	Company Cat#	RRID
Pluripotency Markers			
Rabbit anti-OCT4A	1:200	Cell Signaling Technology, Cat# 2840	RRID:AB_2167691
Rabbit anti-SOX2	1:200	Cell Signaling Technology, Cat# 3579	RRID:AB_2195767
Rabbit anti-NANOG	1:200	Cell Signaling Technology, Cat# 4903	RRID:AB_10559205
Mouse anti-SSEA-4	1:200	Cell Signaling Technology, Cat# 4755	RRID:AB_1264259
Mouse anti-Tra-1-81	1:200	Cell Signaling Technology, Cat# 4745	RRID:AB_2119060
Differentiation Markers			
Rabbit anti-PAX6	1:200	Biolegend, Cat# 901301	RRID:AB_2565003
Rabbit anti-Brachyury	1:200	Cell Signaling Technology, Cat# 81694	RRID:AB_2799983
Rabbit anti-Sox-17	1:200	Cell Signaling Technology, Cat# 81778	RRID:AB_2650582
Secondary antibodies			
Goat anti-rabbit IgG Alexa-fluor 488	1:400	Thermo Fisher Scientific, Cat# A-11034	RRID:AB_2576217
Goat anti-mouse IgG Alexa-fluor 594	1:400	Thermo Fisher Scientific, Cat# A-11032	RRID:AB_2534091

Table 2.3: Trilineage gene-specific primers used for RT-qPCR.

Target	Forward primers (5'-3')	Reverse primers (5'-3')	Product Size (bp)
DLK1	CCCCAAAATGGATTCTGCGAGG	GGTTCTCCACAGAGTCCGTGAA	119
OTX1	TACGCCCTCCTCTTCCTACT	GCATGTGGGTGGTGATGATG	138
HAND1	TCAAGGCTGAACTCAAGAAGG	TGCGTCCTTTAATCCTCTTCTC	120
TBX6	CAGGACAGTGCCCTTCAGCCA	ACAGTGAGAGCAGGAGGTATGG	137
CER1	CAGGACAGTGCCCTTCAGCCA	ACAGTGAGAGCAGGAGGTATGG	142
FOXA2	GGAACACCACTACGCCTTCAAC	AGTGCATCACCTGTTCGTAGGC	134

2.1.5 Karyotyping

Karyotyping was performed as described earlier [10]. Firstly, the IITGi001-A cells (passage 8) were arrested in metaphase by exposure to 200 μg/mL of Colcemid

for 20 minutes. Secondly, the cells were harvested with 0.05% trypsin and treated with hypotonic KCl solution for 12 minutes in a CO_2 incubator at 37 °C. Thirdly, the cells were fixed with Carnoy's fixative and centrifuged at 1000 rpm for 10 minutes at room temperature. These fixed cells were resuspended in 5 mL of modified Carnoy's fixative and spread on a slide, where they were stained according to standard cytogenetics protocol. Images were acquired using the AxioImager A1 (Carl Zeiss, Germany) and G-banded metaphases were analyzed using Ikaros Software (Metasystems, Germany). 40 metaphases were counted per spread.

2.1.6 Analysis of genome integration of plasmids

PCR was performed on genomic DNA isolated from human iPSCs (passage 14) with primers designed against the OCT4, SOX2, and KLF4 regions of the episomal plasmids. Post-harvest, cells were treated with cell lysis solution (Qiagen) for 2 hours at 37 °C followed by treatment with protein precipitate (Qiagen) and centrifuged. The clear supernatant was then mixed well with an equal volume of isopropanol and centrifuged to obtain the DNA pellet. The DNA pellet was then dissolved in a DNA hydration buffer (Qiagen). Genomic DNA was amplified using Emerald (Takara) PCR mix and specific primers (as mentioned in Table 2.4) with the following parameters: denaturation at 98 °C for 30 seconds and 35 cycles at 98 °C for 10 seconds, annealing at 60 °C for 30 seconds, and extension at 72 °C for 30 seconds. The PCR products were analyzed by electrophoresis on 2% agarose gel. The adult human dermal fibroblast (passage 15) DNA was used as a negative control, and 5 ng of pCXLE-hSK, pCXLE-hOCT3/4-shp53-F, and pCXLE-hUL were used as positive controls.

Table 2.4: Primers designed to detect integration of reprogramming plasmids in iPSCs by PCR.

Target	Forward primers (5'-3')	Reverse primers (5'-3')	Product Size (bp)
Oct4 epi-Tg	CATTCAAACTGAGGTAAGGG	TAGCGTAAAAGGAGCAACATAG	124
Klf4 epi-Tg	CCACCTCGCCTTACACATGAAGA	TAGCGTAAAAGGAGCAACATAG	156
L-Myc epi-Tg	GGCTGAGAAGAGGATGGCTAC	TAGCGTAAAAGGAGCAACATAG	954

2.1.7 DNA fingerprinting analysis

DNA from human fibroblasts (passage 15) and fibroblast-derived iPSCs (passage 14) were analyzed for short tandem repeats (STRs) at multiple loci using the Promega GenePrint 10 System STR kit. The loci were amplified and analyzed using ABI 3130 Genetic Analyzer (Applied Biosystems).

2.1.8 Mycoplasma testing

Mycoplasma detection was carried out using MycoAlert Plus Mycoplasma Detection Kit (Lonza) as per the manufacturer's instructions. Briefly, the cells (passage 7) were cultured till they were 80% confluent and the supernatant was collected for mycoplasma testing.

2.2 Results and Discussion

2.2.1 Generation of human iPSCs from adult primary dermal fibroblast cells

The human adult primary dermal fibroblasts were reprogrammed as previously described [100] with minor modifications. The cells were transfected to introduce the episomal vectors expressing OCT3/4, SOX2, KLF4, L-MYC, LIN28, and a p53 shRNA on day 0 (Figure 2.2A). The cells showed typical mesenchymal morphology and were cultured in DMEM containing 10% FBS for 6 days. On day 7, the cells were harvested and re-seeded on a matrigel-coated dish at 0.8×10^4 cells per 35 mm dish in mTeSR^{TM}-E7^{TM} (STEMCELL Technologies) medium. Around day 13, a morphological transition was observed, typical for fibroblast to iPSC reprogramming and referred to as mesenchymal to epithelial transition (Figure 2.2B, C). From day 19 onwards, colonies with typical iPSC morphology, such as a flat colony with packed round cells and large nuclei with scant cytoplasm, were observed (Figure 2.2D). From 1.6×10^4 cells, a total of 14 colonies were picked by mechanically disaggregating them into small clumps without any enzymatic action. Each colony was then seeded in separate matrigel-coated 24 well plates and cultured in (E8^{TM}) medium for 3 passages. After which, they were cultured in StemMACS^{TM} iPS-Brew XF medium for further experiments.

For complete characterization, one of the 14 colonies was selected based on morphology and processed for further characterization. It was named as IIITGi001-A, based on the rules set by hPSC Reg. The name IIITGi001-A stands for Indian Institute of Technology Guwahati iPSC 001 (first donor)-A (donor's first cell line).

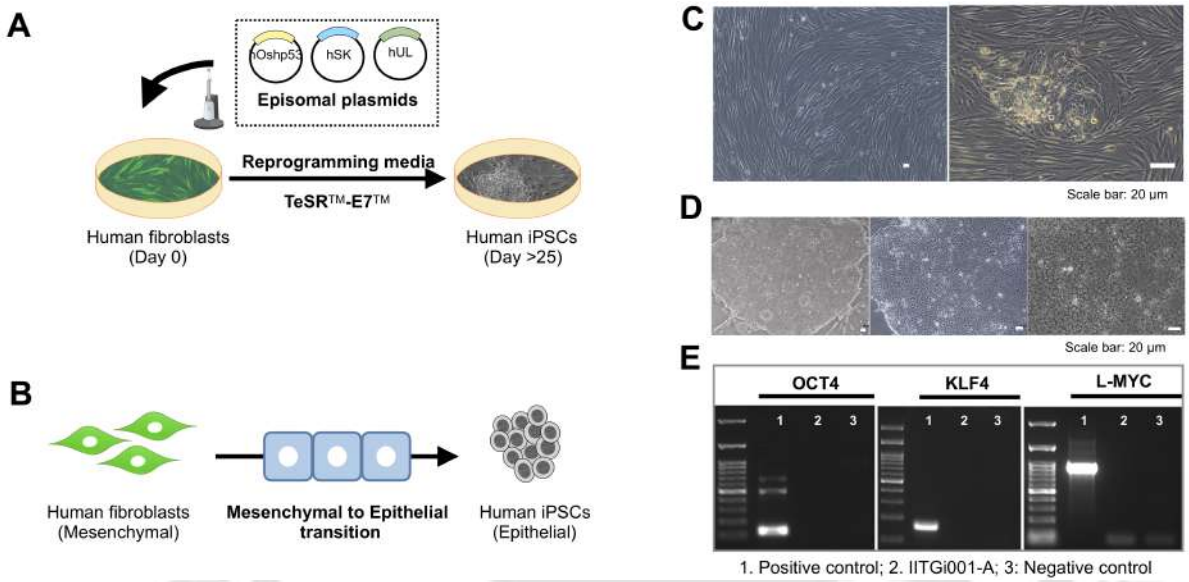


Figure 2.2: Generation of human iPSC line IITGi001-A from human dermal fibroblast cells. (A) Schematic of the protocol followed for the generation of human iPSCs from human fibroblasts using an episomal plasmid combination of pCXLE-hOCT3/4-shp53, pCXLE-hSK and pCXLE-hUL and culturing in reprogramming media TeSR^{TM}-E7^{TM} for 25 days. **(B)** Morphological transitions observed during reprogramming of fibroblasts with typical mesenchymal phenotype to iPSCs with a typical epithelial phenotype. An intermediate mesenchymal to epithelial transition was observed during reprogramming. **(C)** Brightfield image of the morphological transition of fibroblasts to the intermediate mesenchymal to epithelial transition. **(D)** Brightfield image of iPSC colony picked for further characterization at different magnifications. Lower magnification showing the bright border of the colony and the higher magnification showing the high nucleus to cytoplasm ratio of the iPSCs. **(E)** PCR amplification of genomic DNA of the iPSC line to ensure the absence of any episomal plasmid integration. hOCT3/4: human OCT3/4 and shRNA against p53; hSK: human SOX2 and KLF4; hUL: human LIN28 and L-MYC.

2.2.2 Generated human iPSCs were integration-free iPSCs

Firstly, to examine the presence of any episomal vector persisting in the iPSC colony, we extracted the genomic DNA after 10 passages, as reported by previous studies [19, 80] from the tenth passage. Three primers targeting plasmid-specific regions for each of the three reprogramming episomal plasmids were used for genomic PCR amplification. The amplified samples, when analyzed on agarose gel, showed amplification only for the

reaction with the reprogramming plasmid used as template. No amplification was seen for the genomic DNA of the iPSC line IITGi001-A, and the samples were without any template (Figure 2.2E). Therefore, it was confirmed that the iPSC line was free of all three reprogramming plasmid integrations, and the episomal vectors were spontaneously lost over passages after they were introduced.

Given their morphological resemblance and absence of episomal plasmid integration, these iPSCs were further characterized with established pluripotency markers and their ability to give rise to cells of all three germ layers.

2.2.3 Generated human iPSCs expressed pluripotency markers

In general, human iPSCs expressed pluripotency nuclear proteins, namely OCT4, SOX2, and NANOG, and surface antigens SSEA-4 and TRA-1-81 (Figure 2.3A). Further, real-time PCR confirmed the expression of human iPSC-marker genes OCT4, SOX2, NANOG, and REX1 at levels equivalent to or higher than a control iPSC line (Figure 2.3B). Although the expression levels of OCT4, SOX2, and NANOG increased more than one-fold, the expression level of REX1 was less than the expression in the control cells. The presence of these markers indicated strong evidence for the IITGi001-A cell line being pluripotent.

2.2.4 Directed differentiation of the generated iPSCs into all the three germ-layers

To further confirm their ability to give rise to cell types from all three developmental germ layers, namely ectoderm, mesoderm, and endoderm, these iPSCs were differentiated using the STEMdiff^{TM} Trilineage Differentiation Kit as mentioned above. Once the cells were grown in the differentiation-specific media for the desired time, they were either fixed or their RNA was extracted based on the downstream application. The fixed cells were stained with their respective lineage-specific antibodies and showed successful signals for ectoderm-specific marker PAX6, mesoderm-specific marker Brachyury, and endoderm-specific marker SOX17 (Figure 2.4A). This preliminary data was further supported by real-time PCR analysis of the two more markers for each trilineage germ layer. The expression of ectodermal markers DLK1 and OTX1, mesodermal markers HAND1 and TBX6, and endodermal markers CER1 and FOXA2 confirmed the successful

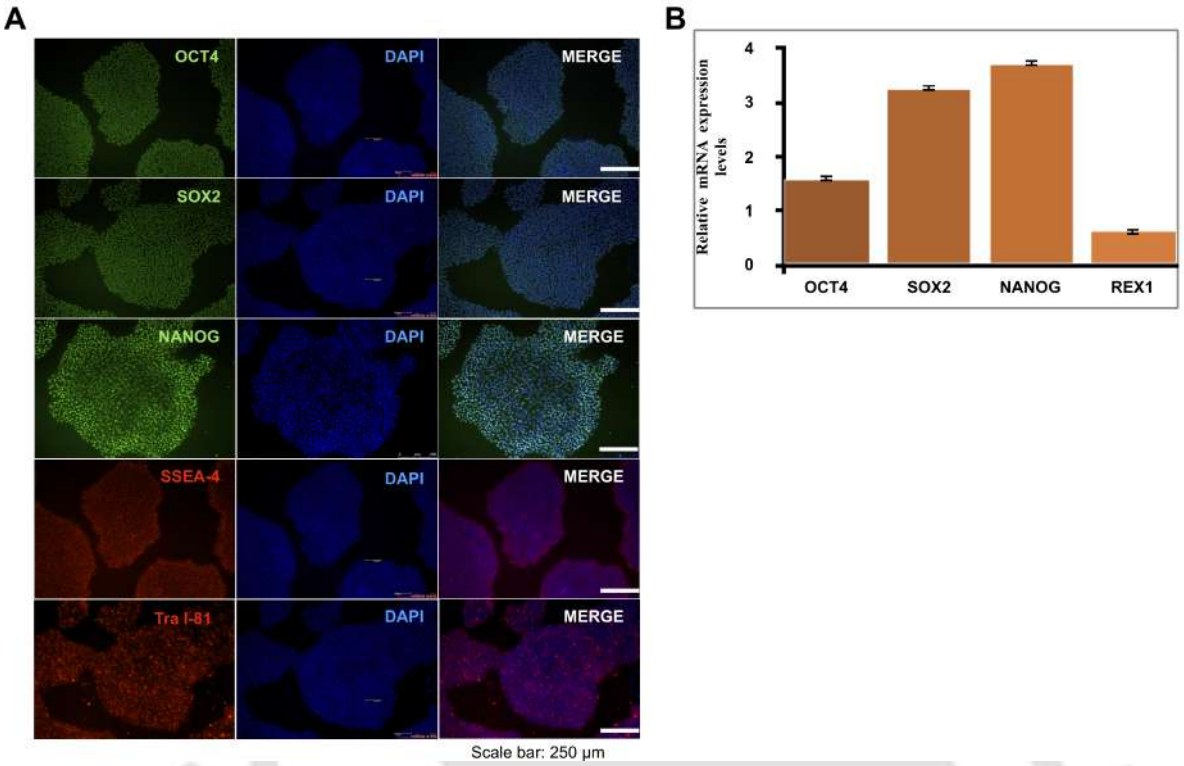


Figure 2.3: IITGi001-A iPSC line showed expression of pluripotency markers. (A) The immunofluorescence assay showed the expression of nuclear markers OCT4, SOX2, and NANOG as well as surface markers SSEA-4 and TRA-1-81. **(B)** Real-time PCR showed the expression of pluripotency markers OCT4, SOX2, NANOG, and REX1 in comparison to an established iPSC line.

trilineage differentiation potential of the cell line IITGi001-A (Figure 2.4B).

2.2.5 Human iPSCs were derived from adult primary dermal fibroblasts and not from cross contamination

PCR of genomic DNA of human iPSC line IITGi001-A was tested for 17 STRs. These 17 STRs were distributed on 10 different loci, namely D5S818, D13S317, D7S820, D16S539, CSF1PO, TH01, D21S11, AMEL, vWA, and TPOX. Results showed that the patterns of all 17 STRs were a complete match between human iPSC IITGi001-A and the starting cell source, the human dermal fibroblasts (Figure 2.5). This confirmed that the iPSC line was derived from fibroblasts and did not result from cross-contamination with other pluripotent cell lines in culture.

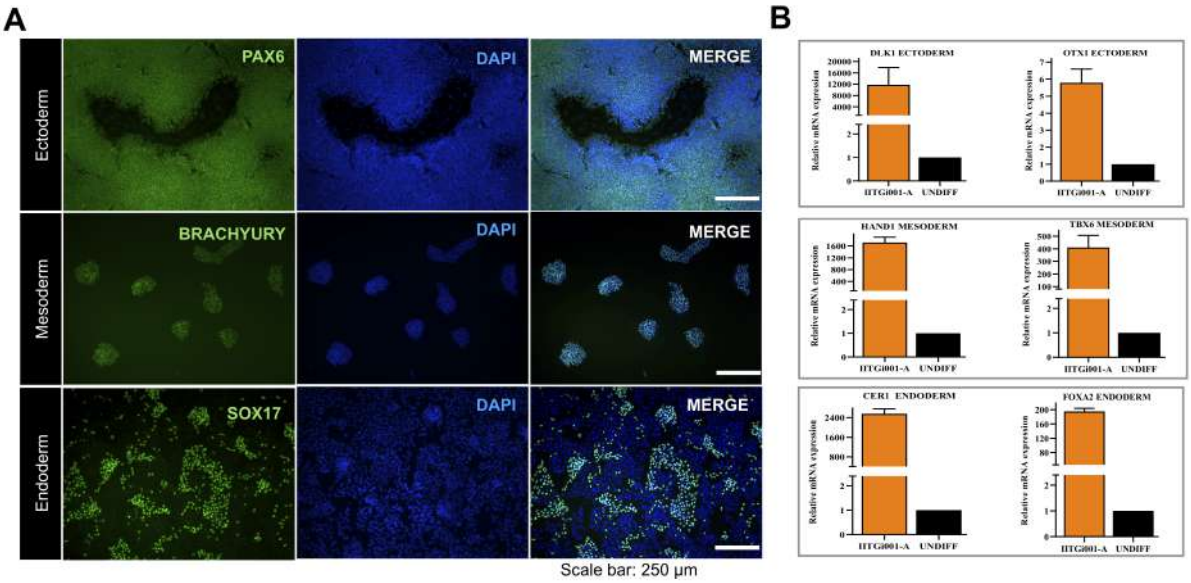


Figure 2.4: Trilineage differentiation of the iPSC line IITGi001-A. (A) Immunofluorescence assay showed the expression of nuclear markers PAX6 for ectodermal lineage, Brachyury for mesodermal lineage, and SOX17 for endodermal lineage. **(B)** Real-time PCR showed the expression of ectodermal markers DLK1 and OTX1, mesodermal markers HAND1 and TBX6, and endodermal markers CER1 and FOXA1 in comparison to control. The mRNA expression levels in differentiated cells were plotted against the expression levels shown by undifferentiated IITGi001-A cells grown in iPSC media.

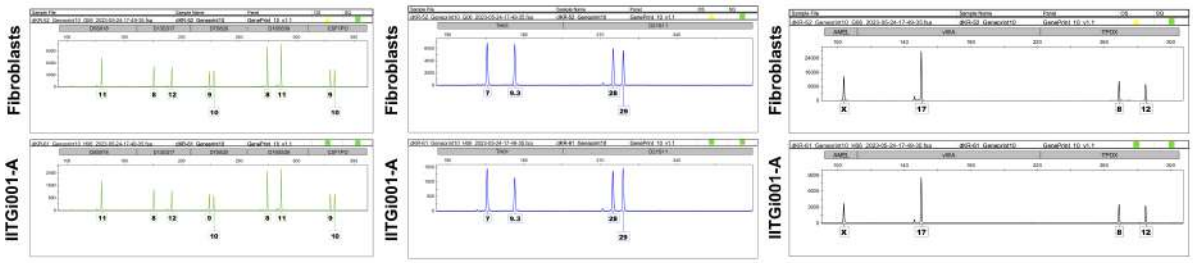


Figure 2.5: Short tandem repeat analysis of the fibroblast cell line and fibroblast-derived iPSC line IITGi001-A

2.2.6 Generated iPSC line IITGi001-A was free of mycoplasma and showed stable karyotype

Confirming the absence of mycoplasma contamination is crucial when establishing a new cell line to ensure the integrity of scientific research. This iPSC line was shown to be free from mycoplasma contamination (Figure 2.6A). Karyotyping of a newly generated iPSC line provides critical information about the genetic stability and integrity of the cells. The number and structure of chromosomes in the iPSCs identified the maintenance

of the typical number and appearance of chromosomes.

Chromosomal abnormalities, such as aneuploidy (an abnormal number of chromosomes), can arise during reprogramming, affecting the cells' differentiation potential and increasing the risk of tumorigenicity. The chromosomal G-band analysis for IITGi001-A showed a normal karyotype of 46, XX (Figure 2.6B), indicating that no chromosomal abnormalities occurred during the reprogramming process.

A

Sample	Reading A	Reading B	Ratio B/A	Conclusion
Positive	42	1623	38.64	Contaminatio
Negative	61	44	0.72	No
IITGi001-A	98	34	0.35	No

Key for B/A ratio
<0.9: No contamination
>0.9-1.2: border line contamination
>1.2: contamination

B

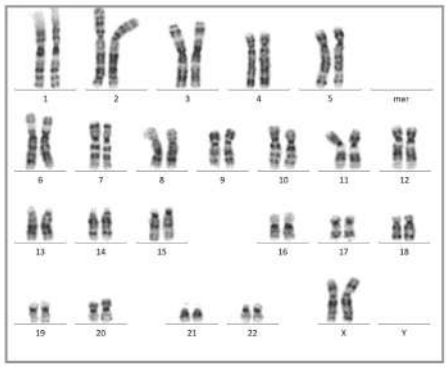


Figure 2.6: iPSC line IITGi001-A was mycoplasma-free and showed a stable karyotype.

2.3 Conclusion

Taken together, our results demonstrated the establishment of a novel iPSC line IITGi001-A. This line was reprogrammed from adult human primary dermal fibroblasts [ATCC (PCS-201-012)] by electroporation using a Neon^{TM} transfection system with oriP/EBNA-1 based reprogramming plasmids expressing OCT3/4, SOX2, KLF4, L-MYC, LIN28, and a p53 shRNA. An embryonic stem cell-like colony was picked based on morphology and then expanded on a matrigel-coated dish. This clone showed the expression of core pluripotency nuclear and surface markers using immunofluorescence staining. Additionally, the expression of pluripotency markers OCT4, SOX2, NANOG, and REX1 was quantified by real-time PCR. Furthermore, trilineage differentiation in vitro demonstrated the expression of ectodermal, mesodermal, and endodermal markers. Genomic PCR confirmed the absence of episomal vector integration in the generated iPSC line. After 14 passages, the iPSCs maintained a normal karyotype (46, XX), and no mycoplasma contamination was observed. Additionally, DNA fingerprinting analysis of the fibroblast DNA and iPSC DNA confirmed the genetic identity of the cell line.

Altogether, we have established a human fibroblast cell-line derived iPSC line and stored a clinically-safe iPSC line.

This integration-free human iPSC line generated in this study can be used as a control cell line and generate mutant iPSC lines by genome editing methods, which will be helpful in basic research, disease modeling, drug screening, and understanding human developmental biology.

CHAPTER 3

GENERATION OF *UTF1* KNOCKOUT HUMAN FIBROBLASTS AND DETERMINING ITS ROLE IN IPSC FORMATION

An article based on this chapter is currently under review as follows:
Raina K, Modak K, Premkumar C, Velayudhan SR, Thummer RP **(2024)**. UTF1 expression is important for the generation and maintenance of human iPSCs.

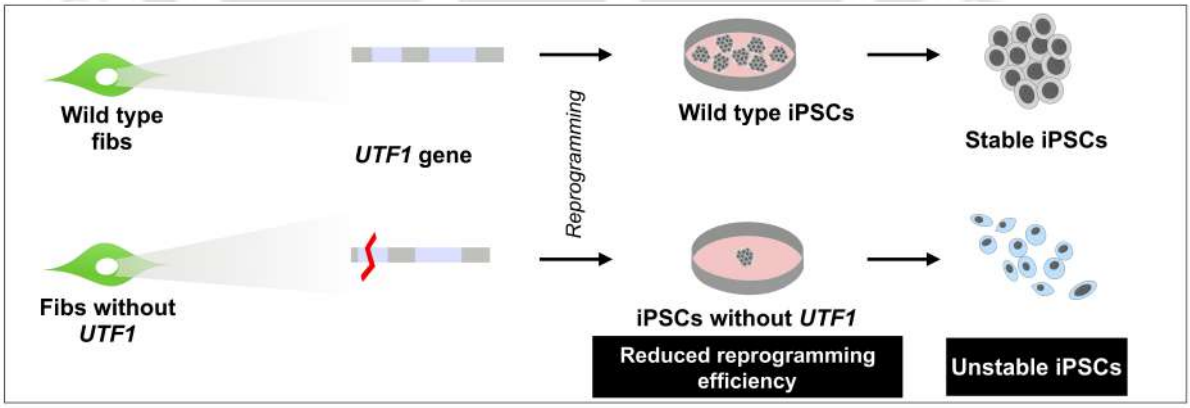


Figure 3.1: Schematic representation of the nucleofection of human fibroblast cells with three episomal plasmids encoding for six reprogramming factors. During the reprogramming process, the mesenchymal phenotype of fibroblasts undergoes a transition to epithelial phenotype, eventually giving rise to iPSCs. Generated iPSCs were further characterized by immunofluorescence microscopy, real-time PCR, and karyotyping.

BRIEF OVERVIEW OF CHAPTER

The first study of iPSC generation included UTF1 (because of its high expression in ESCs) as a reprogramming factor amongst a cocktail of 24 genes that were selected as potential genes for reprogramming the somatic cell's differentiated state to a pluripotent state. It was reported that iPSCs still emerged after the withdrawal of UTF1 from the cocktail of 24 reprogramming factors. Although iPSC colonies did appear, the reprogramming efficiency showed a significant reduction upon removal of UTF1 from the cocktail. The importance of UTF1 was first presented in a 2008 study [165], wherein the addition of UTF1 to the reprogramming cocktail (OCT4, SOX2, KLF4, and c-MYC) observed a drastic increase in the generation of iPSC colonies. Additionally, the presence of UTF1 in the reprogramming cocktail showed a preference for inducing bonafide iPSC colonies over transformed cell clusters, thereby demonstrating its ability to enhance the reprogramming efficiency and hinting at its involvement in defining the pluripotent state.

In this study, we performed knockout experiments targeting the human *UTF1* gene in differentiated cells, namely fibroblasts. To achieve this, we designed dual guide RNAs (gRNAs) and cloned them in a dual gRNAs vector, targeting each of the two *UTF1* exons. The cloned dual gRNAs, along with Cas9 vectors, were employed to generate lentiviruses. Subsequently, fibroblasts were transduced with the generated viruses, aiming to knockout the human *UTF1* gene from fibroblasts prior to reprogramming. The *UTF1* gene-edited cells were then subjected to reprogramming using the episomal vectors as previously described, followed in Chapter 2.

Interestingly, our findings indicated that the absence of UTF1 during the reprogramming process led to a significant reduction in the reprogramming efficiency. Furthermore, the few iPSCs that did emerge exhibited expansion only for the first passage. Beyond passage 1, spontaneous differentiation was observed in *UTF1* knockout fibroblast-derived iPSCs. This underscores the significance of UTF1 in the generation of human iPSCs and in the stability of human iPSCs.

3.1 Materials and Methods

3.1.1 Cell Culture

The adult human primary dermal fibroblasts [ATCC (PCS-201-012)] and lentivirus producer cell line Human Embryonic Kidney 293 T-antigen cells (HEK293T) were cultured in DMEM supplemented with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin and 100 μg/mL streptomycin. All the reagents for cell culture experiments were purchased from Thermo Fisher Scientific. To culture human iPSC colonies picked after reprogramming of fibroblasts, a complete E8^{TM} medium (Gibco) was used with 50 U/mL penicillin and 50 μg/mL streptomycin. The established clones were grown in complete StemMACS iPSC-Brew XF medium (Miltenyi Biotec) with 50 U/mL penicillin and 50 μg/mL streptomycin. The cells were cultured as a monolayer in a humidified incubator at 37 °C in the presence of 5% CO_2. Fibroblasts and HEK293T were passaged at 70–80% confluency, with 0.25% trypsin–EDTA (Gibco) whereas iPSCs were passaged with 0.5 mM EDTA (Invitrogen).

3.1.2 Designing the dual gRNAs targeting the human *UTF1* gene

The human *UTF1* gene has two exons flanking an intron. The full-length gene sequence of the two exons of the human *UTF1* gene were obtained from the UCSC Browser. Exon 1 is a 549 bp sequence and exon 2 is a 476 bp sequence. To design the dual gRNAs of each exon, we used the Broad Institute CRISPRko tool (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). For the best efficiency in gene editing, we selected the gRNA sequences with high on-target efficiency score/on-target rank for each *UTF1* exons. We selected a pair of gRNAs to ensure that the resulting deletion fragment formed after inducing double-strand breaks (DSBs) in both exons is under 100 base pairs in length. We finalized two pairs of gRNAs generating either a 48 bp or 62 bp long deletion fragments for exons 1 and 2, respectively. A 5' 'G' is required for efficient transcription initiation by the U6 promoter; therefore, we selected the gRNAs either starting with 'G' or adding a 'G' at the beginning of the gRNA sequence. When an additional 'G' was added at the beginning of the gRNA forward oligo, a base 'C' was added at the 3' end of the reverse oligo to facilitate proper annealing of the gRNAs. These dual gRNAs were cloned sequentially, and therefore it was ensured that the first gRNAs did not have the restriction site for the enzyme that was used for the second gRNA cloning.

For successful cloning, the restriction site for *Bbs*I (CACCG) was added at the 5' end of the sense guide sequence, while the restriction site for *Sap*I (AAAC) was added at the 5' end of the antisense guide sequence for cloning of the second gRNA, before synthesizing the oligonucleotides.

3.1.3 Generation of dual gRNAs and Cas9 expressing lentiviral vector

As discussed above, a pair of gRNAs (Table 3.1) were designed to target exons 1 or 2 of the human *UTF1* gene. The ordered gRNAs were then annealed and cloned into a modified version of the dual-cassette pKLV2.2-h7SKgRNA5-hU6gRNA5-PGKpuro-BFP-W lentivector (Addgene ID: 72666), generously provided by Kosuke Yusa [137]. In this modified version of the vector, the genes encoding for puromycin (puro) resistance and blue fluorescent protein (BFP) sequences were replaced with the gene coding for EGFP sequence, and therefore, it is hereafter referred to as the pKLV2.2-h7SKgRNA5-hU6gRNA5-PGK-EGFP-W vector. This plasmid was first digested with *Sap*I (New England BioLabs) for cloning the first gRNA followed by a digestion with *Bbs*I (New England BioLabs) for cloning the second gRNA (Figure 3.2A).

Table 3.1: gRNA sequences used for targeting human *UTF1* gene.

Exon 1	**Forward sequence (5'-3')**	**Reverse sequence (5'-3')**
gRNA 1	CACCGGCCTCGCCGCCGCTCCCCG	AAACCGGGGAGCGGCGGCGAGGCC
gRNA 2	CTCGAGCAGATCCGGAAGCTCATG	AACCATGAGCTTCCGGATCTGCT
Exon 2	**Forward sequence (5'-3')**	**Reverse sequence (5'-3')**
gRNA 3	CACCGCCGCCCTGCTGCAGACCCTG	AAACCAGGGTCTGCAGCAGGGCGG
gRNA 4	CTCGCTCGCCGCCCGGCCGCCCCG	AACCGGGGCGGCCGGGCGGCGAG

3.1.4 Generation of dual gRNAs and Cas9 expressing lentiviruses and their transduction

Second-generation lentiviral vectors were co-transfected with the transfer vectors in HEK293T cells. The lentiviral envelope plasmid pMD2.G (Addgene ID: 12259) and the packaging plasmid psPAX2 (Addgene ID: 12260) (gifts from Didier Trono) were co-transfected with the transfer vector (pKLV2.2-h7SKgRNA5-hU6gRNA5-PGK-EGFP-W

or TransEdit Cas9 EFS Blast) either encoding for the dual gRNAs or Cas9, using the jetOPTIMUS transfection reagent (Polyplus), following the manufacturer's protocols. The viral supernatants were collected at 48, 60, and 72 hours, pooled and filtered through a 0.45 μm filter. These viral harvests were stored for transduction of fibroblasts at -80 °C till further use.

The human fibroblasts were co-transduced with the viruses generated from the TransEdit Cas9 EFS Blast lentiviral vector (TransEDIT plasmid TECC 1002) along with either the empty or the cloned dual gRNAs lentiviral vector. The empty vector was introduced to generate the control fibroblasts, while the cloned dual gRNAs vector was introduced to target the human *UTF1* gene to generate the knockout fibroblasts. After 48 hours, the transduced fibroblasts were flow-sorted followed by blasticidin selection (7 μg/mL) for one week to establish stable dual gRNAs and Cas9 expressing fibroblasts (Figure 3.3A).

3.1.5 Flow cytometry and sorting

Cells were cultured and transduced as mentioned earlier. After 48 hours of transduction with lentiviral dual gRNAs, fibroblasts were harvested in their respective culture medium and sorted for the expression of EGFP using BD AriaIII flow cytometer (BD Biosciences) (Figure 3.3A). The analysis of transduction efficiency was analysis was carried out using the FlowJo software (BD Biosciences). The EGFP-positive cells were cultured and expanded for further experiments and cryopreservation.

3.1.6 SDS-PAGE and Western blotting

Whole-cell lysates were harvested using RIPA cell lysis buffer (150 mM sodium chloride, 1% (v/v) NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 25 mM of Tris, and 10 mg/mL PMSF) containing Halt^{TM} Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific) for immunoprecipitation assay. The protein concentrations of the samples were estimated using the Bradford assay kit (Bio-Rad) with bovine serum albumin as a standard. The protein samples were incubated with Laemmli buffer at 95 °C for 5 minutes, resolved on 12% SDS-polyacrylamide gel electrophoresis (PAGE), and transferred to a nitrocellulose membrane. The blots were blocked with 5% non-fat dry milk in 1X TBST buffer for 2 hours at room temperature. This was followed by their incubation with the Guide-it^{TM} anti-Cas9 primary antibody (Takara

Bio, Cat# 632607) in a dilution of 1:1000 and anti-actin Ab-5 (BD Biosciences, Cat# 612657) overnight at 4 °C. Subsequently, the Cas9 and the actin blots were incubated for two hours at room temperature with anti-rabbit (Invitrogen, Cat# 31460) and anti-mouse horseradish peroxidase-conjugated secondary antibodies, respectively, in a dilution of 1:5000. The immunoblots were developed in the presence of a chemiluminescence substrate (Bio-Rad), and the image was recorded using a ChemiDoc^{TM} XRS+ molecular imager equipped with Image Lab^{TM} software (Bio-Rad). The quantification of the Western blots was performed using the ImageJ (1.53) software.

3.1.7 Deletion detection by genomic PCR and agarose gel electrophoresis

To check the deletion of targeted exons, the EGFP and Cas9-encoding cells were cultured in the presence and absence of doxycycline (DOX). Next, the control fibroblasts (with the empty gRNA lentiviral vector and the Cas9 vector) and the *UTF1* knockout fibroblasts (with the dual gRNAs lentiviral vector and the Cas9 vector) were harvested. Post-harvest, the cells were lysed with cell lysis solution (Qiagen) for 2 hours at 55 °C, followed by exposure to protein precipitate (Qiagen), and subsequently it was centrifuged. The clear supernatant was then mixed thoroughly with an equal volume of isopropanol and centrifuged again to extract the DNA pellet. The obtained DNA pellet was then dissolved in DNA hydration buffer (Qiagen) and quantified using a NanoDrop One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Scientific).

PCR was performed on the extracted genomic DNA with primers flanking the intended *UTF1* deletion site for either exons 1 or 2 (Table 3.2) in combination with the EmeraldAmp MAX HS PCR Master Mix (Takara). For exon 1, primers were designed to amplify a 191 bp PCR product in control cells and a 143 bp PCR product in *UTF1* knockout cells. Similarly, for exon 2, primers were designed to amplify a 298 bp PCR product in control cells and a 236 bp PCR product in *UTF1* knockout cells.

Table 3.2: Primer sequences flanking the targeted region of the human *UTF1* gene.

Exon	**Forward primers (5'-3')**	**Reverse primers (5'-3')**	**Product Size (bp)**
Exon 1	CGTCCCCACCGAAGTCTG	GCTGGTTCAAGGTCAGCAG	191
Exon 2	CGTCCCCACCGAAGTCTG	TGGTTCAAGGTCAGCAGCT	298

3.1.8 Deletion detection by capillary electrophoresis and Sanger sequencing

For capillary electrophoresis-based detection, one round of PCR was carried out with previously designed primers, that flanked the target region of the human *UTF1* gene. Using the first round of PCR products, a second round of PCR was performed with a 5' FAM labeled forward primer 5'-TTTCCCTACACGACGCTCTTTACACGACGCTCTTCCG ATCT-3', and the same reverse primer was used in the first round of the PCR (Table 3.2). The second round PCR products were denatured in deionized formaldehyde and analyzed by capillary electrophoresis using an ABI 3500 Genetic Analyzer (Thermo Fischer Scientific). The peak sizes and peak heights of the amplified products were determined using the Peak Scanner software (Thermo Fisher Scientific). For Sanger sequencing, the first round of PCR was carried out similarly to genomic PCR. The second round was performed with the PCR products of the first round, using a reaction setup containing fluorescently labeled dideoxynucleotides and a DNA polymerase. The resulting samples were then purified and sequenced using Brilliant Dye^{TM} v3.1 Terminator Cycle Sequencing Kit (Nimagen) on the 3500 Genetic Analyzer. The sequencing results were aligned to the normal sequences using SnapGene to check for deletion of the desired fragment.

3.1.9 Somatic cell reprogramming using episomal vectors

For episomal-mediated reprogramming, control and *UTF1* knockout fibroblasts were transfected as mentioned previously in Chapter 2, Section 2.1.1: Generation of iPSCs from human fibroblasts.

3.2 Results and Discussion

3.2.1 Generation of *UTF1* knockout human fibroblasts using the CRISPR/Cas9 system

We first generated *UTF1* knockout human fibroblasts using the CRISPR/Cas9 gene editing approach. To achieve this, a pair of lentiviral vectors were employed: one vector for constitutive expression of the Cas9 protein and the other to express dual gRNAs. In

the Cas9 vector, the expression of Cas9 gene was under the EF1α-modified promoter called EFS and had a selectable blasticidin resistance gene (Figure 3.2A). In the dual gRNAs vector, the first gRNA was regulated by the human 7SK promoter and the second gRNA was regulated by the human U6 promoter. This was followed by the expression of EGFP under the PGK promoter (Figure 3.1A). We designed two pairs of dual gRNAs: one pair of gRNAs targeting exon 1 and the other pair targeting exon 2. The cloning of gRNAs in the desired vector was confirmed by Sanger sequencing (Figure 3.2B). To induce DSBs to render a non-functional human *UTF1* gene, the lentiviruses encoding for Cas9 protein and gRNAs were produced. These lentiviruses were then transduced into adult human primary dermal fibroblasts to generate *UTF1* knockout cells (Figure 3.3A). After 48 hours, the cells were sorted for EGFP expressing-positive cells (Figure 3.3B), which was followed by selection for blasticidin resistance of Cas9 expressing cells for one week to ensure selection of dual gRNAs and Cas9 co-expressing fibroblasts. The cells were checked for expression of Cas9 protein by Western blotting. A Cas9 band was observed at the expected size (160 kDa), whereas no band was observed in control cells, indicating no endogenous Cas9 expression (Figure 3.3B). These cells were then expanded for 12 passages to obtain *UTF1* knockout fibroblasts. Simultaneously, an empty gRNA lentiviral vector backbone was used to obtain control cells.

To ensure successful gene editing, the *UTF1* gene knockout was confirmed using genomic DNA PCR. This was performed as the *UTF1* gene is not expressed in human fibroblasts [165] and its expression is specifically restricted to pluripotent stem cells, gametes and during early embryonic developmental stages [102, 99, 37, 5, 111]. The dual gRNAs was designed to generate a deletion of 48 or 62 bp fragments targeting exons 1 or 2, respectively, of the human *UTF1* gene (Figure 3.3C).

To confirm the deletion, we used a pair of primers flanking the target site in each exon to amplify the targeted region. When run on agarose gel, PCR products showed two amplified fragments, indicating partial deletion of human *UTF1* gene in fibroblasts (Figure 3.3D). The 143 bp or 236 bp (as shown by a red arrow) indicated the deletion and the 191 or 298 bp indicated the absence of deletion in both, the *UTF1* exons 1 and 2 targets, respectively. Additionally, capillary electrophoresis gave an estimate of the percentage of *UTF1* gene deletion from the fibroblast population. Fibroblasts targeted for exon 1 deletion were 65.2%, and for exon a total of 60.2% deletion was observed (Figure 3.3E). Further, Sanger sequencing confirmed the precise deletion (48 bp in exon 1 and 62 bp in exon 2) of the targeted region in both the *UTF1* exons, indicating

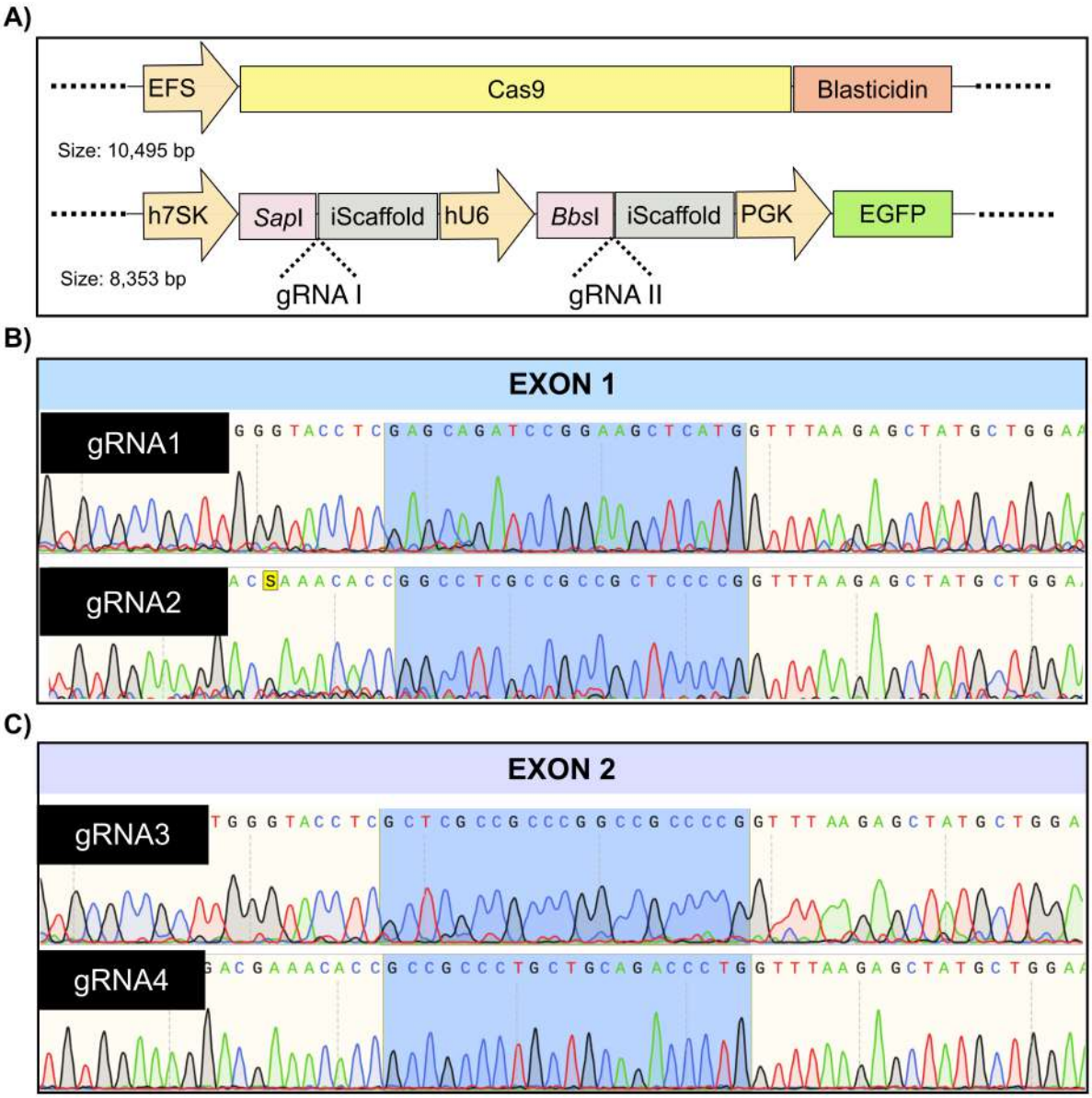


Figure 3.2: CRISPR-Cas9 vectors were used to generate the *UTF1* knockout toolkit. (A) A schematic figure representing the vectors used for Cas9 generation and dual gRNA cloning. **(B)(C)** Confirmation of the dual gRNAs cloned in the gRNA vector targeting two regions on exon 1 and exon 2 of the human *UTF1* gene by Sanger sequencing. EFS: EF-1α Short; h7SK: human 7SK; iScaffold: improved tracrRNA scaffold; hU6: human U6; PGK: Phosphoglycerate kinase; EGFP: Enhanced green fluorescent protein; gRNA: guide RNA

the applicability of the knockout system used (Figure 3.3F). The fibroblasts obtained were a mixed population, with some cells undergoing deletion while others did not, irrespective of the targeted region of the *UTF1* gene. We hypothesize that the partial deletion in human fibroblasts could be because of the presence of the *UTF1* gene within

an inaccessible heterochromatin region in some cells for Cas9 to induce DSBs. Given the lack of UTF1 expression in human fibroblasts [165], we hypothesize that the *UTF1* gene was inaccessible to CRISPR/Cas9 machinery, possibly due to a closed chromatin structure. It has been previously reported that decreased binding of Cas9 protein is observed in the highly condensed portion of DNA [158, 142], which may be due to less accessibility of the Protospacer adjacent motif (PAM) sequence [53]. Also, it is reported that the heterochromatin state of the target sequence can reduce the diffusion of Cas9 [65] and CRISPR-Cas9-mediated mutagenesis by seven-fold [62].

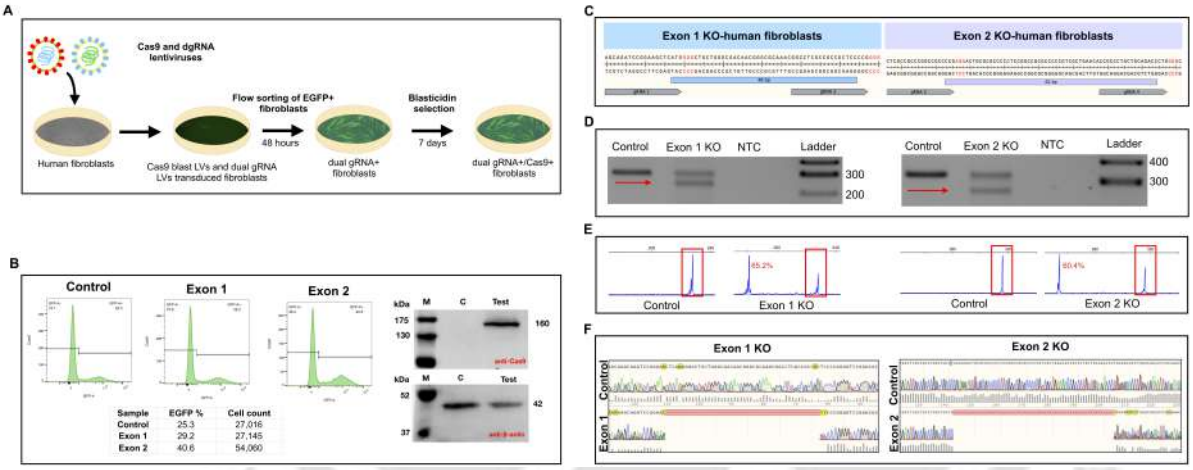


Figure 3.3: Generation of *UTF1* knockout human fibroblasts. **(A)** Schematic representation of the experimental workflow followed for the generation of *UTF1* knockout fibroblasts. **(B)** Western blot confirming the expression of Cas9 protein followed by flow cytometry for EGFP expression for control (empty backbone) and two knockout fibroblast types (exon 1 and exon 2 targeting dual gRNA pairs). **(C)** The size of the deletion fragment is based on the binding site of the dual gRNAs for both exons. **(D)** Results of PCRs using genomic DNA with a forward primer that binds upstream of the first gRNA site and a reverse primer binding downstream of the second gRNA site for control and knockout fibroblasts for both the exons. **(E)** Capillary electrophoresis of the same knockout fibroblasts confirmed a partial deletion of the *UTF1* gene for both exon 1 and exon 2 as peaks matching the size of the position of their respective controls were observed indicated by the presence of a peak near 240 for exon 1 knockout and the presence of a peak near 320 for exon 2 knockout fibroblasts. **(F)** Sanger sequencing for exons 1 and 2 targeted regions indicating their precise deletion. dgRNA: dual guide RNA; M: Protein Marker; C: Control; kDa: kilo Dalton.

3.2.2 *UTF1* deficiency significantly reduced reprogramming efficiency

To gain insights into the role of human UTF1 in the generation of iPSCs, we reprogrammed the previously generated partial *UTF1* knockout fibroblasts using episomal vectors expressing six reprogramming factors, OCT3/4, SOX2, KLF4, L-MYC, LIN28, and a p53 shRNA (Figure 3.4A) [101]. *UTF1* knockout fibroblasts transduced with Cas9 and empty gRNA lentiviral vector backbone were used as a control. iPSC-like colonies emerged in both the knockout conditions (exon 1 knockout and exon 2 knockout) and the control cells (empty gRNA lentiviral vector backbone) of fibroblasts used for reprogramming. The iPSC-like colonies were identified based on morphological characteristics (Figure 3.4B). These colonies exhibited distinct iPSC morphology, characterized by small, compact cells with large nuclei. The iPSC-like colonies emerged around day 19 in control cells whereas the colonies appeared around day 21 for the knockout conditions. The number of iPSC-like colonies from both the knockout conditions and the control were quantified from three independent experiments. A total of 12, 9, and 13 iPSC colonies emerged from control fibroblasts against 3, 3, and 1 for knockout cells, each emerging from 7,700 fibroblasts from three independent experiments. Importantly, the results showed that the number of iPSC-like colonies generated from *UTF1* knockout fibroblasts were significantly lower than those generated from empty vector backbone fibroblasts, used as a control (Figure 3.4C). The iPSC-like colonies were then picked from the control and the two knockout conditions on day 24 and day 26, respectively, and expanded for further analysis.

While it is a limitation of our study, that complete *UTF1* knockout fibroblast cells were not generated using genome editing. Also, we were interested in understanding if the knockout of the human *UTF1* gene would affect the reprogramming process or alter the characteristics of the generated iPSCs. For this, we used the human fibroblasts which showed a mixture of *UTF1* knockout and wild-type cells for iPSC formation, and it was striking to observe a reduced reprogramming efficiency in three independent reprogramming events, confirming a uniform effect. The number of colonies generated was significantly lower (~2 colonies) than in control cells (~11 colonies). Consistent with our results, a study had reported a significant decline in reprogramming efficiency of mouse fibroblasts in the absence of *Utf1* gene [118]. Another study had successfully generated iPSCs from *Utf1* null mouse fibroblasts; however, this study did not report the

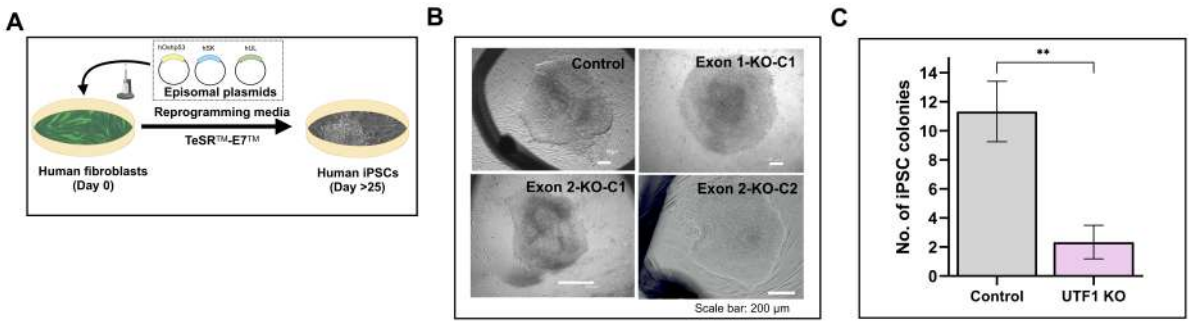


Figure 3.4: Reprogramming of *UTF1* knockout human fibroblasts to iPSCs. (A) A schematic representation depicting the reprogramming protocol followed for generating human iPSCs. **(B)** Morphology of iPSC colonies which were picked up after reprogramming on day 24 (control) and day 26 (knockout). **(C)** Effect on reprogramming efficiency in the absence of UTF1 in fibroblasts compared to control. hOshp53: human OCT3/4 and p53 shRNA; hSK: human SOX2 and KLF4; hUL: human L-MYC and LIN28; KO: knockout; C1: Clone 1; C2: Clone 2.

observation in the context of reprogramming efficiency [99]. Furthermore, we hypothesize that the effect on reprogramming would have been more pronounced had the *UTF1* gene been completely knocked out from the starting population of human fibroblasts.

3.2.3 Confirmation of deletion of the human *UTF1* gene in the fibroblast-derived iPSCs

Next, the genomic DNA was isolated from the control and the knockout conditions to confirm the presence or absence of the *UTF1* gene in the fibroblast-derived iPSC-like cells. To determine this, PCR was performed on the genomic DNA isolated from passage 2 iPSC-like cells. The experimental design to confirm the deletion of the *UTF1* gene in human knockout iPSCs was the same as performed in Figure 3.2C. PCR products when analyzed on agarose gel showed only one smaller band size of 143 or 236 bp (as indicated by arrow) in comparison to a larger band size of 191 or 298 bp in iPSC-like cells derived from empty gRNA lentiviral vector backbone transduced fibroblasts (Figure 3.5A). Further, capillary electrophoresis data showed a 100% deletion for all three fibroblast-derived *UTF1* knockout iPSCs (Figure 3.5B). These results confirmed that the iPSC-like colonies indeed emerged from *UTF1* knockout fibroblasts. Additionally, the deletion (48 bp for exon 1 and 62 bp for exon 2) at the desired region was confirmed for both the knockout conditions by Sanger sequencing with control (Figure 3.5C).

The genomic DNA extracted from the human iPSC colonies generated in our study showed the deletion of the targeted region of the *UTF1* gene. RT-PCR further

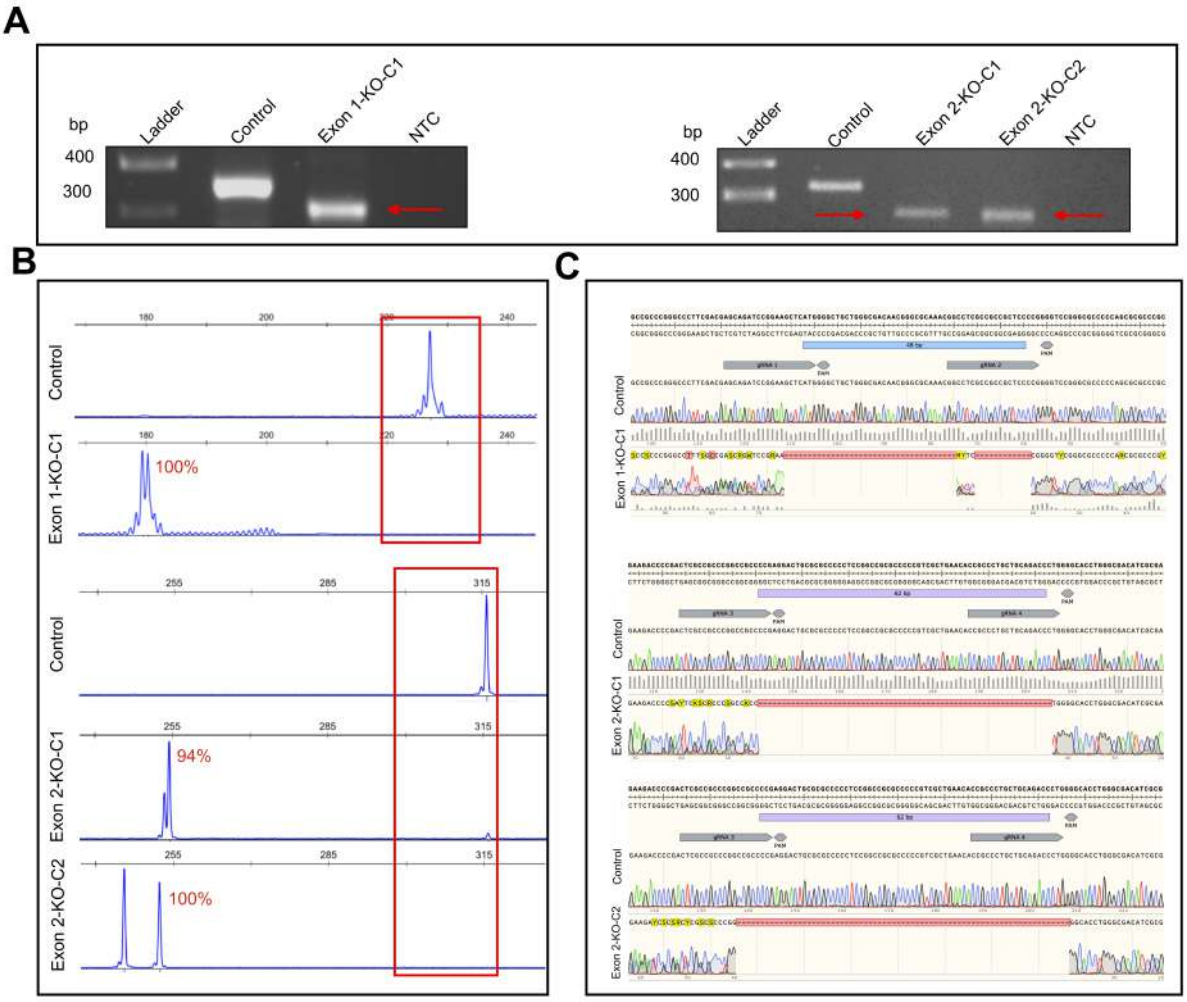


Figure 3.5: Confirmation of *UTF1* deletion in fibroblast-derived iPSCs. (A) Agarose gel electrophoresis analysis of PCR amplified products of the genomic DNA of control and knockout iPSCs generated from *UTF1* knockout fibroblasts. A forward primer that binds upstream of the first gRNA site and a reverse primer binds downstream of the second gRNA site for exon 1 and exon 2 knockout fibroblast-derived iPSCs. **(B)** Capillary electrophoresis of the same knockout clones confirmed the absence of *UTF1* control amplicon in the knockout clones, indicated by the absence of a peak near 220 for exon 1 knockout and the absence of the control peak near 315 for exon 2 knockout fibroblast-derived iPSCs. **(C)** Sanger sequencing for exon 1 knockout clone 1 and exon 2 knockout clone 2 and knockout clone 3 targeted regions indicating their precise deletion. bp: base pair; KO: knockout; NTC: no template control; gRNA: guideRNA; PAM: Protospacer adjacent motif.

confirmed the absence of UTF1 expression from three independent clones. PCR confirmed the absence of UTF1 transcripts, ensuring that iPSCs colonies emerged from *UTF1* knockout fibroblasts. Various studies have highlighted that during the reprogramming of fibroblasts to iPSCs, the chromatin undergoes a closed-to-open chromatin configuration

[40, 72]. This modification in chromatin structure potentially facilitates Cas9 accessibility for targeting the *UTF1* gene in iPSCs, in contrast to the limited access observed in human fibroblasts. Supportingly, the limited iPSCs colonies generated during reprogramming showed *UTF1* knockout even though the starting cells were a mixed population.

3.2.4 Instability of the *UTF1* knockout fibroblast-derived human iPSCs

At passage 1, the cells showed an ideal iPSC-like morphology for both the knockout conditions and the control iPSCs (Figure 3.6A). However, the iPSC-like colonies of both the *UTF1* knockout conditions underwent spontaneous differentiation from passage 2 onwards (Figure 3.6B). These iPSC-like colonies were highly unstable to keep them in culture for expansion as they lost their self-renewal potential. This phenotype was observed in both the knockout conditions, indicating the importance of UTF1 in maintaining iPSC identity. To ensure that this phenotype was a consequence of the absence of UTF1 transcripts, reverse transcriptase PCR was performed. These results further confirmed the absence of UTF1 expression in human *UTF1* knockout fibroblast-derived iPSCs (Figure 3.6C). The expression levels of the core pluripotency factors OCT4 and SOX2 showed a decline in expression for the *UTF1* Exon 1-KO-C1 and Exon 2-KO-C2. However, Exon 2-KO-C1 showed comparable expression levels with the control cells (Figure 3.6C). The cells were frozen at passage 1 for further characterization experiments. Upon revival of these cells, the control cells were revived successfully; however the knockout cells could not be revived even after multiple attempts. The few iPSC clones generated in the absence of UTF1 showed instability upon expansion and supportingly a decline in the expression of core-pluripotency factors OCT4 and SOX2. Additionally, their inability to be revived from liquid nitrogen hint at the importance of UTF1 in establishing and maintaining pluripotency.

3.3 Conclusion

In conclusion, our study highlighted the generation of *UTF1* knockout human fibroblasts using the CRISPR/Cas9 system, resulting in a partial deletion of the *UTF1* gene. This is likely due to its localization within the heterochromatin. Notably, UTF1 deficiency led to a significant reduction in reprogramming efficiency, with *UTF1* knockout

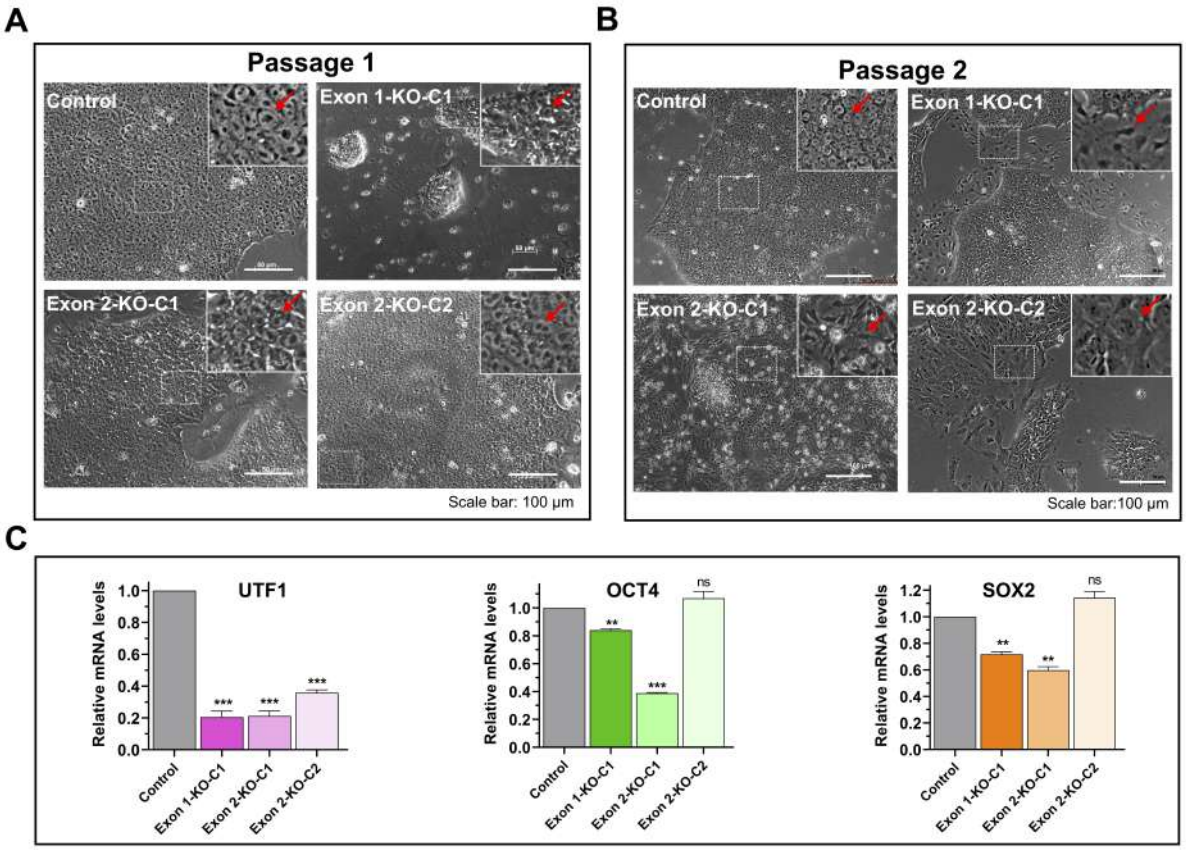


Figure 3.6: Instability of *UTF1* knockout fibroblast-derived iPSCs. (A) The morphology of picked-up clones in passage one showed stable iPSC-like morphology. **(B)** Spontaneous differentiation of *UTF1* knockout colonies at passage two. **(C)** Change in transcript levels of UTF1 and the core pluripotency factors, OCT4 and SOX2, in the control iPSC line and the knockout clones. KO: knockout. The quantitative values show the mean fold change±S.D. Statistical analysis was performed using a two-way ANOVA (n=3). ∗P<0.05, ∗P<0.01, ∗P<0.001, ns:non-significant.

fibroblasts producing fewer iPSC colonies than control fibroblasts. Interestingly, the few iPSC colonies that did emerge exhibited complete deletion of the *UTF1* gene, supporting the hypothesis that *UTF1*, when located in euchromatin rather than heterochromatin (as in fibroblasts), could facilitate the generation of *UTF1* knockout iPSCs. Furthermore, iPSCs generated from *UTF1* knockout fibroblasts demonstrated spontaneous differentiation beyond passage 2, suggesting a role for UTF1 in establishing and maintaining iPSCs. These findings highlight the crucial role of UTF1 in reprogramming human fibroblasts and underscore its importance in maintaining iPSCs.

CHAPTER 4

GENERATION OF HUMAN *UTF1* KNOCKOUT iPSCs AND DETERMINING ITS ROLE IN MAINTENANCE OF STEM CELL CHARACTERISTICS

An article based on this chapter is currently under review as follows:
Raina K, Modak K, Premkumar C, Velayudhan SR, Thummer RP **(2024)**. UTF1 expression is important for the generation and maintenance of human iPSCs.

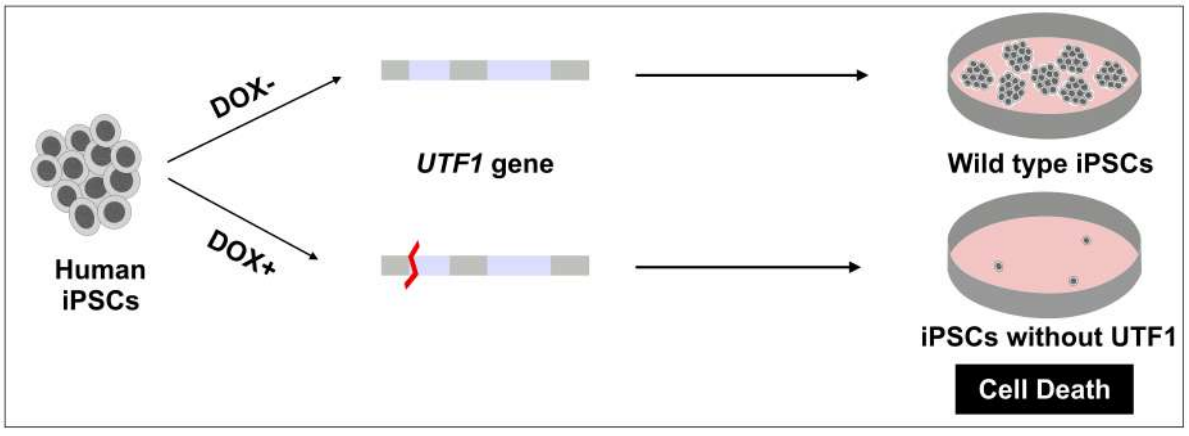


Figure 4.1: A schematic representation of the experimental design followed to determine the effect of the absence of *UTF1* in iPSCs. The cell viability declined in the absence of UTF1 expression.

BRIEF OVERVIEW OF CHAPTER

Assessing the importance of UTF1 in the context of maintaining pluripotency was crucial to gaining insights into its role in iPSCs. To address this, we demonstrated the knockout of the human *UTF1* gene using the same dual gRNAs as previously used for fibroblasts in Chapter 3. To generate these *UTF1* knockout iPSCs, we used an iPSC line with an inducible Cas9 cassette integrated into the genome at the safe harbor site AAVS1. These iCas9 iPSCs were transduced with lentiviruses generated from the dual gRNAs. The gRNAs expressing EGFP iPSCs were sorted and expanded to obtain cells expressing the desired dual gRNAs. Cas9 production was activated by the addition of DOX to the cell culture media to generate DSBs. In the presence of Cas9 protein and active dual gRNAs synthesis, the *UTF1* knockout iPSCs were generated.

This chapter then deals with the effects on the maintenance of iPSCs over 10 days in the absence of UTF1 expression. It was interesting to observe that in the absence of UTF1, iPSCs showed a loss in viability from day 2 of DOX addition. Drastic loss of cell viability was observed over the next 9 days, with very few cells remaining by day 6. This loss in cell viability occurred through apoptosis, as confirmed by the expression of pro-apoptotic markers. Further, to ensure that the changes in cellular viability resulted from the loss of UTF1, genomic DNA was checked for the desired deletion in exon 2 of the *UTF1* gene. A partial deletion was observed on day 5; however, a complete deletion was observed on day 10. Also, at the protein level, a gradual loss in UTF1 expression was observed on days 2 and 4 post-induction, and a complete absence was confirmed by day 6. Further, we also explored the effect of the absence of UTF1 on the core pluripotency factors OCT4 and SOX2, both of which showed a decline in their expression. In conclusion, this chapter highlights the importance of human UTF1 in the maintenance of iPSCs.

4.1 Materials and Methods

4.1.1 Cell Culture

The established iPSC line TJC8iCas9 was cultured in complete StemMACS^{TM} iPSC-Brew XF medium (Miltenyi Biotec) with 50 U/mL penicillin and 50 μg/mL streptomycin. TJC8 was derived from the unique identification code for the clone, and iCas9 indicated the DOX inducible Cas9 cassette integrated at the safe-harbor site AAVS1 [130]. The cells were cultured as a monolayer in a humidified incubator at 37 °C in the presence of 5% CO_2. iPSCs were passaged with 0.5 mM EDTA (Invitrogen). For induction of Cas9 expression, the cells were cultured with 0.5 μg/mL DOX.

4.1.2 Designing the dual gRNAs targeting the human *UTF1* gene

A pair of dual gRNAs targeting exon 2 of the human *UTF1* gene were designed as previously mentioned in Chapter 3, Section 3.1.2: Designing the dual gRNAs targeting the human *UTF1* gene.

4.1.3 Generation of dual gRNAs expressing lentiviral vector

A pair of gRNAs targeting exon 2 of the human *UTF1* gene were cloned in a dual gRNAs vector as mentioned in Chapter 3, Section 3.1.3: Generation of dual gRNAs and Cas9 expressing lentiviral vector. We used one pair of gRNAs targeting exon 2 of the human *UTF1* gene. The same dual gRNAs vector was previously used to generate *UTF1* knockout in fibroblasts, as mentioned in Chapter 3.

4.1.4 Generation of dual gRNAs expressing lentiviruses and their transduction

Second-generation lentiviral vectors were co-transfected with the transfer vectors in HEK293T cells, as mentioned in Chapter 3, Section 3.1.4: Generation of dual gRNAs and Cas9 expressing lentiviruses and their transduction. Briefly, the lentiviral envelope plasmid pMD2.G (Addgene ID: 12259) and the packaging plasmid psPAX2 (Addgene ID: 12260) (gifts from Didier Trono) were co-transfected with the transfer vector (pKLV2.2-h7SKgRNA5-hU6gRNA5-PGK-EGFP-W) encoding for the dual gRNAs, using the jetOPTIMUS transfection reagent (Polyplus). The viral supernatants were collected,

pooled, and filtered through a 0.45 μm filter. This viral harvest was concentrated for the transduction in iPSCs, unlike fibroblasts, where an unconcentrated viral harvest was used for transduction. Concentration was achieved by incubating the viral supernatant with a lentivirus concentrator solution (40% w/v Polyethylene Glycol and 1.2 M sodium chloride) for 4 hours at 4 °C, followed by centrifugation at 14,000 rpm for 2 hours. The supernatant was discarded, and the viral pellet was dissolved in DMEM and frozen at -80 °C for further use.

The human iPSCs were transduced with the concentrated viruses generated from the dual gRNAs lentiviral vector targeting the human *UTF1* gene. After 48 hours, the transduced iPSCs were flow-sorted, followed by blasticidin selection (5 μg/mL) for one week to establish stable dual gRNAs expressing iPSC line.

4.1.5 Flow cytometry and sorting

Cells were cultured and transduced, as mentioned earlier. After 48 hours of transduction with lentiviral dual gRNAs, iPSCs were harvested in StemMACS iPSC-Brew XF medium using TrypLE^{TM} Express (Invitrogen) to obtain single cells for flow-sorting based on the expression of EGFP using BD AriaIII flow cytometer (BD Biosciences). The sorted cells were collected and cultured in StemMACS^{TM} iPSC-Brew XF medium containing 1X RevitaCell^{TM} (Gibco) for 24 hours. Transduction efficiency analysis was carried out using the FlowJo software (BD Biosciences). After 24 hours, the EGFP-positive cells were cultured in iPSC media and expanded for further experiments and cryopreserved.

4.1.6 Immunofluorescence staining and microscopy

To detect pluripotent stem cell surface marker TRA-1-60, human iPSCs were cultured on Matrigel-coated wells with and without DOX. After 6 days, the cells were washed thrice with PBS and then fixed with 4% paraformaldehyde for 20 minutes at room temperature. Following this, the cells were permeabilized and blocked with 0.1% Triton^{TM} X-100 (Sigma-Aldrich) in a buffer containing 1% bovine serum albumin and 5% FBS for 45 minutes at room temperature. The cells were then incubated with TRA-1-60 primary antibody at a dilution of 1:500 (Cell Signaling Technology, Cat# 4746) in blocking buffer at 4 °C overnight, followed by Alexa Fluor^{TM} 594 secondary antibody at a dilution of 1:500 (Invitrogen, Cat# A11032) incubation for 2 hours at room temperature. The nuclei were stained with 10 μg/mL DAPI (Invitrogen) for 5 minutes at room temperature.

Images were captured using a Leica DMi8 Motorized Fluorescence Microscope equipped with LAS F software.

4.1.7 SDS-PAGE and Western blotting

Whole-cell lysates were harvested, and the lysates were resolved on 12% SDS-PAGE and transferred to a nitrocellulose membrane as mentioned previously in Chapter 3, Section 3.1.5: SDS-PAGE and Western blotting. The blots were blocked with 5% non-fat dry milk in 1X TBST buffer for 2 hours at room temperature. The immunoblotting was carried out using rabbit anti-UTF1 at a 1:1000 dilution (Bioss, Thermo Fisher Scientific, Catalog No. 12207-R), rabbit anti-OCT4 at a 1:1000 dilution (Cell Signaling Technology, Cat# 2840), rabbit anti-SOX2 at a 1:1000 dilution (Santa Cruz Biotechnology, Cat# 3579), rabbit anti-Bax2 at a 1:2500 dilution (Cell Signaling Technology, Cat# 5023), anti-apoptosis cocktail at a 1:250 dilution (Abcam, Cat# ab136812) and rabbit anti-β-Actin at a 1:1000 dilution (BD Biosciences, Cat# 612657) for the primary antibodies overnight at 4 °C. Subsequently, the blots were incubated for two hours at room temperature with horseradish peroxidase-conjugated secondary anti-rabbit IgG antibody (Invitrogen, Cat# 31460) at a 1:5000 dilution or the secondary anti-apoptosis cocktail antibody at a 1:100 dilution (Abcam, Cat# ab136812). The immunoblots were developed in the presence of a chemiluminescence substrate (Bio-Rad), and the images were recorded using a ChemiDoc^{TM} XRS+ molecular imager equipped with Image Lab^{TM} software (Bio-Rad). The Western blots were quantified using ImageJ (1.53) software.

4.1.8 Deletion detection by genomic PCR and agarose gel electrophoresis

To check the deletion of targeted exons, the EGFP-positive TJC8iCas9 cells were cultured in the presence and absence of DOX. Next, these DOX+ (Cas9 being expressed) and DOX- (Cas9 not expressed) iPSCs were cultured over 10 days, and both the DOX+ and DOX- cells were harvested for genomic DNA at day 5 and day 10. Post-harvest, the cells were processed as previously mentioned in Chapter 3, Section 3.1.6: Deletion detection by genomic PCR and agarose gel electrophoresis to extract their genomic DNA.

PCR was performed using the extracted genomic DNA with primers flanking the intended *UTF1* deletion site on exon 2. The primer sequences are mentioned in

Chapter 3, Table 3.2. The PCR was performed in combination with the EmeraldAmp MAX HS PCR Master Mix (Takara). In the *UTF1* knockout iPSCs, the amplicon size was expected to be a 236 bp PCR product, whereas the control cells would give a 298 bp PCR product. The products were run on 2% agarose gel for analysis.

4.1.9 Deletion detection by capillary electrophoresis and Sanger sequencing

Capillary electrophoresis and Sanger sequencing for the genomic DNA of TJC8iCas9 cells with the dual gRNAs vector was processed as mentioned in Chapter 3, Section 3.1.7: Deletion detection by capillary electrophoresis and Sanger sequencing.

4.2 Results and Discussion

4.2.1 Generation of *UTF1* knockout human iPSCs using the CRISPR/Cas9 system

In our earlier experiment, we could not generate stable iPSC lines from *UTF1* knockout fibroblasts. This intrigued us to investigate its role in maintaining iPSC identity in a well-characterized iPSC line used in an earlier study [82]. To achieve this, we generated *UTF1* knockout iPSCs using the same lentiviruses (encoding for Cas9 protein and the dual gRNAs) produced to knockout the *UTF1* gene in fibroblasts. The absence of UTF1 expression resulted in the sudden decline in viability of these iPSCs, making it difficult to have adequate cells for further analysis (data not shown). To circumvent this, we then used an iPSC line that encoded for Cas9 upon induction with DOX [130] to regulate the *UTF1* knockout in iPSCs (Figure 4.2A). These iPSCs were transduced with the lentiviruses carrying the lentiviral vector with dual gRNAs to generate *UTF1* knockout iPSCs (Figure 4.2B). Parallelly, iPSCs were also transduced with an empty dual gRNAs lentiviral vector expressing EGFP, serving as a control. The iPSCs were expanded for 48 hours post-transduction, and then EGFP-positive iPSCs were sorted, indicating successful integration and expression of the dual gRNAs vector (Figure 4.2C). These dual gRNAs-expressing iPSCs were then induced with DOX (0.5 μg/mL), resulting in the expression of Cas9. Co-expression of the Cas9 protein and dual gRNAs resulted in the deletion of the targeted region of the human *UTF1* gene.

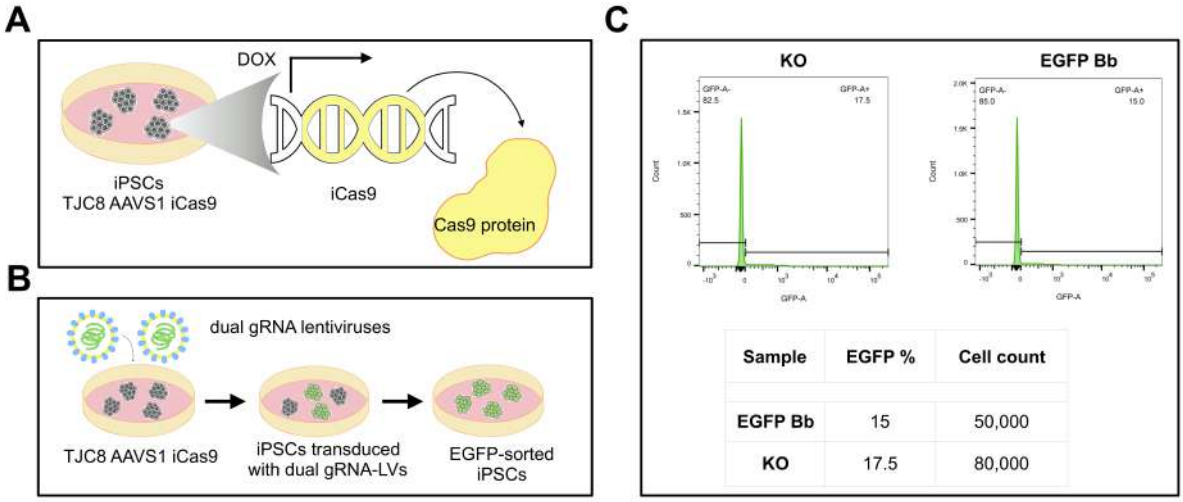


Figure 4.2: Generation of *UTF1* knockout human iPSCs. (A) Schematic representation of the working of the iPSC line TJC8iCas9. Upon induction with DOX, the stably integrated Cas9 vector at the AAVS1 site generated the Cas9 protein. **()B)** A schematic of the workflow followed for transduction and selection of dual gRNA expressing iPSCs. **(C)** Flow cytometry for EGFP expression in control (empty backbone) and knockout iPSCs (exon 2 targeting dual gRNA pair). DOX: Doxycycline; TJC8 AAVS1 iCas9: TJC8 specifies clone number, AAVS1 specifies the integration site, and iCas9 specifies that the Cas9 expression is inducible; gRNA: guide RNA; LV: lentiviruses; KO: knockout; EGFP Bb: Enhanced Green Fluorescent Protein Backbone.

4.2.2 *UTF1* knockout in human iPSCs resulted in loss of cell viability

The EGFP-sorted dual gRNAs expressing iPSCs were cultured either in the presence or absence of DOX. In the presence of DOX, Cas9 and dual gRNAs were co-expressed and this resulted in the deletion of the targeted *UTF1* region, whereas the absence of DOX resulted in the retention of the intact *UTF1* gene because of the lack of Cas9 protein. It was observed that *UTF1* knockout cells resulted in a sudden decline in the viability of these iPSCs upon DOX induction from day 2 onwards (data not shown), with very few cells remaining in culture by day 6 (Figure 4.3A). This drastic decline in the number of iPSCs by day 6 was depicted by the expression of pluripotent stem cell surface marker TRA-1-60 and the dual gRNAs vector by EGFP expression in an entire well (Figure 4.3B). The nuclei of the remaining cells were also stained by DAPI (Figure 4.3B). This indicated the inability of iPSCs to compensate for the gradual loss of UTF1 expression. In contrast, the dual gRNAs expressing iPSCs in the absence of Cas9 (DOX-) and iPSCs expressing the empty dual gRNAs vector in the presence of Cas9 (DOX+) showed a confluent well with iPSCs expressing the pluripotency marker TRA-1-60, and

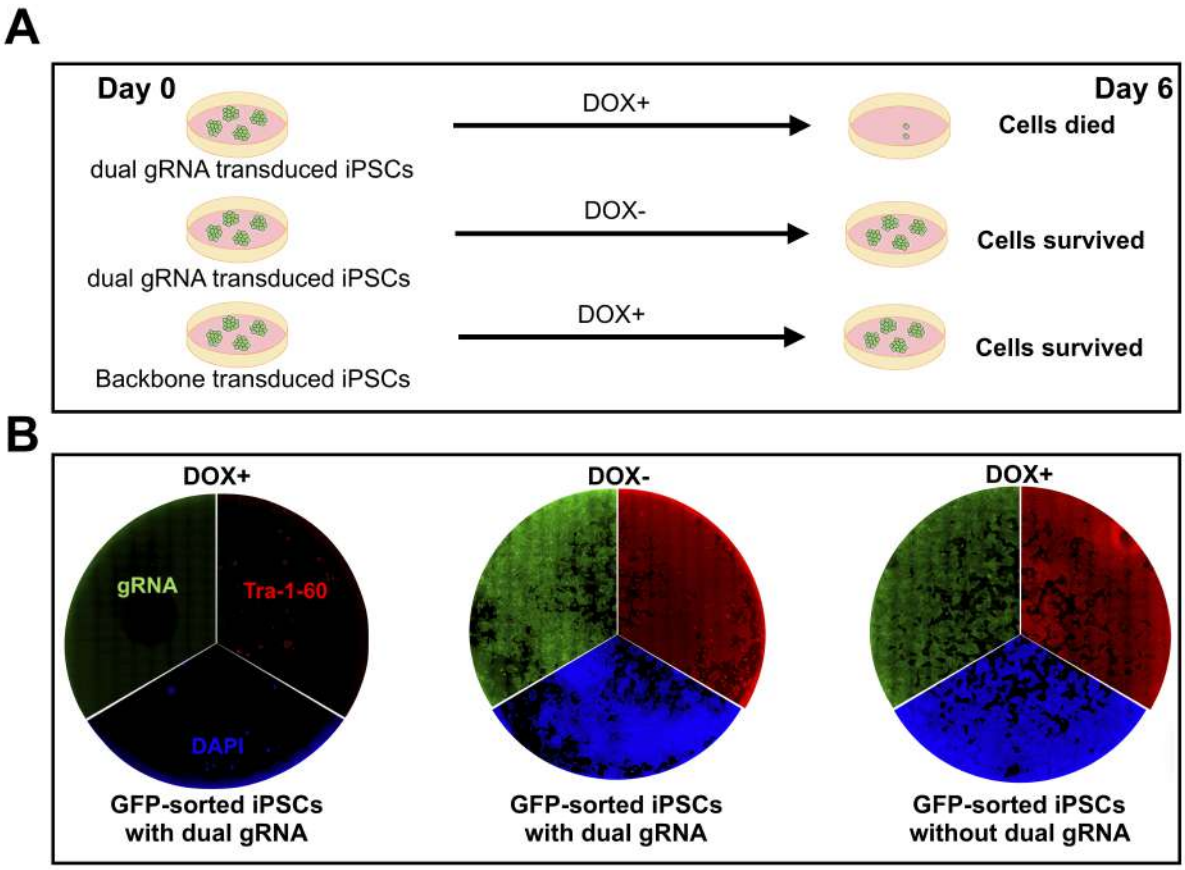


Figure 4.3: Loss of cell viability of human iPSCs upon *UTF1* knockout. (A) A schematic representation depicting the experimental design followed to study the effect of the absence and presence of UTF1 on cell viability. GFP-sorted iPSCs (*UTF1* knockout) lose cell viability upon DOX induction (induced Cas9 expression) on day 6. The control cells, GFP-sorted iPSCs with dual gRNA (DOX-) and GFP-sorted iPSCs with empty backbone vector (DOX+) cells, did not show loss of cell viability on day 6. **(B)** Immunofluorescence assay of the entire well of a 12-well plate of the above-mentioned cell types with TRA-1-60 antibody, dual gRNAs EGFP expression, and DAPI staining showing loss of iPSC viability in *UTF1* knockout well only. DOX: Doxycycline; gRNA: guide RNA.

the dual gRNAs vector indicated by EGFP expression. The nuclear stain DAPI was also co-related with EGFP expression and TRA-1-60 staining (Figure 4.3B).

Additionally, to ensure that the iPSCs lost their viability upon knockout of *UTF1*, the dual gRNAs expressing EGFP-positive iPSCs were seeded with inducible Cas9 iPSCs in a ratio of 1:1 (Figure 4.4A). This mixed iPSC population was cultured either in the presence or absence of DOX. In the presence of DOX, the deletion of the targeted *UTF1* region occurred only in EGFP-expressing cells as they co-expressed dual gRNAs and Cas9. These cells were eventually devoid of the iPSC population, indicating the loss

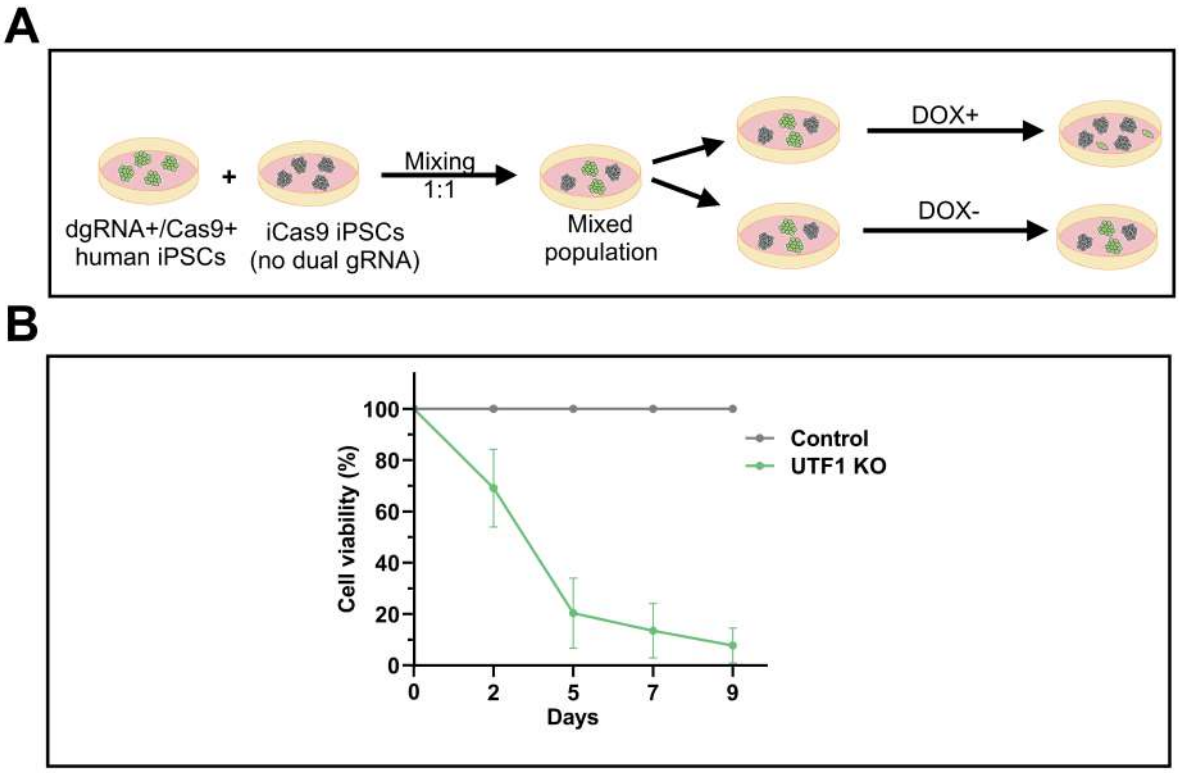


Figure 4.4: Selective viability loss in human iPSCs following *UTF1* knockout. (A) A schematic layout representing the experimental plan. iPSCs co-expressing dual gRNA and Cas9 were cultured in a 1:1 ratio with only Cas9 expressing iPSCs to study the effect of UTF1 on cell viability over 9 days. Cells (mixed population) were cultured in the presence (wild type *UTF1*) and absence (*UTF1* knockout) to study its effect on cell viability. **(B)** A decline in iPSC viability was observed from day 2 through day 9 for DOX+ population. DOX: Doxycycline; gRNA: guide RNA.

of cell viability. This was concluded by an increased ratio of EGFP-negative untransduced cells over 9 days. Flow cytometry analysis showed a drastic decline in EGFP expressing cells from day 2 onwards, with negligible EGFP-expressing cells remaining in culture by day 9 (Figure 4.4B). The well was repopulated by the dual gRNAs-negative iPSCs. The absence of DOX resulted in the retention of the 1:1 ratio of the EGFP expressing iPSCs to the untransduced cells, as indicated by the flow cytometry analysis till day 9 (Figure 4.4B). These results indicated that the loss of iPSC identity was because of the absence of UTF1 expression.

4.2.3 Confirmation of deletion of the human *UTF1* gene in iPSCs

To verify the deletion in the human *UTF1* gene, we performed PCR on genomic DNA samples extracted from day 5 and day 10 of control (DOX-) and *UTF1* knockout (DOX+) iPSCs. The analysis of PCR products from the day 5 genomic DNA sample (DOX+) resulted in two bands, one of 236 bp and the other of 298 bp, indicating partial deletion in these iPSCs. However, the analysis of PCR products from the day 10 genomic DNA sample (DOX+) resulted in only one band of 236 bp (shown by a red arrow), indicating complete deletion of the targeted region in the human *UTF1* gene (Figure 4.5A). Further, capillary electrophoresis supported this observation, with a 44% deletion observed for DOX+ cells on day 5 and a 100% deletion observed for DOX+ cells on day 10 (Figure 4.5B). The control peaks are highlighted in a red box. At the molecular level, the absence of UTF1 was demonstrated by extracting total protein lysates from EGFP-positive iPSCs co-expressing dual gRNAs and Cas9. Samples were collected for Western blots from DOX+ day 2, day 4, and day 6 iPSCs and DOX- day 6 iPSCs. Western blot analysis of these samples showed a gradual decline in the UTF1 protein from day 2 onwards, and its absence was observed in day 6 DOX+ lysate, indicating successful knockout of the human *UTF1* gene (Figure 4.5C).

The knockout of the human *UTF1* gene in iPSCs, showed a complete deletion of the targeted *UTF1* region, unlike what was previously observed in our study with fibroblasts. UTF1 is expressed in iPSCs and is, therefore, easily accessible to the CRISPR/Cas9 machinery. Further, various studies have highlighted that during the reprogramming of fibroblasts to iPSCs, the chromatin undergoes a closed-to-open chromatin configuration [40, 72]. This modification in chromatin structure potentially facilitates Cas9 accessibility for targeting the *UTF1* gene in iPSCs, in contrast to the limited access observed in human fibroblasts.

4.2.4 Decline in expression levels of the core pluripotency markers

The observation that knockout of *UTF1* in iPSCs resulted in loss of cell viability prompted us to investigate the underlying genotypic alterations. Consequently, we delved deeper into the impact of UTF1 deficiency on the pluripotency network. Therefore, we analyzed the change in expression in the core pluripotency factors, OCT4 and SOX2.

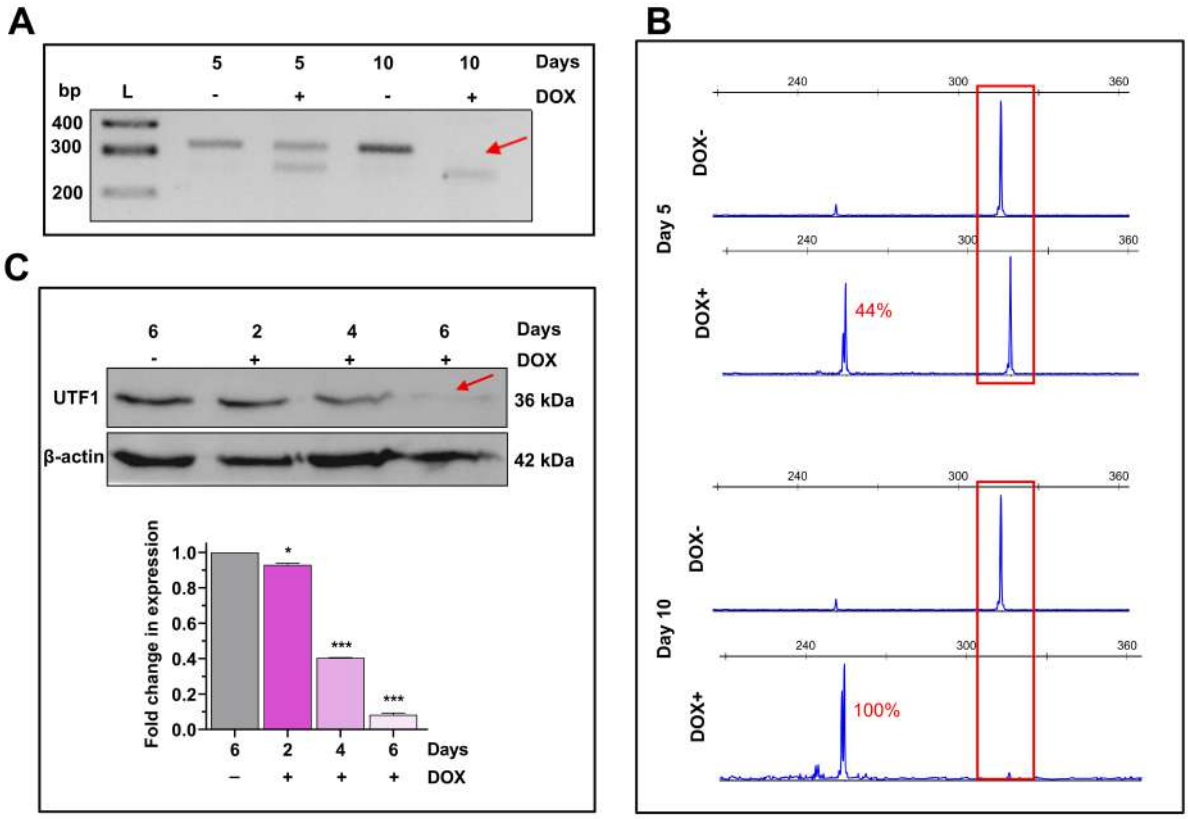


Figure 4.5: Confirmation of deletion of the human *UTF1* gene in iPSCs. (A) Results of genomic DNA PCR using a forward primer that binds upstream of the first gRNA site and a reverse primer binding downstream of the second gRNA site for control (DOX-) and *UTF1* knockout (DOX+) iPSCs at day 5 and day 10 post-DOX induction (Cas9 expression). **(B)** Capillary electrophoresis of the DOX+ iPSCs confirmed a partial *UTF1* knockout with 44% of the amplicons showing deletion on day 5 and a complete *UTF1* knockout with 100% amplicons showing deletion on day 10. **(C)** Western blotting results of GFP-sorted iPSCs with dual gRNA in the presence and absence of DOX show a gradual decline in UTF1 expression for DOX+ cells over 6 days (Western Blot: Top and Image J quantification: Bottom). bp: base pair; L: Ladder; DOX: Doxycycline. The quantitative values show the mean fold change±S.D. Statistical analysis was performed using a two-way ANOVA (n=3). ∗P<0.05, ∗P<0.01, ∗P<0.001.

The effect of the complete absence of UTF1 on core stem cell-specific proteins was investigated using Western blot analysis. The results of day 6 DOX+ and DOX- cells showed a significant decline in the expression of the core pluripotency markers, OCT4 and SOX2, indicating the importance of UTF1 in regulating the pluripotency network (Figure 4.6).

It is well established that the core pluripotency factors OCT4 and SOX2 regulate the expression of UTF1 [128, 98, 97, 5]. Our study provides insight into the effect of UTF1 on the pluripotency network. The absence of UTF1 negatively affects the

stability of the pluripotent state of human iPSCs. It has been previously reported that the UTF1 protein is stably associated with the chromatin in ESCs [139]. Further, it has also been identified as a key regulator of the bivalency state in pluripotent cells [60]. The lack of UTF1 could deregulate the bivalency state, thereby disrupting the pluripotency network. Thus, we hypothesized that in the absence of UTF1, the observed decline in the expression levels of core pluripotency markers such as OCT4 and SOX2 could be attributed to a disruption in the pluripotency network.

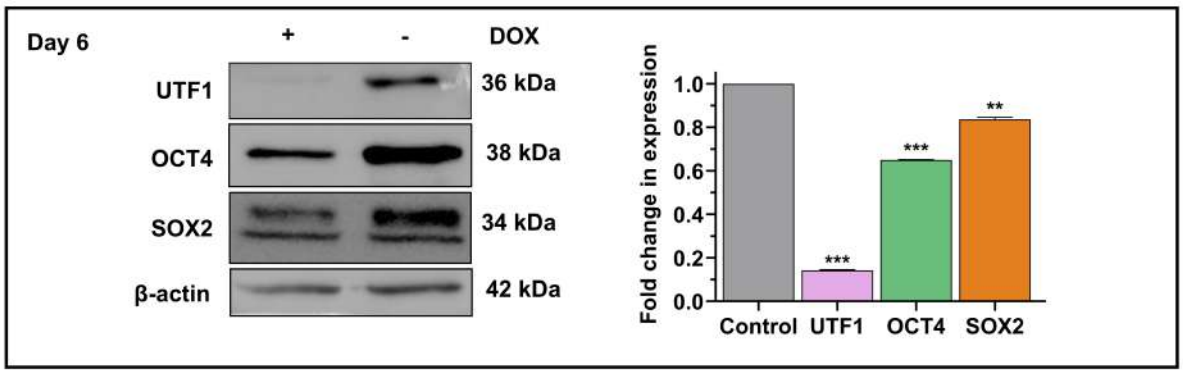


Figure 4.6: Expression of core pluripotency markers in *UTF1* knockout and UTF1 expressing iPSCs. Western blot for pluripotency markers from day 6 protein lysates (Western Blot: Left and Image J quantification: Right). DOX: Doxycycline. The quantitative values show the mean fold change±S.D. Statistical analysis was performed using a two-way ANOVA (n=3). ∗P<0.05, ∗P<0.01, ∗P<0.001.

4.2.5 *UTF1* knockout iPSCs lose cell viability due to the induction of apoptosis

Further, to investigate whether the cells lost viability due to apoptosis, the lysates from day 6 DOX+ and DOX- cells were analyzed. A significant increase in the expression of pro-apoptotic protein Bax2 was observed in cells treated with DOX. Bax2 is known to generate pores in the outer mitochondrial membrane, leading to the release of cytochrome c and activation of caspases [27, 156]. These DOX+ cells showed a significant increase in expression of the apoptotic marker cleaved caspase 3. The activation of procaspase 3 to cleaved caspase 3 has been previously reported to cleave PARP [2]. In our study, we did observe a significant increase in levels of cleaved PARP. Therefore, a considerable increase in the expression of Bax2, cleaved caspase 3, and cleaved PARP in *UTF1* knockout iPSCs (Figure 4.7) confirmed the loss of cell viability through apoptosis.

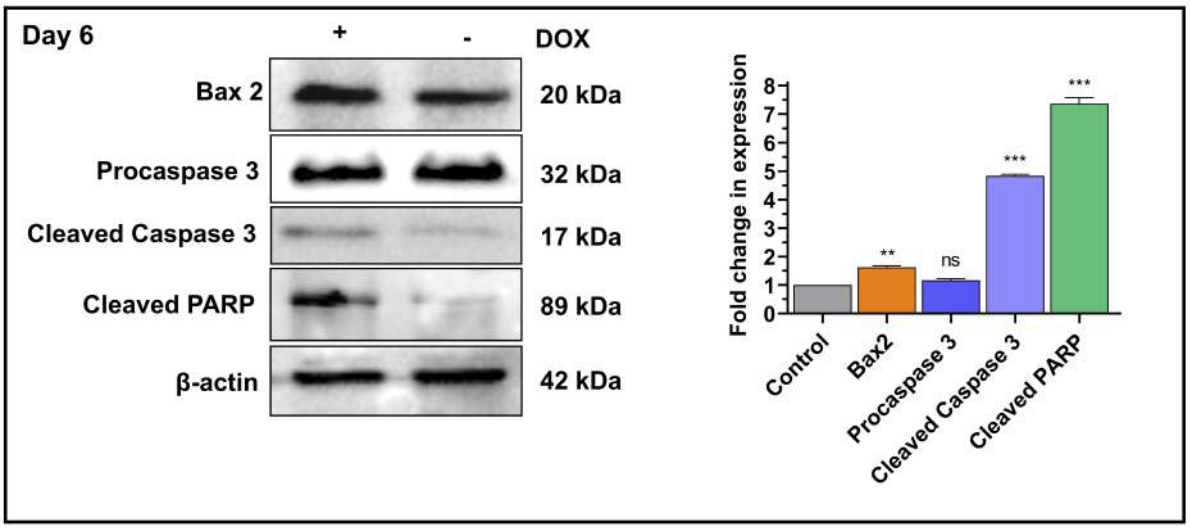


Figure 4.7: Expression of apoptotic markers in *UTF1* knockout and UTF1 expressing iPSCs. Western blot for apoptotic markers from day 6 protein lysates (Western Blot: Left and Image J quantification: Right). DOX: Doxycycline. The quantitative values show the mean fold change±S.D. Statistical analysis was performed using a two-way ANOVA (n=3). ∗P<0.05, ∗P<0.01, ∗P<0.001.

4.3 Conclusion

In summary, our investigation into the generation of *UTF1* knockout human iPSCs via the CRISPR/Cas9 system revealed a notable contrast with fibroblasts, as iPSCs exhibited complete deletion of the *UTF1* gene, likely attributed to its transition from heterochromatin (in fibroblasts) to euchromatin (in iPSCs). Remarkably, UTF1 deficiency in iPSC resulted in a significant decline in their viability from day 2 through day 9. Western blot analysis unveiled a substantial loss of UTF1 expression in iPSCs by day 6 of DOX induction, correlating with Cas9 activation and consequent human *UTF1* gene deletion. Furthermore, this loss of viability was supported by elevated levels of various pro-apoptotic factors, namely Bax2, cleaved Caspase 3, and cleaved PARP. Moreover, the expression of core pluripotency factors OCT4 and SOX2 exhibited a significant reduction by day 6. These findings underscore the crucial role of UTF1 in establishing and maintaining iPSCs, shedding light on its indispensable function in ensuring their viability and pluripotency.

CHAPTER 5

GENERATION OF UTF1 KNOCKDOWN HUMAN iPSCs AND DETERMINING ITS ROLE IN MAINTENANCE OF STEM CELL CHARACTERISTICS

An article based on this chapter is currently under review as follows:

Raina K, Modak K, Premkumar C, Velayudhan SR, Thummer RP **(2024)**. UTF1 expression is important for the generation and maintenance of human iPSCs.

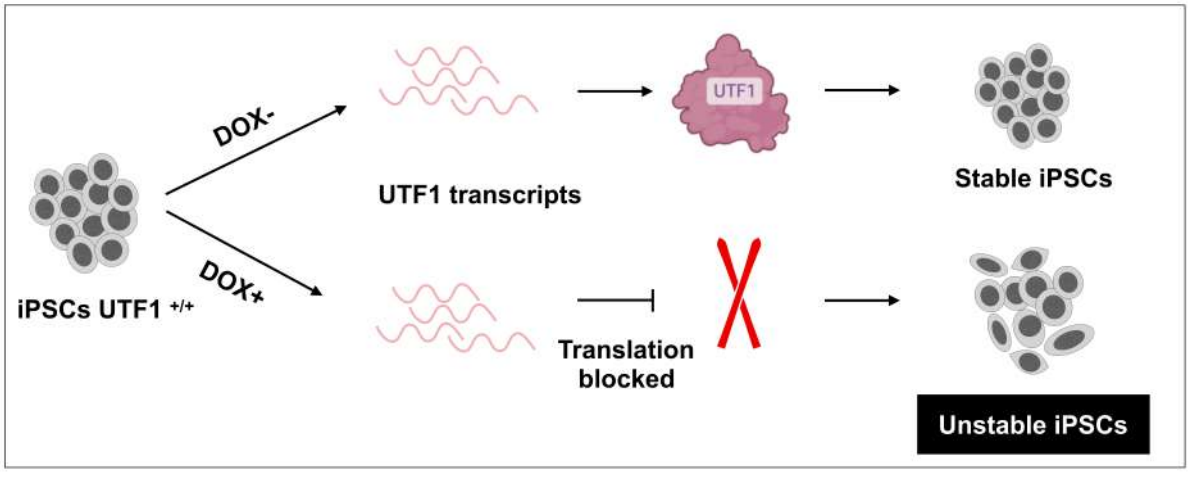


Figure 5.1: A schematic representation of the experimental design followed to study the effect of a decline in the expression of UTF1 in iPSCs.

BRIEF OVERVIEW OF CHAPTER

Considering the significant decrease in viability observed upon UTF1 knockout in iPSCs, we aimed to investigate the implications of downregulation of its expression in an iPSC population. To investigate the same, we demonstrated the generation of a novel short hairpin RNA (shRNA) vector for generating UTF1 knockdown iPSCs. The shRNA against UTF1 was designed and eventually cloned in a DOX inducible vector with a fluorescent marker (ZsGreen), followed by the shRNA against UTF1. Additionally, a puromycin resistance gene was introduced downstream of the endogenous promoter PPP1R12C to select the transfected cells to facilitate selection before inducing shRNA expression. Also, this vector was designed to integrate specifically at the safe harbor site AAVS1 in all the cells. This vector was transfected in human iPSCs via electroporation, and the successfully transfected cells were selected with puromycin. Upon addition of DOX to the culture media, ZsGreen and shRNA against UTF1 were expressed, resulting in the knockdown of human UTF1 in iPSCs.

To ensure the successful knockdown of UTF1 expression, this chapter delves into the decline in UTF1 expression over 8 days. Its expression was observed till day 4, and on day 8, a decrease in UTF1 expression was observed. This reduction in its levels was also visible at the phenotypic levels, wherein the cells showed spontaneous differentiation, which was visible by the change in the typical iPSC morphology. Further, it was observed that most of the cell population was TRA-1-60 negative and ZsGreen positive. Further, we also studied the effect on the core pluripotency factors OCT4 and SOX2, which showed a significant decline in their expression. This further supported the previous phenotypic observation, indicating a considerable change in the pluripotency network. In addition, to ensure that the cells were not undergoing apoptosis similar to the *UTF1* knockout cells, pro-apoptotic markers were analyzed, and no increase in any of these markers was observed. In conclusion, this chapter highlights the importance of sustained human UTF1 expression in maintaining pluripotency in iPSCs.

5.1 Materials and Methods

5.1.1 Cell Culture

The established iPSC line TJC8 [82] was cultured in complete StemMACS^{TM} iPSC-Brew XF medium (Miltenyi Biotec) with 50 U/mL penicillin and 50 μg/mL streptomycin. The cells were cultured as a monolayer in a humidified incubator at 37 °C in the presence of 5% CO_2. iPSCs were passaged with 0.5 mM EDTA (Invitrogen). To select iPSCs having integration of the shUTF1 vector, these cells were cultured in 0.5 μg/mL of puromycin (Gibco) for 4 days. Further, for inducing the expression of shRNA against UTF1, these selected iPSCs were cultured with 0.4 μg/mL DOX.

5.1.2 Generation of a lentiviral vector expressing the shRNA against human *UTF1*

The full-length human UTF1 shRNA sequence was obtained using the Cold Spring Harbor Laboratories database shERWOOD [66] (Table 5.1). The shRNA oligos were ordered and PCR-amplified with primers containing *Xho*I and *EcoR*I restriction sites (Table 5.2). A lentiviral shRNA cloning vector, namely pB-ZmiRE-sPA, was previously developed in the lab (to be published) with a cloning site for shRNA flanked by *Xho*I and *EcoR*I restriction sites. PCR-amplified shRNA oligo and the pB-ZmiRE-sPA vector were digested with *Xho*I and *EcoR*I (New England BioLabs). The digested pB-ZmiRE-sPA plasmid was gel-extracted, and 0.05 pmol of this vector was ligated overnight with the digested 0.1 pmol insert shRNA oligo to obtain the pB-ZmiRE-shUTF1 vector (Figure 5.2A). Next, 2 μL of this ligated product was transformed in XL1-Blue competent cells (Agilent Technologies), and plasmids were extracted from the bacterial clones obtained. The correct shRNA sequence in the pB-ZmiRE-shUTF1-sPA plasmid was confirmed by Sanger sequencing (Figure 5.2B).

Table 5.1: shRNA sequences targeting human *UTF1* gene transcript.

	Primer sequence (5'-3')
shRNA -Forward primer	TGCTGTTGACAGTGAGCGACCGCTACAAGTTCCTTAAAGATAGTGAAGC CACAGATGTATCTTTAAGGAACTTGTAGCGGCTGCCTACTGCCTCGGA
shRNA -Reverse primer	TCCGAGGCAGTAGGCAGCCGCTACAAGTTCCTTAAAGATACATCTGTGG CTTCACTATCTTTAAGGAACTTGTAGCGGTCGCTCACTGTCAACAGCA

Table 5.2: Primer sequences to insert restriction sites flanking the shRNA sequence.

	Primer sequences (5'-3')
for *Xho*I	TGAACTCGAGAAGGTATATTGCTGTTGACAGTGAGCG
for *EcoR*I	TCTCGAATTCTAGCCCCTTGAAGTCCGAGGCAGTAGGC

5.1.3 Generation of an AAVS1 site integrating vector encoding for shRNA against human *UTF1*

Further, the shRNA region from the cloned pB-ZmiRE-shUTF1-sPA plasmid was used to replace the Cas9 region of the pAAVS1-PDi-CRISPRn (Addgene ID: 73500) (gifted by Bruce Conklin) to ensure integration of the shRNA at the safe-harbor AAVS1 site. The pAAVS1-PDi-CRISPRn plasmid was first digested with *Age*I and *Pac*I to remove the Cas9 region, and then the gel was extracted for further cloning (Figure 5.3A). The Kozak sequence, ZsGreen, miR30, shUTF1, and the sPA regions were PCR amplified from the pB-ZmiRE-shUTF1-sPA vector for cloning by Gibson assembly (Figure 5.3B). The PCR product was amplified using a forward primer containing a homologous overhang to the destination plasmid towards the *Pac*I digested site and a reverse primer with a homologous overhang towards the *Age*I digested site. The primer sequences used are mentioned in Table 5.3. For setting up Gibson assembly, the NEBuilder HiFi DNA Assembly Master Mix (New England BioLabs, Cat# E2621L) was used with 0.05 pmol of the digested backbone and 0.1 pmol of the amplified fragment, as per the manufacturer's protocol. Next, 2 μL of this ligated product was transformed in XL1-Blue competent cells (Agilent Technologies) and plasmids were extracted from the bacterial clones obtained. The presence of the correct shUTF1 sequence in the final pAAVS1-ishUTF1-Puro^R plasmid was confirmed by *Hind*III (New England BioLabs) digestion and Sanger sequencing (Figure 5.3C).

Table 5.3: Primer sequences to amplify the desired shUTF1 region.

	Primer sequences (5'-3')
i-shUTF1 -Forward primer	CTCGTAAACTTAAGGTTAATGGCCACCATGGCCCAGTC
i-shUTF1 -Reverse Primer	GGCGATCGATTGCGGCCGCACACACAAAAAACCAACACACAGATC TAATGAAAATAAAGATC

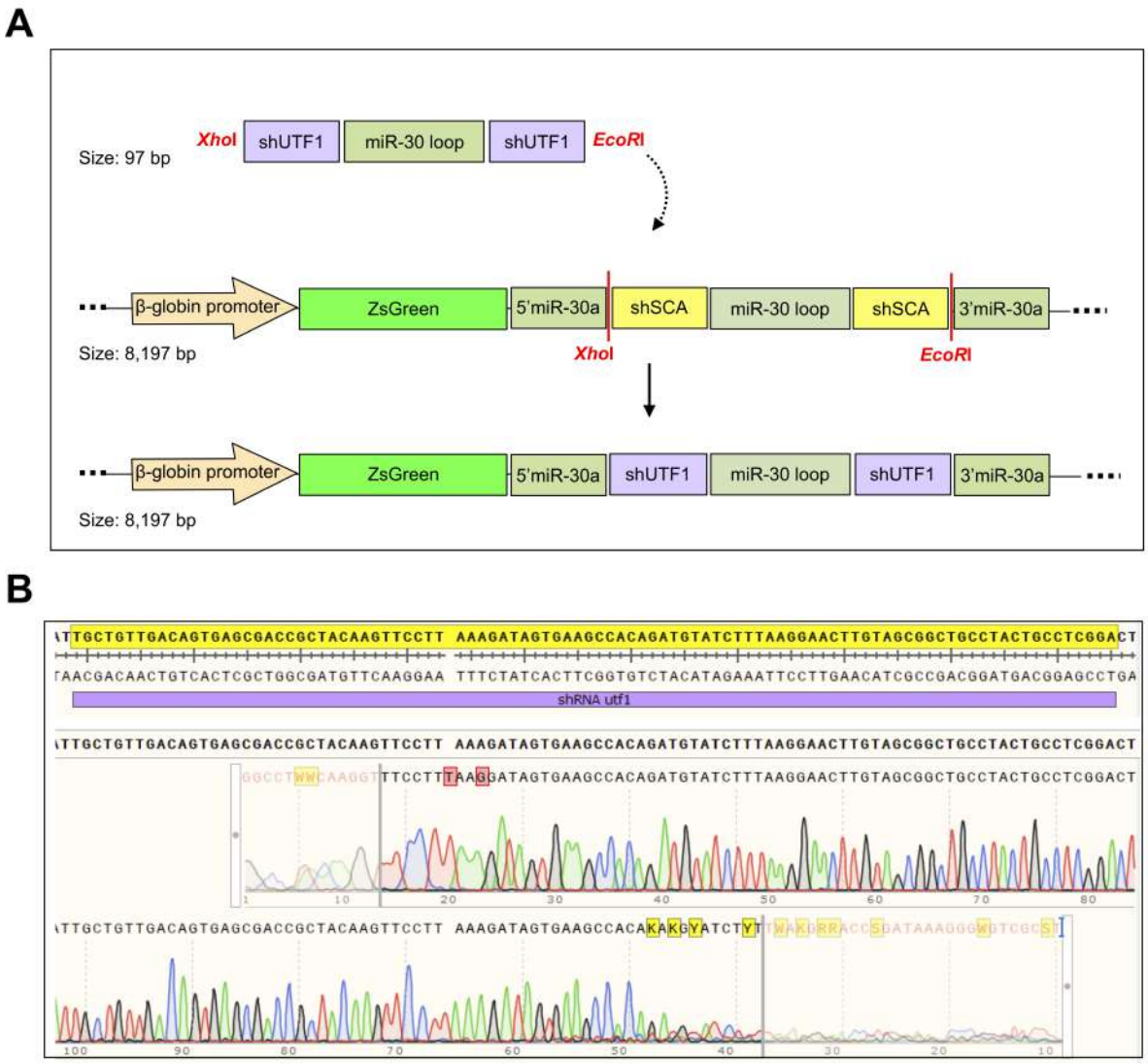


Figure 5.2: Cloning shRNA against human UTF1 in a lentiviral vector. (A) A schematic depicting the cloning of a shRNA against human UTF1 in the lentiviral vector pB-ZmiRE-sPA by restriction enzymes *Xho*I and *EcoR*I. **(B)** Confirmation of the cloned shRNA sequence in the pB-ZmiRE-shUTF1 vector by Sanger sequencing. shUTF1: short hairpin RNA against UTF1; miR: microRNA; ZsGreen: Zoanthus species green; shSCA: short hairpin scaffold.

5.1.4 Transfection of the pAAVS1-ishUTF1-Puro^{R} expression vector in iPSCs

The iPSC line, TJC8 [82], was transfected with the pAAVS1-ishUTF1-Puro^{R} vector along with the plasmids pZT-AAVS1-R1 and pZT-AAVS1-L1 (Addgene ID: 52638 and 52637) (gifts from Mahendra Rao and Jizhong Zou) using a 100 μL neon tip and the Neon^{TM} transfection system (Life Technologies). The electroporation parameters used

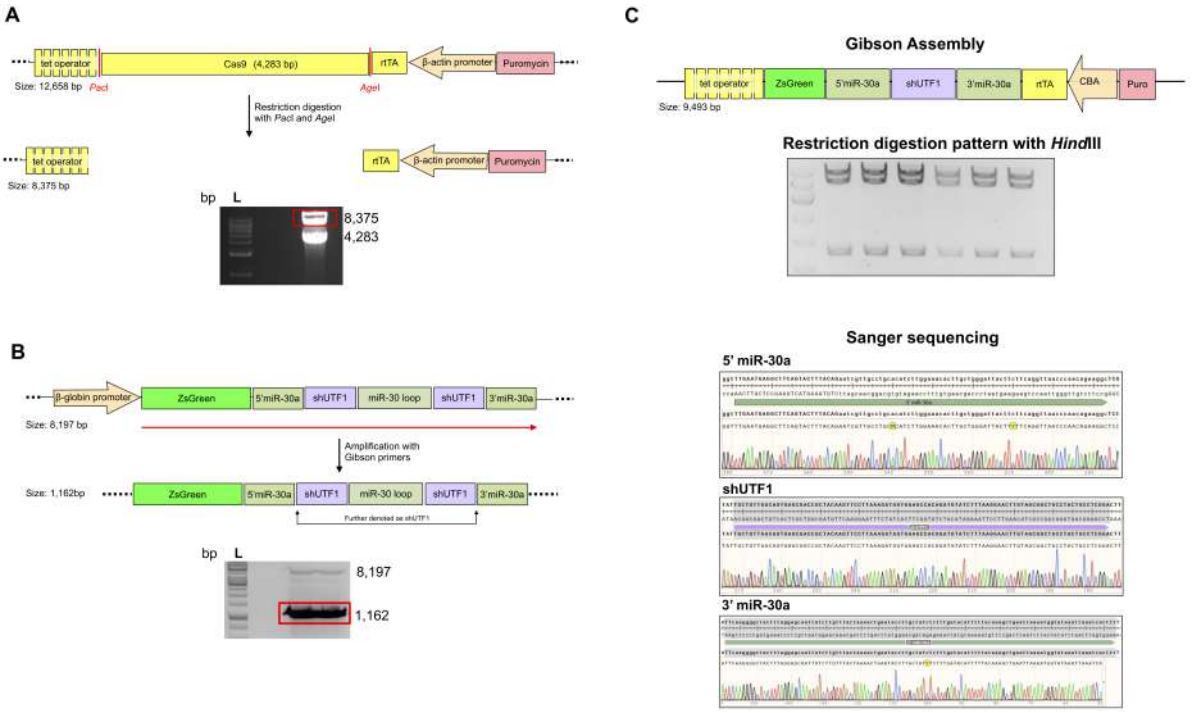


Figure 5.3: Cloning shRNA against human UTF1 in an AAVS1 site integrating vector. (A) A schematic depicting the removal of Cas9 encoding region from the pAAVS1-PDi-CRISPRn vector. Agarose gel electrophoresis separated the Cas9 and the digested backbone (highlighted in red box). **(B)** A schematic depicting the amplification of the desired region from the lentiviral vector pB-ZmiRE-shUTF1. Agarose gel electrophoresis segregated the desired fragment, which was further used for cloning. **(C)** A schematic representation of Gibson assembly of the digested backbone and amplified PCR fragment to obtain the desired pAAVS1-ishUTF1-Puro^R vector. Confirmation of multiple clones by restriction digestion pattern observed with *Hind*III and Sanger sequencing to confirm the cloning of the correct shRNA sequence. rtTA: reverse tetracycline transactivator; bp: base pairs; L: ladder; shUTF1: short hairpin RNA against UTF1; miR: microRNA.

were 1 pulse with a voltage of 1300 V and a length of 30 milliseconds. 1×10^6 cells were pelleted down and then resuspended in 110 μL of electroporation buffer R containing 2.5 μg pAAVS1-ishUTF1-Puro^R or 2.5 μg pAAVS1-GFP-Puro^R, 1 μg pZT-AAVS1-R1 and 1 μg pZT-AAVS1-L1 plasmids. The cell suspension containing the plasmid mix was picked up with the 100 μL neon tip, ensuring no bubbles were formed. The neon tip was then inserted into a neon tube containing electrolytic buffer E2, which acts as a transfection chamber. The cells were then electroporated and seeded into one well of a 12-well Matrigel-coated plate in StemMACS^{TM} iPS-Brew XF medium.

5.1.5 Immunofluorescence staining and microscopy

To detect expression of pluripotency surface marker TRA-1-60, human iPSCs were cultured and processed as previously mentioned in Chapter 4, Section 4.1.6: Immunofluorescence staining and microscopy on Matrigel-coated wells. Briefly, the iPSCs were cultured for 8 days with and without DOX. On day 8, the cells were processed for incubation with the TRA-1-60 primary antibody, followed by the Alexa Fluor^{TM} 594 secondary antibody. The nuclei were stained with 10 μg DAPI (Invitrogen) for 5 minutes at room temperature. Images were captured using the Bio-Rad ZOE Fluorescent Cell Imager or the Leica DMI6000B fluorescence microscope equipped with LAS X software.

5.1.6 SDS-PAGE and Western blotting

Whole-cell lysates were harvested, and the lysates were resolved on 12% SDS-PAGE and transferred to a nitrocellulose membrane as mentioned previously in Chapter 3, Section 3.1.5: SDS-PAGE and Western blotting. The immunoblotting was carried out using anti-UTF1, anti-OCT4, anti-SOX2, anti-Bax2, anti-Apoptosis cocktail, and anti-β-Actin antibodies as previously mentioned in Chapter 4, Section 4.1.7: SDS-PAGE and Western blotting.

5.2 Results and Discussion

5.2.1 Generation of a human AAVS1-ishUTF1-Puro^{R} iPSCs

We utilized a previously reported TALEN-based approach, known for the efficient generation of reporter cell lines [106], to generate iPSCs with a DOX-inducible AAVS1-ishUTF1-Puro^{R} expression cassette integrated at the AAVS1 safe harbor genomic site on human chromosome 19. The donor vector, pAAVS1-ishUTF1-Puro^{R} containing the puromycin resistance gene and a DOX-inducible shUTF1 (ishUTF1) expression cassette, flanked by homologous arms for HDR at the target site, was generated (Figure 5.4A). The expression of the integrated puromycin-resistance gene and shUTF1 at the AAVS1 site was driven by endogenous PPP1RI2C promoter and tetracycline-inducible promoter, respectively (Figure 5.4A). To induce DSBs and enable integration at the AAVS1 site, we electroporated a well-characterized iPSC line TJC8 [82] with TALEN expression plasmids and the pAAVS1-ishUTF1-Puro^{R} vector (Figure 5.4B). The nucleofected cells

were cultured in the presence of puromycin to select the cells with stable integration of the ishUTF1 cassette and puromycin resistance gene near the PPPRI2C promoter (Figure 5.4A,B). The transgene integration at the AAVS1 site in the puromycin-selected cells was first visible by the expression of ZsGreen in the ishUTF1 cells (in the presence of DOX) and CopGFP in control cells (Figure 5.4A).

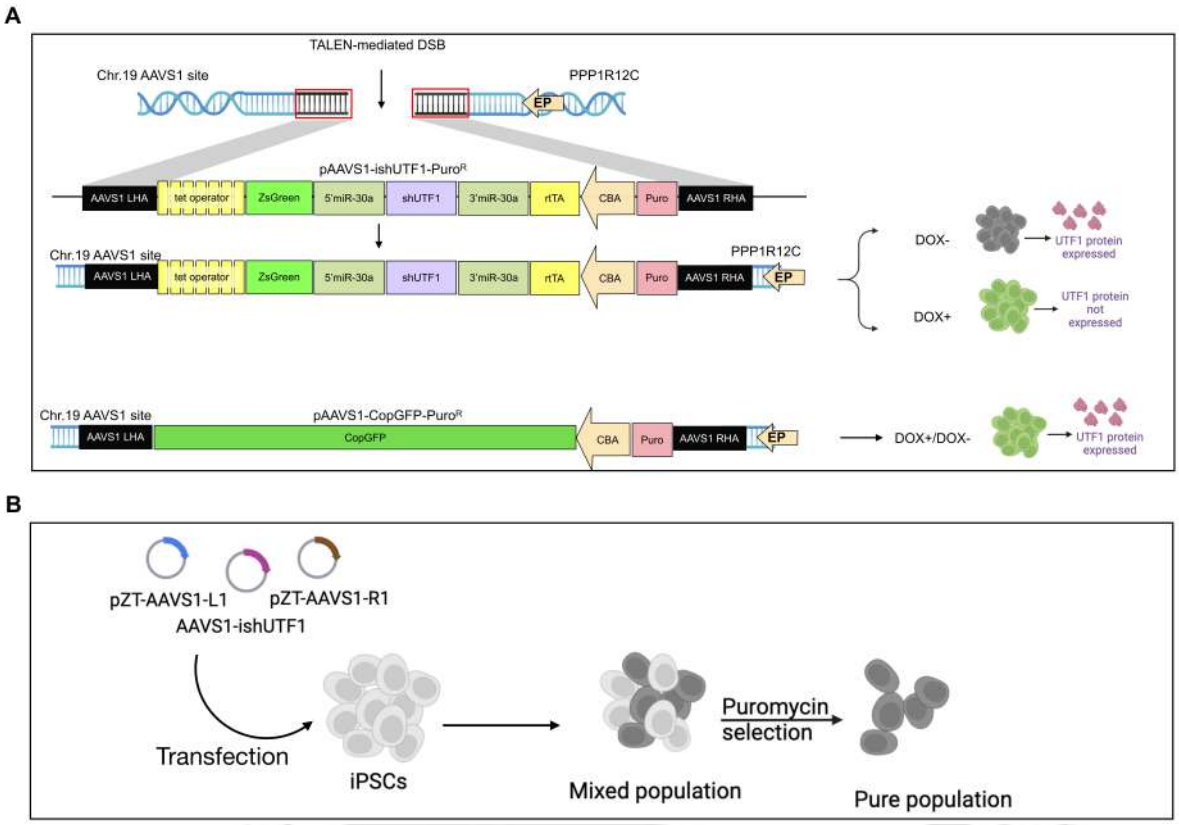


Figure 5.4: UTF1 knockdown in iPSCs. (A) A schematic depicting the integration of DOX-inducible shUTF1 cassette at the AAVS1 safe-harbor site of chromosome 19 by creating a DSB using a TALEN array. This DSB was repaired through the HDR mechanism with the donor plasmid pAAVS1-ishUTF1-Puro^R , which contains the Tet-On-shUTF1 cassette and puromycin resistance gene (Puro^R). The AAVS1 LHA and AAVS1 RHA facilitate the integration at the AAVS1 site. The endogenous PPP1R12C promoter drives the expression of Puro^R , which allows the selection of cells that underwent HDR and, therefore, contain the shUTF1 cassette. The addition of DOX activated the expression of ZsGreen and shUTF1. **(B)** Using the TALEN genome editing technology, a schematic depicting the protocol for generating UTF1 knockdown iPSCs is shown. AAVS1: Adeno-associated virus integration site 1; TALEN: Transcription Activator-Like Effector Nucleases; ZsGreen: Zoanthus species green; RHA: Right Homology Arm; LHA: Left Homology Arm; DOX: Doxycycline; DSB: Double-strand break; CBA: Chicken β-actin; Puro: Puromycin; EP: Endogenous promoter; PPP1R12C: Protein phosphatase 1 regulatory subunit 12C; miR30: microRNA 30; rtTA: reverse tetracycline transactivator; CopGFP: Copepod green fluorescent protein.

5.2.2 Determining the optimal DOX concentration for induction of shRNA

Controlled expression of gene silencing is crucial for regulating the targeted genes in specific cell types. Therefore, we performed experiments to determine the optimal DOX concentration and treatment duration required to achieve maximum ZsGreen expression levels.

When the AAVS1-ishUTF1 iPSCs were cultured in the presence of DOX concentrations ranging from 0.2 to 1 μg/mL for 48 hours, we found that ZsGreen expression could be detected at DOX concentrations as low as 0.2 μg/mL. However, the ZsGreen intensity was more for 0.4 μg/mL, remaining constant even at higher DOX concentrations (Figure 5.5). We concluded that a stable ZsGreen expression could be detected within 24 hours of treating the cells with 0.4 μg/mL of DOX.

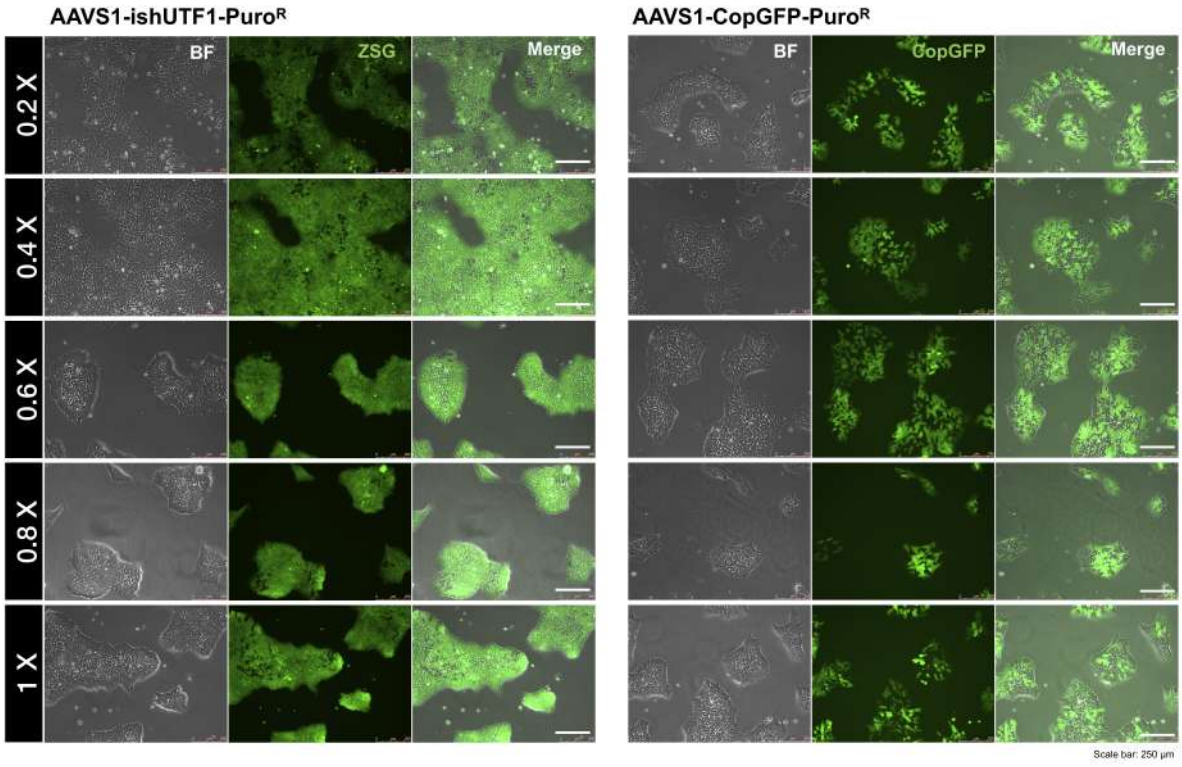


Figure 5.5: The optimal concentration of DOX analyzed for induction of shRNA after 48 hours. AAVS1: AAVS1: Adeno-associated virus integration site 1; shUTF1: short hairpin RNA against UTF1; Puro^{R}: Puromycin resistance.

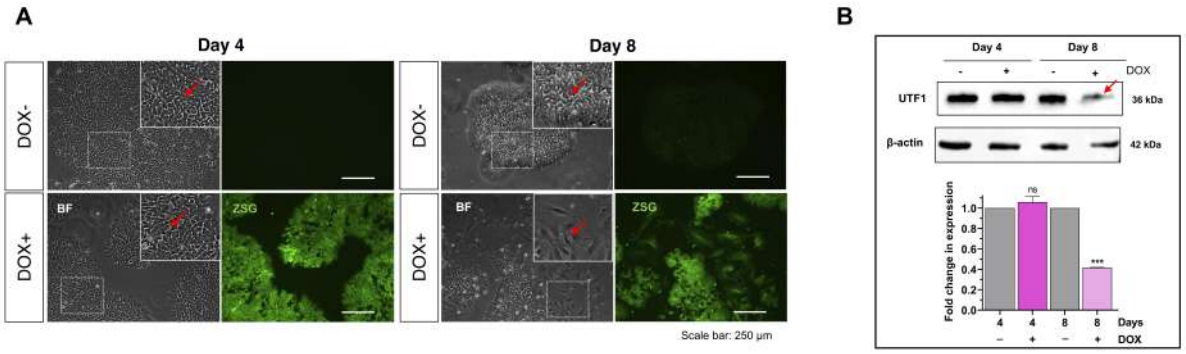


Figure 5.6: The decline in UTF1 levels and its effect on iPSCs. (A) Morphological similarity is seen in shUTF1-activated iPSCs on Day 4. A change in the morphology of iPSCs was observed by day 8 only in DOX+ cells. **(B)** Western blot analysis showing a decline in UTF1 protein expression levels on day eight only in the presence of DOX. Day 4 and DOX- cells for both day 4 and day 8 show comparable UTF1 levels (Western blot: Left and Image J quantification: Right).

5.2.3 Depletion of UTF1 expression results in spontaneous differentiation

The expression of shRNA against UTF1 in DOX+ cells was indicated by the successful expression of the downstream fluorescent marker ZsGreen (Figure 5.6A). The morphology of the DOX+ cells was similar to the DOX- cells on day 4 (inset images marked with red arrow) (Figure 5.6A; left). Interestingly, an apparent change in typical iPSC morphology was observed at day 8 in DOX+ cells (inset images marked with red arrow) but not in the case of DOX- cells (inset images marked with red arrow) ((Figure 5.6A; right). The typical round compact cell morphology was lost to attain more irregular and larger cell morphology. Also, unlike a homogenous, tightly packed-cell population, we observed a more loose and heterogeneous cell population. Further, we checked the expression levels of the surface pluripotency marker TRA-1-60 to determine that the cells were undergoing spontaneous differentiation. It was clearly visible that a change in morphology was also a loss of pluripotency in these cells, indicated by the loss of expression of TRA-1-60 in irregularly shaped ZsGreen expressing iPSCs (Figure 5.7A).

Also, the control cells with the vector AAVS1-CopGFP-Puro^R showed a stable morphology on day 8 in the presence of DOX (Figure 5.7B). Therefore, we concluded that by day 8, the cell population expressing shRNA against UTF1 showed many cells with irregular shapes, indicating a loss of typical iPSC characteristics such as the small and round morphology.

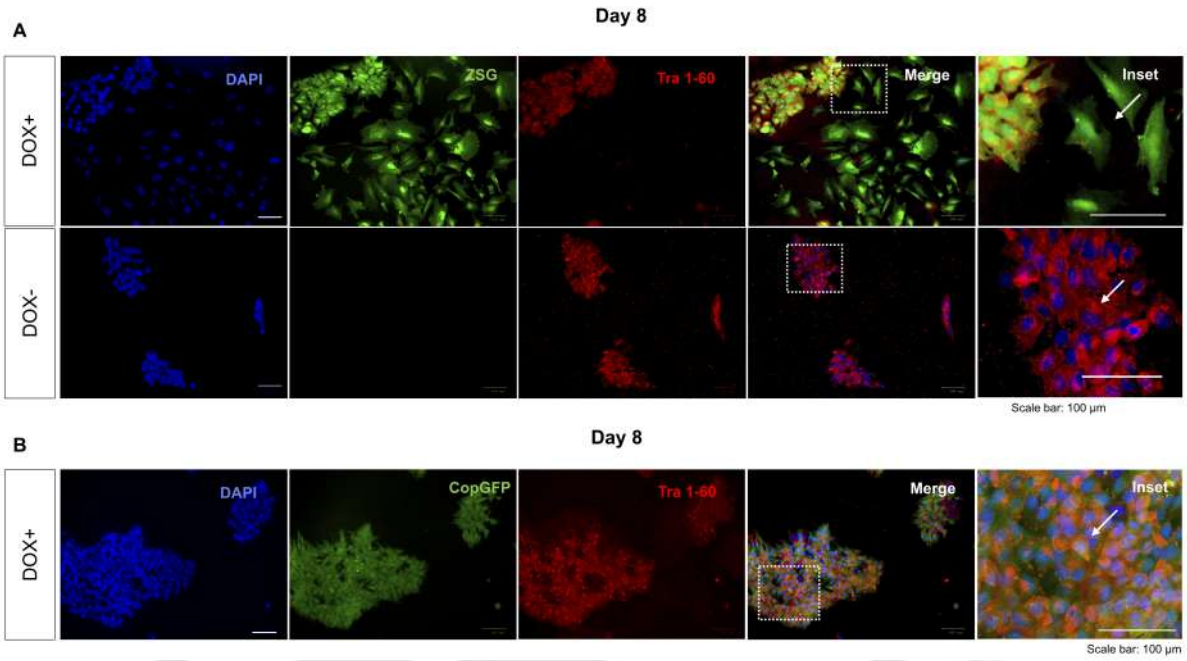


Figure 5.7: Spontaneous differentiation of iPSCs upon depletion in UTF1 levels. (A) Spontaneous differentiation observed in shUTF1-activated iPSCs on Day 8. TRA-1-60 expression is lost in differentiated cells, as indicated by the white arrow. DOX- cells showed expression of TRA-1-60. **(B)** Morphology was maintained in CopGFP (control) iPSCs on Day 8, and TRA-1-60 expression was also observed, as indicated by the white arrow. DOX: Doxycycline; ZsGreen: Zoanthus species green; CopGFP: Copepod green fluorescent protein.

5.2.4 Confirmation of decline in UTF1 expression upon DOX induction

To ensure that the change in cellular morphology was a result of a reduction in UTF1 levels, a Western blot was performed on lysates collected from day 4 and day 8 of control (DOX-) and UTF1 knockdown (DOX+) iPSCs. The Western blot results from day 4 (DOX+) sample showed UTF1 expression similar to the control, indicating no suppression of UTF1 expression. However, the analysis of the day 8 (DOX+) sample showed the absence of UTF1 protein, indicating a reduction in UTF1 expression (Figure 5.6B). Therefore, the loss in iPSC stability was indeed due to a significant decline in the expression of UTF1.

5.2.5 UTF1 knockdown in iPSCs resulted in the decline in expression of core pluripotency markers

The observation that knockdown of UTF1 in iPSCs resulted in differentiation prompted us to investigate the underlying genotypic alterations. Consequently, we delved deeper into the impact of UTF1 deficiency on the pluripotency network. Therefore, we analyzed the effect of this decline in the expression of UTF1 on the core stem cell-specific proteins OCT4 and SOX2 by Western blotting. The results of day 8 DOX+ lysates showed a significant decrease in the expression of the core pluripotency markers, OCT4 and SOX2, compared to the DOX- iPSCs (Figure 5.8A). These findings highlight the critical role of UTF1 in maintaining iPSC pluripotency.

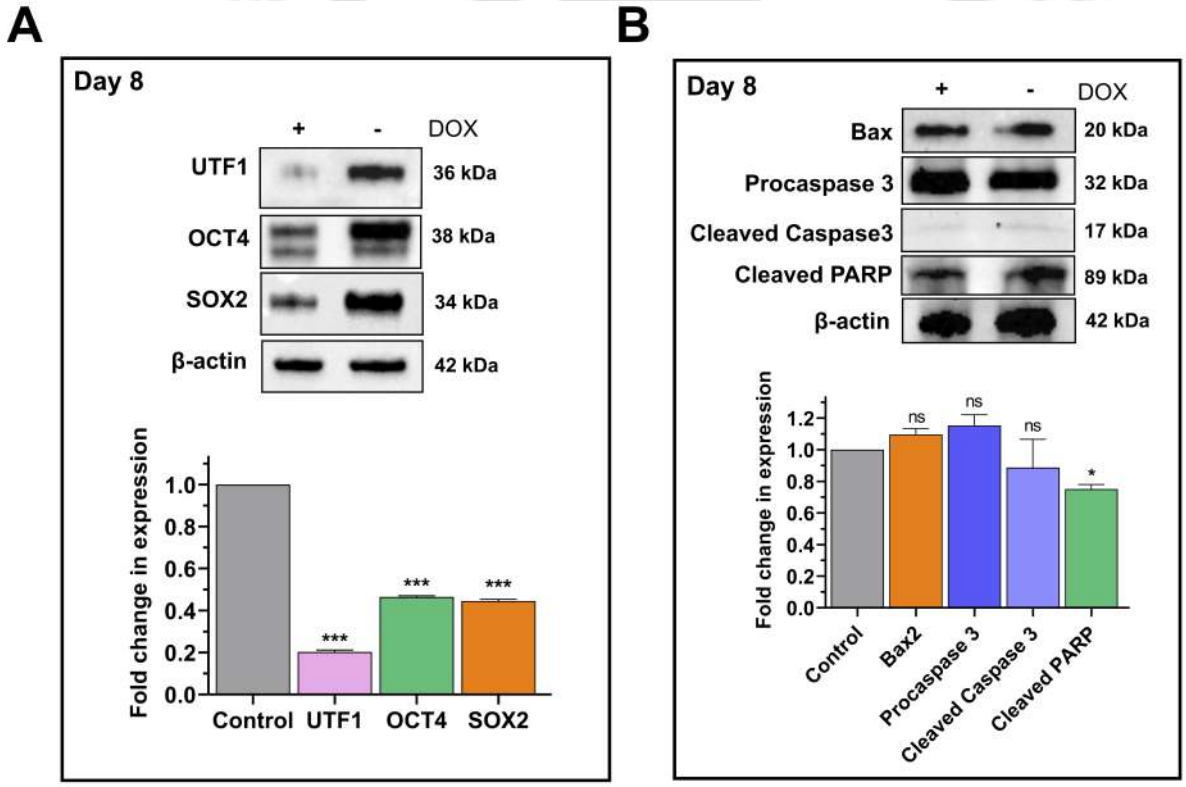


Figure 5.8: Effect of UTF1 knockdown on expression of pluripotency and apoptotic markers. (A) Expression of core pluripotency markers in UTF1 knockdown and UTF1 expressing iPSCs from day 8 protein lysates (Western blot: Left and Image J quantification: Right). **(B)** Expression of apoptotic markers in UTF1 knockdown and UTF1 expressing iPSCs from day 8 protein lysates (Western blot: Left and Image J quantification: Right). DOX: Doxycycline. The quantitative values show the mean fold change±S.D. Statistical analysis was performed using a two-way ANOVA (n=3). $*P<0.05$, $*P<0.01$, $*P<0.001$.

5.2.6 UTF1 knockdown iPSCs did not undergo apoptosis

In contrast to the rapid cell death observed with *UTF1* knockout between day 2 and day 10 post-induction with DOX, cell death was not observed in UTF1 knockdown iPSCs. Further, the lysates from day 8 DOX+ and DOX- UTF1 knockdown cells were analyzed for apoptotic markers to ensure that the cells did not undergo apoptosis. No significant increase in the expression of apoptotic markers, namely Bax2, procaspase3, cleaved caspase3, and PARP was observed, confirming that the cells did not undergo apoptosis (Figure 5.8B).

5.3 Conclusion

In conclusion, this study investigated the importance of human UTF1 in maintaining iPSCs identity. The study revealed that diminishing UTF1 levels by a shRNA system resulted in the instability of human iPSCs. However, unlike a drastic phenotype such as loss of cell viability upon *UTF1* knockout was observed through increase in expression of pro-apoptotic markers, no significant increase was observed in pro-apoptotic markers upon UTF1 downregulation. Remarkably, the loss of UTF1 resulted in spontaneous differentiation in shRNA-expressing cells. Additionally, the cells undergoing spontaneous differentiation were confirmed by TRA-1-60 staining. The control cells showed TRA-1-60 expression, whereas those undergoing spontaneous differentiation were negative for TRA-1-60. Further, to ensure that this phenotype has resulted from the decline in UTF1 expression, a significantly reduced expression level of the human UTF1 protein was observed by Western blotting of day 8 lysates. These observations suggest an essential role of UTF1 in preserving human iPSC identity and stability, emphasizing its importance in maintaining iPSCs.

CHAPTER 6

CONCLUSIONS AND FUTURE PERSPECTIVES

6.1 Conclusions

The work carried out in this thesis begins with the successful generation and characterization of an in-house integration-free human iPSC line, IITGi001-A. Out of 14 iPSC-like clones picked, the clone IITGi001-A was further characterized to confirm its pluripotent identity using different biological assays. In the near future, this pluripotent cell line will be used as a control cell line and create mutant iPSC lines from the same by genome editing methods for basic research, disease modeling, drug screening, and understanding human developmental biology.

To understand the role of human UTF1 in stem cell biology, we have designed and successfully generated a *UTF1* knockout toolbox capable of targeting either of the two human *UTF1* exons to ensure the successful deletion of the *UTF1* gene. We have used this toolkit in our study to generate *UTF1* knockout human fibroblasts and human iPSCs. Notably, while the same knockout kit generated partial *UTF1* knockout in fibroblasts, it resulted in complete deletion of the human *UTF1* gene in iPSCs. We hypothesize that this difference resulted from the lack of UTF1 expression in human fibroblasts, potentially rendering the *UTF1* gene inaccessible to the CRISPR/Cas9 system. Additionally, we have also constructed a shRNA knockdown vector targeting UTF1 expression. The vector has an antibiotic selection marker under an endogenous promoter, providing the benefit of selecting the transfected cells before the induction of shRNA against UTF1. This is an

inducible system and, therefore, offers control over the decline in UTF1 expression over time. This system has been successfully used to suppress UTF1 expression in human iPSCs.

Here, we demonstrated the effect of loss of function studied in establishing and maintaining the iPSC identity. Firstly, we have shown that the targeted deletion of the human *UTF1* gene in fibroblasts exhibited a significant decline in the formation of human iPSCs. Additionally, the few colonies that emerged were unstable beyond the first passage. Secondly, we observed that targeted deletion of the *UTF1* gene in human iPSCs affected their survival significantly, highlighting the importance of the *UTF1* gene in maintaining the iPSC identity. Finally, we observed spontaneous differentiation of iPSCs upon knockdown of human UTF1. These observations collectively demonstrated that UTF1 is important for reprogramming human fibroblasts to iPSCs, and its expression is required to maintain iPSC identity.

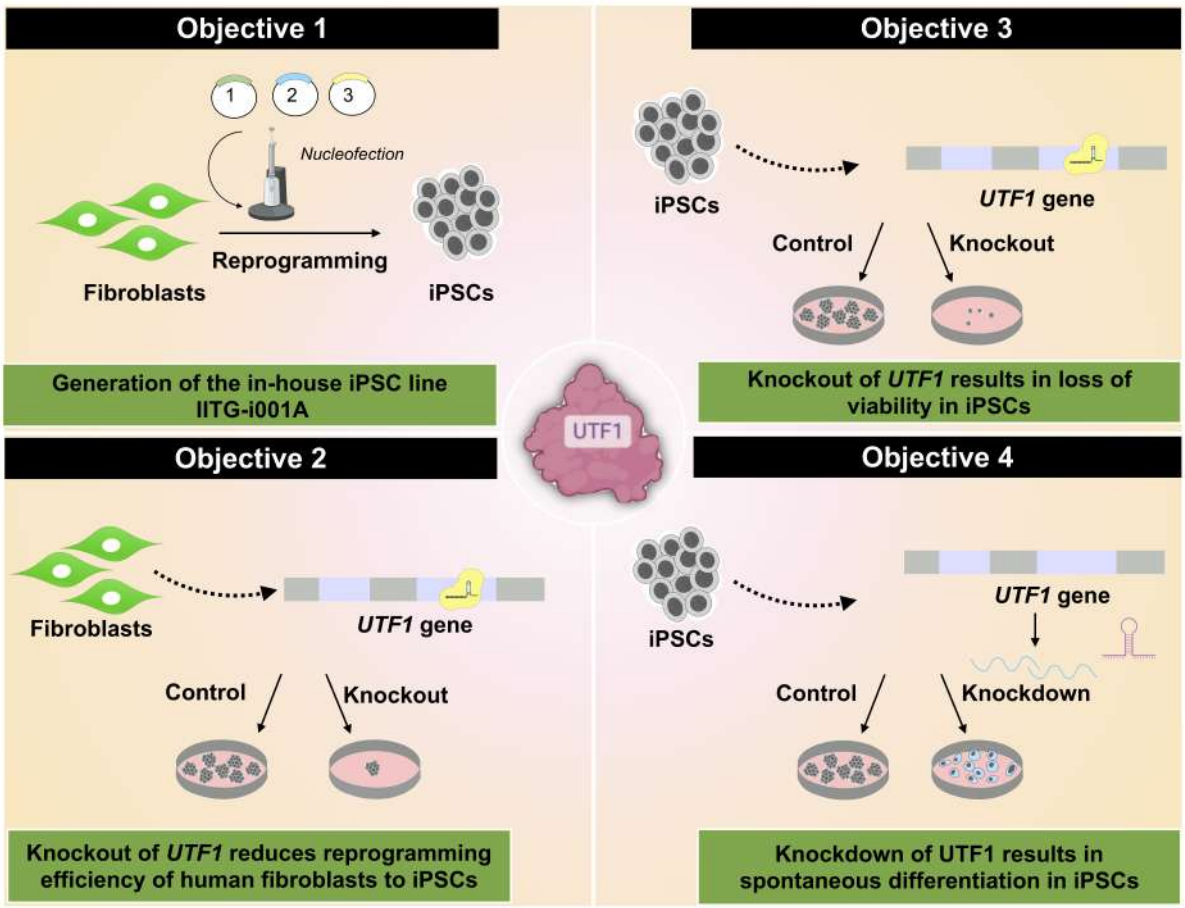


Figure 6.1: Highlights of the work carried out in the thesis.

6.2 Future Perspectives

1. Our in-house characterized stable iPSC line **IITGi001-A** can be used for various **biomedical applications** such as in vitro disease modeling, drug testing and discovery, and personalized regenerative medicine.

2. The human ***UTF1* knockout toolkit** can be used to investigate **its effect** on other desired cell types, including those relevant to **cancer research and human spermatogenesis**.

3. The human **UTF1 knockdown toolkit** comprises an **inducible vector**. This provides **temporal control** over its suppression and can be used to investigate the **stage-specific role of UTF1 in reprogramming**.

4. The knowledge gained regarding the **importance of the *UTF1* gene in the generation and maintenance of iPSCs** can be explored further at the **transcriptome** level to understand its effect on the **pluripotency network**.

5. An inducible human **UTF1 overexpression** vector can be used to investigate the effect of UTF1 at different stages of **reprogramming**. Also, the impact of iPSC stability on UTF1 overexpression and **transcriptome analysis** can further strengthen the understanding of the role of UTF1 in the **human pluripotency network**.

6. Given the high expression of UTF1 in human spermatogonial stem cells, the potential of UTF1 can also be explored in **in vitro fertilization** to better understand spermatogenesis and the fertilization processes.

7. It is widely recognized that murine UTF1 plays a pivotal role in sustaining the poised state of bivalent genes by modulating gene expression via its **interaction with histones**. An intriguing direction for future investigation entails pinpointing the precise core histone(s) responsible for facilitating this interaction. Furthermore, delving into the factors—such as other proteins or modified histones—that regulate this binding would provide deeper insights into the **regulatory mechanisms underlying neurogenesis** and potentially other biological pathways.

8. Studies on the UTF1 protein have revealed that it undergoes post-translational modifications: **phosphorylation and sumoylation**. Understanding the impact of

these modifications is crucial for elucidating the role of UTF1 in cellular functioning. Consequently, investigating mutant forms of UTF1 lacking either sumoylation or phosphorylation can provide valuable insights into its effect on biological processes like **reprogramming, iPSC stability, neurogenesis, and spermatogenesis**.

BIBLIOGRAPHY

[1] M. Afanassieff, Y. Tapponnier, and P. Savatier. Generation of induced pluripotent stem cells in rabbits. *Induced Pluripotent Stem (iPS) Cells: Methods and Protocols*, pages 149–172, 2016.

[2] A. Alam, L. Y. Cohen, S. Aouad, and R.-P. Sékaly. Early activation of caspases during t lymphocyte stimulation results in selective substrate cleavage in nonapoptotic cells. *The Journal of experimental medicine*, 190(12):1879–1890, 1999.

[3] P. Álvarez-Campos, H. García-Castro, E. Emili, A. Pérez-Posada, I. Del Olmo, S. Peron, D. A. Salamanca-Díaz, V. Mason, B. Metzger, A. E. Bely, et al. Annelid adult cell type diversity and their pluripotent cellular origins. *Nature Communications*, 15(1):3194, 2024.

[4] E. Appleton, K. Hong, C. Rodríguez-Caycedo, Y. Tanaka, A. Ashkenazy-Titelman, K. Bhide, C. Rasmussen-Ivey, X. Ambriz-Pena, N. Korover, H. Bai, et al. Derivation of elephant induced pluripotent stem cells. *bioRxiv*, pages 2024–03, 2024.

[5] Q. Bao, A. Morshedi, F. Wang, S. Bhargy, K. Pervushin, W.-P. Yu, and P. Dröge. Utf1 contributes to intergenerational epigenetic inheritance of pluripotency. *Scientific Reports*, 7(1):14612, 2017.

[6] Q. Bao, N. L. Tay, C. Y. Lim, D. H. H. Chua, S. K. Kee, M. Choolani, Y.-H. Loh, S. C. Ng, and C. Chai. Integration-free induced pluripotent stem cells from three endangered southeast asian non-human primate species. *Scientific Reports*, 14(1): 2391, 2024.

[7] A. J. Becker, E. A. McCulloch, and J. E. Till. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. 1963.

[8] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2):315–326, 2006.

[9] S. P. Bharathan, K. V. Manian, S. M. M. Aalam, D. Palani, P. A. Deshpande, M. D.

Pratheesh, A. Srivastava, and S. R. Velayudhan. Systematic evaluation of markers used for the identification of human induced pluripotent stem cells. *Biology Open*, 6(1):100–108, 2017.

[10] S. P. Bharathan, K. Nandy, D. Palani, K. Natarajan, B. George, A. Srivastava, S. R. Velayudhan, et al. Generation of an induced pluripotent stem cell line that mimics the disease phenotypes from a patient with fanconi anemia by conditional complementation. *Stem Cell Research*, 20:54–57, 2017.

[11] S. Bhat, S. P. Kabekkodu, A. Noronha, and K. Satyamoorthy. Biological implications and therapeutic significance of dna methylation regulated genes in cervical cancer. *Biochimie*, 121:298–311, 2016.

[12] A. Bongso and E. H. Lee. Stem cells: from bench to bedside. 2005.

[13] T. C. Bosch. Stem cells in immortal hydra. *Stem Cells: From Hydra to Man*, pages 37–57, 2008.

[14] T. Boveri. Über differenzierung der zellkerne während der furchung des eies von ascaris megalocephala. *Anatomischer Anzeiger*, 2:688–693, 1887.

[15] T. Boveri. Ueber die entstehung des gegensatzes zwischen den geschlechtszellen und den somatischen zellen bei ascaris megalocephala, nebst bemerkungen zur entwicklungsgeschichte der nematoden: Vorgetragen am 15. november 1892. *(No Title)*, 1892.

[16] R. Briggs and T. J. King. Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Proceedings of the National Academy of Sciences*, 38(5): 455–463, 1952.

[17] Y. Buganim, D. A. Faddah, A. W. Cheng, E. Itskovich, S. Markoulaki, K. Ganz, S. L. Klemm, A. Van Oudenaarden, and R. Jaenisch. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, 150(6):1209–1222, 2012.

[18] Y. Buganim, D. A. Faddah, and R. Jaenisch. Mechanisms and models of somatic cell reprogramming. *Nature Reviews Genetics*, 14(6):427–439, 2013.

[19] B.-K. Chou, P. Mali, X. Huang, Z. Ye, S. N. Dowey, L. Resar, C. Zou, Y. A. Zhang, J. Tong, and L. Cheng. Efficient human ips cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. *Cell research*, 21(3):518–529, 2011.

[20] L. S. Corley and M. D. Lavine. A review of insect stem cell types. In *Seminars in cell & developmental biology*, volume 17, pages 510–517. Elsevier, 2006.

[21] J. F. Correa-Vázquez, F. Juárez-Vicente, P. García-Gutiérrez, S. V. Barysch, F. Melchior, and M. García-Domínguez. The sumo proteome of proliferating and neuronal-differentiating cells reveals utf1 among key sumo targets involved in neurogenesis. *Cell Death & Disease*, 12(4):305, 2021.

[22] A.-M. Courtot, A. Magniez, N. Oudrhiri, O. Féraud, J. Bacci, E. Gobbo, S. Proust, A. G. Turhan, and A. Bennaceur-Griscelli. Morphological analysis of human induced pluripotent stem cells during induced differentiation and reverse programming. *BioResearch open access*, 3(5):206–216, 2014.

[23] E. V. Cowdry. *Special Cytology: The Form and Functions of the Cell in Health and Disease; a Textbook for Students of Biology and Medicine...*, volume 2. PB Hoeber, Incorporated, 1928.

[24] S. M. C. de Sousa Lopes, S. van den Driesche, R. L. Carvalho, J. Larsson, B. Eggen, M. A. Surani, and C. L. Mummery. Altered primordial germ cell migration in the absence of transforming growth factor β signaling via alk5. *Developmental biology*, 284(1):194–203, 2005.

[25] M. Déjosez, A. Marin, G. M. Hughes, A. E. Morales, C. Godoy-Parejo, J. L. Gray, Y. Qin, A. A. Singh, H. Xu, J. Juste, et al. Bat pluripotent stem cells reveal unusual entanglement between host and viruses. *Cell*, 186(5):957–974, 2023.

[26] L. C. Dutton, J. Dudhia, D. J. Guest, and D. J. Connolly. Inducing pluripotency in the domestic cat (felis catus). *Stem Cells and Development*, 28(19):1299–1309, 2019.

[27] F. Edlich, S. Banerjee, M. Suzuki, M. M. Cleland, D. Arnoult, C. Wang, A. Neutzner, N. Tjandra, and R. J. Youle. Bcl-xl retrotranslocates bax from the mitochondria into the cytosol. *Cell*, 145(1):104–116, 2011.

[28] A. V. Ereskovsky. *The comparative embryology of sponges*. Springer Science & Business Media, 2010.

[29] M. J. Evans and M. H. Kaufman. Establishment in culture of pluripotential cells from mouse embryos. *nature*, 292(5819):154–156, 1981.

[30] V. Falanga. Stem cells in tissue repair and regeneration. *Journal of Investigative Dermatology*, 132(6):1538–1541, 2012.

[31] S. Fernandes, S. Tembe, S. Singh, S. Vardhan, V. Nair, V. Kale, and L. Limaye. Development and characterization of human ipsc line nccsi004-a from umbilical cord blood (ucb) derived cd34+ cells obtained from donor belonging to indian ethnic population. *Stem Cell Research*, 35:101392, 2019.

[32] C. Ferrario, M. Sugni, I. M. Somorjai, and L. Ballarin. Beyond adult stem cells: dedifferentiation as a unifying mechanism underlying regeneration in invertebrate deuterostomes. *Frontiers in Cell and Developmental Biology*, 8:587320, 2020.

[33] A. Forte, M. T. Schettino, M. Finicelli, M. Cipollaro, N. Colacurci, L. Cobellis, and U. Galderisi. Expression pattern of stemness-related genes in human endometrial and endometriotic tissues. *Molecular medicine*, 15:392–401, 2009.

[34] A. Fukushima, A. Okuda, M. Nishimoto, N. Seki, T.-a. Hori, and M. Muramatsu. Characterization of functional domains of an embryonic stem cell coactivator utf1 which are conserved and essential for potentiation of atf-2 activity. *Journal of Biological Chemistry*, 273(40):25840–25849, 1998.

[35] N. Funayama. The stem cell system in demosponges: suggested involvement of two types of cells: archeocytes (active stem cells) and choanocytes (food-entrapping flagellated cells). *Development genes and evolution*, 223:23–38, 2013.

[36] I. Gallego Romero, B. J. Pavlovic, I. Hernando-Herraez, X. Zhou, M. C. Ward, N. E. Banovich, C. L. Kagan, J. E. Burnett, C. H. Huang, A. Mitrano, et al. A panel of induced pluripotent stem cells from chimpanzees: a resource for comparative functional genomics. *Elife*, 4:e07103, 2015.

[37] C. Galonska, Z. D. Smith, and A. Meissner. In vivo and in vitro dynamics of undifferentiated embryonic cell transcription factor 1. *Stem cell reports*, 2(3): 245–252, 2014.

[38] J. Gao, S. Petraki, X. Sun, L. A. Brooks, T. J. Lynch, C.-L. Hsieh, R. Elteriefi, Z. Lorenzana, V. Punj, J. F. Engelhardt, et al. Derivation of induced pluripotent stem cells from ferret somatic cells. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 318(4):L671–L683, 2020.

[39] J. Garnier, J.-F. Gibrat, and B. Robson. [32] gor method for predicting protein secondary structure from amino acid sequence. In *Methods in enzymology*, volume 266, pages 540–553. Elsevier, 1996.

[40] A. Gaspar-Maia, A. Alajem, F. Polesso, R. Sridharan, M. J. Mason, A. Heidersbach, J. Ramalho-Santos, M. T. McManus, K. Plath, E. Meshorer, et al. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *nature*, 460(7257): 863–868, 2009.

[41] C. Geourjon and G. Deleage. Sopma: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, 11(6):681–684, 1995.

[42] J. Geuder, L. E. Wange, A. Janjic, J. Radmer, P. Janssen, J. W. Bagnoli, S. Müller, A. Kaul, M. Ohnuki, and W. Enard. A non-invasive method to generate induced pluripotent stem cells from primate urine. *Scientific reports*, 11(1):3516, 2021.

[43] M. Gherghiceanu, L. Barad, A. Novak, I. Reiter, J. Itskovitz-Eldor, O. Binah, and L. Popescu. Cardiomyocytes derived from human embryonic and induced pluripotent stem cells: comparative ultrastructure. *Journal of cellular and molecular medicine*, 15(11):2539–2551, 2011.

[44] J. Graumann, N. C. Hubner, J. B. Kim, K. Ko, M. Moser, C. Kumar, J. Cox, H. Schöler, and M. Mann. Stable isotope labeling by amino acids in cell culture (silac) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Molecular & Cellular Proteomics*, 7(4):672–683, 2008.

[45] S. Guenin, M. Mouallif, R. Deplus, X. Lampe, N. Krusy, E. Calonne, K. Delbecque, F. Kridelka, F. Fuks, M. M. Ennaji, et al. Aberrant promoter methylation and expression of utf1 during cervical carcinogenesis. 2012.

[46] J. B. Gurdon. Adult frogs derived from the nuclei of single somatic cells. *Developmental biology*, 4(2):256–273, 1962.

[47] V. Gursky, O. Krasnova, J. Sopova, A. Kovaleva, K. Kulakova, O. Tikhonova, and I. Neganova. How morphology of the human pluripotent stem cells determines the selection of the best clone. 2023.

[48] E. Haeckel. *Natürliche schöpfungsgeschichte*. Deutsches Textarchiv (Kernkorpus), 1868.

[49] E. Haeckel. Anthropogenie, 1st edn (leipzig: Wilhelmengelmann). 1874.

[50] E. H. P. A. Haeckel. *Anthropogenie*. W. Engelmann, 1877.

[51] V. Haecker. Die kerntheilungsvorgänge bei der mesodermund entodermbildung von cyclops. *Archiv für mikroskopische Anatomie*, 39(1):556–581, 1892.

[52] R. Heidstra and S. Sabatini. Plant and animal stem cells: similar yet different. *Nature Reviews Molecular Cell Biology*, 15(5):301–312, 2014.

[53] J. M. Hinz, M. F. Laughery, and J. J. Wyrick. Nucleosomes inhibit cas9 endonuclease activity in vitro. *Biochemistry*, 54(48):7063–7066, 2015.

[54] N. D. Holland and I. M. Somorjai. Serial blockface sem suggests that stem cells may participate in adult notochord growth in an invertebrate chordate, the bahamas lancelet. *EvoDevo*, 11:1–15, 2020.

[55] T. W. Holstein. The hydra stem cell system–revisited. *Cells & Development*, page 203846, 2023.

[56] J. Hong, H. Kim, C. Park, M. Son, K. W. Lee, and Y. J. Kim. Identification of undifferentiated embryonic cell transcription factor 1 as a potential substrate of carboxyl-terminal domain small phosphatases. *Journal of the Korean Chemical Society*, 59(2):188–193, 2015.

[57] K. Hu, L. Li, Y. Liao, and M. Liang. Lncrna gm2044 highly expresses in spermatocyte and inhibits utf1 translation by interacting with utf1 mrna. *Genes & genomics*, 40:781–787, 2018.

[58] T. Humphreys, K. Weiser, A. Arimoto, A. Sasaki, G. Uenishi, B. Fujimoto, T. Kawashima, K. Taparra, J. Molnar, N. Satoh, et al. Ancestral stem cell reprogramming genes active in hemichordate regeneration. *Frontiers in ecology and evolution*, 10:769433, 2022.

[59] N. Ivanova, R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I. R. Lemischka. Dissecting self-renewal in stem cells with rna interference. *Nature*, 442(7102):533–538, 2006.

[60] J. Jia, X. Zheng, G. Hu, K. Cui, J. Zhang, A. Zhang, H. Jiang, B. Lu, J. Yates, C. Liu, et al. Regulation of pluripotency and self-renewal of escs through epigenetic-threshold modulation and mrna pruning. *Cell*, 151(3):576–589, 2012.

[61] D. T. Jones. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology*, 292(2):195–202, 1999.

[62] E. M. Kallimasioti-Pazi, K. Thelakkad Chathoth, G. C. Taylor, A. Meynert, T. Ballinger, M. J. Kelder, S. Lalevée, I. Sanli, R. Feil, and A. J. Wood. Heterochromatin delays crispr-cas9 mutagenesis but does not influence the outcome of mutagenic dna repair. *PLoS Biology*, 16(12):e2005595, 2018.

[63] S. D. Kasowitz, M. Luo, J. Ma, N. A. Leu, and P. J. Wang. Embryonic lethality and defective male germ cell development in mice lacking utf1. *Scientific reports*, 7(1): 17259, 2017.

[64] E. Kingham and R. O. Oreffo. Embryonic and induced pluripotent stem cells: understanding, creating, and exploiting the nano-niche for regenerative medicine. *ACS nano*, 7(3):1867–1881, 2013.

[65] S. C. Knight, L. Xie, W. Deng, B. Guglielmi, L. B. Witkowsky, L. Bosanac, E. T. Zhang, M. El Beheiry, J.-B. Masson, M. Dahan, et al. Dynamics of crispr-cas9 genome interrogation in living cells. *Science*, 350(6262):823–826, 2015.

[66] S. R. Knott, A. R. Maceli, N. Erard, K. Chang, K. Marran, X. Zhou, A. Gordon, O. El Demerdash, E. Wagenblast, S. Kim, et al. A computational algorithm to

predict shrna potency. *Molecular cell*, 56(6):796–807, 2014.

[67] S. M. Kooistra, R. P. Thummer, and B. J. Eggen. Characterization of human utf1, a chromatin-associated protein with repressor activity expressed in pluripotent cells. *Stem cell research*, 2(3):211–218, 2009.

[68] S. M. Kooistra, V. van den Boom, R. P. Thummer, F. Johannes, R. Wardenaar, B. M. Tesson, L. M. Veenhoff, F. Fusetti, L. P. O'Neill, B. M. Turner, et al. Undifferentiated embryonic cell transcription factor 1 regulates esc chromatin organization and gene expression. *Stem cells*, 28(10):1703–1714, 2010.

[69] D. M. Kristensen, J. E. Nielsen, N. E. Skakkebaek, N. Graem, G. K. Jacobsen, E. R.-D. Meyts, and H. Leffers. Presumed pluripotency markers utf-1 and rex-1 are expressed in human adult testes and germ cell neoplasms. *Human reproduction*, 23(4):775–782, 2008.

[70] L. G. Lajtha. Stem cell concepts. *Differentiation*, 14(1-3):23–33, 1979.

[71] A. I. Laskowski and P. S. Knoepfler. Myc binds the pluripotency factor utf1 through the basic-helix-loop-helix leucine zipper domain. *Biochemical and biophysical research communications*, 435(4):551–556, 2013.

[72] D. Li, X. Shu, P. Zhu, and D. Pei. Chromatin accessibility dynamics during cell fate reprogramming. *EMBO reports*, 22(2):e51644, 2021.

[73] M. Li and J. C. I. Belmonte. Ground rules of the pluripotency gene regulatory network. *Nature Reviews Genetics*, 18(3):180–191, 2017.

[74] M. Li and J. C. Izpisua Belmonte. Deconstructing the pluripotency gene regulatory network. *Nature cell biology*, 20(4):382–392, 2018.

[75] W. Li, W. Wei, S. Zhu, J. Zhu, Y. Shi, T. Lin, E. Hao, A. Hayek, H. Deng, and S. Ding. Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. *Cell stem cell*, 4(1):16–19, 2009.

[76] C.-H. Lin, C.-H. Yang, and Y.-R. Chen. Utf1 deficiency promotes retinoic acid-induced neuronal differentiation in p19 embryonal carcinoma cells. *The international Journal of Biochemistry & Cell Biology*, 44(2):350–357, 2012.

[77] A. Liu, L. Cheng, J. Du, Y. Peng, R. W. Allan, L. Wei, J. Li, and D. Cao. Diagnostic utility of novel stem cell markers sall4, oct4, nanog, sox2, utf1, and tcl1 in primary mediastinal germ cell tumors. *The American journal of surgical pathology*, 34(5): 697–706, 2010.

[78] H. Liu, F. Zhu, J. Yong, P. Zhang, P. Hou, H. Li, W. Jiang, J. Cai, M. Liu, K. Cui, et al. Generation of induced pluripotent stem cells from adult rhesus monkey

fibroblasts. *Cell stem cell*, 3(6):587–590, 2008.

[79] M. Liu, L. Zhao, Z. Wang, H. Su, T. Wang, G. Yang, L. Chen, B. Wu, G. Zhao, J. Guo, et al. Generation of sheep induced pluripotent stem cells with defined dox-inducible transcription factors via piggybac transposition. *Frontiers in Cell and Developmental Biology*, 9:785055, 2021.

[80] A. A. Mack, S. Kroboth, D. Rajesh, and W. B. Wang. Generation of induced pluripotent stem cells from cd34+ cells across blood drawn from multiple donors with non-integrating episomal vectors. *PloS one*, 6(11):e27956, 2011.

[81] V. Magliocca, S. Petrini, T. Franchin, R. Borghi, A. Niceforo, Z. Abbaszadeh, E. Bertini, and C. Compagnucci. Identifying the dynamics of actin and tubulin polymerization in ipscs and in ipsc-derived neurons. *Oncotarget*, 8(67):111096, 2017.

[82] K. V. Manian, S. P. Bharathan, M. Maddali, V. M. Srivastava, A. Srivastava, and S. R. Velayudhan. Generation of an integration-free ipsc line (cscri005-a) from erythroid progenitor cells of a healthy indian male individual. *Stem Cell Research*, 29:148–151, 2018.

[83] G. R. Martin. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences*, 78(12):7634–7638, 1981.

[84] U. Marx, F. Schenk, J. Behrens, U. Meyr, P. Wanek, W. Zang, R. Schmitt, O. Brüstle, M. Zenke, and F. Klocke. Automatic production of induced pluripotent stem cells. *Procedia CIRP*, 5:2–6, 2013.

[85] M. A. Melone, M. Giulano, T. Squillaro, N. Alessio, F. Casale, E. Mattioli, M. Cipollaro, A. Giordano, and U. Galderisi. Genes involved in regulation of stem cell properties: studies on their expression in a small cohort of neuroblastoma patients. *Cancer biology & therapy*, 8(13):1300–1306, 2009.

[86] T. S. Mikkelsen, J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B. E. Bernstein, R. Jaenisch, E. S. Lander, and A. Meissner. Dissecting direct reprogramming through integrative genomic analysis. *Nature*, 454(7200):49–55, 2008.

[87] J. Miyoshi and Y. Takai. Structural and functional associations of apical junctions with cytoskeleton. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1778(3): 670–691, 2008.

[88] A. Morshedi, M. Soroush Noghabi, and P. Dröge. Use of utf1 genetic control elements as ipsc reporter. *Stem Cell Reviews and Reports*, 9:523–530, 2013.

[89] M. Mouallif, A. Albert, M. Zeddou, M. M. Ennaji, P. Delvenne, and S. Guenin. Expression profile of undifferentiated cell transcription factor 1 in normal and cancerous human epithelia. *International Journal of Experimental Pathology*, 95 (4):251–259, 2014.

[90] W. E. Müller. The stem cell concept in sponges (porifera): metazoan traits. In *Seminars in cell & developmental biology*, volume 17, pages 481–491. Elsevier, 2006.

[91] M. Myohara. What role do annelid neoblasts play? a comparison of the regeneration patterns in a neoblast-bearing and a neoblast-lacking enchytraeid oligochaete. *PLoS One*, 7(5):e37319, 2012.

[92] R. Nagasaka, M. Matsumoto, M. Okada, H. Sasaki, K. Kanie, H. Kii, T. Uozumi, Y. Kiyota, H. Honda, and R. Kato. Visualization of morphological categories of colonies for monitoring of effect on induced pluripotent stem cell culture status. *Regenerative Therapy*, 6:41–51, 2017.

[93] T. Nakagawa, K. Matsusaka, K. Misawa, S. Ota, K. Takane, M. Fukuyo, B. Rahmutulla, K.-i. Shinohara, N. Kunii, D. Sakurai, et al. Frequent promoter hypermethylation associated with human papillomavirus infection in pharyngeal cancer. *Cancer letters*, 407:21–31, 2017.

[94] N. Neuhaus, J. Yoon, N. Terwort, S. Kliesch, J. Seggewiss, A. Huge, R. Voss, S. Schlatt, R. Grindberg, and H. Schöler. Single-cell gene expression analysis reveals diversity among human spermatogonia. *MHR: Basic science of reproductive medicine*, 23(2):79–90, 2017.

[95] M. Nishimoto, A. Fukushima, A. Okuda, and M. Muramatsu. The gene for the embryonic stem cell coactivator utf1 carries a regulatory element which selectively interacts with a complex composed of oct-3/4 and sox-2. *Molecular and cellular biology*, 1999.

[96] M. Nishimoto, A. Fukushima, S. Miyagi, Y. Suzuki, S. Sugano, Y. Matsuda, T.-a. Hori, M. Muramatsu, and A. Okuda. Structural analyses of the utf1 gene encoding a transcriptional coactivator expressed in pluripotent embryonic stem cells. *Biochemical and Biophysical Research Communications*, 285(4):945–953, 2001.

[97] M. Nishimoto, S. Miyagi, T. Katayanagi, M. Tomioka, M. Muramatsu, and A. Okuda. The embryonic octamer factor 3/4 displays distinct dna binding specificity from those of other octamer factors. *Biochemical and biophysical research*

communications, 302(3):581–586, 2003.

[98] M. Nishimoto, S. Miyagi, T. Yamagishi, T. Sakaguchi, H. Niwa, M. Muramatsu, and A. Okuda. Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the utf1 locus. *Molecular and cellular biology*, 25(12):5084–5094, 2005.

[99] M. Nishimoto, M. Katano, T. Yamagishi, T. Hishida, M. Kamon, A. Suzuki, M. Hirasaki, Y. Nabeshima, Y.-i. Nabeshima, Y. Katsura, et al. In vivo function and evolution of the eutherian-specific pluripotency marker utf1. *PLoS One*, 8(7): e68119, 2013.

[100] K. Okita, Y. Matsumura, Y. Sato, A. Okada, A. Morizane, S. Okamoto, H. Hong, M. Nakagawa, K. Tanabe, K.-i. Tezuka, et al. A more efficient method to generate integration-free human ips cells. *Nature methods*, 8(5):409–412, 2011.

[101] K. Okita, T. Yamakawa, Y. Matsumura, Y. Sato, N. Amano, A. Watanabe, N. Goshima, and S. Yamanaka. An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem cells*, 31(3):458–466, 2013.

[102] A. Okuda, A. Fukushima, M. Nishimoto, A. Orimo, T. Yamagishi, Y. Nabeshima, M. Kuro-o, Y.-i. Nabeshima, K. Boon, M. Keaveney, et al. Utf1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *The EMBO journal*, 17(7):2019–2032, 1998.

[103] J. Organista-Nava, Y. Gómez-Gómez, and P. Gariglio. Embryonic stem cell-specific signature in cervical cancer. *Tumor Biology*, 35:1727–1738, 2014.

[104] G. Pantazis, P. N. Harter, D. Capper, P. Kohlhof, M. Mittelbronn, and J. Schittenhelm. The embryonic stem cell factor utf1 serves as a reliable diagnostic marker for germinomas. *Pathology*, 46(3):225–229, 2014.

[105] A. Pappenheim. Ueber entwickelung und ausbildung der erythroblasten. *Archiv für pathologische Anatomie und Physiologie und für klinische Medicin*, 145:587–643, 1896.

[106] Y. Pei, G. Sierra, R. Sivapatham, A. Swistowski, M. S. Rao, and X. Zeng. A platform for rapid generation of single and multiplexed reporters in human ipsc lines. *Scientific reports*, 5(1):9205, 2015.

[107] K. Pfannkuche, A. Fatima, M. K. Gupta, R. Dieterich, and J. Hescheler. Initial colony morphology-based selection for ips cells derived from adult fibroblasts is substantially improved by temporary utf1-based selection. *PloS one*, 5(3):e9580,

2010.

[108] D. Pfister, K. De Mulder, V. Hartenstein, G. Kuales, G. Borgonie, F. Marx, J. Morris, and P. Ladurner. Flatworm stem cells and the germ line: developmental and evolutionary implications of macvasa expression in macrostomum lignano. *Developmental biology*, 319(1):146–159, 2008.

[109] V. V. Pillai, P. P. Koganti, T. G. Kei, S. Gurung, W. R. Butler, and V. Selvaraj. Efficient induction and sustenance of pluripotent stem cells from bovine somatic cells. *Biology open*, 10(10):bio058756, 2021.

[110] C. Potten and B. Lord. Stem cells: their identification and characterization. *Edinburgh: Churchill Livingston*, pages 200–232, 1983.

[111] K. Raina, C. Dey, M. Thool, S. Sudhagar, and R. P. Thummer. An insight into the role of utf1 in development, stem cells, and cancer. *Stem Cell Reviews and Reports*, 17:1280–1293, 2021.

[112] M. Ramalho-Santos and H. Willenbring. On the origin of the term "stem cell". *Cell stem cell*, 1(1):35–38, 2007.

[113] J. Ramon-Mateu, S. T. Ellison, T. E. Angelini, and M. Q. Martindale. Regeneration in the ctenophore mnemiopsis leidyi occurs in the absence of a blastema, requires cell division, and is temporally separable from wound healing. *BMC biology*, 17: 1–25, 2019.

[114] S. M. Rhind, J. E. Taylor, P. A. De Sousa, T. J. King, M. McGarry, and I. Wilmut. Human cloning: can it be made safe? *Nature Reviews Genetics*, 4(11):855–864, 2003.

[115] J. C. Rink. Stem cell systems and regeneration in planaria. *Development genes and evolution*, 223:67–84, 2013.

[116] D. J. Rodda, J.-L. Chew, L.-H. Lim, Y.-H. Loh, B. Wang, H.-H. Ng, and P. Robson. Transcriptional regulation of nanog by oct4 and sox2. *Journal of Biological Chemistry*, 280(26):24731–24737, 2005.

[117] R. Sablowski. Stem cells in plants and animals. *Nature education*, 3(4):9, 2010.

[118] B. A. Schwarz, M. Cetinbas, K. Clement, R. M. Walsh, S. Cheloufi, H. Gu, J. Langkabel, A. Kamiya, H. Schorle, A. Meissner, et al. Prospective isolation of poised ipsc intermediates reveals principles of cellular reprogramming. *Cell stem cell*, 23(2): 289–305, 2018.

[119] C. Schwarzer, M. Siatkowski, M. J. Pfeiffer, N. Baeumer, H. Drexler, B. Wang, G. Fuellen, and M. Boiani. Maternal age effect on mouse oocytes: new biological

insight from proteomic analysis. *Reproduction*, 148(1):55–72, 2014.

[120] J. Semmler, M. Lehmann, K. Pfannkuche, M. Reppel, J. Hescheler, and F. Nguemo. Functional expression and regulation of hyperpolarization-activated cyclic nucleotide-gated channels (hcn) in mouse ips cell-derived cardiomyocytes after utf1-neo selection. *Cellular Physiology and Biochemistry*, 34(4):1199–1215, 2014.

[121] J. M. Slack. What is a stem cell? *Wiley Interdisciplinary Reviews: Developmental Biology*, 7(5):e323, 2018.

[122] D. Solter. Mammalian cloning: advances and limitations. *Nature Reviews Genetics*, 1(3):199–207, 2000.

[123] D. Solter. From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *Nature Reviews Genetics*, 7(4):319–327, 2006.

[124] M. Stauske, I. Rodriguez Polo, W. Haas, D. Y. Knorr, T. Borchert, K. Streckfuss-Bömeke, R. Dressel, I. Bartels, M. Tiburcy, W.-H. Zimmermann, et al. Non-human primate ipsc generation, cultivation, and cardiac differentiation under chemically defined conditions. *Cells*, 9(6):1349, 2020.

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

[126] K. Takahashi and S. Yamanaka. A developmental framework for induced pluripotency. *Development*, 142(19):3274–3285, 2015.

[127] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, 131(5):861–872, 2007.

[128] S. M. Tan, S. T. Wang, H. Hentze, and P. Dröge. A utf1-based selection system for stable homogeneously pluripotent human embryonic stem cell cultures. *Nucleic acids research*, 35(18):e118, 2007.

[129] W. W. Tang, S. Dietmann, N. Irie, H. G. Leitch, V. I. Floros, C. R. Bradshaw, J. A. Hackett, P. F. Chinnery, and M. A. Surani. A unique gene regulatory network resets the human germline epigenome for development. *Cell*, 161(6):1453–1467, 2015.

[130] V. Thamodaran, S. Rani, and S. R. Velayudhan. Gene editing in human induced pluripotent stem cells using doxycycline-inducible crispr-cas9 system. In *Induced Pluripotent Stem (iPS) Cells: Methods and Protocols*, pages 755–773. Springer, 2021.

[131] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones. Embryonic stem cell lines derived from human blastocysts. *science*, 282(5391):1145–1147, 1998.

[132] R. Thummer, L. Drenth-Diephuis, and B. Eggen. Constitutive gfp-utf1 expression interferes with es and ec cell differentiation. *The Journal of Stem Cell Research and Therapy*, 2(127):1–7, 2012.

[133] R. P. Thummer, L. J. Drenth-Diephuis, K. E. Carney, and B. J. Eggen. Functional characterization of single-nucleotide polymorphisms in the human undifferentiated embryonic-cell transcription factor 1 gene. *DNA and Cell Biology*, 29(5): 241–248, 2010.

[134] J. E. Till and E. A. McCulloch. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation research*, 14(2):213–222, 1961.

[135] J. E. Till, E. A. McCulloch, and L. Siminovitch. A stochastic model of stem cell proliferation, based on the growth of spleen colony-forming cells. *Proceedings of the National Academy of Sciences*, 51(1):29–36, 1964.

[136] S. Tiozzo, F. D. Brown, and A. W. De Tomaso. Regeneration and stem cells in ascidians. *Stem cells: From Hydra to man*, pages 95–112, 2008.

[137] K. Tzelepis, H. Koike-Yusa, E. De Braekeleer, Y. Li, E. Metzakopian, O. M. Dovey, A. Mupo, V. Grinkevich, M. Li, M. Mazan, et al. A crispr dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. *Cell reports*, 17(4):1193–1205, 2016.

[138] M. P. van Bragt, H. L. Roepers-Gajadien, C. M. Korver, J. Bogerd, A. Okuda, B. J. Eggen, D. G. de Rooij, and A. M. van Pelt. Expression of the pluripotency marker utf1 is restricted to a subpopulation of early a spermatogonia in rat testis. *Reproduction*, 136(1):33–40, 2008.

[139] V. van den Boom, S. M. Kooistra, M. Boesjes, B. Geverts, A. B. Houtsmuller, K. Monzen, I. Komuro, J. Essers, L. J. Drenth-Diephuis, and B. J. Eggen. Utf1 is a chromatin-associated protein involved in es cell differentiation. *The Journal of cell biology*, 178(6):913–924, 2007.

[140] D. Van Hoof, J. Muñoz, S. R. Braam, M. W. Pinkse, R. Linding, A. J. Heck, C. L. Mummery, and J. Krijgsveld. Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell stem cell*, 5(2):214–226, 2009.

[141] V. Vanni, F. Caicci, A. Peronato, G. Martello, D. Asnicar, F. Gasparini, L. Ballarin, and L. Manni. New permanent stem cell niche for development and regeneration

in a chordate. *BioRxiv*, pages 2023–05, 2023.

[142] S. A. Verkuijl and M. G. Rots. The influence of eukaryotic chromatin state on crispr–cas9 editing efficiencies. *Current opinion in biotechnology*, 55:68–73, 2019.

[143] K. von Kopylow, H. Staege, W. Schulze, H. Will, and C. Kirchhoff. Fibroblast growth factor receptor 3 is highly expressed in rarely dividing human type a spermatogonia. *Histochemistry and cell biology*, 138:759–772, 2012.

[144] C. H. Waddington. Genetic assimilation of the bithorax phenotype. *Evolution*, pages 1–13, 1956.

[145] S. Wakao, M. Kitada, Y. Kuroda, F. Ogura, T. Murakami, A. Niwa, and M. Dezawa. Morphologic and gene expression criteria for identifying human induced pluripotent stem cells. *PloS one*, 7(12):e48677, 2012.

[146] P. Wang, J. Li, R. W. Allan, C. C. Guo, Y. Peng, and D. Cao. Expression of utf1 in primary and metastatic testicular germ cell tumors. *American journal of clinical pathology*, 134(4):604–612, 2010.

[147] A. Weismann. *Die Continuität des Keimplasma's als grundlage einer Theorie der Vererbung*. Verlag von Gustav Fischer, 1885.

[148] F. D. West, S. L. Terlouw, D. J. Kwon, J. L. Mumaw, S. K. Dhara, K. Hasneen, J. R. Dobrinsky, and S. L. Stice. Porcine induced pluripotent stem cells produce chimeric offspring. *Stem cells and development*, 19(8):1211–1220, 2010.

[149] M. J. Wheelock, Y. Shintani, M. Maeda, Y. Fukumoto, and K. R. Johnson. Cadherin switching. *Journal of cell science*, 121(6):727–735, 2008.

[150] D. J. Whitworth, D. A. Ovchinnikov, and E. J. Wolvetang. Generation and characterization of lif-dependent canine induced pluripotent stem cells from adult dermal fibroblasts. *Stem cells and development*, 21(12):2288–2297, 2012.

[151] D. J. Whitworth, D. A. Ovchinnikov, J. Sun, P. R. Fortuna, and E. J. Wolvetang. Generation and characterization of leukemia inhibitory factor-dependent equine induced pluripotent stem cells from adult dermal fibroblasts. *Stem cells and development*, 23(13):1515–1523, 2014.

[152] D. J. Whitworth, I. J. Limnios, M.-E. Gauthier, P. Weeratunga, D. A. Ovchinnikov, G. Baillie, S. M. Grimmond, J. A. M. Graves, and E. J. Wolvetang. Platypus induced pluripotent stem cells: the unique pluripotency signature of a monotreme. *Stem Cells and Development*, 28(3):151–164, 2019.

[153] A. J. Williamson, D. L. Smith, D. Blinco, R. D. Unwin, S. Pearson, C. Wilson, C. Miller, L. Lancashire, G. Lacaud, V. Kouskoff, et al. Quantitative proteomics

analysis demonstrates post-transcriptional regulation of embryonic stem cell differentiation to hematopoiesis. *Molecular & Cellular Proteomics*, 7(3):459–472, 2008.

[154] I. Wilmut, A. E. Schnieke, J. McWhir, A. J. Kind, and K. H. Campbell. Viable offspring derived from fetal and adult mammalian cells. *Nature*, 385(6619):810–813, 1997.

[155] E. B. Wilson. *The cell in development and inheritance*. Macmillan, 1900.

[156] K. G. Wolter, Y.-T. Hsu, C. L. Smith, A. Nechushtan, X.-G. Xi, and R. J. Youle. Movement of bax from the cytosol to mitochondria during apoptosis. *The Journal of cell biology*, 139(5):1281–1292, 1997.

[157] J. C. Wong, M. M. Jack, Y. Li, and C. O'Neill. The epigenetic bivalency of core pancreatic β-cell transcription factor genes within mouse pluripotent embryonic stem cells is not affected by knockdown of the polycomb repressive complex 2, suz12. *Plos one*, 9(5):e97820, 2014.

[158] X. Wu, D. A. Scott, A. J. Kriz, A. C. Chiu, P. D. Hsu, D. B. Dadon, A. W. Cheng, A. E. Trevino, S. Konermann, S. Chen, et al. Genome-wide binding of the crispr endonuclease cas9 in mammalian cells. *Nature biotechnology*, 32(7):670–676, 2014.

[159] X.-L. Wu and P.-S. Zheng. Undifferentiated embryonic cell transcription factor-1 (utf1) inhibits the growth of cervical cancer cells by transactivating p27 kip1. *Carcinogenesis*, 34(7):1660–1668, 2013.

[160] S. Wunderlich, M. Kircher, B. Vieth, A. Haase, S. Merkert, J. Beier, G. Göhring, S. Glage, A. Schambach, E. C. Curnow, et al. Primate ips cells as tools for evolutionary analyses. *Stem cell research*, 12(3):622–629, 2014.

[161] C. Xu, Y. Zhou, and W. Chen. Expression of undifferentiated embryonic cell transcription factor-1 (utf1) in breast cancers and their matched normal tissues. *Cancer Cell International*, 14:1–4, 2014.

[162] C.-S. Yang, K.-Y. Chang, and T. M. Rana. Genome-wide functional analysis reveals factors needed at the transition steps of induced reprogramming. *Cell reports*, 8 (2):327–337, 2014.

[163] T. Yoshino, U. Bickham, and C. Bayne. Molluscan cells in culture: primary cell cultures and cell lines. *Canadian journal of zoology*, 91(6):391–404, 2013.

[164] J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, et al. Induced pluripotent stem cell lines derived from human somatic cells. *science*, 318(5858):1917–1920, 2007.

[165] Y. Zhao, X. Yin, H. Qin, F. Zhu, H. Liu, W. Yang, Q. Zhang, C. Xiang, P. Hou, Z. Song, et al. Two supporting factors greatly improve the efficiency of human ipsc generation. *Cell stem cell*, 3(5):475–479, 2008.

[166] M. J. Zoran. Regeneration in annelids. *eLS*, 2010.

[167] T. P. Zwaka. Keeping the noise down in es cells. *Cell*, 127(7):1301–1302, 2006.

[168] V. Zywitza, S. Frahm, N. Krüger, A. Weise, F. Göritz, R. Hermes, S. Holtze, S. Colleoni, C. Galli, M. Drukker, et al. Induced pluripotent stem cells and cerebral organoids from the critically endangered sumatran rhinoceros. *Iscience*, 25(11), 2022.

APPENDIX A

PUBLICATIONS

Publications from PhD Thesis

1. **Raina K**, Modak K, Premkumar C, Velayudhan SR, Thummer RP **(2024)**. UTF1 expression is important for the generation and maintenance of human iPSCs. Manuscript is under review.

2. **Raina K**, Joshi G, Modak K, Premkumar C, Priyanka S, Rajesh P, Velayudhan SR, Thummer RP **(2023)**. Generation and characterization of induced pluripotent stem cell line IITGi001-A derived from adult human primary dermal fibroblasts. *Stem Cell Research*. DOI: 10.1016/j.scr.2023.103159.

3. Dey C, **Raina K**, Thummer RP **(2023)**. Production of a Bioactive Recombinant Human UTF1 Protein from *E. coli.*. *Healthcare Research and Related Technologies. NERC 2022*. DOI: 10.1007/978-981-99-4056-1_4.

4. **Raina K**, Dey C, Thool M, Sudhagar S, Thummer RP **(2021)**. An insight into the role of UTF1 in development, stem cells, and cancer. *Stem Cell Reviews and Reports*. DOI: 10.1007/s12015-021-10127-9.

5. Dey C*, **Raina K***, Thool M*, Adhikari P, Haridhasapavalan KK, Sundaravadivelu PK, Venkatesan V, Gogoi R, Sudhagar S, Thummer RP **(2021)**. Auxiliary pluripotency-associated genes and their contributions in the generation of induced pluripotent stem cells. *Molecular Players in IPSC Technology, Volume 12*. Elsevier: Academic Press. ISBN: 9780323900591. *Equal contribution. DOI: 10.1016/B978-0-323-90059-1.00007-5.

6. Dey C*, **Raina K***, Haridhasapavalan KK*, Thool M*, Sundaravadivelu PK, Adhikari P, Gogoi R, Thummer RP **(2021)**. An Overview of Reprogramming Ap-

proaches to Derive Integration-free Induced Pluripotent Stem Cells for Prospective Biomedical Applications. *Recent Advances in iPSC Technology, Volume 5 in Advances in Stem Cell Biology.* Elsevier: Academic Press. ISBN: 9780128222317. *Equal contribution. DOI: 10.1016/B978-0-12-822231-7.00011-4.

7. Haridhasapavalan KK*, **Raina K***, Dey C*, Adhikari P, Thummer RP **(2020)**. An Insight into Reprogramming Barriers to iPSC Generation. *Stem Cell Reviews and Reports.* *Equal contribution. DOI: 10.1007/s12015-019-09931-1.

8. Sundaravadivelu PK, **Raina K**, Thool M, Ray A, Joshi JM, Kaveeshwar V, Sudhagar S, Lenka N, Thummer RP **(2021)**. Tissue-restricted Stem Cells as Starting Cell Source for Efficient Generation of Pluripotent Stem Cells: An Overview. *Cell Biology and Translational Medicine, Volume 15 in Advances in Experimental Medicine and Biology.* Springer Nature. DOI: 10.1007/5584_2021_660.

9. Ray A*, Joshi JM*, Sundaravadivelu PK*, **Raina K**, Lenka N, Kaveeshwar V, Thummer RP **(2021)**. An overview on promising somatic cell sources utilized for the efficient generation of induced pluripotent stem cells. *Stem Cell Reviews and Reports.* *Equal contribution. DOI: 10.1007/s12015-021-10200-3.

Publications from other collaborative research

1. Subhendu SS, Bora A, Golder A, **Raina K**, Haridhasapavalan KK, Thummer RP **(2022)**. Gelatin-PVA-AgNPs triad composite as wound healing hydrogel with wounded skin surface protective efficiency. *SSRN Electronic Journal.* DOI: 10.2139/ssrn.4219683.

2. Banerjee K, Bhattacherjee D, **Raina K**, Thummer RP, Bhabak KP **(2022)**. Benzimidazole-based ionic and non-ionic organoselenium compounds: innovative synthetic strategies, structural characterization and preliminary anti-proliferative activities. *New Journal of Chemistry.* DOI: 10.1039/d2nj01322c.

3. Bhattacherjee D, **Raina K**, Mandal T K, Thummer RP, Bhabak KP **(2022)**. Targeting Wnt/β-catenin Signaling Pathway in Triple-negative Breast Cancer by Benzylic Organotrisulfides: Contribution of the Released Hydrogen Sulfide towards Potent Anti-Cancer Activity. *Free Radical Biology and Medicine.* DOI: 10.1016/j.freeradbiomed.2022.08.029.

4. Haridhasapavalan KK, Sundaravadivelu PK, Bhattacharyya S, Harsh Ranjan S, **Raina K**, Thummer RP **(2021)**. Generation of cell-permeant recombinant human transcription factor GATA4 from *E. coli*. *Bioprocess and Biosystems Engineering.* DOI: 10.1007/s00449-021-02516-8.

5. Dey S, Patel A, **Raina K**, Pradhan N, Biswas O, Thummer RP, Manna D **(2020)**. Stimuli-Responsive Anticancer Drug Delivery System with Inherent Antibacterial Activities. *Chemical Communications.* DOI: 10.1039/C9CC08834B.

Conferences/Workshops attended

1. Annual Meet by **“International Society for Stem Cell Research"**, ABoston, Massachusetts, USA.
   [*June, 2023*]

2. International symposium on **“Bone Marrow failure”** organized by Christian Medical College, Vellore.
   [*April, 2022*]

3. Training program in **“Generation and Maintenance of human iPS cells”** organized by the Department of Biotechnology (DBT), Government of India and The Centre for iPS Cell Research and Application (CiRA), Kyoto University, Japan.
   [*November, 2019*]

4. Workshop on **“CRISPR-Cas9 based targeted genome editing”** organized by CSIR-Institute of genomics and Integrated biology (IGIB), New Delhi.
   [*April, 2019*]

5. International symposium on **“Regulatory epigenomics: From large data to useful models”** organised by India | EMBO, MGM Beach Resort, Chennai.
   [*March, 2019*]

Awards

1. **Travel Grant** for “International Society for Stem Cell Research” annual meet by Science and Engineering Research Board, Department of Science and Technology (SERB-DST), Government of India.
   [*June, 2023*]

2. **Travel Grant** for “Training program in generation and maintenance of human iPS cells” by the Department of Biotechnology (DBT), Government of India and The Centre for iPS Cell Research and Application (CiRA), Kyoto University, Japan.
   [*Novebmer, 2019*]