Elucidating the role of human ZSCAN4 in the maintenance of induced pluripotent stem cells

A Thesis

Submitted in Partial Fulfilment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Pradeep Kumar Sundaravadivelu

Under the supervision of

Dr. Rajkumar P. Thummer

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

May, 2026

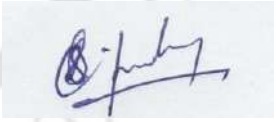
Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

DECLARATION

I do hereby declare that the content embodied in this thesis entitled "**Elucidating the role of human ZSCAN4 in the maintenance of induced pluripotent stem cells**" 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 of 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.



Date: 01/05/2026

Pradeep Kumar Sundaravadivelu
(206106110)

Indian Institute of Technology Guwahati
Department of Biosciences and Bioengineering

CERTIFICATE

It is certified that the work described in this thesis entitled "**Elucidating the role of human ZSCAN4 in the maintenance of induced pluripotent stem cells**", by **Mr. Pradeep Kumar Sundaravadivelu** (Roll No. 206106110) 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
(Thesis Supervisor)

Date: 01/05/2026

"கற்றது கை மண் அளவு, கல்லாதது உலகளவு."

- ஔவையார்

“What is known is but a handful of sand; what is unknown is the size of the world.”

- Avvaiyar

Acknowledgement

"யாதும் ஊரே யாவரும் கேளிர்."
-கணியன் பூங்குன்றனார்
"To us, all towns are our own, and all people are our kin."
-Kaniyan Pungundranar

These profound words by the Sangam poet Kaniyan Pungundranar serve as the perfect preamble to this section. My doctoral journey has been a testament to this spirit of universal kinship. Scientific research is rarely a solitary path; it is a collaborative tapestry woven with the support of mentors, colleagues, and friends who transcend geographical and institutional boundaries. As I reflect on the years spent in the laboratory and the field, I realize that this thesis is not merely a personal achievement, but a collective result of the kindness and intellectual generosity of a global community. It is with a heart full of gratitude that I acknowledge the "kinsmen" who guided me through this transformative journey.

My deepest and most sincere gratitude is first and foremost reserved for my supervisor, **Dr. Rajkumar P. Thummer**, Boss of the Laboratory for Stem Cell Engineering and Regenerative Medicine. His unwavering guidance and profound passion for scientific inquiry have been the primary reason for my own love for this field. I am particularly grateful for the collegial and supportive environment he fosters within the workspace, which allowed me to conduct my research with ease. The intellectual freedom he afforded me throughout my tenure was instrumental in my professional growth and helped me achieve significant milestones in the laboratory. Beyond his academic mentorship, he is one of the most humble and grounded individuals I have ever encountered. His perseverance, patience, and ability to remain motivated during challenging times serve as a blueprint for my own future career. I consider myself exceptionally fortunate to have had him as a supervisor, and I will remain forever grateful for his transformative impact on my journey as a researcher.

I want to thank **Dr. Shirisha Nagotu** from Department of Biosciences and Bioengineering (BSBE), for her motivation and constant support throughout this journey. I would like extend this gratitude to the members of my doctoral committee, **Prof. Manish Kumar** (Chairperson) and **Dr. Anil Mukund Limaye** of the Department of BSBE and **Prof. Lal Mohan Kundu** of the Department of Chemistry for their invaluable suggestions. The scientific discussions that we had during my progress review every year has not only polished my thesis work but also fundamentally shaped my development as a scientist.

I would extend my thanks to current Head of the Department of BSBE, **Prof. Utpal Bora** and other former HODs, **Prof. Rakhi Chaturvedi**, and **Prof. Latha Rangan** for providing the departmental facilities for carrying out my research work. I would also like to acknowledge all the technical staffs of the department for their help and support, especially **Mrs. Prarthana Swargari**, for all the scientific discussions that we had. Also, I would like to acknowledge Prime Minister's Research Fellowship (PMRF) program and Ministry of Education, Government of India for the financial support and research funding during my tenure.

My journey at IIT Guwahati has been a truly enriching experience that I will always cherish. At the heart of this journey was my first research mentor, **Dr. Krishna Kumar Haridhasapavalan**, to whom I owe an immense debt of gratitude. As my first guide in the laboratory, he painstakingly taught me the fundamental principles of research from the ground up. Beyond his role as a mentor, he has been a constant, brotherly figure in my life, offering unwavering support to this day. He is truly one of the primary architects of my scientific identity, and I am profoundly grateful for the role he played in moulding me into the researcher I am today.

Science can be solitary - but I was lucky enough never to feel alone. **Dr. Chandrima Dey, Dr. Gloria Narayan, Dr. Madhuri Thool**, and **Dr. Khyati Raina** made our laboratory feel less like a workspace and more like home. From the very first day, they wrapped me in a warmth that is rare to find in academic settings. Our brainstorming sessions were more than intellectual exercises - they were the kind of conversations that stay with you, quietly reshaping the way you think and the scientist you are becoming I carry their mentorship not just in my research, but in my heart. I would like to extend this thanks to the current lab members that has helped me during this journey, **Ms. Atreyee Borthakur, Mr. Naveen Raj, Mr. Harsh Vikram Singh, Ms. Ayushi, Mr. Dhaya Shankaran, Mr. Subramanian V, Ms. Shilpendra,** and **Ms. Priyanka Gupta**, whose friendship and support eased the inevitable hardships of this journey. Also, I would like to thank the former members of the lab **Ms. Akriti Agrawal, Mr. Allan Blessing Harrison Raj, Ms. Ronima KR,** and **Mr. Kanish Nareen** for their valuable assistance. I am especially indebted to **Mr. Allan Blessing Harrison Raj** and **Ms. Priyanka Gupta**, whose efforts significantly accelerated the progress of my work and I will forever be thankful for their contributions. Furthermore, I am deeply grateful to **Ms. Ronima KR**, her constant encouragement and presence were vital in navigating the more challenging phases of my doctoral tenure. I would also thank the OBCAL lab members, **Dr. Nayan Moni Deori, Dr. Pracheeta Deb, Dr. Riddhi Banerjee, Ms. Atchaya SR, Ms. Selva Parameshwari S, Ms.**

Tanveera Rounaque Sarhadi, Ms. Isha Kashyap, Ms. Ankita Adhikary, Ms. Dharshani Chopade, and **Mr. Soham Basu** for being supportive throughout this journey.

This journey is incomplete without the friends outside of my work as they transformed this place into second home for me. I would like to thank elder members first, **Dr. Shanmugam** and **Mr. Giridhar** for not being the boring, elder ones. My friends, **Dr. Thirukumaran, Mr. Bharathi, Ms. Jenny, Ms. Sujisha, Ms. Stuthi, Mr. Naveen Raj,** soon-to-be doctors, **Mr. Thilagaraj** and **Mr. Kannan** brought more laughter into my days than I could have asked for. Extending this gratitude further, **Mr. Ragavan, Mr. Harish, Mr. Askkar, Mr. Jeeva, Mr. Vishwa** became a steady, joyful presence through it all, and the newest additions to the gang**, Mr. Ajay, Mr. Sushanth, Mr. Thilakaraj, Mr. Solmon, Ms. Harini, Ms. Vaishnavi, Mr. Hariharan, Ms. Mohanapriya (X2), Ms. Saraswathi, Ms. Dharshini, Mr. Kishore,** and **Ms. Prarthana** arrived just in time to fill the final chapter with fresh energy and warmth. I would like to thank the Tamil Sangam, IIT Guwahati for bringing the native festivals and delicious food here in Guwahati. Also, a special thanks to my friends **Mr. Rohan Pal** and **Mr. Amay** for the memorable rantings, get-togethers and celebrations that we had during this journey. Special mention to **Mr. Sai Ganapathy, Mr. Santhosh** and **Mr. Sowmi** - old friends who never needed catching up with, thank you for being the kind of constants that make change feel safe.

Last but certainly not least, I dedicate this achievement to my family, who have been my greatest source of strength. I am eternally grateful to my parents, **Mr. Sundaravadivelu P** and **Mrs. Geetha S**, for their endless sacrifices and for instilling in me the value of perseverance; their unconditional love has been the bedrock of my life. My sister, **Mrs. Pavithra**, and my brother-in-law, **Mr. Rajkumar**, have been my pillars of support, offering encouragement whenever the journey became taxing. A special mention goes to my nephew, **Mithran RP**, and the two newest blessings in our family, one-month-old **Rudhran** and **Aadhiran**, whose innocence and presence brought much-needed joy and perspective during the final stages of this work. Without the collective prayers and unwavering belief of my family, I would not have reached this milestone. They are the true force behind my success.

Signing off

Pradeep Kumar Sundaravadivelu

CONTENTS

Page No.

List of figures Page no.

List of Tables

Page no.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Chapter 1

Introduction and Review of Literature

Parts of Chapter 1 have been published in the below mentioned manuscripts:

- **Sundaravadivelu PK**, Borthakur A, Agrawal A, Sen T, Thummer RP (**2024**). An overview on the generation of integration-free induced pluripotent stem cells and their subsequent differentiation into cardiomyocytes and β-cells. In Advances in Stem Cell Research: An Indian Perspective. Eds Prof. Rama Shanker Verma. https://doi.org/10.1142/9789811294891_0005 (World Scientific).
- Thool M*, **Sundaravadivelu PK***, Sudhagar S, Thummer RP (**2022**) A Comprehensive Review on the Role of ZSCAN4 in Embryonic Development, Stem Cells, and Cancer. Stem Cell Reviews and Reports. Jun 23:1-7. https://doi.org/10.1007/s12015-022-10412-1 (Springer Nature) *Equal contribution.

Overview

Chapter 1 outlines the fundamental concepts of stem cell biology and highlights the emergence of induced pluripotent stem cells (iPSCs) as a transformative advancement in the field. iPSCs are generated through the reprogramming of terminally differentiated somatic cells by the introduction of defined transcription factors or alternative strategies that activate a pluripotency-associated gene expression program. Owing to their pluripotent nature, iPSCs hold immense potential for applications in disease modeling, drug discovery, regenerative medicine, and understanding basic developmental biology. Despite their promise, several critical challenges must be addressed before the clinical translation of iPSC-based technologies. Among these, genomic instability and the accumulation of genetic abnormalities during reprogramming and prolonged culture remain major concerns. Over the past decade, multiple strategies have been developed to enhance the genomic integrity of iPSCs. One such approach involves the inclusion of ZSCAN4, a transcription factor known to play a crucial role in early embryonic development and stem cell biology. In murine systems, incorporation of ZSCAN4 alongside the core Yamanaka factors has been shown to significantly improve both the efficiency of reprogramming and the genomic stability of resulting iPSCs. ZSCAN4 has been implicated in telomere maintenance, genomic integrity, and transient activation of a unique transcriptional state associated with pluripotency. While the functional role of ZSCAN4 has been extensively characterized in mouse embryonic stem cells (ESCs) and reprogramming models, its role in human cellular reprogramming and maintenance of human iPSCs remains insufficiently understood. This gap in knowledge represents an important area of investigation in the field of stem cell biology. The advent of precise genome editing technologies, particularly CRISPR-Cas9, has enabled targeted manipulation of gene function, providing a powerful platform to dissect the roles of specific factors in defined cellular contexts. In this study, CRISPR-Cas9-based approaches have been employed to modulate ZSCAN4 expression in mammalian cells, with the objective of elucidating its functional role in the maintenance and stability of iPSCs. Accordingly, this chapter provides an overview of ZSCAN4 biology, with particular emphasis on its role in cellular reprogramming and pluripotency. It also outlines the key methodologies employed in this study and defines the research gap addressed in this thesis.

1.1 Understanding stem cells

The term "stem cells" was first coined in late 19th century by German biologist Ernst Haeckel as "*stammzelle*" to describe them as unicellular ancestor from which all multicellular life descended. Later, in 1890s, scientists, Theodor Boveri and Valentin Hacker used the term "stem cells" to describe cells that are present inside an embryo. This was the birth of modern-day definition, where stem cells are defined as primitive, undifferentiated cells present in an organism defined by two cardinal properties. First is self-renewal, where these cells can replicate themselves while maintaining an undifferentiated state. Second is potency, where these cells possess the potential to differentiate into specialized cell types. These properties are regarded as the stemness of these cells. Based on the potency of the stem cells, they can be classified into four distinct types. Firstly, totipotent stem cells, possessing the highest level of potency. These totipotent stem cells can give rise to all embryonic and extra-embryonic lineages (e.g. placenta and umbilical cord). These cells have the potential to form a whole new organism given appropriate conditions. In human embryonic development the period corresponding to zygote genome activation is considered as totipotent (Xu and Liang, 2022). This includes the zygote that forms after fertilization, and the cells present in 2-cell, 4-cell, and 8-cell stage. Secondly, pluripotent stem cells, the cells can differentiate into cells belonging to all three germ layers, namely ectoderm, mesoderm and endoderm. These cells can virtually give rise any cell type present in an organism except for extra embryonic cell types. The inner cell mass [source of Embryonic Stem Cells (ESCs)] present in the blastocyst stage of embryonic development represents the gold standard of pluripotency. Thirdly, multipotent stem cells, which can differentiate into a limited subset of closely related cellular lineages. Adult stem cells, which are the only stem cells present in adult human body, are multipotent in nature. Examples of multipotent stem cells are Hematopoietic stem cells (HSCs), Mesenchymal stem cells (MSCs), Neural Stem Cells, and other adult stem cells present in an adult human body. HSCs are present in bone marrow, which has the potential to differentiate into various types of blood cells. MSCs are present in multiple locations in the human body and can differentiate into bone cells, fat cells, cartilage, and other cell types. NSCs are primarily found in specific regions of the central nervous system, such as the subventricular zone and the hippocampus, and have the potential to differentiate into major neural cell types, including neurons, astrocytes, and oligodendrocytes. Fourth, unipotent stem cells have the most restricted differentiation potential, giving rise to only a single cell

type. One such example is spermatogonial stem cells (SSCs), present in testes that can differentiate into sperm cells. These SSCs can maintain their own population (self-renewal) and can also differentiate to sperm cells (differentiation ability) because of which these are also categorized as stem cells.

The field of stem cell biology is rapidly evolving, thus exploring the molecular mechanisms underlying the self-renewal and differentiation capabilities of stem cells is vital for their subsequent biomedical applications. ESCs being pluripotent, are an obvious choice of study to gain an in-depth understanding. These cells have a wide spectrum of applications in various domains, including drug development and toxicity testing, disease modeling, personalized regenerative therapy, understanding human developmental biology, and other biomedical applications. Despite their significant therapeutic potential, the clinical translation of human ESCs has been substantially hindered by major bioethical concerns. The derivation of human ESCs traditionally requires the destruction of a human blastocyst, a process that raises ethical issues regarding the moral status of embryonic life. Consequently, regulatory authorities and ethics committees in several countries have imposed bans or stringent restrictions on research involving human ESCs.

1.2 The emergence of iPSCs

Driven by the ethical and practical limitations associated with the derivation of human ESCs, researchers sought to induce pluripotency artificially, *in vitro*, to terminally differentiated somatic cells. This pursuit originated from a fundamental question in developmental biology, whether the potential for differentiation is governed by the genetic material present in the nucleus or the biochemical environment of the cytoplasm. The initial exploration of this concept dates back to 1952, when Robert Briggs and Thomas J. King performed pioneering nuclear transplantation experiments. By transferring a nucleus from a *Rana pipiens* blastocyst into an enucleated oocyte, they successfully generated functional embryos that progressed into viable, healthy tadpoles (Briggs and King, 1952). However, because the donor nucleus was derived from an early-stage embryo, it remained unclear if the observed pluripotency was an inherent trait of the nucleus or a result of cytoplasmic induction.

This ambiguity was resolved in 1962 by the pioneering work of John B. Gurdon, who demonstrated that nuclei from terminally differentiated intestinal epithelial cells of

Xenopus laevis could support full development following nuclear transfer (Gurdon, 1962). Despite the differentiated state of the donor nucleus, its transplantation into an enucleated egg resulted in the development of viable tadpoles (Gurdon, 1962). This landmark study demonstrated that the somatic genome retained the potential to become totipotent and can be induced by embryonic cytoplasmic factors. This methodology, termed as somatic cell nuclear transfer (SCNT), established the foundational principle that the differentiated state is reversible. These findings provided the essential theoretical framework for the eventual discovery of iPSCs by demonstrating that terminally differentiated cells could be returned to a state of developmental plasticity.

The feasibility of SCNT in a mammalian system was definitively established in 1997. Wilmut and colleagues successfully extrapolated principles established by Gurdon's amphibian models to *Ovis aries* (sheep), demonstrating that the genomic plasticity of differentiated cells was conserved across higher taxa (Wilmut et al., 1997). In this seminal study, a nucleus derived from an udder cell was transferred into an enucleated oocyte. Following successful nuclear integration, the reconstructed embryo was implanted into a surrogate mother, resulting in the birth of Dolly, the first mammal cloned from an adult somatic cell (Wilmut et al., 1997). Later, similar studies were conducted, where a somatic cell was fused with ESC leading to a hybrid cell, which proved to be pluripotent *in vivo* (Cowan et al., 2005; Tada et al., 2001). Collectively, these milestones yielded two pivotal conclusions. Firstly, both unfertilized oocytes and established ESC lines possess the necessary cytoplasmic machinery to induce pluripotency in somatic nuclei. Secondly, the reprogramming process is robust enough to support the development of viable offsprings in higher mammals. While SCNT provided a revolutionary platform for developmental biology and disease modeling, it was hampered by significant drawbacks. The process was technically demanding, suffered from extremely low efficiency, and most importantly, relied on the consumption of oocytes and the creation of embryos intended for destruction. These factors led to significant bioethical concerns, resulting in stringent regulatory bans on the use of SCNT-derived human embryos for research. This necessitated the search for a more accessible, non-embryonic method to achieve cellular reprogramming.

The discovery that oocytes and ESCs can reprogram the somatic epigenome led to the hypothesis that pluripotency is driven not by the entire cellular milieu, but by a defined set of molecular factors specifying the pluripotent state. It was reasoned that identifying

and overexpressing the key transcription factors enriched in ESCs could sufficiently drive the dedifferentiation of somatic nuclei.

The transition from theoretical possibility to experimental reality required the identification of specific transcription factors capable of overwriting somatic cell identity. In a landmark 2006 study, Kazutoshi Takahashi and Shinya Yamanaka initially performed a large-scale functional screen of 24 candidate genes cloned into retroviral vectors. Through a process of elimination, they identified a core set of four indispensable factors, namely Oct3/4, Sox2, Klf4, and c-Myc, universally referred to as the Yamanaka factors (Takahashi and Yamanaka, 2006). Using these Yamanaka factors, they reprogrammed murine fibroblasts into a pluripotent state. These cells were termed as iPSCs, exhibited the hallmark morphological, functional and epigenetic properties similar to ESCs (Takahashi and Yamanaka, 2006).

The following year, this area of research achieved a second major milestone when two independent research groups successfully extended this technology to human cellular systems. Prof. Shinya Yamanaka and team generated human iPSCs by reprogramming human fibroblasts using the Yamanaka factors, OCT4, SOX2, KLF4, and c-Myc (Takahashi et al., 2007). In a concurrent study, Prof. James A. Thomson and his colleagues identified a partially different cocktail of transcription factors to generate human iPSCs from somatic cells. Out of 14 factors investigated, the reprogramming factors identified in this study were OCT4, SOX2, NANOG, and LIN28 (OSNL), often termed as the Thomson factors (Yu et al., 2007). Unlike the Yamanaka approach, the Thomson group employed a lentiviral-mediated delivery system for genomic integration. These breakthrough studies provided a virtually unlimited source of patient-specific pluripotent cells, effectively bypassing the ethical and immunological barriers associated with the use of embryonic-derived cells. Further, these seminal studies established the foundational protocols for cellular reprogramming, effectively launching a new paradigm in stem cell biology centered on the derivation and therapeutic application of iPSCs.

1.3 Major roadblocks in the generation of iPSCs

While the initial derivation of human iPSCs was a monumental achievement, several inherent limitations have hindered their integration into clinical therapeutics and personalized medicine. The pioneering studies by Yamanaka and Thomson labs utilized

retroviral and lentiviral vectors (Takahashi et al., 2007; Yu et al., 2007). Even though these methods have higher efficiency, they randomly integrate into the host genome, posing a significant risk of insertional mutagenesis. Such permanent genomic alterations can disrupt endogenous gene function or lead to aberrant activation of neighboring genes, rendering the resulting iPSCs unsuitable for clinical applications. To resolve this, a number of non-integrative approaches have been investigated for their potential to generate clinical grade, *bonafide* iPSCs, which are briefly discussed elsewhere (Borgohain et al., 2019; Dey et al., 2021a; Haridhasapavalan et al., 2019).

Another major safety concern arises from the nature of reprogramming cocktail itself. Several core factors including c-Myc, Klf4, and Lin28 are potent oncogenes. Their overexpression can trigger the undesirable activation of oncogenic pathways within the generated iPSCs, potentially leading to tumorigenesis or teratoma formation (Aguirre et al., 2023). Moreover, the efficiency of the reprogramming using this combination of factors is very low (~0.01%), considering these factors were delivered using one of the most efficient viral-based vectors (Raab et al., 2014). Further, a significant proportion of the resulting colonies often manifest as “pre-iPSCs”, cells that have initiated but failed to complete the reprogramming process (Hirata et al., 2012; Jiang et al., 2013; Maekawa et al., 2011; Zhao et al., 2008). The presence of oncogenic factors coupled with the stress of reprogramming process, often results in poor genomic stability (Hirata et al., 2012; Jiang et al., 2013). These stability issues present a formidable hurdle for the development of safe, standardized cell-based therapies.

As previously established, generating iPSCs with stable genome remains one of the most significant challenges hindering their application in therapeutics. To mitigate this, extensive research has focused on optimizing the reprogramming cocktail through the addition of supplementary factors. Inclusion of factors like UTF1 (Zhao et al., 2008; Raina et al., 2025), GLIS1 (Maekawa et al., 2011), ZSCAN4 (Hirata et al., 2012; Jiang et al., 2013) alongside core Yamanaka factors have demonstrated to be effective in maintaining the genomic stability, thereby enhancing the quality of the generated iPSCs. Given its unique ability to regulate chromosomal stability and facilitate the transition from a “pre-iPSC” to a fully stable pluripotent state, ZSCAN4 served as the primary focal point of this thesis.

1.4 ZSCAN4: The savior to generate genomically stable iPSCs

ZSCAN4 (Zinc and SCAN domain containing 4) is a zinc finger transcription factor, which enhances the reprogramming efficiency when included along with the original Yamanaka cocktail of reprogramming factors. This gene is expressed exclusively in the two-cell stage, in *in vitro* cultured ESCs, and during early embryonic development, where its expression is essential for embryo implantation (Falco et al., 2007; Zalzman et al., 2010). In addition, the expression of ZSCAN4 is found to be essential for telomere length maintenance in both ESCs and iPSCs (Wang et al., 2012; Zalzman et al., 2010), which prevented senescence in these cells, enabling them to self-renew indefinitely. As mentioned earlier, inclusion of ZSCAN4 during reprogramming resulted in iPSCs having better genomic stability (Hirata et al., 2012; Jiang et al., 2013). ZSCAN4 is also reported to have a role in the regulation of cancer stem cells (Portney et al., 2020). To date, studies have shown the mechanistic role of ZSCAN4 in mouse ESCs and in the generation of mouse iPSCs but the role of ZSCAN4 in the generation and maintenance of human iPSCs is not known. Therefore, this thesis focuses on elucidating the functional role of ZSCAN4 in the generation of human iPSCs and their maintenance.

1.4.1 ZSCAN4 gene

ZSCAN4 is a protein-coding gene (Table 1.1) and is present only in mammals (Falco et al., 2007). The mouse genome comprises of nine members of this gene (*Zscan4a*, *Zscan4b*, *Zscan4c*, *Zscan4d*, *Zscan4e*, *Zscan4f*, and three pseudogenes), whereas the human genome contains only one member. In mice, all nine members of the gene have multi-exonal organization and are firmly grouped in an 850 kb region on chromosome 7. Three out of these nine genes in mice are pseudogenes, namely *Zscan4-ps1*, *Zscan4-ps2* and *Zscan4-ps3*. Members of mouse *Zscan4* gene, *Zscan4c, Zscan4d* and *Zscan4f* encodes full-length open reading frames (ORFs) with 506 amino acids (aa) having more than 94% similarity and are jointly called *Zscan4*, indicating these genes are likely to share the same function (Edelstein and Collins, 2005; Falco et al., 2007). In humans, *ZSCAN4* is present on chromosome 19 spanning a length of 27 kb (Edelstein and Collins, 2005; Falco et al., 2007).

Table 1.1. Physicochemical parameters of human ZSCAN4 sequence.

Parameters	**Values**
Protein-coding DNA Sequence	
Number of nucleotides	1302 bp
% GC content	43.9%
Nucleotide composition	Adenine (A): 33.41% (435 bases) Thymine (T): 22.73% (296 bases) Guanine (G): 20.89% (272 bases) Cytosine (C): 22.96% (299 bases)
Protein	
Number of amino acids	433
Molecular weight	48.95 kDa
Theoretical pI	6.47
Total number of negatively charged residues	Asp+Glu: 52
Total number of positively charged residues	Arg+Lys: 47

1.4.2 ZSCAN4 Protein

Human ZSCAN4 protein has a length of 433 amino acids (aa). Analysis of this sequence using InterPro database (Apweiler et al., 2001; Blum et al., 2021) revealed that the protein consists of two characteristic domains. One of the two critical domains is the SCAN domain at the N-terminal region of the protein (InterPro Id: IPR003309), which is between 44-126 residues, belonging to SCAN domain superfamily (Fig. 1.1). SCAN acquired its name from initials of four proteins that have this domain in common (**S**RF-ZBP, **C**Tfin-51, **A**W-1, **N**umber 18 cDNA) (Williams et al., 1995). This domain is often referred to as the leucine-rich region (Edelstein and Collins, 2005; Falco et al., 2007; Pengue et al., 1994). In vertebrates, the SCAN domain is a highly conserved domain having a unique motif that is involved in protein-protein interaction (Edelstein and Collins, 2005; Storm et al., 2009). Deletion of this domain diminishes its property of self-association,

indicating ZSCAN4 can form dimers or oligomers after self-interaction (Dan et al., 2017). In case of mouse, Zscan4c, Zscan4d and Zscan4f are all 506 aa in length and all of them have their SCAN domain from 37 to 119 residues. The other two members, Zscan4b and Zscan4e, are 505 amino acids in length. They have their SCAN domain from 53 to 86 residues. Limited information is available about the domain organisation and the sequence of Zscan4a in the databases to date.

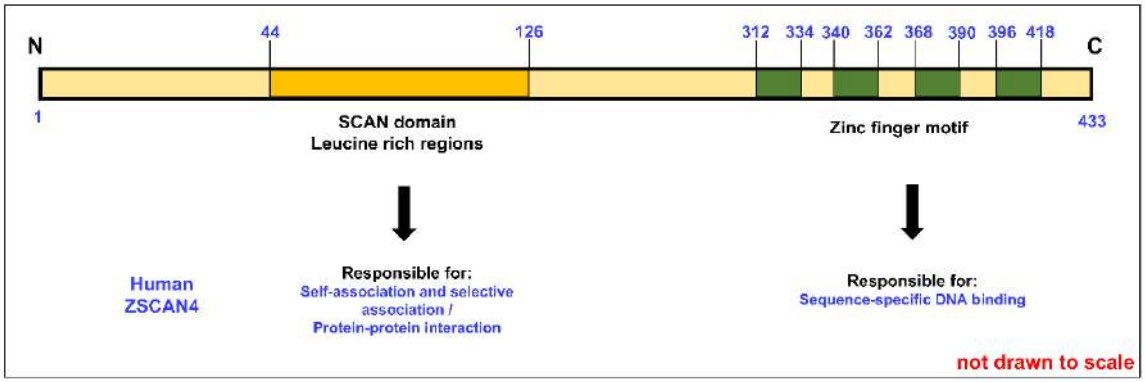


Fig. 1.1. Pictorial representation of domains present in human ZSCAN4 protein. The SCAN domain is present towards the N-terminal region of the protein, and four zinc finger domains present towards the C-terminal end of the protein.

The second characteristic domain is the zinc finger domain (total four in number; InterPro Id: IPR013087), which is present at the C-terminal region of the human ZSCAN4 protein. This domain belongs to the zinc finger C_2H_2 superfamily and is shown in Fig. 1.1. In case of mouse Zscan4c, Zscan4d and Zscan4e, the first zinc finger motif is located at position 395-417 aa, second at 424-446 aa, third at 452-474 aa, and fourth at 480-503 aa. The C_2H_2 motifs of the zinc finger domain are primarily entailed in binding to a specific fragment of DNA and interacting with several other cellular factors to regulate the transcription of its target genes (Edelstein and Collins, 2005). Each finger is a short stretch of approximately 20-30 aa comprising of two conserved cysteines and histidines. The zinc ion part of the finger assists in proper folding to form compact structures. The structures comprising of α-helix includes the conserved histidines, whereas β-turn includes the conserved cysteines (Pavletich and Pabo, 1991). Intriguingly, Zscan4f was found to have another isoform of 77 aa (UniProt ID: L7N266) (Kumpfmueller, 2011). Especially, these three members of mouse Zscan4 family (*Zscan4a*, *Zscan4b,* and *Zscan4e*) code for truncated proteins; expressed at basal level in the late two-cell stage and ESCs (Falco et al., 2007). However, the exact function of these truncated proteins remains elusive. The three-

dimensional structure of ZSCAN4 protein is not reported to date. The DNA binding activity of zinc finger domain is highly specific, which can be utilized for designing zinc finger nucleases. These zinc finger nucleases can be used for genetic manipulation at precise sites within the genome of various organisms, which is discussed later in this chapter (Perez et al., 2008; Urnov et al., 2010).

The three-dimensional and secondary structure of ZSCAN4 protein is not reported to date. Therefore, the secondary structure was predicted using bioinformatics online tools, namely GOR4 (Garnier et al., 1996), SOPMA (Geourjon and Deléage, 1995) and PSIPRED (Buchan and Jones, 2019; Jones, 1999). The three-dimensional structure (Fig. 1.2A) was obtained from Alphafold 3 database (Jumper et al., 2021; Varadi et al., 2024). The prediction from all these servers revealed that the protein majorly consists of random coils and other structures with a significant contribution of α-helices and a small amount of β-sheets (Fig. 1.2B). The presence of random coils or intrinsically disordered regions in a transcription factor is a characteristic feature of a transcription factor necessary for its functionality (Garza et al., 2009). Moreover, analysis of the amino acid composition of ZSCAN4 protein revealed the presence of glutamic acid and aspartic acid in significantly larger amounts (Table 1.1), which might also be the reason for the dominance of random coils and others structures in the predicted secondary structure of human ZSCAN4 protein (Smith et al., 1996).

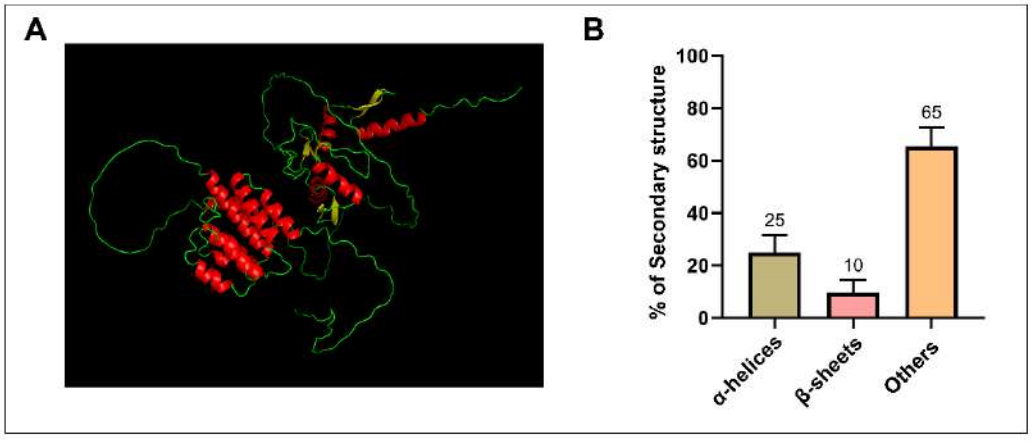


Fig. 1.2. The three-dimensional and secondary structure content of human ZSCAN4 protein. **(A)** The three-dimensional structure of the human ZSCAN4 protein was obtained from Alphafold 3 database and was visualized using BIOVIA discovery studio visualizer. **(B)** The amount of different secondary structures present in human ZSCAN4 protein predicted using different web servers, namely GOR4, PSIPRED, SOPMA, Alphafold 3.

In order to study the conservancy among ZSCAN4 protein from different species belonging to vertebrata sub-phylum, the orthologs of ZSCAN4 were obtained from the NCBI ortholog database. We found about 75 ZSCAN4 orthologs, which subsequently, were aligned using NCBI Constraint-based Multiple Alignment Tool (COBALT) (Papadopoulos and Agarwala, 2007). Using the WebLogo online tool (Crooks et al., 2004; Schneider and Stephens, 1990), the obtained multiple sequence alignment data was visualized (Fig. 1.3). Interestingly the whole SCAN domain is not conserved. Instead, it can be noticed that a part of SCAN domain at the beginning and a small part at the end of the domain is conserved among the orthologs of ZSCAN4 protein (indicated by a red solid line). In a similar manner, only certain parts of four zinc finger domains are conserved among the orthologs (indicated by a red dotted line) (Fig. 1.3).

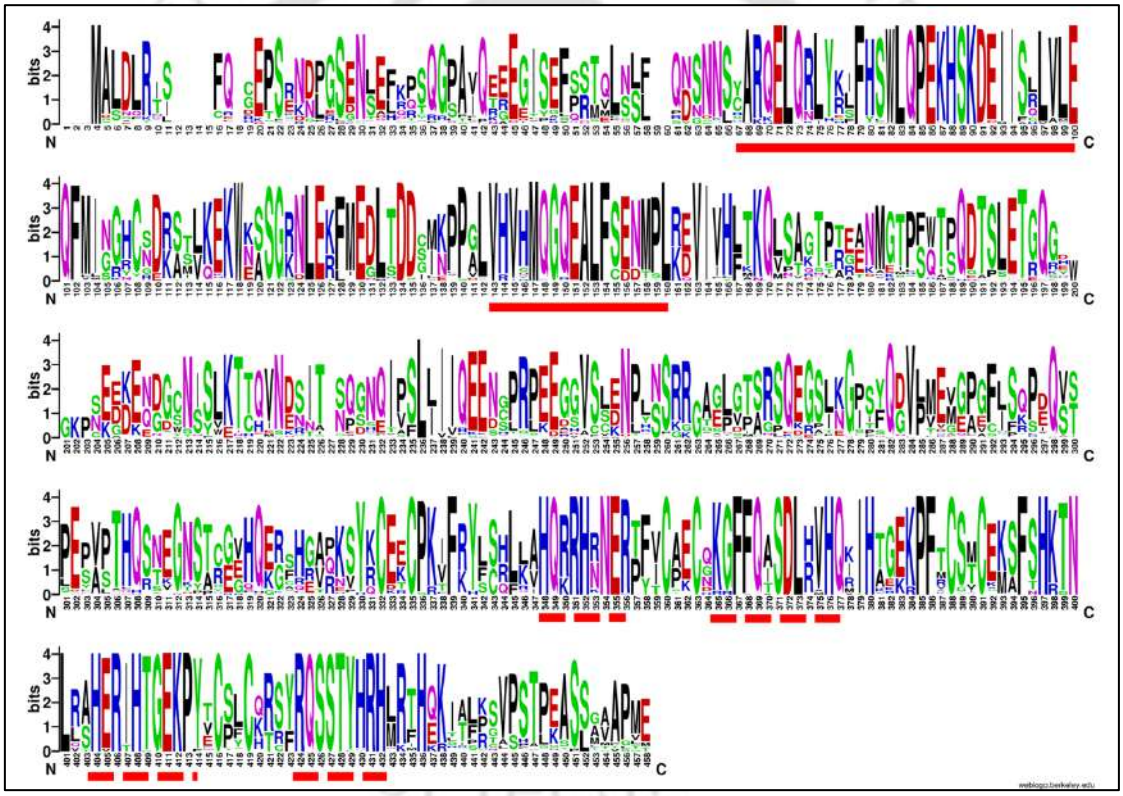


Fig. 1.3. Conservancy among ZSCAN4 orthologs. ZSCAN4 orthologs belonging to different species of vertebrata sub-phylum were obtained from NCBI ortholog database, aligned using NCBI's COBALT online tool and visualized using WebLogo online server. The height of each amino acid corresponds to its conservancy among the orthologs. The most conserved region in SCAN domain is indicated by a red solid line and in Zinc finger domains by a red dotted line.

Further, a phylogenetic tree was also constructed by the neighbour joining method with bootstrap 1000 iterations using Molecular Evolutionary Genetics Analysis (MEGA) version X software (Kumar et al., 2018). The formation of different clades represents evolutionary changes that took place among these species. The values at the nodes are bootstrap values, which shows the probability of formation of that particular clade in terms of percentage. Fig. 1.4 shows that most of the bootstrap values are well above 60, showing the high consistency of the tree. The tree was inferred with *Homo sapiens* as the reference point. It is evident that there are three distinct evolutionary clusters in the tree, denoted as A, B, and C (Fig. 1.4). Cluster A itself can be divided into three distinct groups denoted as A1, A2, and A3. The species in A1 belong to order carnivora, A2 belong to bats, and A3 belong to ungulates. Similarly, cluster B can be divided into two groups, B1 and B2. B1 consists of species from order rodents and some primates and B2 from order primates (apes and monkeys). *Homo sapiens* falls under group B2. The closest species to humans is *Gorilla gorilla*. Zscan4 protein family from mice falls under cluster C, which interestingly is completely distinct and separated from rest of the orthologs. In agreement with this, pairwise global alignment between ZSCAN4 proteins from mice and human reveals only 47% sequence similarity.

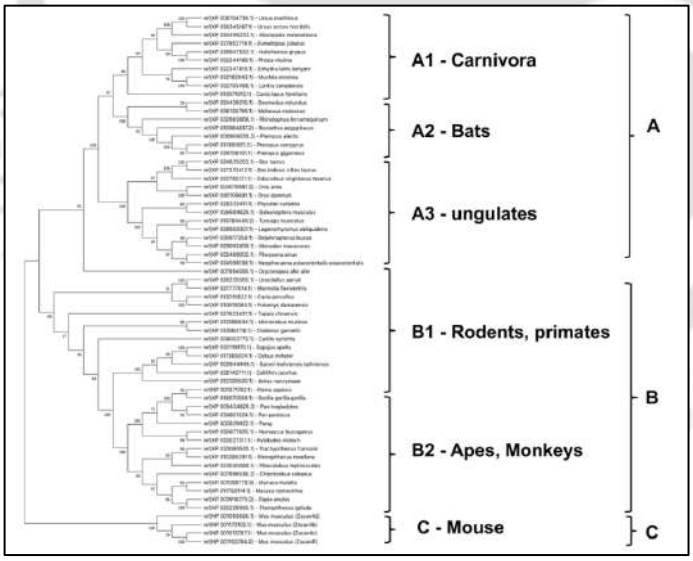


Fig. 1.4. Phylogenetic relationship among the orthologs of ZSCAN4. ZSCAN4 orthologs belonging to different species of vertebrata sub-phylum were obtained from NCBI ortholog database, aligned using NCBI's COBALT online tool. Consequently, the phylogenetic tree was constructed using MEGA X software using neighbor joining method with 1000 bootstrap iterations. The tree itself is grouped into 3 major clades depending on their class and order, having their own subgroups. The numbers at the node represent the bootstrap values.

Although the function of majority of SCAN-Zinc finger proteins is obscure, certain SCAN-Zinc finger proteins (ZNF274, Zfp369, and Zfp496) have been involved in the transcriptional regulation of specific growth factors (Nerve growth factor: Neurotrophin receptor interacting factor) and some of the auxiliary genes [alpha2(XI) collagen gene, tyrosine hydroxylase gene, c-myb, myeloperoxidase, CD34, and LDOC1] associated with cell differentiation and survival (Edelstein and Collins, 2005). Hence, ZSCAN4, which is one of SCAN4-Zinc finger proteins is speculated to have a role in cell survival and differentiation (Edelstein and Collins, 2005). Moreover, it is reported that half-life of Zscan4 protein is about 6 hours in mouse ESCs (Storm et al., 2014) and about 8.3 hours in cancer cells (Portney et al., 2018). It is reported that ZSCAN4 protein is degraded by proteasomal-mediated ubiquitination at lysine 48 instead of autophagy (Portney et al., 2018). Since ZSCAN4 expression is tightly regulated, this is crucial for maintenance of optimal level of protein inside the cell and clearance from the cell after its degradation (Portney et al., 2018).

In addition, the protein interacting partners of ZSCAN4 protein were analyzed using STRINGv11 database (Szklarczyk et al., 2019), which is depicted in Fig. 1.5. A protein essential for two-cell stage formation (Hendrickson et al., 2017), DUX4 was found to interact with ZSCAN4. It was reported that ZSCAN4, MBD3L2, LEUTX are direct targets of DUX4 (Oliva et al., 2019), and they were predicted to be interacting with one another. Moreover, DUX4 also regulates PRAMEF1 and TRIM43 (Ferreboeuf et al., 2014), which were also predicted to be interacting with ZSCAN4. The mouse ESCs express Zscan4 as a transient burst from time to time, which is referred as Z4 event (Akiyama et al., 2015; Falco et al., 2007). The gene TMEM92 is one of the genes reported to be upregulated specifically

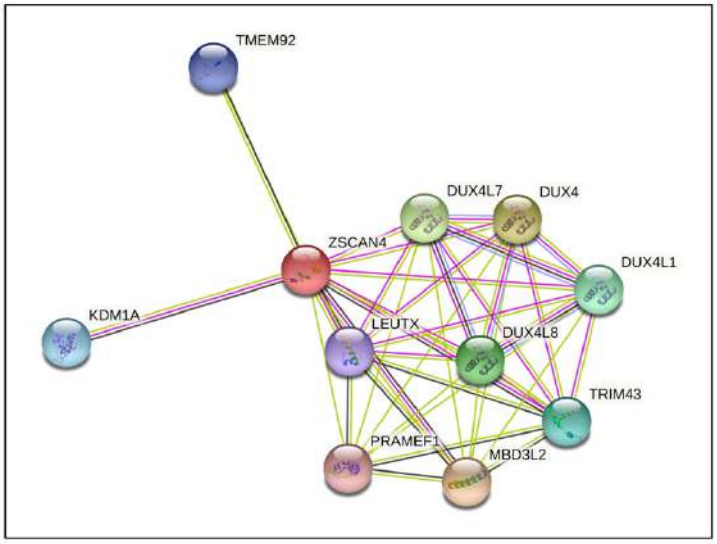


Fig. 1.5. Protein interacting partners of human ZSCAN4. The proteins that are interacting or predicted to be interacting with human ZSCAN4 protein were obtained from STRING v11 database.

during this Z4 event (Akiyama et al., 2015), which is predicted to be interacting with ZSCAN4. Furthermore, mass spectrometry data suggests that KDM1A, a chromatin repressing complex interacts directly with ZSCAN4, regulating certain areas of heterochromatin (Akiyama et al., 2015). Apart from these proteins, experimental evidences suggest that ZSCAN4 closely interacts with shelterin complex proteins, namely Trf1 and Rap1, which is speculated to be working together with ZSCAN4 in regulating telomere length (Lee and Gollahon, 2014; Lee and Gollahon, 2015).

1.4.3 Role of ZSCAN4 in embryonic development and germ cells

Zscan4 is exclusively expressed at the late two-cell stage of preimplantation embryos (Falco et al., 2007). Therefore, *Zscan4* transcripts are absent in oocytes, fertilized eggs, and in other preimplantation stages, indicating its high specificity and comparatively short half-lives. Moreover, Zscan4 expression is not observed in blastocysts, such as inner cell mass and early blastocyst outgrowth. Approximately, 90% of transcripts of the two-cell stage of mouse embryos are *Zscan4d*. Additionally, Zhang and colleagues reported expression of Zscan4 in 2-cell stage with a sharp spike in three-four cell stage of preimplantation (Zhang et al., 2006). The expression of Zscan4 is almost none before two-cell stage and after four-cell stage, once again proving a tight regulation in its expression (Zhang et al., 2006). In humans and porcine models, ZSCAN4 is expressed in two-cell and four-cell stage with sharp spike at the six-eight cell stage of embryos that is concurrent with the major tide of zygotic genome activation (ZGA) (Li et al., 2023; Vassena et al., 2011).

As mentioned earlier, a study revealed that *Zscan4* is expressed in ESCs (Falco et al., 2007). The authors also showed that *Zscan4* expression is absent in trophoblast stem cells, neural stem/progenitor cells and embryonal carcinoma cells such as P19 and F9 cells (Falco et al., 2007). Interestingly, apart from ESCs, a study showed that Zscan4 is expressed in a tiny portion of pancreas of adult mice (Ko et al., 2013).

In mouse embryonic development, the two-cell stage is one of the most important stages as the cell undergoes a drastic change in its epigenetic profile (Akiyama et al., 2015). Furthermore, it coincides with ZGA, which is a transition event, where the maternally

acquired transcripts are depleted and the zygotic genome gets activated through *de novo* transcription (Falco et al., 2007; Hamatani et al., 2004). ZGA is one of the foremost and vital events in the developmental stage of animals and major burst occurs solely in the late two-cell stage (Hamatani et al., 2004). A multitude of studies were conducted for the identification of novel genes that are expressed during ZGA and specifically during two-cell stage (*Eif1a*, *U2afbp-rs*, *Hprt*, *Pdha1*, *Prps*, *Odc*, *Cox7c* and *Zscan4*) (Falco et al., 2007; Hamatani et al., 2004; Ko et al., 2000; Wang and Latham, 1997). *Zscan4* was identified to be one of the genes and is solely expressed during ZGA, indicating its utility as a potential marker for ZGA studies (Falco et al., 2007). Beyond murine models, transient ZSCAN4 expression has also been reported during ZGA in porcine parthenogenetic embryos, suggesting a conservative role for ZSCAN4 in early embryogenesis (Li et al., 2023).

To decipher the function of Zscan4 during implantation stage of embryonic development, siRNA-mediated knockdown of Zscan4 was carried out. Knockdown revealed a delay in progression from two-cell to four-cell stage (Fig. 1.6). After the transient arrest, embryos attain the blastocyst stage. However, these blastocysts were functionally abnormal, leading to failure in implantation or proliferation in blastocyst outgrowth culture used for derivation of ESCs, pinpointing that Zscan4 is crucial for preimplantation development (Falco et al., 2007). Further exploration of functional aspect of Zscan4 uncovered its role in protection against DNA damage. Loss of Zscan4 in mouse embryos elevated DNA damage in these cells, demonstrating the crucial role of Zscan4 in DNA damage response (Srinivasan et al., 2020). It was reported that Zscan4 helps in maintaining the DNA stability by binding to $(CA)_n/(TG)_n$ rich regions in the genome that are prone to recombination and torsional strain during ZGA (Akiyama et al., 2024; Srinivasan et al., 2020). The role of Zscan4 against DNA damage is evident not only in mouse ESCs (Srinivasan et al., 2020; Tsai et al., 2023), but also in mouse iPSCs, which is discussed elsewhere (Hirata et al., 2012; Jiang et al., 2013). Since the expression spike of ZSCAN4 in both mouse and human embryonic development coincides with ZGA, it is most likely that the functionality of ZSCAN4 in maintaining DNA stability will be conserved in human embryonic development as well.

Apart from mouse ESCs and during ZGA in embryonic development, there are several reports stating the expression of Zscan4 in adult cells. In mice, it is reported that

Zscan4 protein is expressed in adult germ cells (ovaries and testes). Also, it is reported that Zscan4 expression was evident in spermatocytes and sertoli cells of testes, whereas, in females, it was observed in the germinal vesicle stage of oocytes. Specifically, Zscan4 acts

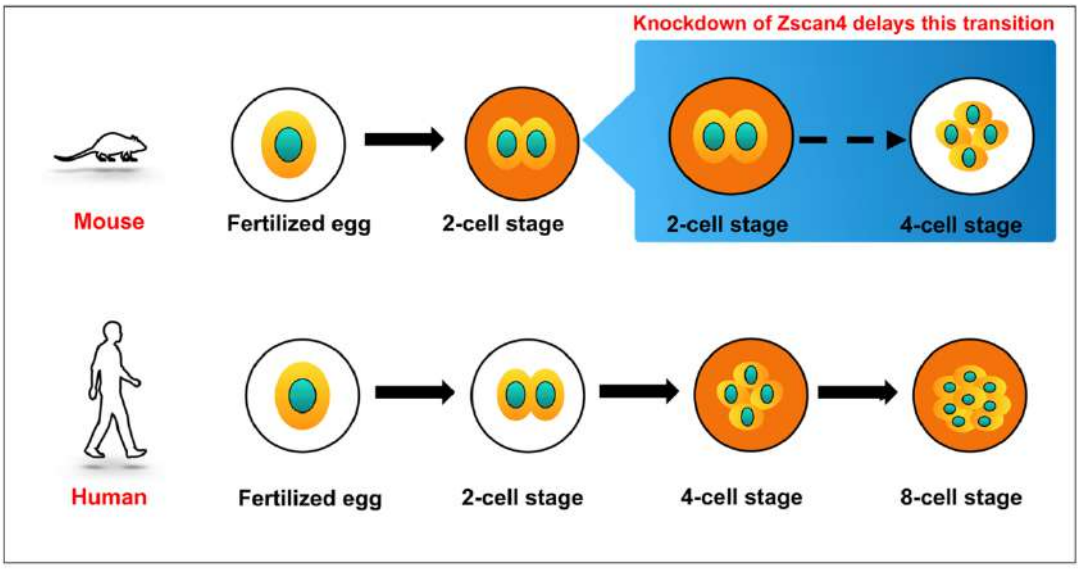


Fig. 1.6. Expression of ZSCAN4 during embryonic development. In mouse, Zscan4 is expressed at late 2-cell stage and is highlighted in orange. However, knockdown of Zscan4 delays the 2-cell stage to 4-cell stage transition. In case of humans, ZSCAN4 is expressed at 4-8 cell stage and is absent in blastocysts.

as an essential regulator to maintain centromere integrity in mouse oocyte meiosis (Choi et al., 2025). Thus, Zscan4 may have a pivotal role not only in the preimplantation stage of embryos but also in germline formation (Ishiguro et al., 2017a). In the human (adult) pancreas, a few ZSCAN4$^+$ cells are present in the islets of Langerhans, ducts, oval-shaped cells (stellate cells), and acinar cells (Ko et al., 2013). During acute pancreatitis, ZSCAN4$^+$ cells rises drastically but once inflammation ceases, the cells revert to normalcy (Ko et al., 2013). This signifies, the expression of ZSCAN4 is tightly regulated universally.

1.4.4 Role of ZSCAN4 in ESCs

In mouse ESCs, Zscan4 exhibits a highly heterogeneous “spotted” (spot-in-colony) pattern of expression (Carter et al., 2008). Moreover, *Zscan4* is exclusively expressed in only 5% of undifferentiated colonies of mouse ESCs (Falco et al., 2007). As previously described, its expression is tightly regulated such that, on a daily basis, only 3% of ESCs turned from being Zscan4$^-$ to Zscan4$^+$, while 47% of Zscan4$^+$ ESCs attained Zscan4$^-$ state, contributing to an overall of 5% of ESCs expressing Zscan4 at any given point of time

(Zalzman et al., 2010). Furthermore, *Zscan4d* expression is switched off right after the late two-cell stage, whereas *Zscan4c* and *Zscan4f* (to a little extent) expression is switched on during blastocyst outgrowth culture. Thus, Zscan4 might be regarded as a vital player in the derivation of ESCs (*in vitro*) from the inner cell mass (Carter et al., 2008; Falco et al., 2007). Notably, all cells trigger Zscan4 expression at least once by the end of nine passages (Carter et al., 2008; Falco et al., 2007).

Functionally, an important role of Zscan4 in ESCs is global derepression of both constitutive and facultative heterochromatin (Fig. 1.7). As mentioned earlier, mouse ESCs fluctuates between two transient states i.e. Zscan4$^-$ and Zscan4$^+$ and this alteration (Zscan4$^-$ to Zscan4$^+$) causes immediate activation of heterochromatin derepression that eventually leads to clustering of pericentromeric heterochromatin around the nucleolus (Akiyama et al., 2015). Moreover, this transient outburst of transcription of Zscan4 also leads to epigenetic changes involving histone modifications, especially elevated levels of H3K27ac and reduced levels of DNA methylation, which in turn derepresses heterochromatin, making it more accessible to perform transcription (Akiyama et al., 2015). Notably, Zscan4 forms complexes with both activating chromatin remodeling complexes such as SWI/SNF and HATs, and repressing chromatin remodeling complexes such as LSD1/KDM1A, HDAC1, HDAC2, and NuRD that synchronously come together on heterochromatin (Akiyama et al., 2015; Ishiguro et al., 2017b). Formation of these multiprotein complexes leads to rapid derepression and repression of heterochromatin. The above-mentioned, transient expression of Zscan4 induces transcription of genes due to active conformation of heterochromatin, which can be considered a usual part of mammalian development and tissue homeostasis, but immediately revert to its silent conformation (Akiyama et al., 2015). More importantly, some of these epigenetic changes that occurs due to the derepression of heterochromatin are mostly harmful for the cells. So, the cells try to negate the epigenetic changes affecting their survival. This better explains the reason behind the briefness of this transition phase of ESCs and for not being in Zscan4$^+$ state for a prolonged period (Ko, 2016). Moreover, Zscan4 can also form complexes with KAP1, a central protein involved in regulation of heterochromatin structure (Akiyama et al., 2015). Taken entirely, Zscan4$^+$ cells continuously undergo chromatin remodeling and forming complexes with chromatin remodelers / epigenetic modifiers. Thus, transient expression of Zscan4 is crucial for epigenetic regulation of ESCs.

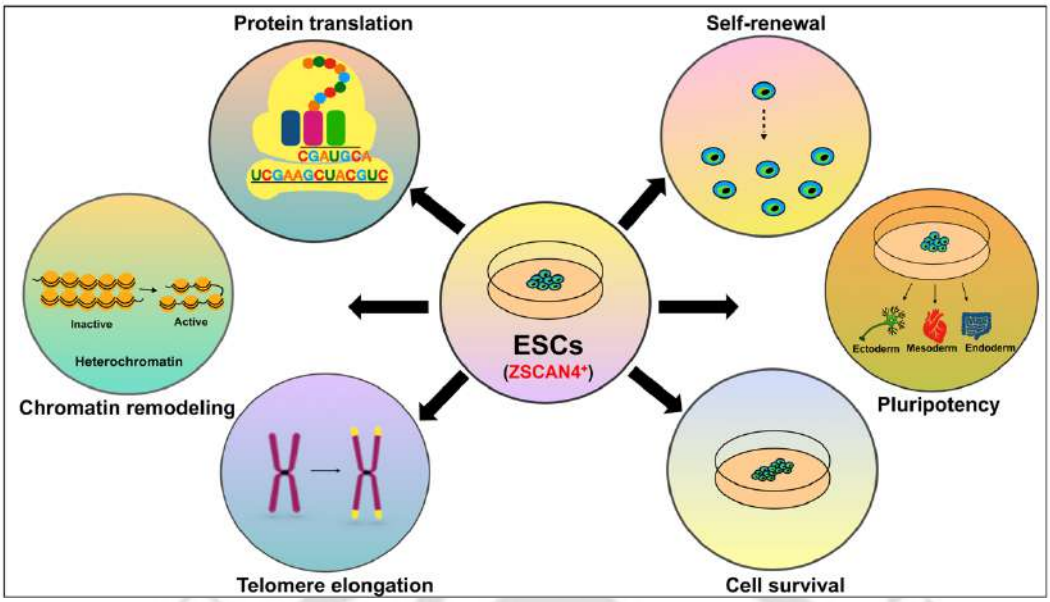


Fig. 1.7. Role of ZSCAN4 in ESCs. ZSCAN4 is expressed transiently in ESCs. It has multidimensional roles such as repression of protein translation, elongation of telomere, remodeling of chromatin structures, enhancing cell survival associated with DNA damage and enhancing the self-renewal and pluripotency of ESCs.

Another critical role of Zscan4 is inhibition of global translation (Fig. 1.7). In two-cell stage embryo, expression of eukaryotic translation initiation factor 1A (*Eif1a*) was increased with the increase in expression of Zscan4, and therefore ZSCAN4 was thought to be involved in the process of translation (Hamatani et al., 2004; Wang et al., 2004; Zeng et al., 2004). Interestingly, another study reported that it is not overexpression of *Eif1a* but *Eif1a*-like genes, which is responsible for global repression of protein translation. Besides, *Eif1a*-like genes are highly expressed not only in two-cell stage embryos but also during ZSCAN4$^+$ state of ESCs (Hung et al., 2013). Importantly, the suppression of protein synthesis was observed in ZSCAN4$^+$ state of ESCs. This indicates that *Eif1a*-like genes and ZSCAN4 may play a role in global translation repression in ESCs.

A study reported that two-cell stage of mouse embryos express both MERVL (murine endogenous retroviruses with leucine tRNA primer) and ZSCAN4 (Eckersley-Maslin et al., 2016; Macfarlan et al., 2011). Similar scenario is also observed in case of mouse ESCs. In MERVL$^+$ZSCAN4$^+$ cells, demethylation of DNA occurs globally. As soon as cells exit this transient phase, the methylation is restored. However, the genomic imprints are lost, causing inhibition of translation of DNA methyltransferases (Dnmts). These

studies emphasize that ZSCAN4 plays a vital role in the repression of global protein synthesis. However, further exploration is required to completely understand the other cellular and metabolic changes taking place during this process (Akiyama et al., 2015; Eckersley-Maslin et al., 2016; Ishiuchi et al., 2015; Macfarlan et al., 2011). The role of ZSCAN4 in DNA demethylation is reported in other models like porcine models as well (Li et al., 2023).

ESCs have two remarkable characteristics, namely self-renewal and pluripotency (Niwa, 2007). In ESCs, overexpression of *Zscan4c* enhances self-renewal. Also, Zscan4 activation is pivotal as ESCs tend to lose their ability of unlimited proliferation (Zalzman et al., 2010). It is reported that ESCs regulates self-renewal via glycogen synthase kinase (GSK) and Phosphatidylinositol 3-Kinase (PI3K) pathway (Bone et al., 2009; Paling et al., 2004). In accordance with this, Zscan4 expression was enhanced with the inhibition of GSK. Importantly, Zscan4 was recognized as a downstream target of PI3K. In addition, P110α is a pivotal catalytic isoform of PI3K; and inhibition of P110α downregulates the expression of Zscan4 (Kumpfmueller, 2011; Storm et al., 2014). Additionally, vitamin A was reported to upregulate the expression of ZSCAN4 to increase self-renewal in mouse ESCs, through the activation of PI3K/AKT cascade (Chen and Khillan, 2008; Chen and Khillian, 2010; Sharova et al., 2016). All these data indicate that activation of PI3K cascade can activate the expression of ZSCAN4 (Kumpfmueller, 2011; Storm et al., 2014). *Zscan4c* also plays a crucial role in maintaining the pluripotency properties of ESCs. During the long-term culturing of ESCs, the potential to develop into an embryo declines gradually (Amano et al., 2013). Hence, rejuvenation of ESCs is required, which can be acquired by manipulation of Zscan4 expression. In ESCs, the potential hike after repeated activation of Zscan4 or exposure of Zscan4 protein increases Zscan4^{+} state that eventually improves the quality of ESCs (Amano et al., 2013). All these studies suggest that Zscan4 has a very significant role in maintaining self-renewal and pluripotency property in mouse ESCs (Fig. 1.7).

One of the main reasons behind the transient expression of Zscan4 is that its continuous expression will lead to hyper-recombination of telomeres, which will eventually cause cellular senescence over time (Dan et al., 2014). In ESCs, Zscan4 is not only involved in the stabilization of the genome but also in the elongation of telomeres (Zalzman et al., 2010). Telomeres are heterochromatin structures with a repetitive stretch of 5'-TTAGGG-

3' (in humans) and specific proteins that protect the chromosome ends from damage and aging (Blasco, 2005). Appropriate length of telomere is indispensable for self-renewal and pluripotency of ESCs (Dan et al., 2017; Dan et al., 2022). Zscan4^{+} (transient) state of ESCs is associated with rapid telomere extension by telomere recombination and upregulation of genes that are unique for meiosis-specific homologous recombination. These genes are co-localized with Zscan4 on telomeres. In ESCs, knockdown of Zscan4 causes shortening of telomeres, increase in karyotypic abnormalities, spontaneous sister chromatid exchange (SCE), and reduced cell proliferation, until cells reach senescence (Zalzman et al., 2010). The impact of knockdown is observed gradually and ultimately leads to apoptosis and depletion of cell proliferation after seven to eight passages. Alternative lengthening of telomeres (ALTs) is an alternative mechanism for telomeres elongation, which is independent of telomerase (Bryan et al., 1995), and depends solely on homologous recombination (Bryan et al., 1997). It is reported that ZSCAN4 plays a vital role in regulation of telomere length via ALT pathway (Dan et al., 2022). Activation of ALT pathway has been observed in mouse ESCs following the deletion of TERC complex (Dan et al., 2022). In these telomerase-deficient cells, significant upregulation of ZSCAN4 has been reported, providing compelling evidence for its involvement in ALT mechanism to regulate telomere length (Dan et al., 2022). Another study reported Ddb1- and Cul4-associated factor 11 (Dcaf11) is required for activation of ZSCAN4 in early embryos and ESCs (Le et al., 2021). Interestingly, Dcaf11 is involved in ALT-mediated telomere lengthening (Le et al., 2021). Hence, ZSCAN4 plays a pivotal role in maintaining telomere elongation, most likely through ALT-mediated pathway.

Furthermore, in mouse ESCs, Zscan4 inhibits DNA methylation at the global level to promote telomere elongation (Dan et al., 2017). Furthermore, ZSCAN4 directly represses Dnmt1 and Uhrf1, which are major units of the DNA methylation complex. Mechanistically, Zscan4 is required for recruitment of Dnmt1 and Uhrf1, and promotion of degradation of both these components via ubiquitination pathway. Blockade of DNA demethylation leads to prohibition of telomere elongation, which is directly linked to expression of Zscan4 (Dan et al., 2017). As stated before, mouse ESCs undergo dramatic epigenetic changes to regulate transient activation and repression of specific heterochromatin regions, a process which might be inherent part of a unique mechanism for the maintenance of genome stability (Akiyama et al., 2015). Hence, ZSCAN4 plays a crucial role in telomere elongation and genome stability.

Intriguingly, lengthening of telomeres by ZSCAN4 follows ALT mechanism, which is independent of telomerase activity as mentioned earlier (Bryan et al., 1995). It is reported that Zscan4 plays a pivotal role in the recruitment of other proteins (SPO11, DMC1, SMC1β and TRF1) to chromatin to facilitate telomere recombination (Zalzman et al., 2010). Furthermore, karyotypic abnormalities such as random chromosome fusions and deletions were observed solely due to knockdown of Zscan4. In ESCs, Zscan4 decreases the chances of non-telomeric SCE and inhibits spontaneous SCE (Zalzman et al., 2010). Moreover, overexpression of Zscan4 induces homologous recombination of genes associated with meiosis. Further, SPO11, an enzyme that facilitates double-strand DNA breaks along with DMC1 (RecA homolog), which is a recombinase is requisite for meiotic recombination (Keeney et al., 1997; Mahadevaiah et al., 2001; Reinholdt and Schimenti, 2005). It has been reported that Zscan4 foci together with SPO11 and DMC1 are located on telomeres. In addition, Zscan4 foci and SPO11 are co-localized in the G2 phase of the cell cycle. Taken together, Zscan4 in ESCs cultures not only helps in inhibition of cell crisis but also in maintenance of normal karyotype even after multiple cell divisions.

Similar to ESCs, parthenogenetic ESCs also showed telomere elongation with increased expression of ZSCAN4. Parthenogenetic ESCs are generated from parthenogenetically activated oocytes (Yin et al., 2014). On other hand, in mouse ESCs cultures, feeders such as mouse embryonic fibroblasts (MEFs) promotes the expression of Zscan4 by the production of Fst1 and BMP4. Thus, indicating the role of feeders in the maintenance of telomeres length and genomic stability and eventually self-renewal and pluripotency of ESCs (Guo et al., 2018).

Under feeder-free milieu, mouse ESCs are expressing Zscan4 in the presence of LIF and serum. However, ZSCAN4 expression drops drastically after the induction of differentiation around day four. ZSCAN10 then regulates the expression of ZSCAN4 and several other genes crucial for differentiation in ESCs (Zhang et al., 2006). Both ZSCAN4 and ZSCAN10 share a common domain and regulates pluripotency in both mouse and human ESCs (Wang et al., 2007; Yu et al., 2009; Zhang et al., 2006). It is reported that retinoic acid can induce differentiation and upregulate ZSCAN4 expression, which contrasts with previous reports (Tagliaferri et al., 2016). Taken together, induction of differentiation in ESCs culture can upregulate or downregulate the expression of Zscan4,

depending on the treatment of differentiation-inducing agents, and requires further detailed investigation.

The identification of interacting partners can be beneficial in deciphering the function associated with ZSCAN4. Firstly, Lysine-specific histone demethylase 1A (LSD1) plays a crucial role in maintaining the right balance of epigenetic profile as already stated, thereby, enhancing the interaction between Zscan4 and other binding partners. Right before telomere elongation, heterochromatic marks at telomeres are erased by LSD1 (Storm et al., 2014; Wang et al., 2007). Also, C-terminal binding protein 2 (Ctbp2) interacts with Zscan4 by the association of Ctbp PXDLS binding motifs that are well-established in ZSCAN4 family and thus can act as a transcriptional repressor (Balasubramanian et al., 2003; Storm et al., 2014; Wang et al., 2007). Lastly, Histidine triad nucleotide-binding protein 1 (HTNT1) associates with Zscan4 and plays a crucial role in cell survival. Oxidative stress and DNA-damaging agents upregulate the expression of Zscan4. Thus, the interaction between HTNT1 and Zscan4 protects the cell from damage thereby increasing cellular viability (Kumpfmueller, 2011). In ESCs, *Tbx3* is also involved in upregulation of ZSCAN4^{+} state. Thus, both Tbx3 and Zscan4 are associated with self-renewal and telomere maintenance in ESCs (Dan et al., 2013). Hence, Zscan4 is reported to interact with a number of partners such as LSD1, Ctbp, HTNT1 and TBX3 for the maintenance of self-renewal of ESCs and cell viability.

Interestingly, *Zscan4* gene is a suitable candidate to be explored as its expression is confined to a subpopulation of ESCs (Paduano et al., 2013). The expression pattern of Zscan4 inside a stem cell colony is uneven and appears towards the center of the colony. Furthermore, cells expressing Zscan4 tend to form clumps (Paduano et al., 2013). The cell location and its pattern inside a colony can provide a hint for better interpretation of cell differentiation, meta-stable profile of cells, and cell analysis helps to hypothesize biological significance of cells exhibiting specific pluripotent-like features (Paduano et al., 2013). Furthermore, this setup can be helpful to decipher the role of Zscan4 in the process of morphogenesis. Thus, Zscan4 has various roles to play in ESCs, and more exploration is required to attain a deeper understanding of its functionality.

Lastly, Zscan4 is also entailed in cell survival and response to DNA damage in ESCs (Fig. 1.7). Reports suggests that ZSCAN4 interacts with PARP1 to promote the repair of double stranded breaks (DSBs) via NHEJ pathway (Tsai et al., 2023). Addition of zeocin

and cisplatin (DNA damage-inducing agents) are reported to increase the expression of Zscan4 in a dose-dependent fashion (Storm et al., 2014). These two compounds are known to induce double-stranded DNA breaks, leading to cell cycle arrest in G2/M. In addition to this, expression of Zscan4 is increased in late S-phase/early G2 of the cell cycle, thereby pinpointing a selective function for Zscan4 during the G2 checkpoint. Other similar studies also show the involvement of ZSCAN4 in DNA damage response during meiotic progression (Choi et al., 2025). Altogether, Zscan4 responds to DNA damage of ESCs and aids in the cell survival.

1.4.5 Role of ZSCAN4 in iPSCs

Studies have reported that ZSCAN4 has a crucial role in reprogramming and maintenance of iPSCs (Fig. 1.8). Similar to ESCs, ZSCAN4 is involved in stabilizing and maintaining length of telomeres in iPSCs (Wang et al., 2012). As mentioned earlier, telomere is a short, repeating DNA sequence at the end of each chromosome having a sequence of 5'-TTAGGG-3' (in humans) (Blasco, 2005). The telomere helps in chromosomal stability and prevents the ends of chromosome from fusing together (De Lange, 2009). Moreover, the telomere shortens with each replicative cycle because of the semi-conservative mechanism of replication (Nandakumar and Cech, 2013). As the telomeres shortens, the cells proceed towards senescence. Since stem cells have the ability to self-renew indefinitely, their telomere length has to be regulated (Hiyama and Hiyama, 2007). Primarily, in iPSCs, the length of telomeres is regulated either by telomerase-dependent or -independent mechanism (Allsopp, 2012). The telomerase is a ribonucleoprotein that has two functional components, namely telomerase reverse transcriptase (tert) and telomerase RNA component (terc) (Lingner et al., 1997; Palm and De Lange, 2008). The telomerase-independent mechanism is called alternate lengthening of telomere, which involves recombination (Cesare and Reddel, 2010). The transcription factor ZSCAN4 is reported to be one of the members involved in ALT mechanism of telomere elongation (Wang et al., 2012; Zalzman et al., 2010). ZSCAN4 regulates the telomere by a recombination mechanism called telomere sister chromatid exchange (T-SCE) (Wang et al., 2012; Zalzman et al., 2010). Studies show that telomerase is dispensable for reprogramming but is essential for maintenance of iPSCs (Fu et al., 2018; Wang et al., 2012). In accordance to this, even in the absence of telomerase, fibroblasts were reprogrammed to iPSCs successfully, with pluripotency characteristics (Wang et al., 2012). Interestingly, in the absence of telomerase, the expression of ZSCAN4 was upregulated in

iPSCs, indicating that the ALT mechanism was activated to regulate telomeres (Wang et al., 2012). This activation of ZSCAN4 for telomere regulation can also be achieved by crotonylation of histones (Fu et al., 2018), or by Dcaf11 as mentioned earlier (Le et al., 2021), which further improves the quality of iPSCs. Moreover, knockdown of both telomerase and ZSCAN4 in iPSCs resulted in shorter telomere length, with increasing passages compared to wild-type iPSCs (Wang et al., 2012). This telomere recombination by ZSCAN4 requires relatively relaxed chromatin and global DNA demethylation (Eckersley-Maslin et al., 2016; Wang et al., 2012). Similar to mouse ESCs, in case of reprogramming also, the global DNA demethylation is brought about by downregulation of Dnmts by activation of ZSCAN4 along with MERVL transcriptional network (Eckersley-Maslin et al., 2016; Wang et al., 2012). Moreover, studies show that ZSCAN4 interacts with Tet2 (a protein belonging to TET family) and promotes DNA demethylation, thereby increasing ZSCAN4-mediated T-SCE (Cheng et al., 2020; Lu et al., 2014). This ZSCAN4-Tet2 interaction is also found to be regulating proteasome activity and cellular metabolic machinery, shifting from oxidative phosphorylation to glycolysis, which is crucial for somatic cell reprogramming to iPSCs (Cheng et al., 2020; Panopoulos et al., 2012).

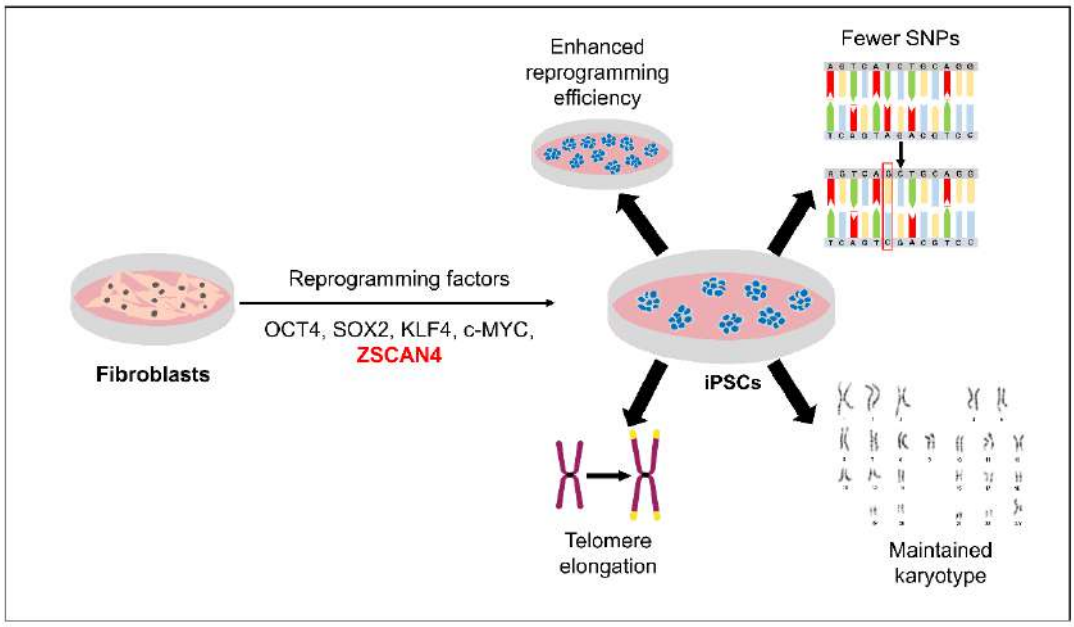


Fig. 1.8. Role of ZSCAN4 in iPSCs. Enumerating the advantages of inclusion of ZSCAN4 in the reprogramming cocktail to generate iPSCs. Inclusion of ZSCAN4 along with core reprogramming factors results in enhanced efficiency of reprogramming, fewer single nucleotide polymorphisms in resulting iPSCs.

Another significant role of ZSCAN4 is enhancing the efficiency of reprogramming of iPSCs. Studies show that inclusion of ZSCAN4 in the reprogramming cocktail along with Yamanaka factors (OSKM), enhanced the reprogramming efficiency significantly (Hirata et al., 2012; Jiang et al., 2013). Also, treatment of MEFs with cell extracts of mouse ESCs overexpressing ZSCAN4 enhanced the efficiency of the reprogramming (Kwon et al., 2015). Moreover, the addition of ZSCAN4 along with OSKM lead to increased genome stability, longer telomeres, better karyotypes, and quality iPSCs, as deduced by tetraploid complementation assays (TCA) compared to OSKM-mediated iPSCs (Hirata et al., 2012; Jiang et al., 2013). Also, fewer single nucleotide variations were reported in iPSCs when ZSCAN4 was included in the reprogramming cocktail (Su et al., 2013).

Telomere regulation is fairly important to generate quality iPSCs that gives higher percentage of live born pups in TCA. Telomerase-dependent regulation mechanism is slower and requires a large number of cell divisions (Siegl-Cachedenier et al., 2007). Inclusion of ZSCAN4 enhances the quality of iPSCs by regulating telomere length rapidly and giving rise to higher percentage of live born pups in TCA (Jiang et al., 2013). Furthermore, during OSKM-mediated reprogramming, the expression of ZSCAN4 is observed in a fraction of cells only during the later stages of reprogramming (Hirata et al., 2012; Jiang et al., 2013). In contrast to other factors that requires 8-10 days of forced expression, expression of ZSCAN4 is only required during the initial stages of reprogramming (first 4 or 7 days) to generate iPSC colonies efficiently (Hirata et al., 2012). This activation of ZSCAN4 can also be brought about by inclusion of small molecules in the reprogramming cocktail (Park et al., 2015). The oncogenic factor c-MYC, in Yamanaka cocktail can be replaced with ZSCAN4, without compromising the reprogramming efficiency (Hirata et al., 2012). Moreover, the inclusion of ZSCAN4 activated early embryonic genes, namely Prame16, Trim43a, Trim43b, Tcstv1 and many more during reprogramming (Hirata et al., 2012). For this tight regulation to be possible, an efficient regulatory system is essential both at gene and protein level. At gene level, ZSCAN4 is negatively regulated by Rif1 (Dan et al., 2014) and at protein level by E3 ubiquitin ligase RNF20 (Portney et al., 2018). Thus, ZSCAN4 assists in telomere regulation in a telomerase-independent manner and also helps in the generation of quality, genetically stable iPSCs during reprogramming with high efficiency.

1.4.6 Role of ZSCAN4 in Cancer

The role of ZSCAN4 in maintaining telomeres, which paves way for the indefinite self-renewal capability of stem cells is emphasized in earlier sections. Cancer cells are also similar to stem cells in this aspect. Therefore, shortly after the role of ZSCAN4 was identified, researchers were keen to deduce its role in the immortality of cancer cells. Various functions of ZSCAN4 in cancer are depicted in Fig. 1.9. The first study in 2014 studied the expression of ZSCAN4 in telomerase positive and ALT pathway cancer cells and found out that ZSCAN4, indeed, has a role in telomere maintenance of cancer cells in a telomerase-independent manner (Lee and Gollahon, 2014). It has been further confirmed that ZSCAN4 has a pivotal role to play in lengthening of telomere in non-seminoma testicular germ cell tumors (Sun et al., 2019), most likely through an indirect interaction with TERF1 (Lee and Gollahon, 2015). Moreover, it has been found that ZSCAN4 interacts with two of the shelterin complexes, Rap1 and Trf1, to perform this function (Lee and Gollahon, 2014; Lee and Gollahon, 2015). The shelterin complex is a six-subunit complex protein that plays a significant role in telomere maintenance (De Lange, 2009). Furthermore, it has been found that the expression of ZSCAN4 is upregulated in response to DNA damage induced by genotoxic chemotherapeutic drugs in cancer cells

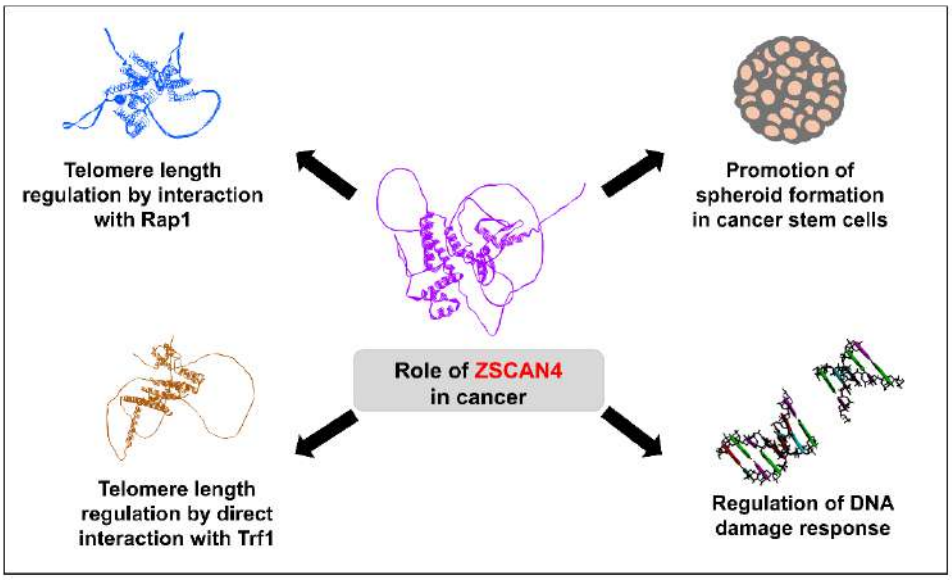


Fig. 1.9. Role of ZSCAN4 in cancer. Demonstration of various roles of ZSCAN4 in cancer cells. ZSCAN4 is involved in regulating telomere length in case of cancer by interaction with two of the shelterin complexes trf1 and rap1. Apart from this, ZSCAN4 is also reported to enhance spheroid formation in a cancer stem cell and regulates DNA damage response (The three-dimensional structures of Trf1, Rap1 and ZSCAN4 shown in the figure are obtained from Alphafold database).

(Zhang et al., 2018). This upregulation is also essential for sustaining senescence-associated secretory phenotype, which does not favor long term survival of cancer patients (Zhang et al., 2018). On top of all these, ZSCAN4 alters the epigenetic state and facilitates chromatin remodeling in cancer stem cells (CSCs) to promote CSC phenotype (Portney et al., 2020). Also, it enables spheroid formation of CSCs and enhances expression of some of the CSC (BMI1 and CD44) markers, enumerating the usage of ZSCAN4 as a potential marker for CSCs (Portney et al., 2020). Moreover, ZSCAN4 also enhances expression of certain pluripotency-associated genes in cancer cells, namely OCT4, SOX2, KLF4, and NANOG, by facilitating a functional histone hyperacetylation of histone H3 (Portney et al., 2020). All these studies suggest that ZSCAN4 has a prominent role in maintaining the tumorigenic population of CSCs and can be a likely therapeutic target in cancer.

The multifaceted regulatory role of ZSCAN4 is extensively studied in murine models and malignant cancer cell lines as discussed in very detail in chapter 1. Extensive literature suggests that ZSCAN4 vital roles in the regulation of transcription, translational inhibition, chromatin remodeling, reprogramming enhancer, pluripotency preservation, telomeres maintenance, genomic stabilization, cell differentiation, and survival in stem cells. Despite its proven efficacy as a reprogramming enhancer in mouse systems, specifically in generating genomically stable iPSCs, there is a notable absence of data regarding ZSCAN4 in human reprogramming or human iPSCs. One of the major roadblocks to clinical translation of iPSCs is the genomic instabilities associated with them. Given that ZSCAN4 mitigates these instabilities in murine models, it represents a high-priority candidate for investigating similar or potentially superior roles in human cells. To elucidate the functional necessity of ZSCAN4 in human systems, loss-of-function analysis is required. Precise genome editing serves as an efficient tool for the definitive correlation between a gene and its phenotypic and genotypic output. Therefore, this study utilizes the CRISPR-Cas9 genome editing toolbox to generate homozygous ZSCAN4 knockout human cells, specifically human iPSCs. This approach aims to provide a comprehensive understanding of the vital role of the gene in human cellular homeostasis and reprogramming efficiency.

1.5 Genome editing

Genome editing is a powerful technique that gained popularity after the advent of nuclease-based, target-specific genome editing methods. Traditionally, targeted gene inactivation or modification was achieved through homologous recombination. However, its low efficiency limited its practical application (Gaj et al., 2013). Eventually, the idea of introducing double strand breaks (DSBs) in the genome using targeted nucleases emerged, enabling more efficient and precise genome modification (Gaj et al., 2013; Kim et al., 2009). Once a DSB is introduced in the genome, cell's machinery activates the DNA damage response (DDR) to repair the DSB using one of the two pathways, namely homology directed repair (HDR) or non-homologous end joining (NHEJ) (Fig. 1.10). The HDR pathway leverages the sister chromatid as a repair template to facilitate highly precise and intentional genomic modifications (Yang et al., 2020). In laboratory settings, a donor plasmid is provided having the insert of interest flanked by specific left and right homologous arms is required (Fig. 1.10), which will be inserted in the target site where DSB was made (Gaj et al., 2013; Kim et al., 2009). This is called gene insertion, and it has wider range of applications in genome editing (Kim et al., 2009; Yang et al., 2020). NHEJ is an error prone pathway that seals the repair by joining the ends with little or no homology, leading to insertions or deletions (Fig. 1.10) (Kim et al., 2009; Yang et al., 2020). This is called gene inactivation or gene knockout (Kim et al., 2009). It is reported that frequency of HDR pathway is very low in comparison to NHEJ pathway. Moreover, in case of a DSB, mammalian cells prefer NHEJ pathway to repair the DSB over HDR pathway (Yang et al., 2020). This is because, NHEJ pathway is active throughout the cell cycle except in mitosis in contrast to HDR pathway whose activity is restricted only during S/G2 phase. Further repairing the DSB via NHEJ pathway is quicker than the HDR pathway (Yang et al., 2020).

In order to study and understand the vitality of a gene present in an organism, it is permanently eliminated or inactivated using the above-mentioned nuclease-based genome editing methods. There are various established methods of these nuclease-based genome editing methods that are both specific to the target and efficient. Among the many methods, three widely used and explored methods are Transcription Activator-Like Effector Nucleases (TALENs), Zinc Finger Nucleases (ZFNs), and Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR associated protein 9 (CRISPR-Cas9).

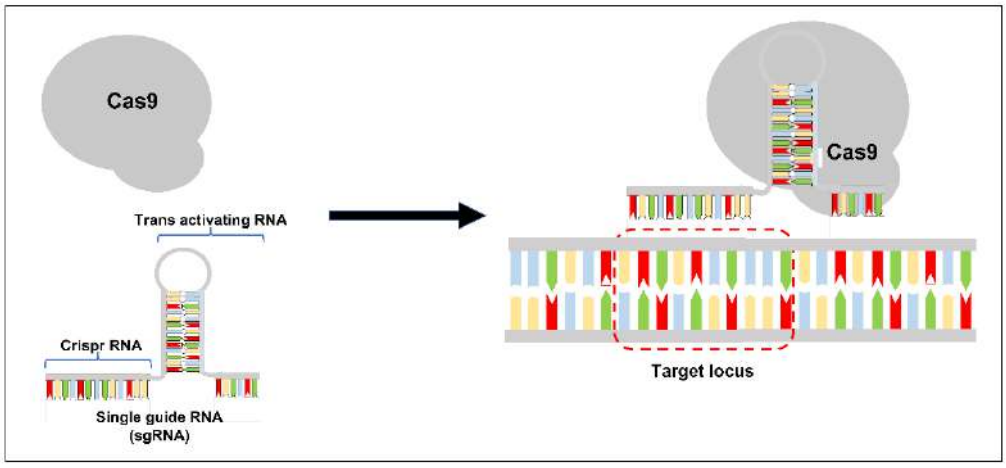


Fig. 1.10. Schematic depicting CRISPR/Cas9 toolbox. The schematic showing assembly of sgRNA and Cas9 nuclease to induce a double stranded break at the target locus.

1.5.1 Zinc Finger Nucleases

ZFNs are artificial, programmable nucleases that can be used to modify the region of interest in the genome (Miller et al., 2007; Urnov et al., 2005). ZFNs has two different components that aid in its specific, targeted activity. Firstly, it has a zinc-finger protein consisting of a C2H2 zinc finger DNA binding domain, which binds to the specific target region in the genome. Secondly, it has a nuclease domain of *Fok*I restriction enzyme, which introduces a double stranded break in the target region (Ochiai and Yamamoto, 2017). *Fok*I must dimerize for its nuclease activity. Therefore, two ZFN monomers are required to perform the target specific nuclease activity (Kim et al., 2009). Each zinc finger can recognize and bind to a 3-bp sequence in the major groove of the DNA making its activity specific (Gaj et al., 2013; Khan, 2019; Ochiai and Yamamoto, 2017). Generally, three to six zinc fingers are used identify and bind to a target region of 9 to 18 bp (Gaj et al., 2013; Kim et al., 2009). Although ZFNs can target the specific region in the genome with high specificity (Gaj et al., 2013; Kim et al., 2009), there are certain disadvantages associated with this technique. As mentioned earlier, Zinc fingers can recognize only a 3-bp sequence, limiting its suitability to target any region in the genome (Kim et al., 2009). Further, designing of ZFNs for a particular target area required an extensive screening process that is both time consuming and costly. Moreover, off-target effect of ZFNs is a major concern leading to unwanted modifications in regions other than the target region (Kim et al., 2009).

1.5.2 Transcription Activator-Like Effector Nucleases

Another widely used method to modify the genome is TALENs. TALENs are very much similar to ZFNs where it contains a DNA-binding protein domain derived from a plant pathogenic *Xanthomonas* species and the nuclease domain of *Fok*I enzyme (Gaj et al., 2013; Kim et al., 2009). TALEs are naturally occurring proteins, which are composed of repetitive arrays of 33 to 35 amino acids. The amino acid present in 12th and 13th position in each repeat can recognize and bind to a specific nucleotide (Kim et al., 2009). These two amino acids are called repeat-variable di-residues (RVDs) (Gaj et al., 2013). By modifying these RVDs to different combinations, all the four nucleotides can be targeted, making the TALENs specific to the target of interest (Gaj et al., 2013; Kim et al., 2009). Four RVDs, namely Asn-Ile, Asn-Gly, Asn-Asn, His-Asp can target the four nucleotides adenine, thymine, guanine, cytosine, respectively (Kim et al., 2009). Although TALENs are an efficient, highly specific genome editing method, there some limitations associated with them. To enhance the specificity of the TALENs, upto 20 tandem TALEs are constructed that can recognize and bind to a 20 bp target region (Kim et al., 2009). The designing and construction of DNA sequences that code for these TALEs is a labor intensive and costly process. Further, all the 20 repeats of TALEs will have a similar, homologous DNA coding sequence, which tends to recombine with each other in the cells (Kim et al., 2009). Moreover, there is a requirement of the presence of thymine at the 5' end of the target region for the TALENs to bind that makes it applicability limited (Gaj et al., 2013; Kim et al., 2009).

1.5.3 Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR associated protein 9

CRISPR-Cas9 is an RNA-based genome editing method that gained traction after the revolutionary study published in 2012 (Jinek et al., 2012). The CRISPR-Cas9 system is adapted from adaptive immunity of lower prokaryotes like bacteria and archaea. Briefly, whenever a specific bacterium encounters a foreign DNA like plasmid or invading phages, it fragments the foreign DNA or the genetic material from the phages and inserts these short fragments in their own genome. These small fragments are inserted in a region in the genome called CRISPR array, which contains identical direct repeats, and the fragments are inserted in between these repeats that are called as spacers (Barrangou et al., 2007; Doudna and Charpentier, 2014; Jinek et al., 2012). These spacer regions are novel, highly conserved regions in phages' genome that serves as a memory in adaptive immunity for

future phage attacks (Jinek et al., 2012). This phase is regarded as the adaptation phase. Subsequently, the spacer regions are transcribed within the bacteria to generate precursor CRISPR RNA (pre-crRNA) that are then processed to generate CRISPR RNAs (crRNAs). This processing of pre-crRNA to crRNA is triggered by trans-activating crRNA (tracrRNA) that is complementary to the pre-crRNA. The mature crRNA and tracrRNA forms a complex by base pairing with each other (Fig. 1.11) and this dual RNA complex is essential for the recruiting and activating the nuclease (Doudna and Charpentier, 2014; Gaj et al., 2013; Jinek et al., 2012). This phase is called the expression and processing phase, and it is specific for Type II CRISPR/Cas systems. Other types of CRISPR/Cas system adapt different mechanisms (Jinek et al., 2012). Third and final phase is called the interference phase where the invading phage or foreign DNA is cleaved with the help of the recruited nuclease (Gaj et al., 2013; Jinek et al., 2012). The crRNA-tracrRNA complex then associates itself with Cas9 nuclease. This whole complex then searches for the foreign DNA or phage genome, which is complementary to the crRNA (Jinek et al., 2012). Once complementary region for the crRNA is identified, Cas9 nuclease makes a DSB in the target exactly 3 bps upstream of a site called protospacer adjacent motif (PAM) (Doudna and Charpentier, 2014; Gaj et al., 2013; Jinek et al., 2012). PAM is the site present adjacent to the protospacer in the target DNA, which is the complementary region to the crRNA. PAM site is very crucial for the recognition and binding of the crRNA-tracrRNA-Cas9 complex. In *Streptococcus pyogenes*, the PAM is a 5' NGG 3' consensus sequence, and Cas9 introduces a double-stranded break (DSB) approximately 3 base pairs upstream PAM sequence (Gaj et al., 2013; Jinek et al., 2012). This target sequence-specific activity of the RNA-nuclease complex protects the bacteria from foreign DNA and the invading phages by integrating a fragment of foreign DNA or phage genome in its own genome creating a memory effect for future. This is very much similar to adaptive immune system in mammals.

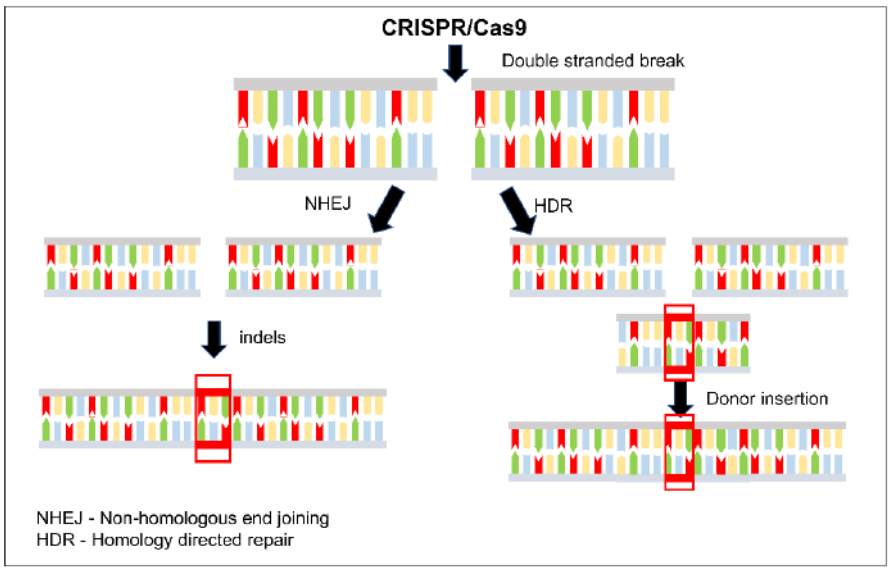


Fig. 1.11. Illustration of repair of double stranded breaks. The DSB induced by CRISPR/Cas9 toolbox is being repaired via one of the two DNA repair pathways, non-homologous end joining or homology-directed repair, which requires a donor DNA.

1.6 Research gap

The regulatory capacity of ZSCAN4 in the generation and maintenance of mouse iPSCs is partially elucidated. But, in case of human iPSCs, its functional role remains uncharacterized. Given the robust data highlighting the potential of ZSCAN4 in mouse ESCs and early reports in mouse iPSC generation, bridging this gap for human iPSCs is the utmost need of the hour. Hence, this study aims to delineate the role of human ZSCAN4 in iPSC maintenance. Primary focus of this thesis is to investigate whether human ZSCAN4 maintains functional conservation with mouse ZSCAN4 in the generation of iPSCs, specifically seeking to define the exact molecular mechanism involved. To provide a comprehensive analysis of ZSCAN4, we have adopted a dual-pronged experimental strategy. Firstly, we have generated recombinant human ZSCAN4 protein, which can be utilized in time- and dose-dependent manner for reprogramming human fibroblasts to generate human iPSCs. This will aid in monitoring its impact on the kinetics and efficiency of the reprogramming process. Also, this approach avoids the risks of genomic integration associated with the viral vectors while providing insights into the direct role of protein in human iPSC generation. Secondly, a loss-of-function analysis using CRISPR-Cas9 genome editing toolbox. By functionally inactivating *ZSCAN4* gene, we can deduce its essentiality

for genomic stability and self-renewal in human iPSCs. Based on the literature review and the research gap, we have designed the objectives for this thesis as follows.

1.7 Objectives

- Generation of cell permeant recombinant human ZSCAN4 fusion protein
- Generation and characterization of human iPSC line
- Knockout of human *ZSCAN4* gene using CRISPR-Cas9 toolbox
- Characterization of *ZSCAN4* knockout iPSCs

Chapter 2

Generation of cell permeant recombinant human ZSCAN4 fusion protein

This chapter has been accepted for publication as follows:

- **Sundaravadivelu PK**, Augustine ABHR, Thummer RP (**2026**) Generation of cell permeant recombinant ZSCAN4 protein for various biological applications, In Recent Developments in Biotechnology, Proceedings of Research and Industrial Conclave - Synergy 2025, Springer Nature, Singapore (**Accepted**)

Overview

ZSCAN4 is a transcription factor essential for early embryonic development and telomere length regulation in pluripotent stem cells (PSCs), where it supports self-renewal and genomic stability. Incorporation of this factor into the transcriptional cocktail for mouse induced Pluripotent Stem Cell (iPSC) reprogramming has been shown to enhance reprogramming efficiency, maintain genomic stability, and induce cancer stem cell-like phenotypes. However, a better understanding of the role of ZSCAN4 in iPSCs and cancer cells is essential. Our study aims in the generation of purified recombinant human ZSCAN4 protein having cell-permeant ability, to prospectively be used for biological applications in a time- and dose-dependent manner. The recombinant ZSCAN4 fusion protein was produced in *Escherichia coli* (*E. coli*), and purified under native purification conditions. Additionally, two truncated ZSCAN4 functionally impaired variants, lacking key domains, were generated as putative negative controls. Subsequently, the characterization of purified ZSCAN4 proteins using Western blotting and Circular Dichroism (CD) spectroscopy was performed. Additionally, the influence of various factors like storage buffer, temperature, and the presence of animal-derived products like fetal bovine serum, on the stability and solubility of the purified ZSCAN4 protein was carried out. In addition, we have validated the transduction ability of the purified proteins in mammalian cells using immunofluorescence. This purified recombinant ZSCAN4 protein toolbox with transduction ability enables prospective systematic investigation of its functions in telomere regulation, cell proliferation, stem cell generation and maintenance, and genome stability, offering insights into both stem cell biology and cancer research.

2.1 Introduction

The transcription factor Zinc and SCAN domain containing 4 (ZSCAN4) is recognized as a pivotal factor exerting multifaceted roles across crucial biological processes, including early embryonic development, generation, and maintenance of pluripotent stem cells, and cancer stem cells. ZSCAN4 exhibits an extremely tight and temporally restricted expression pattern during early embryogenesis. In mice, it is reported to be exclusively expressed in the two-cell stage (Falco et al., 2007), while human embryonic development shows ZSCAN4 expression in the four to eight cell stage (Vassena et al., 2011). This transient, highly regulated expression profile highlights its critical, developmental-stage-specific function. Within embryonic stem cells (ESCs), ZSCAN4 expression remains confined to approximately 5% of the cell population at any given time in culture (Zalzman et al., 2010). The inclusion of ZSCAN4 during the reprogramming of mouse fibroblasts to iPSCs has shown to yield considerable benefits. ZSCAN4 not only enhanced reprogramming efficiency drastically but also was instrumental in ensuring the genomic stability of the resulting iPSCs, significantly reducing the occurrence of single nucleotide polymorphisms (SNPs) (Hirata et al., 2012; Jiang et al., 2013). One of the key functions of ZSCAN4 lies in telomere length maintenance, a process critical for the indefinite self-renewal characteristic of PSCs and cancer cells. ZSCAN4 is identified as one of the key players in regulating telomere length via telomerase independent pathway, also known as alternate lengthening of telomeres (ALT) (Wang et al., 2012; Zalzman et al., 2010). Functional studies have shown that knockdown of ZSCAN4 led to telomere shortening in ESCs. Furthermore, ZSCAN4 contributed to the maintenance of cancer stem cell phenotype (Portney et al., 2020). Mechanistically, ZSCAN4 has been observed to interact with components of the shelterin complex, the primary protein machinery at the telomeres (Lee and Gollahon, 2014; Lee and Gollahon, 2015). Previous reports strongly suggested that ZSCAN4 interacted with these shelterin complex proteins through direct or indirect modulation (Lee and Gollahon, 2014; Lee and Gollahon, 2015).

The production of recombinant proteins has profoundly advanced biotechnology, yielding numerous biologically active proteins for various biological applications (Borgohain et al., 2018). Recombinant proteins offer a highly valuable and safe approach for applications like cell reprogramming and cancer therapy, as the purified protein can be directly delivered to target cells in a time- and dose-dependent manner without any

modification to the genome (Borgohain et al., 2019; O'Malley et al., 2009; Sommer and Mostoslavsky, 2013). However, attaining highly pure, stable protein remains a challenge due to certain hurdles in heterologous expression. Given the significant roles of ZSCAN4 in development, pluripotency, and telomere biology, an in-depth understanding of the mechanism of action of ZSCAN4 is essential. Therefore, the primary goal of this chapter is to overcome the existing hurdles in heterologous expression and establish an optimized protocol for generating highly pure, stable recombinant ZSCAN4 protein, and its necessary negative controls. Prospectively, this purified ZSCAN4 protein toolbox will be utilized to precisely investigate the mechanistic role of ZSCAN4 in key cellular processes.

2.2 Materials and Methods

2.2.1 Cloning and confirmation of recombinant constructs

The coding sequence of human *ZSCAN4* was obtained from the NCBI RefSeq database (NM_152677.2). This sequence was codon-optimized using GeneOptimizer web server, and the optimization was analyzed using GenScript Rare Codon Analysis (GRCA) and Graphical Codon Usage Analyzer 2.0 (GCUA) *in silico* tools for maximal translational efficiency in *Escherichia coli* (*E. coli*). Three distinct codon-optimized fusion tags, namely nuclear localization sequence (N), transactivator of transcription (T), and octa-histidine (H), were fused to either the 5' or the 3' end of the optimized sequence to generate two different fusion gene constructs, namely HTN-*ZSCAN4* and *ZSCAN4*-NTH. Furthermore, two more constructs were designed by deleting two important domains in ZSCAN4 protein. The constructs, HTN-*ZSCAN4*-ΔDB, in which the DNA binding domain is deleted with tags on the 5' end, and HTN-*ZSCAN4*-ΔSD, in which the SCAN domain is deleted with tags on the 5' end, were generated. All these fusion gene constructs were synthesized by GenScript Biotech Corporation, Nanjing, China. Subsequently, the synthesized constructs were cloned in pET28a(+) expression vector. The cloning was confirmed using restriction digestion, and the integrity of the cloned sequences were verified by Sanger sequencing (Eurofins Genomics India Pvt. Ltd., Karnataka, India).

2.2.2 Soluble expression analysis

The total and soluble expression levels of the fusion proteins were analyzed. The expression of fusion proteins was induced in small culture volumes at an OD_{600} of ~0.5 using 0.25 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubation at 37 °C for 2 h. Subsequently, the cell pellets were harvested, resuspended in resuspension buffer,

lysed using ultrasonication, and the lysates were clarified by centrifuging at 10,000 rpm for 30 mins at 4 °C to separate out the soluble fraction from the insoluble fraction. All these samples were further analyzed using SDS-PAGE and Western blotting.

To maximize the amount of soluble expression, expression parameters such as cell density during induction, inducer concentration, post induction incubation temperature and time were comprehensively screened and optimized. The different values screened for these parameters are given in Table 2.1.

Table 2.1. List of expression parameters screened, along with the values screened for each parameter.

Expression parameters	Values screened
Inducer concentration (mM)	0.05, 0.1, 0.25, 0.5
Cell density during induction (OD_{600})	~0.5, ~1.0, ~1.5
Post-induction incubation time (h)	2, 4, 6, 8
Post-induction incubation temperature (°C)	37, 18

2.2.3 Purification of recombinant proteins

The purification of these recombinant fusion proteins was performed using immobilized metal ion affinity chromatography. Using the optimized expression parameters, the expression was induced, the cells were lysed, and the soluble fraction was separated. The soluble fraction was incubated with Ni-NTA resin (Bio-Rad) overnight. Subsequently, the column was washed sequentially with three different wash buffers having increasing concentrations of imidazole. Finally, the protein was eluted using the elution buffer. The samples collected during purification were analyzed using SDS-PAGE and Western blotting. The eluted proteins were desalted to remove the salts and imidazole using PD-10 desalting columns (Cytiva). The desalted proteins were aliquoted, flash frozen using liquid nitrogen and were stored in -80 °C until further use. The composition of buffers used in purification are given in Table 2.2.

Table 2.2. Composition of the buffers used in purification.

Buffers (pH 7.0)	Phosphate Buffer (PB)	Glycerol	Sodium chloride	Imidazole
Resuspension buffer	20 mM	20%	450 mM	40 mM
Wash buffer I	20 mM	-	450 mM	100 mM
Wash buffer II	20 mM	-	450 mM	200 mM
Wash buffer III	20 mM	-	450 mM	300 mM
Elution buffer	20 mM	20%	450 mM	1000 mM

2.2.4 SDS-PAGE and Western blotting

Protein concentrations were measured using Bradford assay (Bradford reagent; Bio-Rad) using bovine serum albumin (BSA; HiMedia) as a standard. The absorbance values were measured using a multi-plate reader (Multiskan GO, Thermo Scientific).

For Coomassie staining, the samples were resolved in 12% SDS-polyacrylamide gel. Subsequently, the gels were fixed using gel fixing solution [50% ethanol (Merck Millipore), 10% glacial acetic acid (Merck Millipore) in deionized water] for 1 h, washed using gel washing solution [50% methanol (Merck Millipore), 10% glacial acetic acid (Merck Millipore) in deionized water] for 1 h. Then, the gels were stained using Coomassie brilliant blue G-250 (Merck Millipore; 0.4% in gel washing solution) for 10 mins, destained using gel washing solution for 30 mins, and the gels were imaged using the ChemiDoc XRS+ system (Bio-Rad).

For Western blotting, the proteins resolved in the polyacrylamide gel were electro-transferred to nitrocellulose membrane (Bio-Rad) using the Trans Blot Turbo system (Bio-Rad). After transfer, the membrane was washed using 1X Tris-buffered saline (TBS), and the transfer was confirmed using Ponceau S staining (HiMedia). After washing the membrane with 1X TBS, it was blocked using 5% skimmed milk in 1X TBST (0.1% Tween-20 in 1X TBS), incubated with respective primary antibody (Table S3) diluted in 5% BSA in 1X TBST overnight at 4 °C. Next, the membrane was incubated with corresponding HRP-conjugated secondary antibody (Table S3) diluted in 5% skimmed milk in 1X TBST for 1 h at room temperature. Finally, the blot was developed using enhanced chemiluminescence substrate (Bio-Rad) and imaged using the ChemiDoc XRS+ system (Bio-Rad).

2.2.5 Characterization of purified proteins

Next, the purified proteins were verified for their identity using Western blot with α-ZSCAN4 antibody (Novus Biologicals). Further, the secondary structure of the purified protein was characterized using far-UV CD spectroscopy in J-1500 spectropolarimeter (Jasco, Japan). The spectrum was obtained as an average of 5 accumulations from 190 nm to 250 nm, using a 1 mm quartz cuvette at a scan rate of 100 nm/min with a data integration time of 1 s. The buffer in which protein was eluted (20 mM PB) was used for blank subtraction. The obtained spectrum was normalized using the molarity of the purified protein to yield delta epsilon ($\Delta\boldsymbol{\varepsilon}$) values. A graph was plotted with wavelength on the x-axis and $\Delta\boldsymbol{\varepsilon}$ on the y-axis using GraphPad Prism 8 software.

2.2.6 Stability of the purified recombinant protein

The stability of the purified protein was investigated under standard cell culture conditions. The purified protein was diluted to a final concentration of 400 nM in the appropriate cell culture media according to experimental requirement. The prepared protein transduction media was incubated at 37 °C in the presence of 5% CO_2 in a humidified atmosphere. The samples were collected at different time points and were analyzed using Western blotting.

2.2.7 Cell culture

The MDA-MB-231 cancer cells were procured from NCCS Pune, India. The cells were cultured in growth media containing DMEM (GIBCO) supplemented with 10% fetal bovine serum (FBS; GIBCO) and 1% antibiotic-antimycotic solution (GIBCO). The cells were cultured until 80-90% confluency and then passaged using Trypsin-EDTA solution (GIBCO) at a split ratio of 1:5. The cells were seeded according to experimental requirements.

2.2.8 MTT assay

The MDA-MB-231 cells were seeded in a 96-well plate at a seeding density of 4000 cells/well on day 0. After 24 h, the cells were treated with growth media containing 20 mM PB and 20% glycerol buffer in a volume that approximately corresponds to 400 nM of purified HTN-ZSCAN4 protein. The treatment was given for different time periods (6 h, 12 h, and 24 h) as per experimental requirement. After the treatment, the media was removed, and the cells were incubated with MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] media containing 0.25 mg/ml of MTT (Merck Millipore)

dissolved in 10% 1X phosphate buffered saline (PBS) in DMEM for 2 h at 37 °C. Post-incubation, the MTT media was removed, the formazan crystals were dissolved using dimethyl sulfoxide (DMSO; HiMedia) and the absorbance values were measured at 570 nm with 650 nm as reference wavelength using a multi-plate reader (Multiskan GO, Thermo Scientific).

2.2.9 Cellular transduction and immunofluorescence

The MDA-MB-231 cells were treated with 400 nM of purified recombinant protein at 50 to 60% confluency. The treatment was done for 6 h, and subsequently, the cells were washed with 1X PBS and fixed with 4% paraformaldehyde for 15 min at room temperature. Subsequently, the cells were permeabilized using 0.1% Triton-X-100 in 1X PBS for 15 min at room temperature, followed by a blocking solution containing 0.5% BSA, 0.15% glycine, and 10% normal goat serum in 1X PBS for 1 h at room temperature. The cells were then incubated overnight at 4 °C with respective primary antibody (Table S3) diluted in blocking solution, followed by respective fluorophore-conjugated secondary antibody (Table S3) diluted in blocking solution for 1 h at room temperature. Finally, the nucleus was counterstained with Hoechst 33342 (Cell Signaling Technology; 1:10,000 in 1X PBS) for 5 min at room temperature, and then the cells were observed and imaged using an inverted fluorescence microscope (ZOE™ Fluorescent Cell Imager, Bio-Rad).

2.3 Results

2.3.1 Codon optimization and cloning

The coding sequence of *ZSCAN4* before codon optimization was evaluated for translational efficiency in *E. coli* using GRCA and GCUA. Initial assessment revealed the presence of codons that might hamper the expression of ZSCAN4 in *E. coli* (Fig. 2.1A). Specifically, analysis using GRCA revealed almost 11% of codons (Fig. 2.1A; *left*; red bars before the green dotted line) are low frequency codons, which will impede heterologous expression in *E. coli*. Moreover, the codon adaptation index for the non-optimized sequence was 0.6, which was suboptimal for expression. Therefore, to mitigate these concerns, the coding sequence of *ZSCAN4* was optimized using GeneOptimizer *in silico* tool. Analyzing the codon-optimized sequence of *ZSCAN4* using GRCA tool, revealed that all the suboptimal codons were replaced with optimal, synonymous codons that increased the codon quality (Fig. 2.1A; *left*; blue bars), which will lead to maximum expression in *E.*

coli. Moreover, the codon adaptation index also increased to 0.94, which was more optimal for robust expression in *E. coli*.

Similarly, the non-optimized *ZSCAN4* sequence was analyzed using GCUA, which showed the presence of non-optimal, rare codons in the coding sequence whose relative adaptiveness value was less than 30, posing as potential bottlenecks likely to diminish the expression of ZSCAN4 in *E. coli* (Fig. 2.1A; *right*; magenta bars). Therefore, these codons and other codons that might hamper the expression were replaced with the synonymous, high-frequency codons for optimal expression GeneOptimizer *in silico* tool. Consequently, analyzing the codon-optimized sequence with GCUA revealed better relative adaptiveness value for all codons (more than 30), which will lead to maximum expression of ZSCAN4 in *E. coli* (Fig. 2.1A; *right*). These results from multiple computational platforms indicate significant enhancement in the quality of the codons, suggesting high-level heterologous expression of ZSCAN4 in *E. coli*.

In this study, three different functional fusion tags were incorporated, namely octahistidine (HIS; H) to facilitate affinity chromatography-based purification, transactivator of transcription (TAT; T), a cell penetrating peptide to assist intracellular delivery of recombinant protein, and a nuclear localization sequence (NLS; N) to ensure

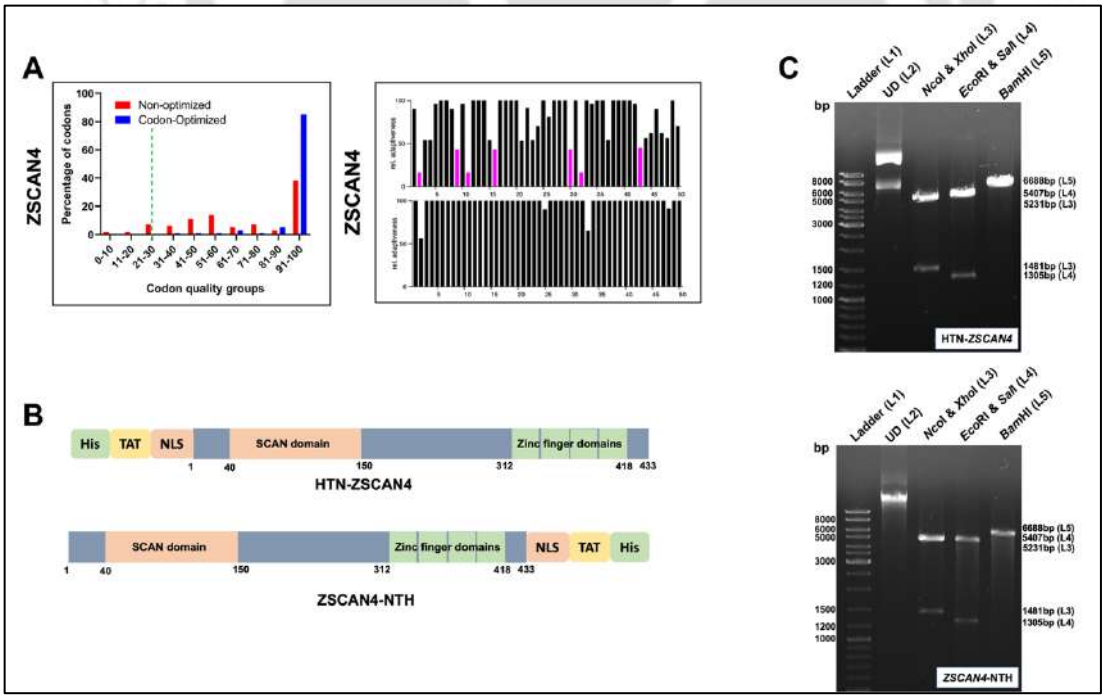


Fig. 2.1. Codon optimization, schematic of gene constructs, and confirmation of cloning. A) Analysis of the quality of codons before and after codon optimization using GRCA (*left*) and GCUA (*right*). **B)** Schematic representing the genetic construct for HTN-

ZSCAN4 (*top*) and *ZSCAN4*-NTH (*bottom*; not drawn to scale). **C)** Confirmation of cloning of HTN-*ZSCAN4* (*top*) and *ZSCAN4*-NTH (*bottom*) using restriction digestion analysis. bp: basepairs; UD: undigested.

localization of the recombinant protein inside the nucleus. Therefore, the codon-optimized sequences along with the fusion tags, HTN-*ZSCAN4* (with the fusion tags at 5' end) and *ZSCAN4*-NTH (with the fusion tags at 3' end) were synthesized (Fig. 2.1B). Subsequently, these constructs were cloned in pET28a(+) expression vector using respective restriction enzymes. The cloning was confirmed by performing restriction digestion, which showed successful cloning of the insert along with the tags in the vector (Fig. 2.1C). Further, this was verified by DNA sequencing (data not shown).

2.3.2 Influence of expression parameters on the soluble expression

The total and soluble expression of the two fusion proteins was examined by transforming the cloned recombinant constructs (pET28a(+)-HTN-*ZSCAN4* and pET28a(+)-*ZSCAN4*-NTH) in *E. coli* BL21(DE3) strain and inducing their expression using IPTG. The SDS-PAGE and Western blotting images showed that both fusion proteins were overexpressed in the BL21(DE3) strain. Notably, it was observed that in case of both the fusion proteins, HTN-ZSCAN4 and ZSCAN4-NTH, the amount of proteins present in the soluble fraction was relatively low compared to the insoluble fraction (Fig. 2.2). Next, various expression parameters were screened in order to attain maximal soluble expression of the fusion proteins. The expression parameters like inducer concentration, cell density

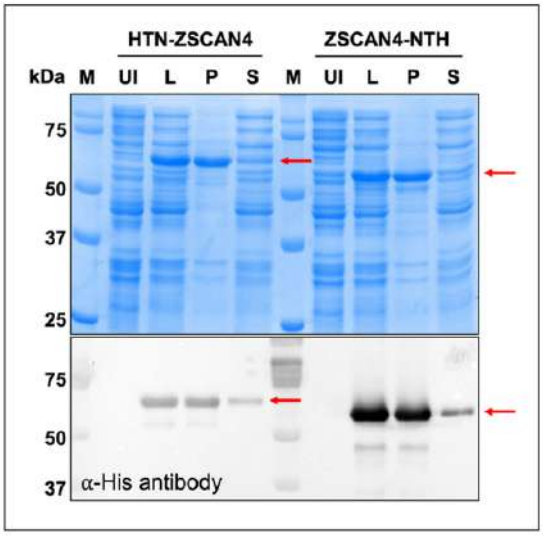


Fig. 2.2. Soluble expression analysis. Soluble expression of HTN-ZSCAN4 and ZSCAN4-NTH fusion proteins analyzed using SDS-PAGE and Western blotting. M: marker; UI: uninduced; L: whole cell lysate; P: insoluble pellet fraction; S: soluble supernatant fraction.

at the time of induction (OD_{600}), post induction incubation temperature, and incubation time were screened.

Firstly, the influence of inducer concentration on the total and soluble expression of HTN-ZSCAN4 and ZSCAN4-NTH was analyzed (Fig. 2.3). From the results, it was evident that inducer concentration influenced the overall expression in both fusion proteins. As the IPTG concentration increased, total expression of both HTN-ZSCAN4 and ZSCAN4-NTH increased (Fig. 2.3). Interestingly, as the IPTG concentration was increased, there was no significant difference observed in the case of soluble expression, as was evident from the Western blot (Fig. 2.3; *bottom*). Even though increasing the IPTG concentration enhanced the total expression of ZSCAN4, it did not have any effect on soluble expression. Therefore, the IPTG concentration of 0.05 mM was chosen as the optimal IPTG concentration for maximal soluble expression of HTN-ZSCAN4 and ZSCAN4-NTH fusion proteins.

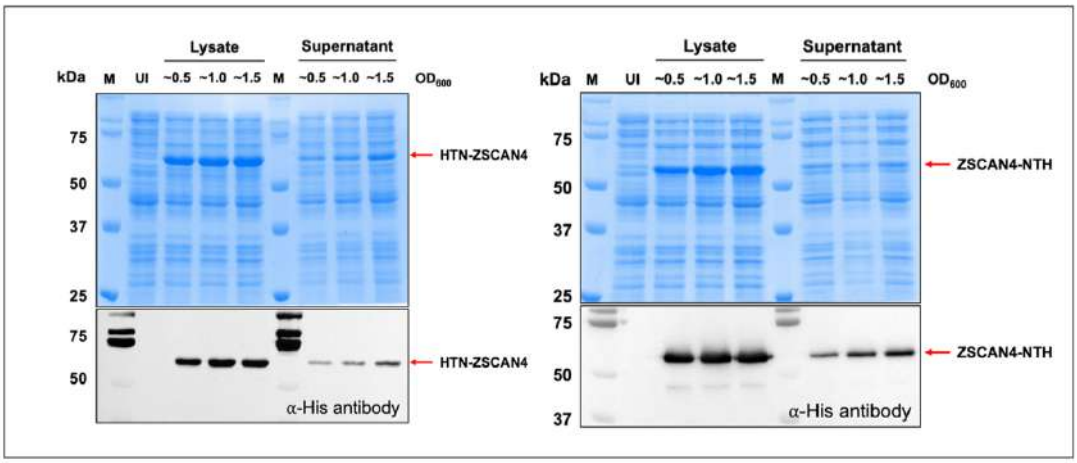


Fig. 2.3. Optimization of inducer concentration. A range of inducer (IPTG) concentrations was screened for maximal soluble expression of HTN-ZSCAN4 (*left*) and ZSCAN4-NTH (*right*) fusion proteins, analyzed using SDS-PAGE and Western blotting. M: marker; UI: uninduced.

Next, the effect of cell density (OD_{600}) at the time of induction on the soluble expression of ZSCAN4 fusion proteins was studied. From Figure 2.4, it was observed that as the OD_{600} increased, the amount of total and soluble expression increased for both the

fusion proteins. It was evident from the Western blotting that when the cultures were induced at OD_{600} ~1.5, the total and soluble expression of HTN-ZSCAN4 and ZSCAN4-NTH was maximum (Fig. 2.4). Therefore, inducing at OD_{600} ~1.5 was chosen as the optimal cell density for the maximal soluble expression of ZSCAN4 fusion proteins.

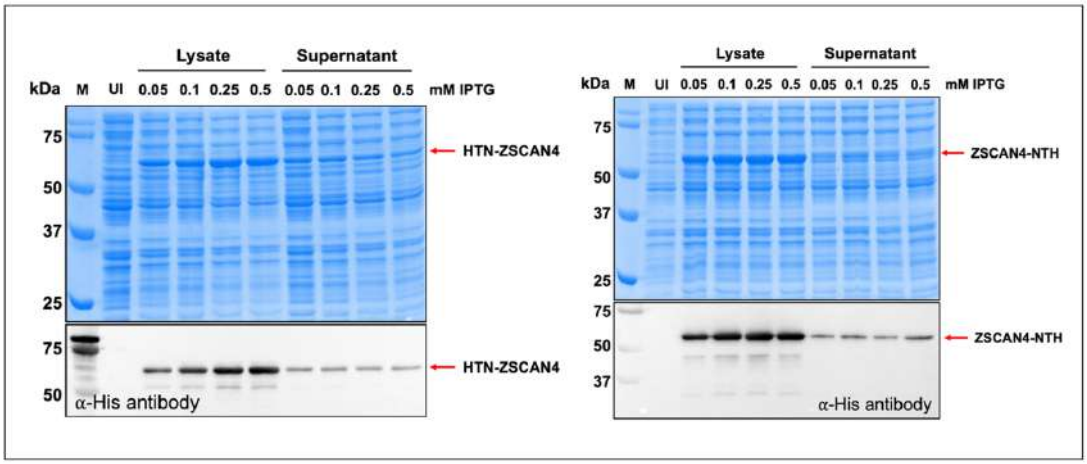


Fig. 2.4. Optimization of optical density. A range of values for optical density (OD_{600}) at the time of induction were screened for maximal soluble expression of HTN-ZSCAN4 (*left*) and ZSCAN4-NTH (*right*) fusion proteins analyzed using SDS-PAGE and Western blotting. M: marker; UI: uninduced.

Subsequently, the effect of post-induction incubation time on the soluble expression of ZSCAN4 fusion proteins was analyzed. The bacterial culture after induction was incubated for different time periods and then analyzed for total and soluble expression. Fig. 2.5 showed that increasing the incubation time did not affect the overall or the soluble expression of both ZSCAN4 fusion proteins. Notably, in the case of ZSCAN4-NTH,

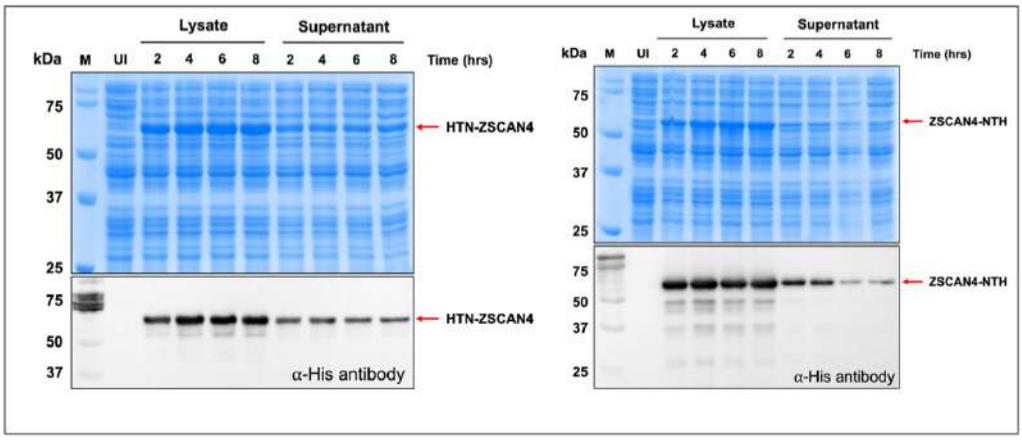


Fig. 2.5. Optimization of post-induction incubation time. Different time periods for post-induction incubation time were screened for maximal soluble expression of HTN-ZSCAN4 (*left*) and ZSCAN4-NTH (*right*) fusion proteins analyzed using SDS-PAGE and Western blotting. M: marker; UI: uninduced.

soluble expression decreased beyond 4 h of incubation time (Fig. 2.5), even though there was no change in the overall expression. These results showed that an incubation time of 2 h was sufficient to produce maximal soluble expression of HTN-ZSCAN4 and ZSCAN4-NTH fusion proteins.

Finally, the effect of incubation temperature on the soluble expression of fusion proteins was studied. Two different temperatures for incubation, 37 ºC and 18 ºC, were analyzed for their effect on the soluble expression. In case of ZSCAN4 fusion proteins, reducing the incubation temperature to 18 ºC significantly hampered its expression (Fig. 2.6). Both Coomassie and Western blotting results showed that the overall and soluble expression declined (Fig. 2.6) when the temperature was reduced to 18 ºC. Therefore, 37 ºC was chosen as the optimal induction temperature for future experiments. Different values that were screened for the expression parameters, and the final optimized values are listed in Table 2.3.

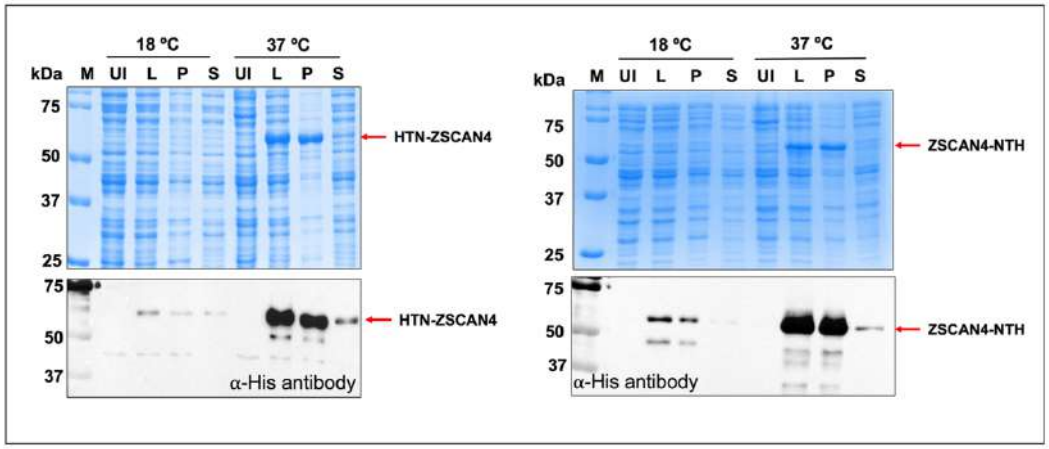


Fig. 2.6. Optimization of post-induction incubation temperature. Two different incubation temperatures were screened for maximal soluble expression of HTN-ZSCAN4 (*left*) and ZSCAN4-NTH (*right*) fusion proteins analyzed using SDS-PAGE and Western blotting. M: marker; UI: uninduced; L: whole cell lysate; P: insoluble pellet fraction; S: soluble supernatant fraction.

Table 2.3. List of various expression parameters optimized, their screened values and their final optimized values.

Expression parameters	Values screened	Optimal value
Inducer concentration (mM)	0.05, 0.1, 0.25, 0.5	0.05
Cell density during induction (OD_{600})	~0.5, ~1.0, ~1.5	~1.5
Post-induction incubation time (h)	2, 4, 6, 8	2
Post-induction incubation temperature (°C)	37, 18	37

2.3.3 Influence of salt on the secondary structure of purified ZSCAN4 fusion protein

In our study, Ni-NTA resin has been used, where the octahistidine tag will interact with the Ni^{2+} ions immobilized by NTA. The bound proteins will be eluted out using imidazole, which competitively binds to Ni^{2+} ions and releases the bound protein molecules. The fig. 2.7 shows the purification of HTN-ZSCAN4 and ZSCAN4-NTH fusion proteins. From the yield analysis by Bradford method, it was evident that the yield of the HTN- ZSCAN4 fusion protein (260.6 µg/ml) was relatively higher compared to ZSCAN4-NTH fusion protein (124.7 µg/ml). Therefore, from here on, all other experiments were carried out using HTN-ZSCAN4 fusion protein only unless mentioned otherwise. Subsequently, the secondary structure of the purified fusion proteins was analyzed by CD spectroscopy. The CD spectrum showed that the purified fusion proteins (HTN-ZSCAN4 and ZSCAN4-NTH) did not maintain the secondary structure (Fig. 2.7A and B; *bottom*).

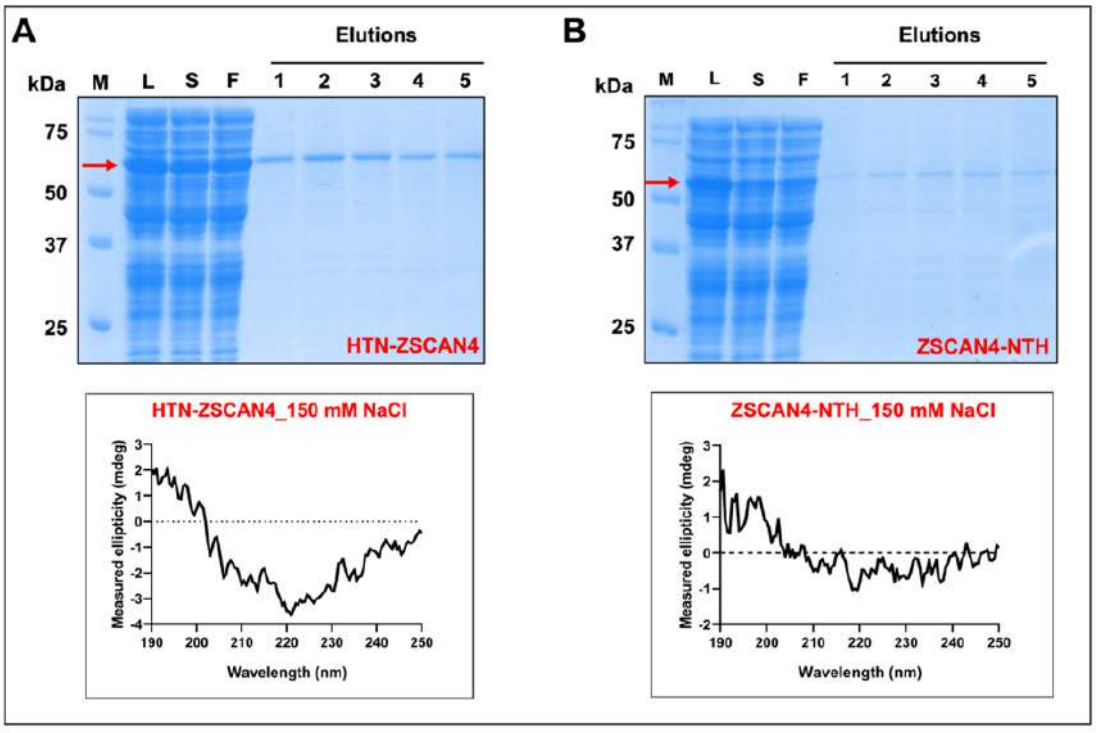


Fig. 2.7. Elution profile and secondary structure analysis of ZSCAN4 fusion proteins. Elution profile and secondary structure analysis of the purified HTN-ZSCAN4 **(A)** and ZSCAN4-NTH **(B)** using CD spectroscopy under lower salt concentration purification buffers, with wavelength plotted on the x-axis and measured ellipticity plotted on the y-axis. M: marker; L: whole cell lysate; S: soluble supernatant fraction; F: flow-through fraction.

The spectra showed random spikes throughout the range of wavelengths analyzed (190 - 240 nm) rather than showing specific peaks and dips (Fig. 2.7A and B; *bottom*). This indicates that the purified fusion proteins did not have a stable secondary structure.

Therefore, a range of different salt (NaCl) concentrations in the purification buffers was screened, ranging from 0 mM to 600 mM, to produce a structurally stable purified HTN-ZSCAN4 protein (Fig. 2.8A). Fig. 2.8A shows the elution profile of HTN-ZSCAN4 protein purified with different salt concentration containing buffers. Subsequently, all these purified proteins were analyzed for their structural stability using CD spectroscopy. Fig. 2.8B shows the compiled CD spectra of HTN-ZSCAN4, purified using buffers containing different salt concentrations. A trend was observed in spectra, as the concentration of salt was increased, the purified proteins had a more stable secondary structure (Fig. 2.8B). When the HTN-ZSCAN4 protein was purified with buffers containing 0, 150 or 300 mM of salt, a flat, spiky spectrum was observed, indicating an unstable secondary structure. Notably, when the salt concentration in the buffers was increased to 450 mM or 600 mM, a stable CD spectrum was observed (Fig. 2.8B).

From these results, it can be concluded that the usage of purification buffers with salt concentrations of 450 mM and 600 mM is optimal for purifying a structurally stable HTN-ZSCAN4 protein. Interestingly, it can be observed that as the salt concentration was increased, the amount of non-specific proteins eluting out along with the HTN-ZSCAN4 fusion protein, also increased (Fig. 2.8A).

Therefore, to reduce the non-specific proteins eluting out along with HTN-ZSCAN4, the amount of imidazole in the purification buffers (all wash buffers and elution buffer) was doubled. The imidazole concentration in the wash buffers was modified to 100, 200, and 300 mM, and finally, the HTN-ZSCAN4 protein was eluted out using 1 M imidazole. Fig. 2.8C shows the elution profile of HTN-ZSCAN4 purified with a higher imidazole concentration buffer using 450 mM and 600 mM salt concentration buffers. The

results revealed that increasing the concentration of imidazole indeed helped in reducing the amount of non-specific proteins eluting out (Fig. 2.8C). Although imidazole helped in reducing non-specific proteins, the elution profile of ZSCAN4 protein purified using 600 mM of salt shows an observable amount of non-specific proteins compared to its

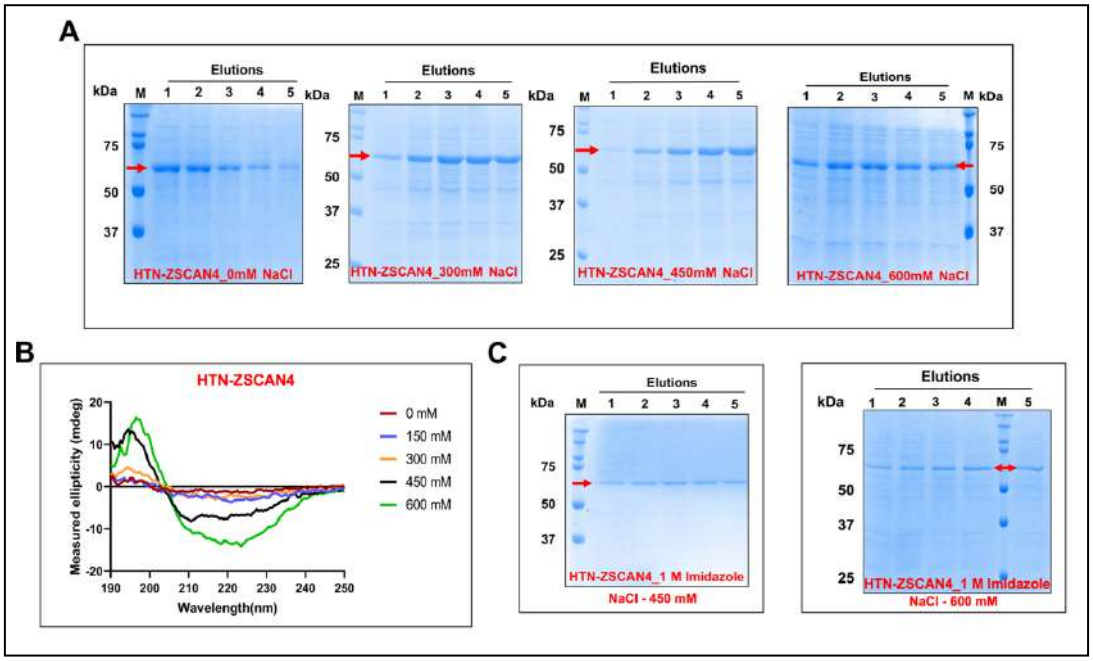


Fig. 2.8. Optimization of buffer conditions for purification and stable secondary structure of HTN-ZSCAN4. A) Optimization of NaCl concentration for the purification of HTN-ZSCAN4 protein. **B)** Analysis of secondary structure of HTN-ZSCAN4 protein purified using a gradient of concentrations of NaCl by performing CD spectroscopy with wavelength plotted on the x-axis and measured ellipticity plotted on the y-axis. **C)** Elution profile of HTN-ZSCAN4 protein purified using 1M imidazole. M: marker.

counterpart (Fig. 2.8C). Therefore, a salt concentration of 450 mM and higher imidazole concentrations (100, 200 and 300 mM in wash buffers and 1 M in elution buffer) were chosen as optimal for the efficient purification of HTN-ZSCAN4 fusion protein.

Lastly, the HTN-ZSCAN4 protein purified under optimal conditions, was characterized by Western blotting using α-histidine and α-ZSCAN4 antibodies and Far-UV CD spectroscopy (Fig. 2.9). From the Western blotting using α-ZSCAN4 antibody, it is evident that the purified protein is indeed ZSCAN4 confirming its identity (Fig. 2.9A). Moreover, the CD spectrum indicates that the purified protein has maintained a stable secondary structure under the optimized purification conditions (Fig. 2.9B).

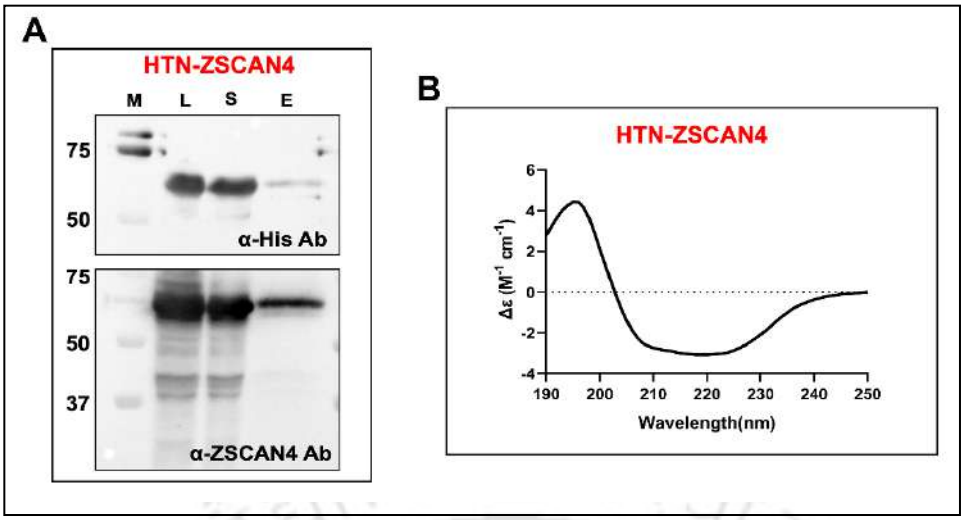


Fig. 2.9. Biochemical and biophysical characterization of purified HTN-ZSCAN4 protein. A) The purified HTN-ZSCAN4 was verified for its identity by performing Western blotting using α-His antibody (*top*) and α-ZSCAN4 antibody (*bottom*). **B)** CD spectroscopy of purified HTN-ZSCAN4 fusion protein with wavelength plotted on the x-axis and Δε plotted on the y-axis. M: marker; L: whole cell lysate; S: soluble supernatant fraction; E: elution fraction.

2.3.4 Production of ΔZSCAN4 fusion proteins

Similar to wild-type ZSCAN4 fusion proteins, two more fusion proteins were generated. Firstly, ZSCAN4 in, which DNA-binding domain (Zinc finger domain) was deleted, represented as HTN-ZSCAN4-ΔDB (depicted in Fig. 2.10A), and secondly, ZSCAN4 protein in, which the SCAN domain was deleted, represented as HTN-ZSCAN4-ΔSD (depicted in Fig. 2.10A). These fusion proteins will serve as a negative control (non-functional proteins) in future experiments. These gene constructs were generated and cloned in pET28a(+) expression vector. Cloning was confirmed by restriction digestion (Fig. 2.10B) and further confirmed by DNA sequencing (data not shown).

Subsequently, using the optimized expression parameters, soluble expression of these ΔZSCAN4 fusion proteins was analyzed using SDS-PAGE and Western blotting (Fig. 2.11A). These results demonstrated that these two ΔZSCAN4 fusion proteins have soluble expression (Fig. 2.11A), and therefore can be purified from the soluble fraction under native conditions. Next, these HTN-ZSCAN4-ΔDB and HTN-ZSCAN4-ΔSD fusion proteins were purified using the optimized purification parameters. The SDS-PAGE analysis shows that these proteins have been purified with maximum purity (Fig. 2.11B).

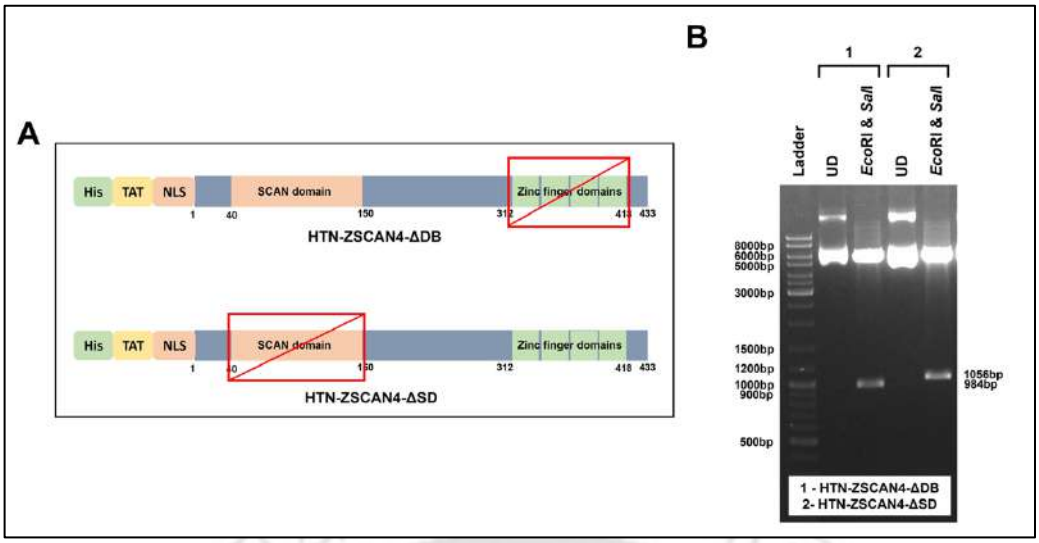


Fig. 2.10. Schematic and confirmation of cloning of the *ΔZSCAN4* gene constructs. A) Schematic depicting the genetic constructs of HTN-*ZSCAN4*-ΔDB (top) and HTN-*ZSCAN4*-ΔSD (bottom; not drawn to scale). **B)** Confirmation of cloning of HTN-*ZSCAN4*-ΔDB and HTN-*ZSCAN4*-ΔSD using restriction digestion analysis. bp: basepair; UD: undigested.

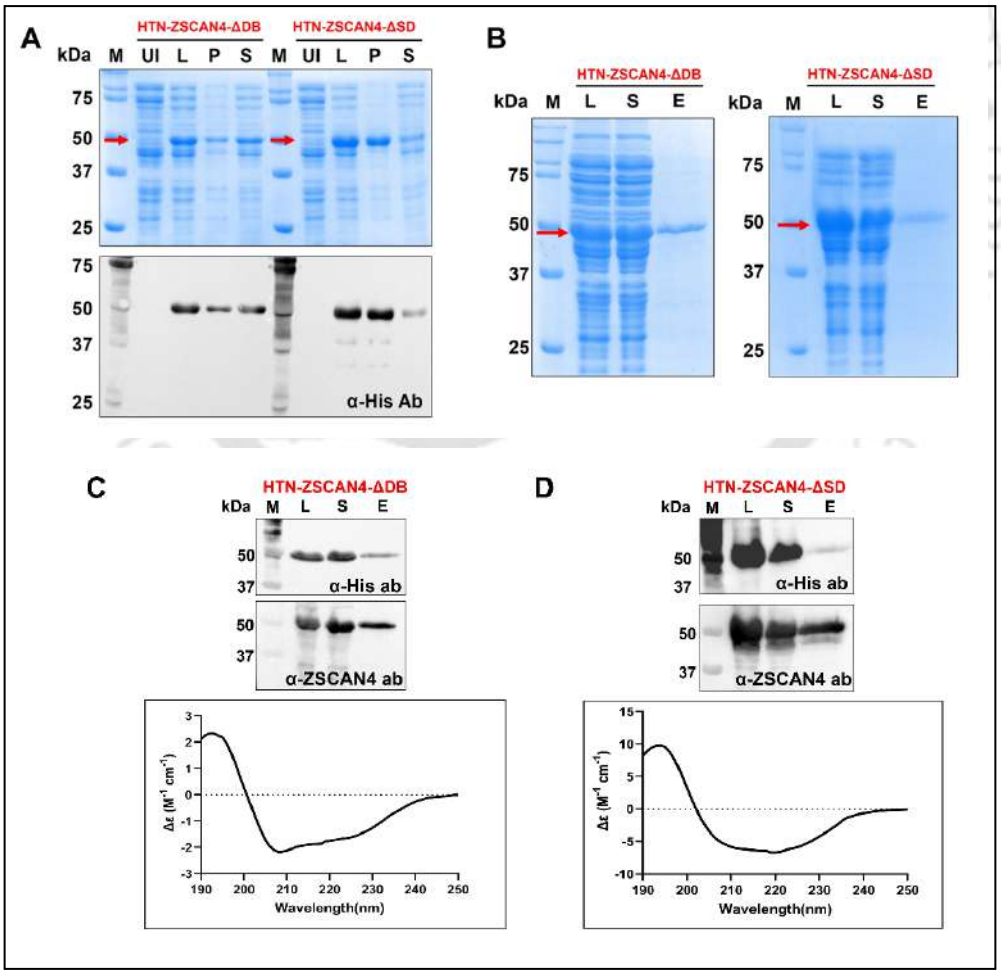


Fig. 2.11. Soluble expression analysis, purification and characterization of ΔZSCAN4 proteins. A) Soluble expression analysis of HTN-ZSCAN4-ΔDB and HTN-ZSCAN4-ΔSD. **B)** Native purification of HTN-ZSCAN4-ΔDB and HTN-ZSCAN4-ΔSD. **C)** Characterization of purified HTN-ZSCAN4-ΔDB protein using Western blotting (*top*) and CD spectroscopy (*bottom*). **D)** Characterization of purified HTN-ZSCAN4-ΔSD protein using Western blotting (*top*) and CD spectroscopy (*bottom*). M: marker; UI: uninduced; L: whole cell lysate; P: insoluble pellet fraction; S: soluble supernatant fraction; E: elution fraction.

Further, these purified proteins were characterized using CD spectroscopy and Western blotting, as described previously. The Western blot authenticated the identity of the purified ΔZSCAN4 fusion proteins (Fig. 2.11C&D), and the CD spectra showed that the purified ΔZSCAN4 proteins have maintained their secondary structure (Fig. 2.11C&D).

2.3.5 Effect of cell culture nutritional supplements on the stability of purified ZSCAN4 fusion protein

Next, the stability of purified HTN-ZSCAN4 protein under standard cell culture conditions was examined. To accomplish this, 400 nM of purified protein was resuspended in DMEM in the presence of 10% FBS. Subsequently, the stability of the protein was analyzed using Western blotting. The figure shows that the purified HTN-ZSCAN4 protein did not migrate properly in the presence of the media supplement, FBS (Fig. 2.12A). An intense thick band can be observed in the blot instead of a distinct band (Fig. 2.12A). Even a reduction in the amount of FBS did not incur any change in its migration pattern (data not shown). Since the purified HTN-ZSCAN4 protein did not migrate properly in the presence of FBS, other alternate nutritional supplements, namely knockout serum replacement (KOSR) and Albumax, were used to check the stability. Surprisingly, even in the presence of KOSR and Albumax, the purified HTN-ZSCAN4 protein did not migrate properly. Similar to FBS, an intense thick band with a distinct smear running downwards was observed when purified protein was resuspended in media containing KOSR and Albumax in the blot (Fig. 2.12A).

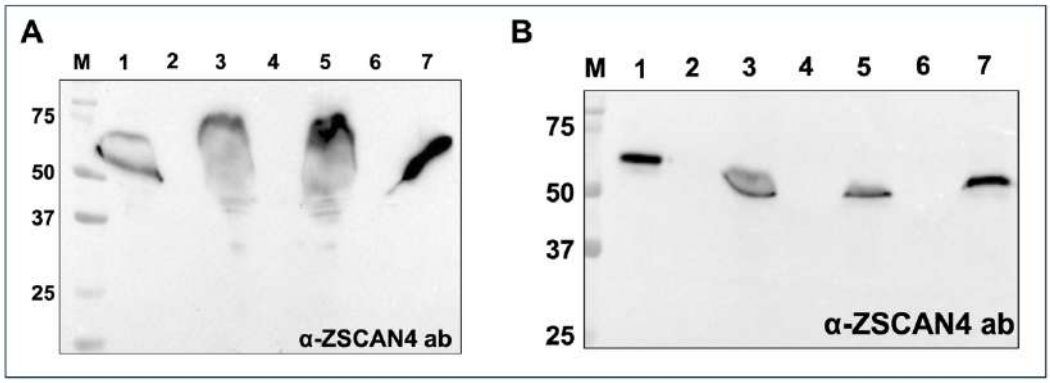


Fig. 2.12. Analysis of the influence of cell culture nutritional supplements on the stability of purified HTN-ZSCAN protein. A) Analysis of the stability of HTN-ZSCAN4 purified protein in the presence of different nutritional supplements under standard cell culture conditions. M: marker; 1: 400 nM of HTN-ZSCAN4 in the presence of DMEM and 10% FBS; 2: DMEM and 10% FBS; 3: 400 nM of HTN-ZSCAN4 in the presence of DMEM and 10% KOSR; 4: DMEM and 10% KOSR; 5: 400 nM of HTN-ZSCAN4 in the presence of DMEM and 10% Albumax; 6: DMEM and 10% Albumax; 7: 400 nM of HTN-ZSCAN4 in the DMEM. **B)** Analysis of the stability of HTN-ZSCAN4 purified protein in presence of complete media under standard cell culture conditions. M: marker; 1: 400 nM of HTN-ZSCAN4 in Essential 8 media; 2: Essential 8 media; 3: 400 nM of HTN-ZSCAN4 in Stembrew media; 4: Stembrew media; 5: 400 nM of HTN-ZSCAN4 in Amniomax; 6: Amniomax media; 7: 400 nM of HTN-ZSCAN4 in DMEM.

Therefore, alternate complete media, which do not require any nutritional supplements, were analyzed for their compatibility with purified HTN-ZSCAN4. Essential 8 media and Stembrew media used for stem cell culture and Amniomax media used for amniocytes culture were analyzed. It was found that purified HTN-ZSCAN4 was stable in the presence of these complete media and no smear was observed, like it did in the presence of FBS (Fig. 2.12B). From the results it is evident that the purified protein is stable in all the three complete media (Fig. 2.12B). To analyze the stability of the purified HTN-ZSCAN4 protein at various time points, the purified protein, HTN-ZSCAN4 diluted in Amniomax and Stembrew was incubated under standard cell culture conditions for 24 h and 48 h. Subsequently, the stability was investigated using Western blotting. The results demonstrated that the HTN-ZSCAN4 protein was stable for at least 24 h in both Amniomax and Stembrew media (Fig. 2.13).

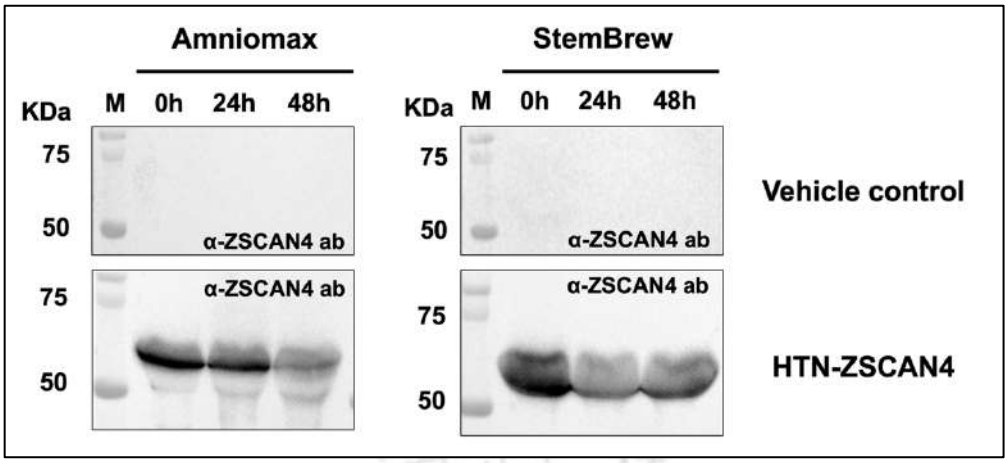


Fig. 2.13. Stability analysis of HTN-ZSCAN4 purified protein under standard cell culture conditions. Analysis of the stability of HTN-ZSCAN4 purified protein in Amniomax and Stembrew media for various time points under standard cell culture conditions.

2.3.6 Aggregation of the purified protein greatly hampers its nuclear localization

Next, we examined the cell transduction and nuclear translocation ability of the purified HTN-ZSCAN4 protein. The MDA-MB-231 cells were treated with purified HTN-ZSCAN4. Normally, the protein transduction media will be filtered to remove any aggregated protein. Alongside treating the cells with filtered protein transduction media, the cells were also treated with unfiltered protein transduction media to analyze the extent of aggregates and the influence it has on the cell transduction ability. The fig. 2.14A shows cells treated with both unfiltered and filtered protein media, stained using the α-ZSCAN4 antibody. The top panel shows the control cells, and a faint fluorescent signal is observed, indicating low levels of endogenous expression of ZSCAN4 in MDA-MB-231 cells (Fig. 2.14A). The middle panel shows the cells treated with unfiltered protein. On observing the merged images on the middle panel, it is evident that the majority of the protein did not localize to the nucleus, instead it was localized in the cytoplasm (Fig. 2.14A). The bottom panel shows the cells treated with filtered protein media, which shows localization of HTN-ZSCAN4 in the nucleus (Fig. 2.14A; *inset*). In the case of HTN-ZSCAN4 (filtered) transduced cells, an increased fluorescent signal was observed compared to the control cells. This confirmed that the soluble HTN-ZSCAN4 protein remaining after filtration successfully transduced into the cells and localized in the nucleus (Fig. 2.14A). These results indicate that the purified protein undergoes extensive aggregation under the current storage conditions, which significantly reduces its cellular uptake and subsequent nuclear localization.

To further confirm the formation of aggregates in the purified protein, a far-UV CD spectroscopy was performed with the purified protein, before and after removal of aggregates by centrifugation. The CD spectra before removal of aggregates was similar to the previously mentioned results (Fig. 2.9B&2.14B). Interestingly, the spectra that was obtained after removal of aggregates shows an almost flat spectra indicating very negligible, or no protein present in the solution (Fig. 2.14B). Moreover, measuring concentration of the protein sample after removal of aggregates using Bradford method showed undetectable levels of protein was present in the sample. Taken together, these results suggest that the majority of the purified protein is aggregated in the present buffer and storage conditions.

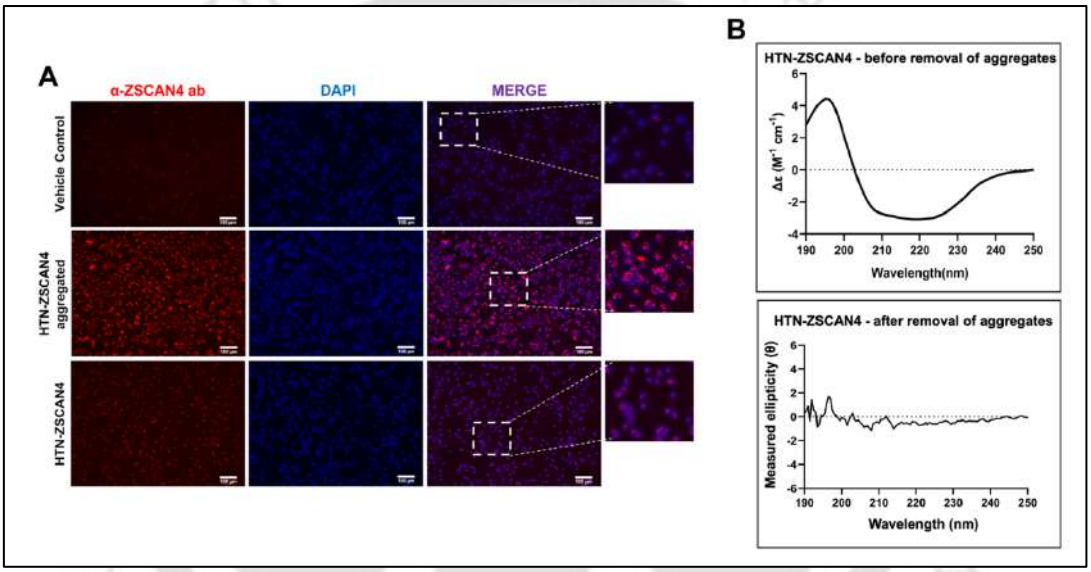


Fig. 2.14. Influence of aggregation on cell transduction and secondary structure. A) Immunofluorescence analysis of HTN-ZSCAN4 protein transduction (filtered and unfiltered) in MDA-MB-231 cells. **B)** CD spectroscopy of HTN-ZSCAN4 purified protein before and after the removal of aggregates by centrifugation, with wavelength plotted on the x-axis and measured ellipticity or Δε plotted on the y-axis. Scale bar: 100 µm.

2.3.7 Inclusion of glycerol in protein storage buffer greatly enhanced the solubility of the purified protein

To demonstrate the effect of glycerol on mammalian cells, MDA-MB-231 cancer cells were treated with different glycerol buffers for three different time points, and subsequently, an MTT assay was performed. The results show that the cells were more than 90% viable, when treated with 20 mM PB with 10% glycerol and 20 mM PB with 20% glycerol for 6 h (Fig. 2.15A). Treating the cells for longer time periods or with buffers containing glycerol beyond 20% affected the survival rate significantly (below 90%; Fig. 2.15A). Since, for future experiments, repeated protein treatment will be required, survival

rate below 90% will end up creating more stress to the cells on repeated treatment. Therefore, we chose 20 mM PB with 10% glycerol and 20 mM PB with 20% glycerol for the storage of protein and a treatment time of 6 h as the optimal conditions for future experiments. To analyze the amount of soluble protein present in these new buffer conditions, a Bradford assay was performed. The results revealed that in the absence of glycerol, almost no protein was left in soluble fraction (Fig. 2.15B). Notably, in the presence of 10% and 20% glycerol, ~20% and ~80% of protein was observed in the soluble fraction (Fig. 2.15B). Therefore, for further experiments, purified protein was eluted out in 20 mM PB in the presence of 20% glycerol and stored for further use.

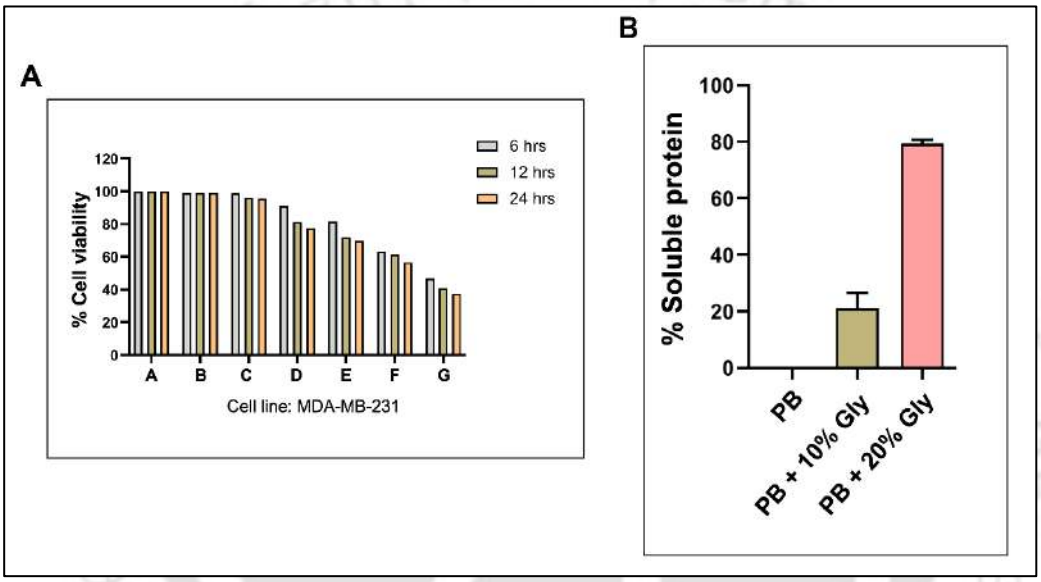


Fig. 2.15. Analysis of the effect of glycerol on cell viability and protein solubility. A) Cell viability assay to assess the effect of buffers containing different percentages of glycerol for various time points in MDA-MB-231 cells. A: growth media; B: 20 mM PB; C: 20 mM PB and 10% glycerol; D: 20 mM PB and 20% glycerol; E: 20 mM PB and 30% glycerol; F: 20 mM PB and 40% glycerol; G: 20 mM PB and 50% glycerol. **B)** Analysis of the amount of soluble protein present stored in buffers containing different percentages of glycerol by Bradford assay. PB: 20 mM phosphate buffer.

2.3.8 Transduction of purified recombinant fusion proteins

Now that we have a stable, purified protein, we next transduced the MDA-MB-231 cells with the purified protein (stored in 20 mM PB with 20% glycerol) and analyzed the transduction efficiency. Immunofluorescence results revealed that the HTN-ZSCAN4 protein has transduced inside the cell and localized in the nucleus (Fig. 2.16A). On comparing the transduction efficiency of the purified protein stored with and without glycerol, it is evident that protein stored in 20% glycerol transduces more efficiently than

its counterpart (Fig. 2.16A). Quantification of fluorescence signals from immunofluorescence images revealed that HTN-ZSCAN4 stored in 20% glycerol achieved approximately twice the transduction efficiency compared to the protein stored without glycerol (Fig. 2.16B). Similarly, the purified domain deleted proteins (HTN-ZSCAN4-ΔDB and HTN-ZSCAN4-ΔSD) were also transduced in MDA-MB-231 cells and analyzed. The fluorescent images showed that both of these proteins, HTN-ZSCAN4-ΔDB and HTN-ZSCAN4-ΔSD, have successfully transduced inside the cells and localized in the nucleus (Fig. 2.16C).

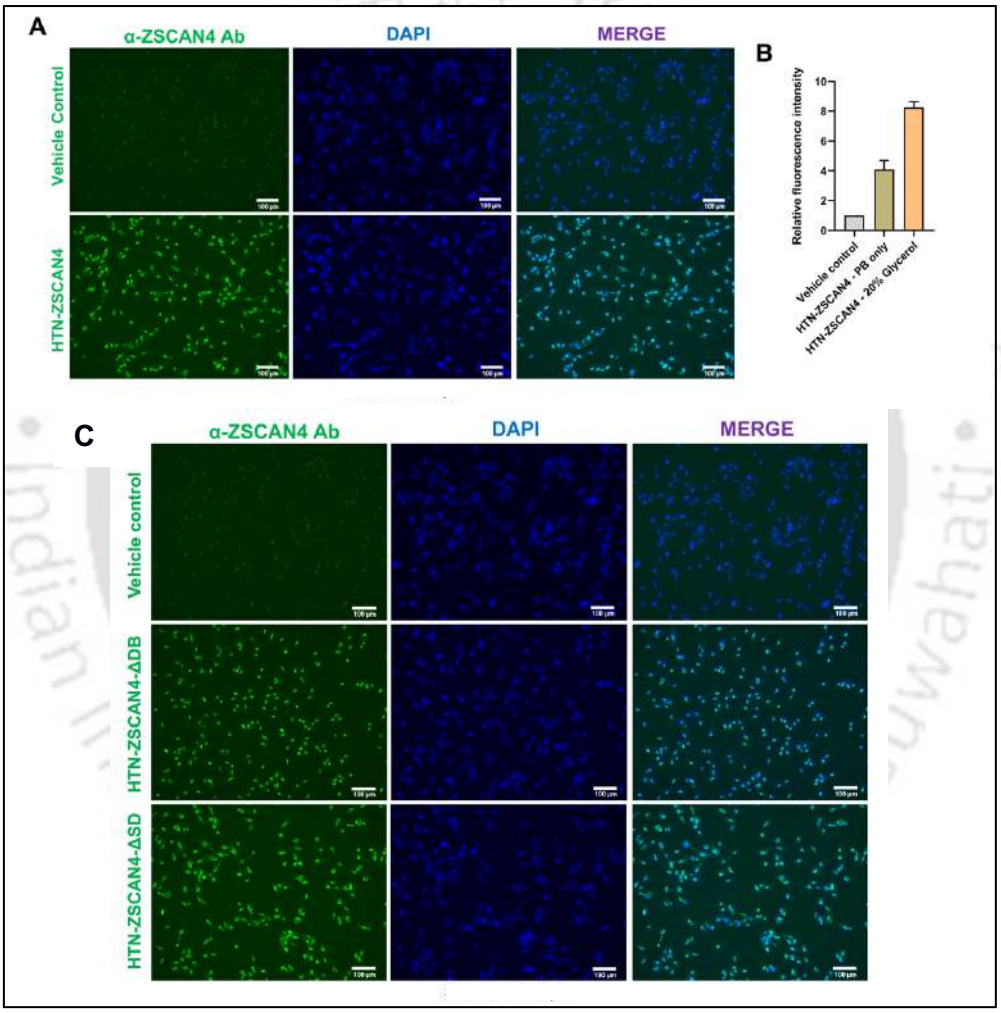


Fig. 2.16. Intracellular transduction of ZSCAN4 fusion proteins. A) Immunofluorescence analysis of HTN-ZSCAN4 protein transduction stored in 20 mM PB and 20% glycerol in MDA-MB-231 cells. **B)** Relative fluorescence intensity analysis of (A). **C)** Immunofluorescence analysis of HTN-ZSCAN4-ΔDB and HTN-ZSCAN4-ΔSD protein transduction stored in 20 mM PB and 20% glycerol in MDA-MB-231 cells. Scale bar: 100 µm.

2.4 Discussion

Bacterial systems are widely used for recombinant protein expression and purification for a variety of reasons, such as ease of culture, well-known genetics and so forth (Borgohain et al., 2018; Khow and Suntrarachun, 2012). Therefore, in this study, the bacterial expression host (*E. coli*) was chosen to express the human ZSCAN4 protein. Codon bias is one of the significant barriers that affects (Borgohain et al., 2018; Khow and Suntrarachun, 2012) the efficiency of expression during heterologous expression (Burgess-Brown et al., 2008). To overcome this, codon optimization was performed for the coding sequence of *ZSCAN4*. Codon optimization involves synonymous replacement of the rare codons with the preferred codons for *E. coli*. Subsequently, maximum expression of the recombinant ZSCAN4 protein can be achieved in *E. coli*. Several studies have shown the importance of codon optimization and its effect on the expression of recombinant proteins (Burgess-Brown et al., 2008; Maertens et al., 2010). Following codon optimization, various expression parameters were screened and optimized to obtain maximal soluble expression. In this study, the maximum amount of soluble expression is vital since the fusion protein is being purified from the soluble fraction to maintain the native conformation. Previous studies from our lab have shown the influence of various expression parameters on the overall expression of recombinant proteins (Dey et al., 2023, 2022, 2021b; Thool et al., 2023, 2021).

From the optimization results, it is evident that inducer (IPTG) concentration did not significantly influence the soluble expression of ZSCAN4 fusion proteins. In contrast, the growth phase of the bacterial culture at the time of induction was found to be a highly influential parameter. Several studies have reported the influence of the phase of growth of the bacterial culture during the time of induction on the expression of recombinant proteins (Galloway et al., 2003; García-Fraga et al., 2015; Ou et al., 2004; San-Miguel et al., 2013). All these studies report induction at different growth phases, like early log phase (San-Miguel et al., 2013), mid-log phase (García-Fraga et al., 2015) and late log phase (Galloway et al., 2003; Ou et al., 2004) induces optimal recombinant protein expression. These studies suggest that the growth phase of the culture during induction has a significant impact on the protein expression and is highly specific to each protein. In case of ZSCAN4 fusion proteins, inducing the cultures during late log phase (OD_{600} ~1.5) resulted in maximum soluble expression of these proteins. Previous studies have also reported similar results

where expression is maximum when the cultures are induced at late log phase (Galloway et al., 2003; Ou et al., 2004).

Screening of incubation time revealed that it had no significant effect on the expression of ZSCAN4 fusion proteins. Notably, in the case of ZSCAN4-NTH fusion protein, increasing the incubation time beyond 4 h reduced the soluble expression. Numerous studies have suggested that lower incubation temperature after induction, greatly aided in enhancing the soluble expression of recombinant protein (San-Miguel et al., 2013; Sørensen and Mortensen, 2005). There are several reports from our group, which have revealed the same in case of some recombinant proteins (Haridhasapavalan et al., 2023, 2022, 2021a). In case of ZSCAN4 fusion proteins, reducing the incubation temperature did not enhance the soluble expression rather it completely hampered the overall expression. This is in accordance with one of our previous studies (Haridhasapavalan et al., 2021b). Collectively, these results underscore that protein-specific optimization of expression parameters is vital to achieve maximal overall and soluble expression of recombinant proteins.

Following optimization of expression parameters, ZSCAN4 fusion proteins were purified using immobilized metal ion chromatography. Subsequently, purified proteins were characterized using far UV-CD spectroscopy. Far-UV CD spectroscopy is a routinely used technique to study the secondary structure of proteins (Greenfield, 2007; Whitmore and Wallace, 2008). The spectra will give characteristic peaks and dips depending on the secondary structure content of the protein (Greenfield, 2007; Whitmore and Wallace, 2008). The initial results revealed that the purified ZSCAN4 fusion proteins did not have a stable secondary structure. Numerous studies have shown that the concentration of salt present in the buffer can have a significant effect on the solubility and the structural stability of the purified proteins (Achouri et al., 2012; Boivin et al., 2013; Choi and Ma, 2005; Qi et al., 2019). Further, it has been demonstrated that salt can influence protein conformation by influencing hydrogen bonding in the system (Qi et al., 2019) and electrostatic and hydrophobic interactions (Damodaran and Kinsella, 1982). Therefore, various salt (NaCl) concentrations in purification buffers were screened for their influence on maintaining the secondary structure of the purified protein. Consequently, all these purified HTN-ZSCAN4 proteins were analyzed for their secondary structure using CD spectroscopy. The CD spectra revealed a stable secondary structure at 450 mM and 600 mM of NaCl

concentration, but the structure is disrupted as the salt concentration decreases. This is in agreement with previous studies showing that lowering the salt molarity can induce conformational changes and disrupt secondary structure, whereas higher salt concentrations promote structural stability (Achouri et al., 2012; Bhat et al., 2018). One of our previous studies had reported the influence of salt on the soluble expression of recombinant protein (Haridhasapavalan et al., 2021a). Interestingly, it was evident from the gel images that the salt concentration in purification buffers has a direct correlation with the amount of non-specific proteins eluting out. This can be attributed to salting out, where high salt concentration promoted protein precipitation, leading to elution of non-specific, weakly bound proteins (Choi and Ma, 2005). To eliminate these non-specific proteins from the eluate fraction, the amount of imidazole used in purification buffers was doubled. This was done to increase the stringency of the wash steps, leading to the removal of weakly-bound proteins, thereby yielding pure HTN-ZSCAN4 fusion protein.

Next, using the optimized expression and purification parameters established for the full-length ZSCAN4 protein, the domain-deleted, functionally impaired variants of ZSCAN4 proteins (ΔZSCAN4 proteins) were also expressed and purified. To assess the structural consequences of domain deletion, the purified ΔZSCAN4 proteins were analyzed using far UV-CD spectroscopy. It is interesting to note that in case of HTN-ZSCAN4-ΔDB, the spectra showed a negative peak at 208 nm, which is much more intense in comparison to the spectra of HTN-ZSCAN4. This change in the spectra is attributed to the decrease or absence of β-sheets and the presence of more α-helices in the structure (Barrow et al., 1992). According to the structure reported in the Alphafold 3 database, the DNA-binding domain (Zinc finger domain) of ZSCAN4 is the only domain that has β-sheets in its secondary structure. Zinc-finger domain present in ZSCAN4 belongs to C_2H_2 superfamily (Thool et al., 2022), which is characterized by the presence of two β-sheets and one α-helix (β/β/α). Deletion of this domain will lead to loss of those β-sheets in the secondary structure, which is in correlation with the observed CD spectra. Alternatively, there was no significant difference in the spectra of HTN-ZSCAN4-ΔSD in comparison to the wild-type protein. Further, Western blotting with ZSCAN4 antibody verified the identity of the purified proteins.

Subsequently, the stability of HTN-ZSCAN4 purified protein was assessed under standard cell culture conditions, which revealed that the protein is unstable or did not

migrate properly in the presence of media supplements like FBS, KOSR and Albumax. FBS is rich in proteins, hormones, growth factors among other substituents. Since it is extensively rich in proteins, they are commonly used to enrich the growth of mammalian cells. Also, FBS is a commonly used system to mimic the physiological conditions in testing the biocompatibility of metals and their alloys for their application in implants (Gu et al., 2011, 2009; Johnson et al., 2017; Li et al., 2019; Liu et al., 2019; Scheideler et al., 2013; Xu et al., 2022; Yamamoto and Hiromoto, 2009). Although the role of FBS in the degradation of metals and their alloys is controversial and not clearly understood, numerous studies have shown that FBS is capable of degrading certain metals and their alloys, like magnesium (Gu et al., 2009; Johnson et al., 2017) and zinc (Gu et al., 2011; Li et al., 2019). Notably, zinc is degraded at a much higher rate in the presence of FBS compared to other metals and their alloys (Gu et al., 2011; Li et al., 2019). ZSCAN4 is a transcription factor, which is characterized by the presence of four zinc finger domains towards its C-terminal end as its DNA binding motif (Thool et al., 2022). As mentioned earlier, the zinc finger domain in ZSCAN4 belongs to the C_2H_2 superfamily (Thool et al., 2022). This is characterized by a zinc atom coordinately bonded to two cysteine and two histidine residues. Further, the coordinately bonded zinc atom contributes to the structural stability of the zinc finger domain present in the protein (Berg and H-y, 1995; Frankel et al., 1987). Absence of this zinc atom leads to unfolding of this domain, leading to disruption in its structure (Berg and Hy, 1995; Frankel et al., 1987). Based on these findings, we hypothesized that the instability of ZSCAN4 in the presence of FBS might be due to the degradation of the zinc atom present in the zinc finger domain of ZSCAN4 protein by FBS, leading to structural instability. Other recombinant transcription factors purified from our lab, which do not have a zinc finger domain has been proven to be stable in the presence of FBS (Dey et al., 2023, 2022, 2021b; Thool et al., 2023, 2021). The instability observed with KOSR and Albumax is likely due to the presence of common active components, as KOSR is a synthetic analog of FBS. BSA is one of the common components in all three supplements. Therefore, it is safe to assume that BSA might be the reason for the instability of ZSCAN4 protein. But further studies are required to study the zinc-chelation mechanism and identify the specific component responsible for the instability. To overcome this, this study investigated the stability of HTN-ZSCAN4 in the presence of other defined, complete media like Essential 8 media, Stembrew media and Amniomax media. From the results, it is evident that the purified HTN-ZSCAN4 is stable in Stembrew and Amniomax media for

at least 24 h under standard cell culture conditions. These alternatives offer suitable environments for future functional assays requiring a stable ZSCAN4 protein.

Further cell transduction studies revealed that HTN-ZSCAN4 is prone to aggregation under the current buffer and storage conditions. To mitigate this, cryoprotectant like glycerol was included in the storage buffer as its inclusion in the storage buffer is known to enhance the solubility and reduce aggregation (Feng and Yan, 2008; Ghahghaei et al., 2011; Vagenende et al., 2009). Our ultimate application is to transduce these proteins in mammalian cells, and glycerol is known to have a toxic effect on cell viability of mammalian cells (Dinsdale et al., 1992; Wiebe and Dinsdale, 1991). Therefore, it is vital to find out the optimal glycerol buffer [20 mM PB; base buffer with different percentages of glycerol] for the storage of proteins that ensures protein stability without compromising the health and growth of mammalian cells. Hence, an MTT assay was performed to analyze the effect of varying concentrations of glycerol on mammalian cells. Subsequently, using the optimal glycerol buffer, the stability of the purified HTN-ZSCAN4 protein was analyzed. The results revealed that protein is ~80% soluble when stored in 20 mM PB in the presence of 20% glycerol. Finally, all the purified proteins (HTN-ZSCAN4, HTN-ZSCAN4-ΔDB, HTN-ZSCAN4-ΔSD) stored under optimal conditions were transduced into MDA-MB-231 cells. The immunofluorescence results revealed that all three purified proteins (HTN-ZSCAN4, HTN-ZSCAN4-ΔDB, HTN-ZSCAN4-ΔSD) successfully transduced inside the cells and translocated in the nucleus. This validated ZSCAN4 recombinant protein toolbox can be reliably utilized in a time- and dose-dependent manner to precisely investigate the critical role of ZSCAN4 in pluripotent stem cell maintenance, cellular reprogramming, and cancer biology.

2.5 Conclusion

This study successfully established a meticulously optimized protocol for the expression and purification of the recombinant ZSCAN4 fusion protein from *E. coli* under native conditions. Initial screening of expression parameters demonstrated that optimizing these parameters, particularly optical density at the time of induction, is essential to attain maximal soluble expression. Further, this study demonstrated the influence of NaCl concentration in the purification buffers on the generation of a structurally stable protein. Moreover, we determined that HTN-ZSCAN4 fusion protein was unstable in the presence of nutritional supplements like FBS, KOSR and Albumax. This finding hypothesized that

instability might be due to the degradation of zinc ions by FBS, which is critical for the stability of ZSCAN4. Furthermore, we successfully incorporated glycerol in the storage buffer, which was shown to mitigate protein aggregation effectively. ZSCAN4 recombinant protein toolbox generated in this study can be utilized to unfold and understand the mechanistic role of ZSCAN4 in various biological processes. As mentioned previously, ZSCAN4 has a tightly regulated, temporal expression in early embryonic development and in ESCs. The recombinant protein toolbox developed in this study provides the means to control the expression of ZSCAN4 in a time- and dose-dependent manner, paving the way to gain deeper insights. Further, this will enable us to study the role of ZSCAN4 in reprogramming in a stage-specific manner. Given that cancer cells maintain indefinite proliferative capacity largely through telomere length regulation, and ZSCAN4 is a critical regulator of this process, the toolbox offers a valuable approach to study telomere biology in greater detail. Understanding these mechanisms will advance our knowledge of cancer cell self-renewal and may assist in the development of targeted anti-cancer therapies.

Chapter 3

Generation and characterization of human iPSC line

This chapter has been accepted for publication as follows:

- **Sundaravadivelu PK**, Raina K, Borthakur A, Mahto P, Shetty D, Dhamne C, Kaveeshwar V, Martin CA, Radhakrishnan S, Rela M, Thummer RP (2026). Derivation and characterization of the iPSC line IITGi001-B from primary human adult dermal fibroblasts. Stem Cell Research. Feb 13:103931. https://doi.org/10.1016/j.scr.2026.103931 (Elsevier).

Overview

Induced pluripotent stem cells (iPSCs) are *in vitro* generated, artificial stem cells reprogrammed from terminally differentiated somatic cells by ectopic expression of stem cell-specific transcription factors. In this study, adult human dermal fibroblasts were reprogrammed using the Y4 combination of episomal plasmids expressing hOCT3/4, short hairpin RNA against p53, hSOX2, hKLF4, hL-MYC, and hLIN28 to generate iPSCs. Following the formation of colonies, a representative iPSC-like colony was extensively characterized to confirm its pluripotent characteristics and its identity. Morphologically, the established iPSC line exhibited hallmark features, including small, rounded cells with a shiny border and a high nuclear to cytoplasmic ratio. Then, the robust expression of core pluripotent nuclear markers, namely OCT4, SOX2, NANOG, REX1, and ZSCAN4 was validated using RT-qPCR, immunofluorescence, and flow cytometry. Further, the expression of pluripotent surface markers, namely TRA-1-60, TRA-1-81, and SSEA4 was determined using immunofluorescence and flow cytometry. Moreover, the absence of the integration of episomal vectors in the generated iPSC line was verified by performing semi-quantitative PCR using primers specific to the episomal plasmids used for reprogramming. Furthermore, *in vitro* trilineage differentiation capability of the established iPSC line was demonstrated via immunofluorescence and RT-qPCR, which validated the ability of the iPSC line to differentiate into cells belonging to all three germ layers, namely ectoderm, mesoderm, and endoderm. Additionally, the genetic stability of the established iPSC line was confirmed by karyotyping. Lastly, the absence of mycoplasma and the genetic identity of the generated iPSCs were verified using mycoplasma analysis and DNA fingerprinting analysis, respectively. All these results substantiate that the established iPSC line exhibits hallmark characteristics of typical iPSCs. This extensively characterized and established iPSC line can be utilized for various biological and downstream biomedical research applications.

3.1 Introduction

Induced pluripotent stem cells are generated by reprogramming adult somatic cells, through overexpression of stem cell-specific transcription factors. Human iPSCs were first generated in the year 2007 by two independent research groups using retroviral-mediated (Takahashi et al., 2007) and lentiviral-mediated (Yu et al., 2007) approaches. However, the iPSCs generated using these integrative approaches carry significant risks of insertional mutagenesis, rendering them unsafe for use in clinical and therapeutic applications. To mitigate these safety concerns, a number of non-integrative approaches have been developed to generate transgene-free iPSCs (Borgohain et al., 2019; Dey et al., 2021a; Haridhasapavalan et al., 2019). Among these, episomal plasmid-based reprogramming stands out because of its transient presence, integration-free nature, and ease in use (Dey et al., 2021a; Haridhasapavalan et al., 2019). The presence of the trans-acting factor Epstein-Barr Virus Nuclear Antigen 1 (EBNA1) and cis-acting viral origin of replication (oriP) enables episomal plasmids to replicate independently of the target cell's replication machinery while being maintained as an extrachromosomal DNA (Dey et al., 2021a; Haridhasapavalan et al., 2019). This method of reprogramming was first reported in the year 2011, where human dermal fibroblasts (HDF) were reprogrammed to generate iPSCs (Okita et al., 2011). A combination of three plasmids, named the Y4 combination, was used in the study, which proved to be the most efficient combination of factors among the different combinations screened, to generate iPSCs (Okita et al., 2011). Following this, many studies have employed this episomal vector-based approach to generate iPSCs (Meng et al., 2012; Schmitt et al., 2017; Su et al., 2013; Wen et al., 2016), but this study (Okita et al., 2011) remains a pioneering study for this approach.

The selection of the source cell for reprogramming is a critical determinant for reprogramming efficiency and the quality of the iPSC generated (Raab et al., 2014; Ray et al., 2021; Sundaravadivelu et al., 2021). Fibroblasts remain the most commonly used cell source for reprogramming because of the following reasons: (i) cheap and easy handling, (ii) well-established cell line in research, (iii) can be obtained using semi-invasive methods, (iv) easier isolation and expansion, and (v) favourable epigenetic state for reprogramming (Raab et al., 2014). Therefore, in this study, we have utilized adult human dermal fibroblasts (AHDF) to generate integration-free, *bonafide* iPSCs using the Y4 combination of episomal plasmids (Okita et al., 2011). This integration-free approach was chosen to

ensure the production of transgene-free iPSC lines. Further, the generated iPSCs were characterized extensively to demonstrate their pluripotent characteristics and to prove their identity.

3.2 Materials and Methods

3.2.1 Generation of iPSCs from human dermal fibroblasts

AHDF [ATCC (PCS-201–012)] were transfected with Y4 combination of episomal plasmids (Okita et al., 2011), namely pCXLE-hOCT3/4-shp53 (Addgene ID: 27077), pCXLE-hSK (Addgene ID: 27078), and pCXLE-hUL (Addgene ID: 27080) by electroporation. Briefly, $5x10^5$ AHDF cells were electroporated with 1 µg of each plasmid using the Neon™ Electroporation System (Thermo Fisher Scientific). Cells were resuspended in R buffer (Neon Transfection Kit) containing the specific plasmids and electroporated using the following parameters: 1650 V, 3 pulses, and 10 ms per pulse. Immediately after electroporation, cells were plated onto 6-well plates containing pre-warmed fibroblast growth medium (FGM) consisting of DMEM (GIBCO) supplemented with 10% fetal bovine serum (FBS; GIBCO), without antibiotics, and incubated at 37 °C in a humidified atmosphere with 5% CO_2. After 24 h, the cells were replenished with fresh FGM with 1% of penicillin-streptomycin (GIBCO) and subsequently refreshed every 48 h until day 6. On day 7, the cells were reseeded on matrigel (Corning)-coated 6-well plates in FGM with 1% penicillin-streptomycin at the desired ratio (ideally 1:12). The subsequent day, the cells were supplemented with TeSR™-E7™ medium for reprogramming (STEMCELL Technologies), which was replenished every 24 h until morphologically distinct and iPSC-like colonies appeared. iPSC-like colonies appeared between day 22 and 25. The colonies were picked manually and expanded on matrigel-coated plates using Essential 8™ media (GIBCO). Cells were passaged in a split ratio of 1:4 using 0.5 mM EDTA solution (Invitrogen). The generated iPSCs were cryopreserved at multiple passages for downstream characterization and experimental use.

3.2.2 Characterization of the generated iPSCs for the expression of markers

All the experiments were performed with the generated iPSCs of less than passage 10 unless otherwise mentioned. The generated iPSCs were characterized for the expression of key surface and nuclear markers using RT-qPCR, immunofluorescence, and flow cytometry.

3.2.3 Real-Time quantitative PCR (RT-qPCR)

RNA was isolated from cells using TRIzol (Invitrogen) reagent. Subsequently, cDNA was synthesized using a cDNA synthesis kit (Bio-Rad) following the manufacturer's instructions. RT-qPCR was performed using 2X SYBR green supermix (Bio-Rad) in AriaMx quantitative PCR systems (Agilent). The results were normalized using GAPDH as the reference gene. The results thus obtained were visualized and analyzed using AriaMx software (Agilent). The relative expression levels were calculated using the standard $2^{-\Delta\Delta ct}$ method, and the graphs and statistical analyses were done using GraphPad Prism 8 software. The results were plotted as mean±SEM, and they were statistically analyzed using student's t-test with 95% confidence interval.

3.2.4 Immunofluorescence

Immunofluorescence was performed as previously described in section 2.2.8. Briefly, the cells were grown up to 70 to 80% confluency for immunofluorescence. The cells were fixed, permeabilized, blocked and incubated overnight at 4 °C with respective primary antibody (Table S3) diluted in blocking solution. Subsequently, they were incubated in respective fluorophore-conjugated secondary antibody (Table S3) diluted in blocking solution for 1 h at room temperature. Finally, the nucleus was counterstained with Hoechst 33342 (Cell Signaling Technology; 1:10000 in PBS) for 5 min at room temperature, and then the cells were observed and imaged using an inverted fluorescence microscope (ZOE™ Fluorescent Cell Imager, Bio-Rad).

3.2.5 Flow cytometry

The cells were dissociated using TrypLE™ Express Enzyme (GIBCO), washed twice with 1X PBS, and then the cells were subjected to live/dead staining using propidium iodide dye. For surface markers, the cells were incubated using the specific primary antibody, followed by the respective secondary antibody conjugated with Alexa Fluor™ 488 (Table S3). For nuclear markers, the cells were fixed with 4% paraformaldehyde, permeabilized using 0.1% Triton-X-100 in 1X PBS and blocked using 1X PBS containing 0.5% BSA and 10% normal goat serum prior to antibody staining. Data acquisition was performed using a BD FACSLyric™ flow cytometer, with unstained cells used as negative control and single fluorophore-stained samples used for compensation. Flow cytometry data were analyzed and represented using FlowJo™ software.

3.2.6 In vitro trilineage differentiation

Pluripotency of the iPSCs was assessed using an *in vitro* directed trilineage differentiation assay. The iPSCs were differentiated into derivatives of three germ layers, namely ectoderm, mesoderm, and endoderm, using STEMdiff™ Trilineage Differentiation Kit (Cat. No.: 05230; STEMCELL Technologies) according to the manufacturer's instructions. Briefly, iPSCs were harvested as a single-cell suspension and seeded on day 0 at a density of 1.5×10^4 cells per well for ectoderm and endoderm differentiation and 1.0×10^4 cells per well for mesoderm differentiation on matrigel-coated 24-well plates. On day 1, respective media (either ectoderm or mesoderm or endoderm differentiation media) was applied to the cells for lineage-specific differentiation until day 5 (for mesoderm and endoderm) or day 7 (for ectoderm). At the end of the differentiation period, cells were subjected to immunofluorescence and RT-qPCR analysis to assess the expression of lineage-specific markers.

3.2.7 Genomic DNA isolation

The genomic DNA was isolated using the standard phenol:chloroform:isoamyl alcohol (25:24:1) (PCI; Sigma-Aldrich) protocol. Briefly, one volume of PCI was added to the cells, and the mixture was centrifuged at 16,000×g for 5 minutes. The upper aqueous phase containing the genomic DNA was collected. To the aqueous phase, glycogen, ammonium acetate, ethanol, RNAse, and Proteinase K were added and incubated at -20 °C overnight to precipitate the genomic DNA. Subsequently, the precipitated genomic DNA was pelleted by centrifuging at 16,000×g for 30 minutes at 4 °C. Then, the pellet was washed with 70% ethanol, air dried, and resuspended in Tris-EDTA buffer. The isolated genomic DNA was stored at -20 °C until further use.

3.2.8 Semi-quantitative PCR

Semi-quantitative PCR was performed using the isolated genomic DNA. 100 ng of genomic DNA was used per reaction of PCR, and the corresponding episomal plasmids (20 ng) served as a positive control, and nuclease-free water was used as a no-template control. PCR was performed using specific primers (Table S2) with Taq DNA polymerase (New England Biolabs) in a Mastercycler® nexus X2 gradient (Eppendorf). The PCR conditions were as follows: 95 °C for 5 min, 35 cycles of 95 °C for 30s, 54 °C for 30s, 72 °C for 30s, and final extension at 72 °C for 10 min. The PCR products were resolved using 2% agarose

gel and imaged using ChemiDoc™ XRS+ (Bio-Rad). The image was analyzed and processed using Image Lab™ software (Bio-Rad, California, USA).

3.2.9 Karyotyping

Conventional cytogenetic analysis was performed to assess the chromosomal integrity of the generated iPSCs. The iPSCs (passage 10) were arrested in metaphase by treatment with 200 µg/ml of HiKaryoXL Colcemid Solution (HiMedia) for 20 min. Then, the cells were harvested with 0.05% Trypsin-EDTA (HiMedia) and treated with hypotonic KCl solution for 15 min in a water bath at 37 °C. The cells were fixed with Carnoy's fixative and centrifuged at 1000 rpm for 10 min at room temperature. These fixed cells were resuspended in 5 ml of Carnoy's fixative and spread on a slide where they were banded according to standard GTG-banding protocol. Images were acquired using the AxioImager A1 (Carl Zeiss, Germany), and G-banded metaphases were analyzed using Ikaros Software (Metasystems, Germany). A total of 20 metaphase spreads were karyotyped and analyzed.

3.2.10 Short tandem repeats (STR) analysis

Cell line authentication was performed using Short Tandem Repeat (STR) profiling. Cell pellet samples corresponding to the iPSCs (Passage 6) and AHDF (Passage 7) were submitted to MedGenome Labs Ltd. (Bangalore, India) for analysis. STR typing was carried out using a validated multiplex STR assay amplifying 16 STR loci located on autosomal chromosomes (2, 3, 4, 5, 7, 8, 11, 12, 13, 16, 18, 19, and 21) and sex chromosomes (AMEL). PCR amplification of STR loci was followed by capillary electrophoresis, and allele calling was performed based on comparison with an internal allelic ladder. Positive control of DNA was included to confirm assay performance. Data analysis was conducted using standard fragment analysis software, and electropherograms were reviewed for peak quality and allele assignment.

3.2.11 Mycoplasma analysis

Mycoplasma contamination was evaluated in passage 6 iPSCs using a validated PCR-based mycoplasma detection assay (HiMedia Laboratories, India). PCR reactions were prepared in accordance with the manufacturer's instructions using a commercial mycoplasma detection kit (HiMedia). Each assay run included a positive control (mycoplasma DNA), a negative control (nuclease-free water), and the test sample. Amplification was performed in a calibrated thermocycler (Model No.: MB-PM-21), and PCR products were analyzed by 1.5% agarose gel electrophoresis using a 100 bp DNA

ladder. Assay validity was confirmed by the appropriate performance of the positive and negative controls.

3.3 Results

3.3.1 Morphological analysis of IITGi001-B iPSCs

AHDF cells were reprogrammed using the Y4 combination of episomal plasmids (Okita et al., 2011) to generate human iPSCs. The plasmids were electroporated in AHDF cells on day 0. The cells were reseeded on matrigel-coated plates in reprogramming media on day 7. From day 15 to day 20, mesenchymal to epithelial transition (MET) was observed. The MET was characterized by distinct morphological features like smaller, spindle-shaped, closely packed cells in between elongated AHDF cells (Fig. 3.1A). These MET eventually progressed into definitive iPSC-like colonies, which were prominent between day 22 and 25. These colonies exhibited typical iPSC-like morphological features like small, rounded, tightly packed cells with shiny boundary and a high nuclear to cytoplasmic ratio (Fig. 3.1A), similar to ESCs. In total, 25 colonies were observed, out of which 13 colonies were manually picked using a micropipette and expanded. After picking, these colonies maintained the typical iPSC-like morphology (Fig. 3.1B). A representative colony was characterized extensively for expression of pluripotency markers and *in vitro* differentiation capability. The chosen representative iPSC colony was registered in hPSCreg (Stem cell line repository), and the unique stem cell identifier is IITGi001-B (hereafter referred as IITGi001-B).

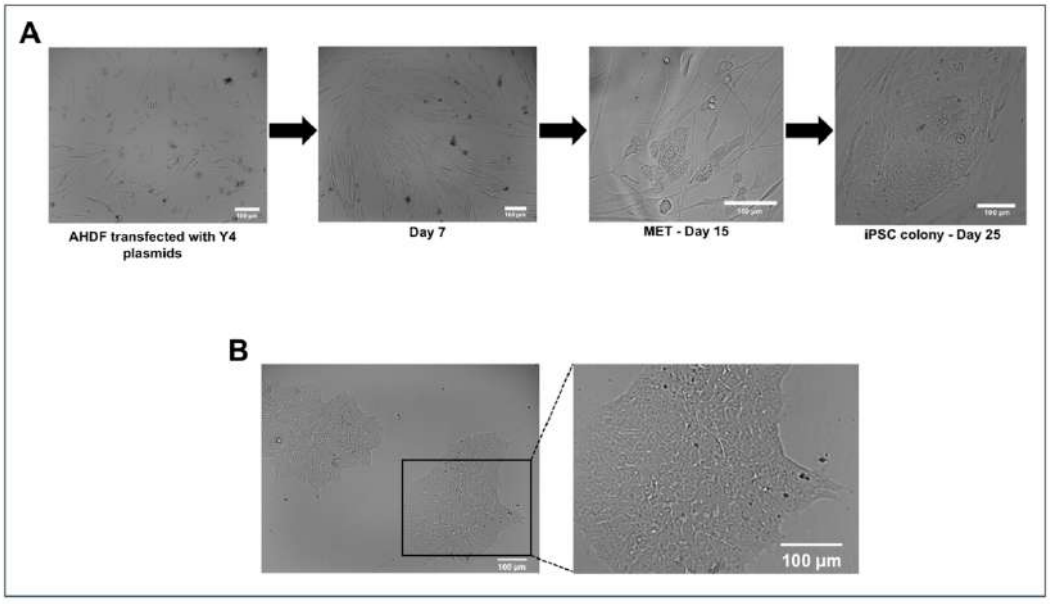


Fig. 3.1. Morphological changes observed during reprogramming. **(A)** Brightfield microscopy images taken at different stages of reprogramming until an iPSC-like colony was formed. **(B)** Brightfield microscopy image showing the morphology of an iPSC colony after picking. Scale bar: 100 µm.

3.3.2 Expression analysis of pluripotency surface and nuclear markers of IITGi001-B iPSCs

The expression levels of core pluripotency nuclear markers, namely OCT4, SOX2, NANOG, and REX1 were analyzed using RT-qPCR to evaluate the molecular identity of the generated IITGi001-B iPSC line. Due to negligible or undetectable expression of these markers, particularly OCT4, NANOG and REX1, in the parental human dermal fibroblasts (AHDF), relative quantification using fold-change ($2^{-}\Delta\Delta Ct$) would yield disproportionately large and biologically uninformative values. Therefore, the graphs are plotted with ΔCt values (normalized to an internal housekeeping gene GAPDH), where lower ΔCt value translates to higher relative expression (higher transcript abundance). The results showed robust expression of all four core pluripotency markers. Notably, OCT4 exhibited highest relative expression (lowest ΔCt value) compared to AHDF (Fig. 3.2). Further, SOX2, NANOG, and REX1 demonstrated higher relative expression, relative to AHDF cells (Fig. 3.2). These data confirm the successful activation of the endogenous pluripotency gene regulatory network.

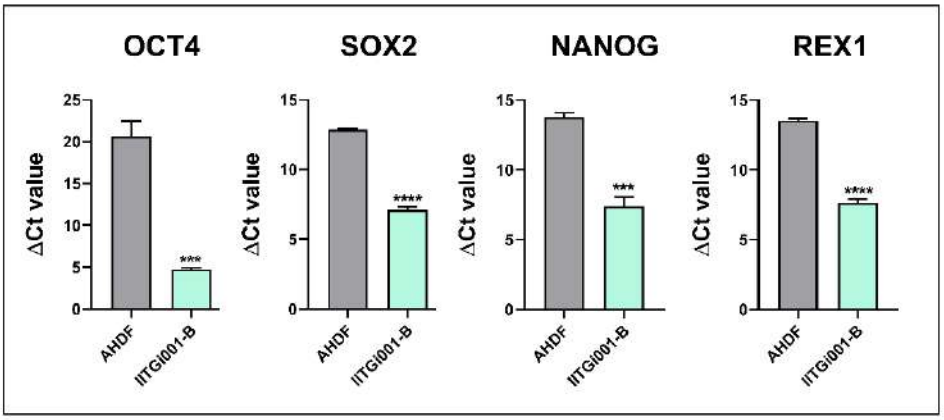


Fig. 3.2. Expression levels of pluripotency markers relative to AHDF cells. RT-qPCR results showing the relative mRNA levels of pluripotency markers compared to AHDF cells. The expression levels were normalized using GAPDH, and the results were plotted as mean±SEM. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

To evaluate the pluripotency of the newly established IITGi001-B line, its nuclear marker expression profile was compared against two validated and established iPSC lines,

namely AAVS1-PDi-Cas9 (hereafter referred as TJC8iCas9; Manian et al., 2018; Thamodaran et al., 2021) and our in-house IITGi001-A (Raina et al., 2023). The quantitative analysis revealed that the generated iPSC line, IITGi001-B, has a comparable transcriptional profile for the nuclear markers, OCT4, SOX2, NANOG, and REX1 relative to TJC8iCas9 and IITGi001-A (Fig. 3.3). In comparison to TJC8iCas9, the generated iPSC line IITGi001-B exhibited slightly elevated expression of OCT4, SOX2, and REX1, while NANOG levels were marginally lower (Fig. 3.3; *left*). In contrast, a comparison with IITGi001-A cell line revealed slightly higher expression of SOX2 but a lower relative expression of OCT4 and NANOG (Fig. 3.3; *right*). The expression level of REX1, a naïve pluripotency marker, remained consistent in both the cell lines (Fig. 3.3; *right*). Collectively, these comparative data confirm that IITGi001-B iPSC line robustly expressed the core pluripotency markers at levels consistent with validated iPSC lines.

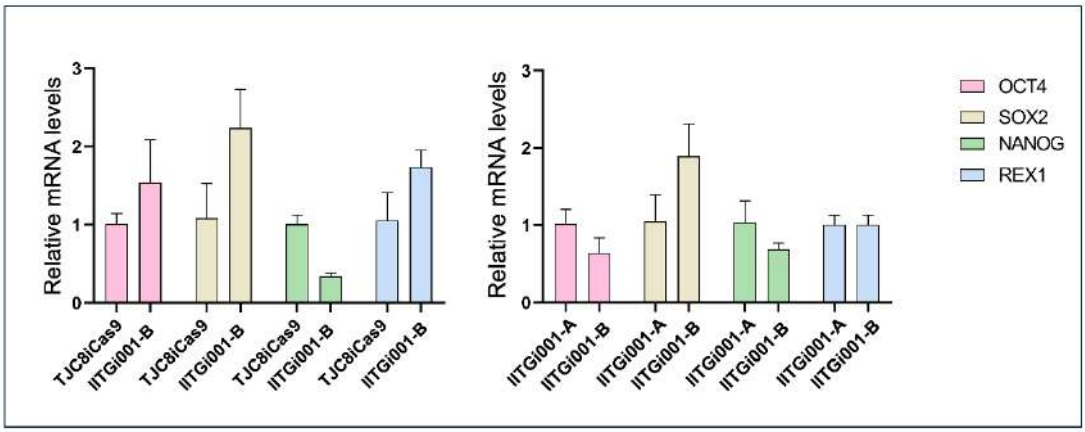


Fig. 3.3. Expression levels of pluripotency markers relative to TJC8iCas9 and IITGi001-A cells. RT-qPCR results showing the relative mRNA levels compared to TJC8iCas9 cells (left) and IITGi001-A cells (right). The expression levels were normalized using GAPDH, and the results are plotted as mean±SEM.

Following transcriptomic validation, the expression of pluripotency markers was examined at the protein level by performing immunofluorescence. Firstly, the expression of characteristic surface markers, namely TRA-1-60, TRA-1-81, and SSEA4 was assessed. These fluorescent images revealed the robust expression of these markers at the surface of the cells (Fig. 3.4; *left*). Subsequently, the expression of nuclear markers was investigated. The immunofluorescence results revealed strong expression of nuclear markers, namely OCT4, SOX2, NANOG, and ZSCAN4 (Fig. 3.4; *right*). The merged images of the markers and Hoechst 33342 nuclear counterstain prove that these markers are localized in the nuclei of the cells (Fig. 3.4; *right*), consistent with their expected subcellular localization as

transcription factors. Collectively, these data demonstrate that the IITGi001-B line consistently expressed the essential nuclear and surface-specific pluripotency markers, authenticating its identity as a *bonafide* iPSC line.

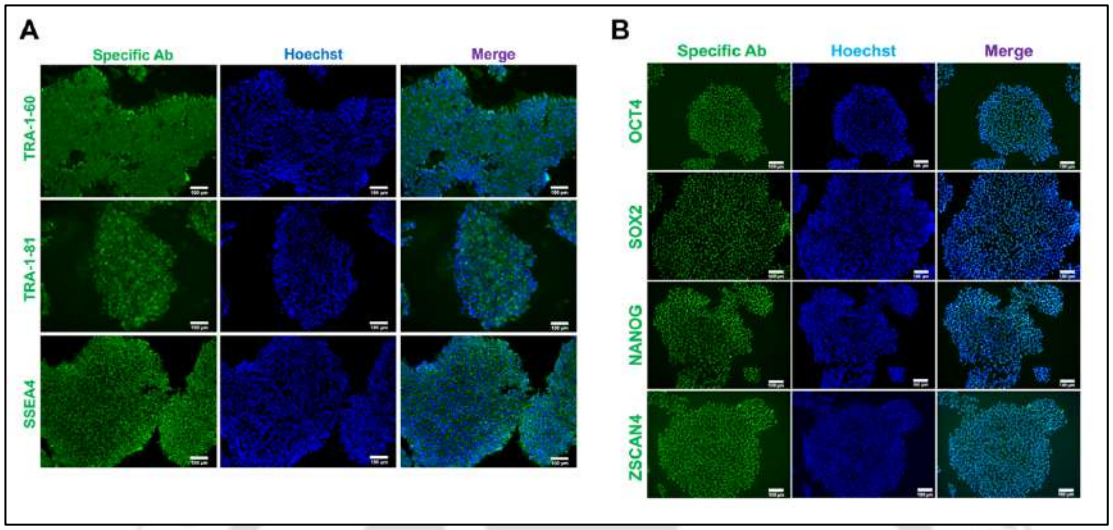


Fig. 3.4. Characterization of IITGi001-B cells for expression of pluripotency markers using immunofluorescence. Immunofluorescence was performed to analyze the expression of surface **(A)** and nuclear markers **(B)** in IITGi001-B cells. The images were visualized in an inverted fluorescence microscope. Scale bar: 100 µm.

Next, the expression of surface and nuclear markers was analyzed using flow cytometry to quantitatively evaluate the homogeneity of the IITGi001-B iPSC line. To ensure data accuracy, cellular viability was first assessed via propidium iodide staining, and the subsequent expression of markers analysis was strictly performed on the gated healthy population. The results demonstrated a highly homogeneous expression profile, with over 90% of the population stained positive for the surface markers, namely TRA-1-60, TRA-1-81, and SSEA4 (Fig. 3.5). Similarly, quantitative assessment of nuclear markers revealed more than 95% of the cells were positive for the expression of OCT4, SOX2, NANOG, and ZSCAN4 (Fig. 3.5).

In summary, the robust and uniform expression of these hallmark surface and nuclear markers at both RNA and protein levels confirmed the successful establishment of a *bonafide* IITGi001-B iPSC line.

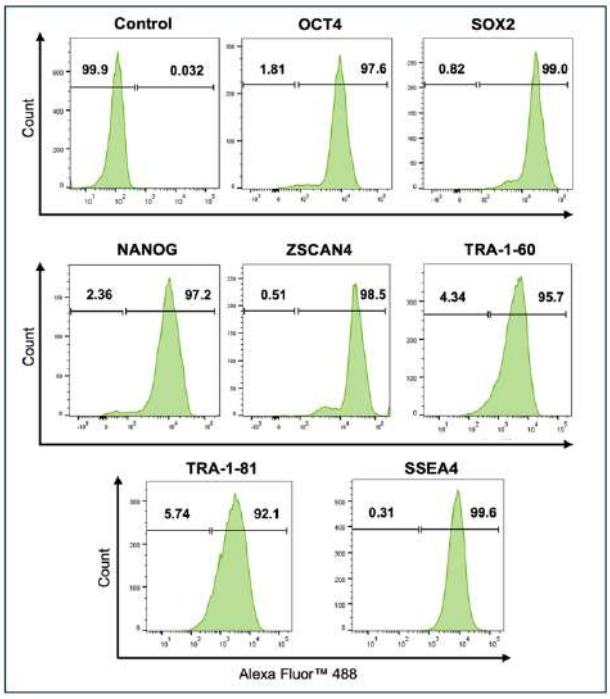


Fig. 3.5. Characterization of IITGi001-B cells for expression of pluripotency markers using flow cytometry. Flow cytometry was performed to analyze the expression of nuclear and surface markers in IITGi001-B cells. The results were analyzed and depicted using FlowJo software.

3.3.3 Assessment of integration of episomal plasmids in the genome of IITGi001-B iPSCs

As mentioned earlier, the Y4 combination episomal plasmids (Okita et al., 2011) were used for the generation of IITGi001-B iPSCs. A semi-quantitative PCR was performed to confirm the transgene-free status of IITGi001-B iPSC line using the genomic DNA as the template. Vector-specific primers targeting each of the three episomal plasmids within the reprogramming cocktail were utilized. Assay reliability was established by using the plasmids as positive controls, which yielded high intensity amplicons at the expected size (Fig. 3.6; *second lane*), thereby verifying the integrity and specificity of the primers used for the analysis. The absence of amplification in the no-template control (Fig. 3.6; *last lane*) further confirmed the specificity of the assay. No detectable bands (Fig. 3.6; *third lane*) were observed in the samples containing IITGi001-B genomic DNA at sizes corresponding to those in the positive control (*second lane*). This validated the integration-free nature of the generated IITGi001-B iPSC line, confirming that the reprogrammed cells have returned to a transgene-free state.

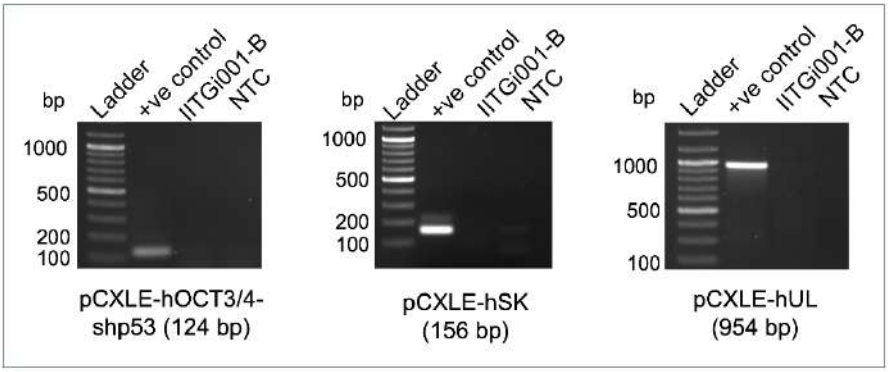


Fig. 3.6. Assessment of transgene-free status of IITGi001-B cells. Semi-quantitative PCR was performed to verify the absence of episomal plasmids in the genome of the cells using plasmid-specific primers. The plasmids were used as positive control and nuclease-free water was used as no-template control. bp: basepairs; NTC: no-template control.

3.3.4 Trilineage differentiation potential of IITGi001-B iPSCs

The IITGi001-B iPSCs were differentiated into three germ layers, namely ectoderm, mesoderm, and endoderm to functionally validate the pluripotent characteristics of the cells. Following the differentiation protocol, the lineage commitment was rigorously assessed through analysis of lineage-specific markers at the RNA and protein levels using immunofluorescence and RT-qPCR.

First, the expression of lineage-specific markers was analyzed after trilineage differentiation by performing immunofluorescence. The fluorescent images revealed robust expression of PAX6 in the cells differentiated towards ectoderm (Fig. 3.7). Similarly, successful commitment to the mesodermal and endodermal lineages was confirmed by the strong expression of Brachyury and SOX17, respectively (Fig. 3.7). These findings demonstrate that the generated IITGi001-B iPSCs have trilineage differentiation potential, confirming its status as a functionally pluripotent stem cell model.

In parallel, expression of additional lineage-specific markers was analyzed by performing RT-qPCR to attain further molecular evidence on the pluripotent characteristics of the generated IITGi001-B iPSCs. Transcript levels were normalized against undifferentiated iPSCs to determine the fold-enrichment for each germ layer. The RT-qPCR results revealed increased relative expression of the lineage-specific markers compared to undifferentiated cells (Fig. 3.8). A strong, robust expression of ectoderm-specific markers, OTX1, OTX2, and DLK1 was observed relative to undifferentiated cells

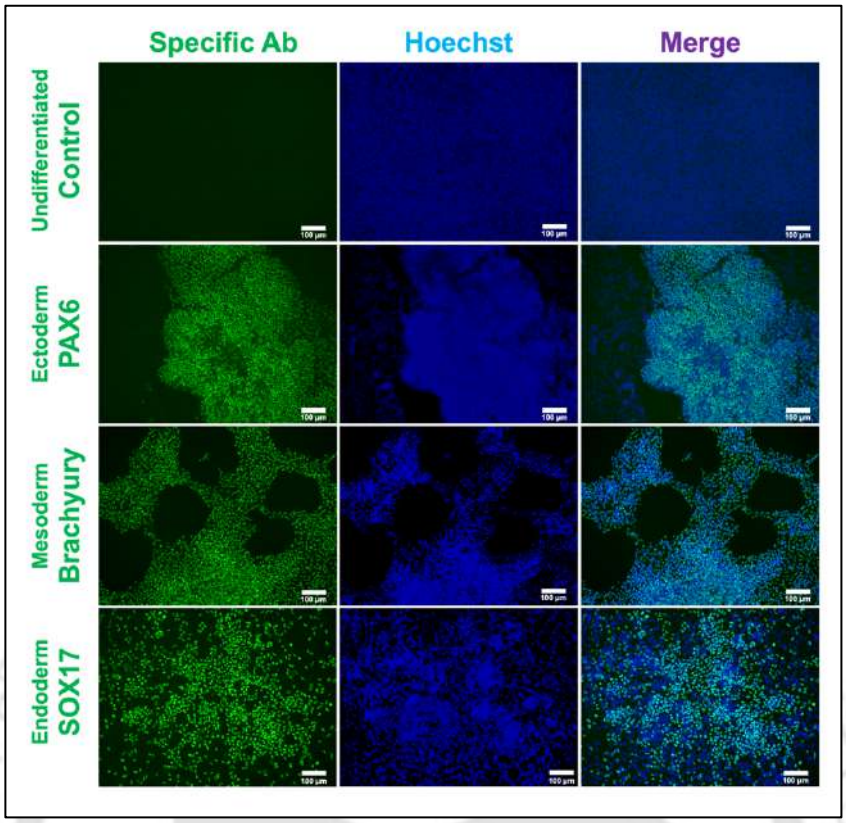


Fig. 3.7. Assessment of *in vitro* trilineage differentiation potential using immunofluorescence. The IITGi001-B cells were induced to differentiate towards all three germ layers, and the expression of lineage-specific markers were analyzed using immunofluorescence. The images were visualized using an inverted fluorescence microscope. Scale bar: 100 µm.

(Fig. 3.8). These data provided strong evidence that the IITGi001-B iPSCs differentiated towards ectoderm. Similarly, cells directed towards mesodermal fate exhibited robust expression of characteristic mesoderm markers, namely CDX2, TBX6, and CXCR4 (Fig. 3.8), compared to undifferentiated cells. These results show that the IITGi001-B iPSC differentiated towards mesoderm lineage. Finally, for endoderm lineage, expression of endoderm-specific markers, namely CER1, EOMES, and FOXA2 was studied. The results revealed strong expression of all the three markers, relative to undifferentiated cells (Fig. 3.8). Collectively, these quantitative data demonstrate that the generated iPSC line, IITGi001-B cells possess the developmental plasticity to differentiate into derivatives of all three germ layers, thereby authenticating their pluripotent characteristics.

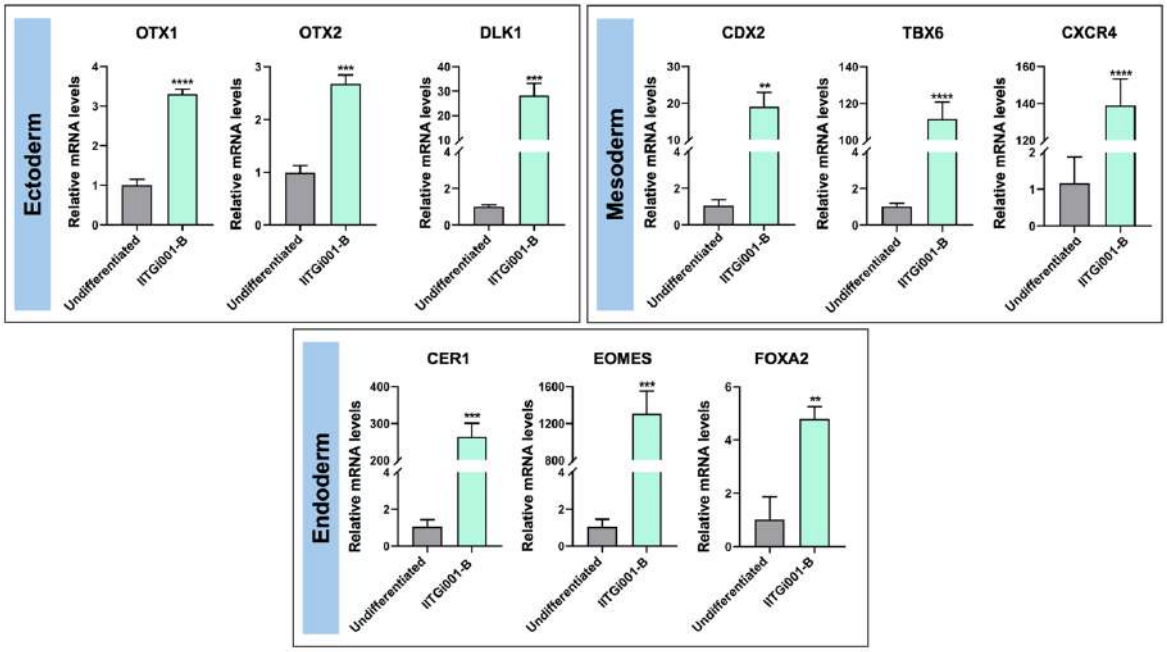


Fig. 3.8. Assessment of *in vitro* trilineage differentiation potential using RT-qPCR. The IITGi001-B cells were induced to differentiate towards all three germ layers, and the expression of lineage-specific markers were analyzed using RT-qPCR. The expression levels were normalized using GAPDH, and the results were plotted as mean±SEM. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

3.3.5 Karyotype analysis of IITGi001-B iPSCs

Karyotyping was performed to assess the genomic integrity of the generated iPSC line, IITGi001-B. Standard GTG-banding analysis of both AHDF cells and IITGi001-B cells revealed a consistent, normal diploid (46, XX) constitution with no detectable chromosomal abnormalities (Fig. 3.9). Detailed GTG-banding analysis across 20 independent metaphase spreads confirmed that the iPSCs maintained full chromosomal integrity. No evidence of numerical or structural abnormalities, including aneuploidy, translocations, deletions, or duplications was observed (Fig. 3.9). These data demonstrate that the episomal reprogramming process did not compromise the cytogenetic profile of the cell line. The preservation of a stable, normal genotype at an early passage further validated the suitability of the IITGi001-B iPSC line for research and downstream applications.

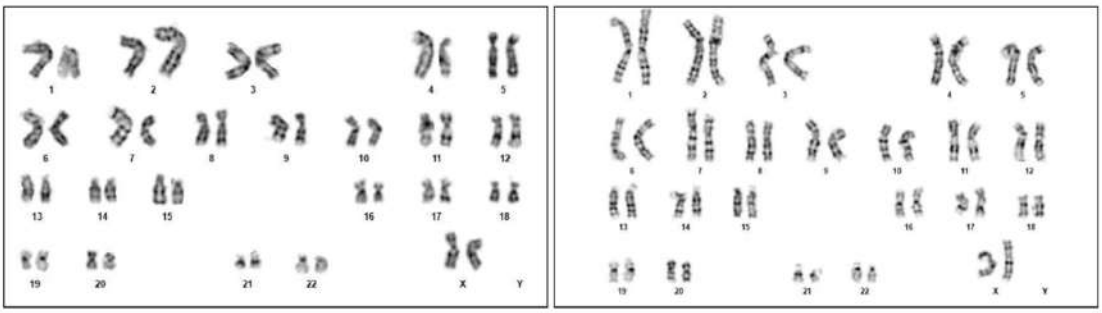
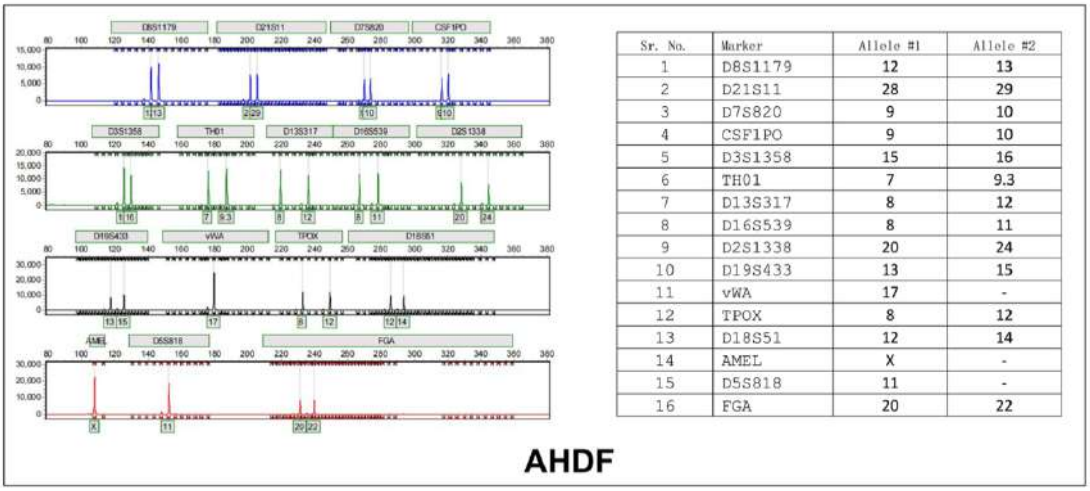


Fig. 3.9. Karyotyping analysis of AHDF and IITGi001-B cells. The genomic stability of the AHDF cells (*left*) and IITGi001-B cells (*right*) was studied by performing G-banded karyotyping. A total of 20 metaphase spreads were analyzed.

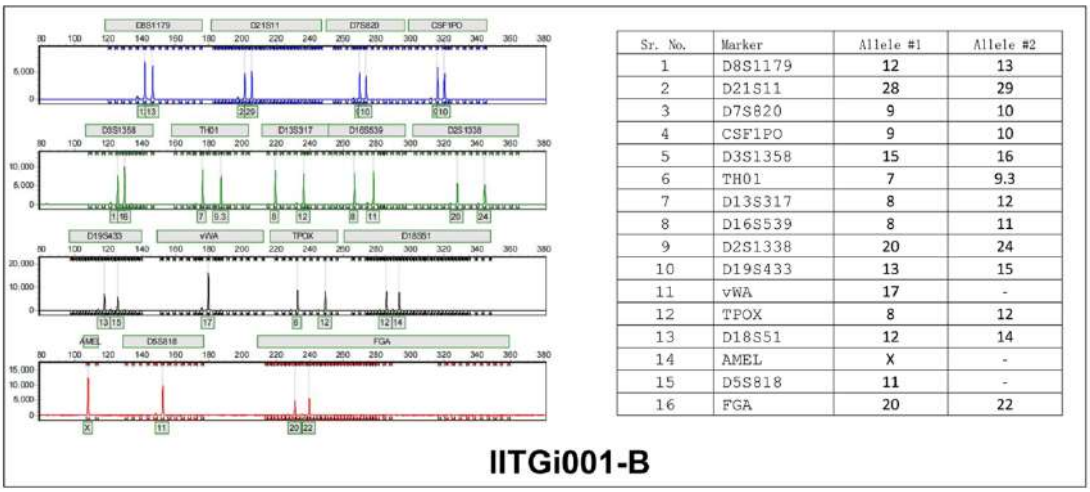
3.3.6 Short Tandem Repeats (STR) analysis of IITGi001-B iPSCs

Short Tandem Repeats (STR) analysis was performed to authenticate that the generated iPSC line IITGi001-B originated from the parental AHDF line. The number of repeats of each marker present in both alleles at specific loci was analyzed with the genomic DNA of both AHDF cells and IITGi001-B cells. The perfect match of the number of repeats of each marker verifies the identity of the generated iPSC line with the parental AHDF cell line. The STR profiling was done in 16 loci, encompassing 15 loci in somatic or asexual chromosomes, namely D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, FGA, and one locus in sex chromosome, AMEL (Fig. 3.10). The target loci were amplified and resolved using capillary electrophoresis. The electrogram showed different intensity peaks for each of the markers, signifying the number of repeats of that specific marker at each allele. The AMEL locus exhibited a single peak corresponding to the X chromosome, identifying the AHDF source as a female origin. A table with the number of repeats of each marker in both alleles has been generated (Fig. 3.10). A comprehensive comparison of the allelic sizes across all 16 loci revealed a 100% match between the IITGi001-B cells and AHDF cells (Fig. 3.10). These identical STR profiles provided definitive evidence of genetic identity and confirmed that the reprogramming process maintained genomic stability without cross-contamination. This rigorous authentication established the IITGi001-B line as a validated tool for biomedical research.



Sr. No.	Marker	Allele #1	Allele #2
1	D8S1179	12	13
2	D21S11	28	29
3	D7S820	9	10
4	CSF1PO	9	10
5	D3S1358	15	16
6	TH01	7	9.3
7	D13S317	8	12
8	D16S539	8	11
9	D2S1338	20	24
10	D19S433	13	15
11	vWA	17	-
12	TPOX	8	12
13	D18S51	12	14
14	AMEL	X	-
15	D5S818	11	-
16	FGA	20	22

AHDF



Sr. No.	Marker	Allele #1	Allele #2
1	D8S1179	12	13
2	D21S11	28	29
3	D7S820	9	10
4	CSF1PO	9	10
5	D3S1358	15	16
6	TH01	7	9.3
7	D13S317	8	12
8	D16S539	8	11
9	D2S1338	20	24
10	D19S433	13	15
11	vWA	17	-
12	TPOX	8	12
13	D18S51	12	14
14	AMEL	X	-
15	D5S818	11	-
16	FGA	20	22

IITGi001-B

Fig. 3.10. Short Tandem Repeats (STR) analysis of IITGi-001B cells. The genetic identity of the IITGi001-B cells was verified by analyzing the STR loci across the genome of AHDF cells (*top*) and IITGi001-B cells (*bottom*), and comparing of the number of repeats of each STR locus in both the alleles.

3.3.7 Confirmation of the absence of mycoplasma in IITGi001-B iPSCs

The absence of mycoplasma was verified using a highly sensitive, PCR-based mycoplasma detection kit. The presence or absence of mycoplasma-specific genomic sequences was visualized using agarose gel electrophoresis following amplification with specific primers. Assay validity was confirmed using a positive control DNA (included in the kit), and a no-template control (water) used as a negative control (Fig. 3.11). No amplification was observed with the genomic DNA of IITGi001-B cells (Fig. 3.11; *last lane*), whereas clear, definitive amplification was observed in the positive control (Fig. 3.11; *second lane*). Based on these findings, the IITGi001-B cell sample was interpreted as negative for mycoplasma contamination, meeting the rigorous biosafety standards required for validated pluripotent stem cell lines.

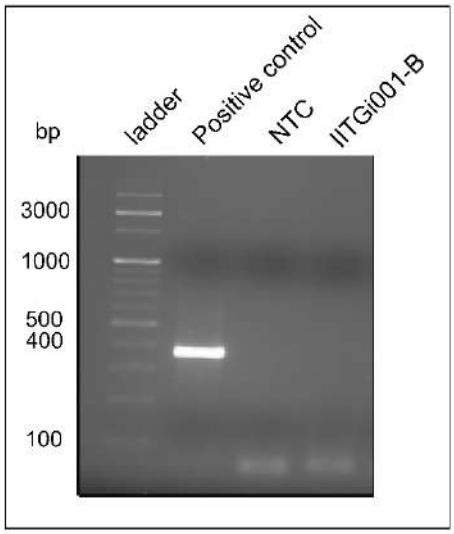


Fig. 3.11. Confirmation of the absence of mycoplasma in IITGi-001B cells. The absence of mycoplasma in IITGi001-B cells was verified using a PCR-based detection kit. The genome of IITGi001-B cells was analyzed for the presence of mycoplasma-specific genomic sequences using specific primers. The experiment was performed with necessary positive control and nuclease-free water was used as no-template control. bp: basepairs; NTC: no-template control.

3.4 Discussion

Episomal plasmids are one of the most efficient and safest ways to generate iPSCs. Since these plasmids possess an autonomous and extrachromosomal replication, they do not integrate with the genome, resulting in the generation of transgene-free iPSCs, free from genetic modifications (Dey et al., 2021a; Haridhasapavalan et al., 2019; Okita et al., 2011). Hence, these iPSCs have a wide range of biomedical applications like disease modelling, drug screening, functional genomic studies, and understanding human developmental biology. Ideally, reprogramming human fibroblasts to iPSCs involves a 25-day protocol. A critical developmental milestone, mesenchymal to epithelial transition (MET), was observed on day 15, followed by the emergence of iPSC colonies on day 20. By day 25, colonies displaying hallmark pluripotent morphology, distinct from the surrounding AHDF cells, were manually picked for further expansion and cryopreservation. In total, 25 iPSC colonies were observed, and subsequently, 13 iPSC colonies were picked and cryopreserved for future use. Since $5x10^5$ AHDF cells were used for reprogramming, leading to a reprogramming efficiency of ~0.005%, which is

comparable to other previously reported studies (Raab et al., 2014; Ray et al., 2021). Subsequently, one of the colonies was chosen and characterized extensively.

Firstly, the pluripotency of the generated iPSC line was authenticated through a multifaceted expression analysis of hallmark nuclear and surface markers. Similar to previous studies, standard surface markers were used for analysis, namely TRA-1-60, TRA-1-81, and SSEA4 (Bos et al., 2026; Jiang et al., 2026; Raina et al., 2023). Similarly, in the case of nuclear markers, standard markers were used for assessment, namely OCT4, SOX2, NANOG, and REX1, which are widely used to verify the pluripotency of stem cells (Bos et al., 2026; Jiang et al., 2026; Raina et al., 2023). Additionally, ZSCAN4 was also used as one of the markers, since it marks the genomically stable iPSCs (Hirata et al., 2012; Jiang et al., 2013). To ensure experimental rigor, we employed a combinatorial analytical approach including RT-qPCR, immunofluorescence, and flow cytometry. All these experiments demonstrated high levels of expression of all the markers, verifying the identity of the generated IITGi001-B iPSC line. In addition, expression of nuclear markers was analyzed relative to established iPSC lines (TJC8iCas9; IITGi001-A) using RT-qPCR. The IITGi001-B cells showed comparable expression of these nuclear markers, consistent with our previous study (Raina et al., 2023), further confirming the identity of the iPSCs. Further, to verify the integration-free status of the generated IITGi001-B iPSCs, PCR was performed using plasmid-specific primers. The PCR results revealed the absence of all three episomal plasmids in the genomic DNA of IITGi001-B cells, demonstrating that the established iPSCs have cleared the exogenous vectors and returned to a transgene-free state.

Furthermore, to evaluate the developmental plasticity of IITGi001-B cells, directed *in vitro* trilineage differentiation was performed, followed by dual-level analysis of expression of lineage-specific markers by using RT-qPCR and immunofluorescence. The results demonstrated the inherent capacity of the iPSC line to differentiate into derivatives of all three germ layers. Similar results were reported in previous studies as well (Bos et al., 2026; Jiang et al., 2026; Raina et al., 2023). STR analysis was performed to confirm that the generated IITGi001-B cells originated from the parental AHDF cells. The results confirmed the identity of the IITGi001-B cells. Further, the genomic integrity of the line was verified using G-banded karyotyping, which revealed a normal karyotype of the generated IITGi001-B cells without any chromosomal abnormalities. The line was also confirmed to be free of mycoplasma contamination. To facilitate broader research utility

and transparency, the IITGi001-B cell line has been officially registered and indexed in the Human Pluripotent Stem Cell Registry (hPSC reg), a cell line repository/bank (https://hpscreg.eu/user/cellline/edit/IITGi001-B). Collectively, all these results demonstrated that the generated IITGi001-B cells are true, *bonafide* iPSCs with true pluripotent ability and normal karyotype. This iPSC line can be used in a wide variety of biomedical and therapeutic applications.

3.5 Conclusion

Adult human dermal fibroblasts were reprogrammed to generate human iPSCs by transfecting Y4 plasmids by electroporation. Following a 25-day protocol, 13 iPSC-like colonies were generated. One of those colonies, designated as IITGi001-B, was picked and characterized extensively. The established IITGi001-B iPSC colony exhibited characteristic morphological features of iPSCs, like circular, compact cells with a shiny boundary, similar to ESCs. The pluripotent identity was validated through robust expression of iPSC surface and nuclear markers, namely TRA-1-60, TRA-1-81, SSEA4, OCT4, SOX2, NANOG, ZSCAN4, and REX1, which was demonstrated using various assays and experiments. Furthermore, the pluripotent characteristics of the generated IITGi001-B iPSCs was authenticated using an *in vitro* trilineage differentiation assay, which confirmed the expression of lineage-specific markers after directed differentiation of cells towards all three germ layers. This demonstrated the pluripotent characteristics of the IITGi001-B cell line. In addition, the integration-free nature of the IITGi001-B iPSCs was confirmed by performing semi-quantitative PCR, confirming the absence of episomal plasmids used for reprogramming. Genomic stability was corroborated by karyotyping analysis, which showed the normal genotype of IITGi001-B cells. Moreover, STR analysis confirmed the identity of the IITGi001-B cells. Absence of mycoplasma was also verified in the IITGi001-B cells. All these results establish IITGi001-B cell line as a *bonafide*, integration-free cell line, which will provide a platform for various biomedical applications like disease modelling, drug screening, functional genomic studies, regenerative medicine and understanding human developmental biology.

Chapter 4

Knockout of human *ZSCAN4* gene using CRISPR-Cas9 toolbox

This chapter has been accepted for publication as follows:

- **Sundaravadivelu PK**, Thummer RP (**2026**) Design and validation of CRISPR-Cas9-mediated knockout of ZSCAN4 gene in mammalian cells, In Recent Developments in Biotechnology, Proceedings of Research and Industrial Conclave - Synergy 2025, Springer Nature, Singapore (**Accepted**)

Overview

The CRISPR-Cas9 is a powerful and an efficient genome-editing tool to precisely edit or modify specific target regions in the genome of a wide range of organisms. Among all the genome editing methodologies available, CRISPR-Cas9 system has proven to be the most efficient with minimal off-target effects. Several factors can significantly influence its editing efficiency. In particular, the eukaryotic chromatin state of the target locus, guide RNA (gRNA) potency, and clonal selection strategies play a critical role in determining the success of targeted gene disruption. This study aims to assess the effect of these factors and optimize an efficient CRISPR-Cas9 system to knockout of the target gene of interest, *ZSCAN4*. To achieve this, we have used lentiviral-mediated transduction of Cas9 and dual gRNA construct to knockout *ZSCAN4* in human cells. Firstly, the effect of the transduction ratio of lentiviruses on the efficiency of CRISPR-Cas9 to knockout *ZSCAN4* gene was studied. The results show that increasing the transduction ratio did not have any effect on the deletion efficiency. Further, the influence of eukaryotic chromatin state on the efficiency of CRISPR-Cas9 toolbox was studied by using cells having different chromatin state. Moreover, we used a global DNA demethylating agent (5-Aza-dC) known to relax chromatin structure, and assessed its effect on the capability of the CRISPR-Cas9 toolbox to knockout *ZSCAN4* gene. This treatment significantly improved gene editing efficiency, highlighting the importance of chromatin accessibility in facilitating Cas9-mediated cleavage. Finally, we demonstrated that single cell expansion is essential to obtain homozygous clonal populations, ensuring biallelic and uniform gene knockout. Together, our findings highlight key factors that influence CRISPR-Cas9-mediated gene editing and provide an optimized framework to improve knockout efficiency in mammalian systems.

4.1 Introduction

CRISPR-Cas9 is a powerful and highly precise tool to edit or modify specific target regions in the genome of any organism with high precision (Doudna and Charpentier, 2014). Numerous studies have been published to date demonstrating the gene editing capability of CRISPR-Cas9 toolbox. It is widely regarded as the most efficient genome editing technology currently available that exhibits relatively low off-target effects compared to other genome editing methodologies (Gaj et al., 2013; Jinek et al., 2012; Khan, 2019). However, the efficacy of the CRISPR-Cas9 system is significantly influenced by several biological factors. These factors include viral transduction ratio used for the delivery of gRNA and Cas9, eukaryotic chromatin state (euchromatin or heterochromatin) of the target locus, potency of the gRNA, and clonal selection. These factors collectively play a very crucial role in determining the efficiency of gene editing. In this chapter, our aim is to attain homozygous, uniform knockout of our gene of interest in the cells by systematically assessing the influence of these factors on the knockout efficiency of CRISPR-Cas9.

We have chosen *ZSCAN4* as our target gene to knockout in human cells. ZSCAN4 is a transcription factor that has a crucial role in early embryonic development, generation, and maintenance of induced pluripotent stem cells (iPSCs), and cancer stem cells. ZSCAN4 expression is reported to be restricted to the two-cell stage during mouse embryonic development (Falco et al., 2007). In embryonic stem cells (ESCs), ZSCAN4 exhibits a heterogeneous expression pattern, where it is expressed only in 5% of cells at any given time, highlighting its tight regulation in expression (Zalzman et al., 2010). Its inclusion during mouse iPSC generation enhances efficiency and genomic stability, and it also promotes cancer stem cell phenotypes (Hirata et al., 2012; Jiang et al., 2013; Portney et al., 2020). Various other roles of ZSCAN4 in different cell types are concisely explained elsewhere (Thool et al., 2022).

Hence, we aim to generate homozygous human *ZSCAN4* knockout cells using CRISPR-Cas9 toolbox. This will provide deeper insight into the role of ZSCAN4 in the reprogramming paradigm and its contribution to the maintenance of self-renewal and pluripotency in iPSCs. Therefore, this chapter investigates the influence of the aforementioned factors on the efficiency of CRISPR-Cas9-based knockout of human *ZSCAN4* gene, and thereby establish an optimized protocol to generate homozygous

ZSCAN4 knockout in mammalian cells. Subsequently, these *ZSCAN4*$^{-/-}$ cells will be utilized for further studies to understand various biological roles of ZSCAN4 in human cells.

4.2 Materials and Methods

4.2.1 Cell culture

In this study, HEK293T cells, MDA-MB-231 cells, and adult human primary dermal fibroblasts (AHDF; ATCC Cat. No.: PCS-201-012; kind gift from Prof. RV Shaji) were used. These cells were cultured in DMEM (GIBCO) supplemented with 10% fetal bovine serum (FBS; GIBCO) and 1% antibiotic-antimycotic solution (GIBCO) at 37 °C with 5% CO_2 in a humidified atmosphere. The cells were passaged at 80 to 90% confluency using 0.05% Trypsin-EDTA (GIBCO). The iPSCs used in this study, AAVS1-PDi-Cas9 (a kind gift from Prof. RV Shaji), have a doxycycline-inducible Cas9 cassette integrated at the AAVS1 site, which is explained elsewhere (Thamodaran et al., 2021). The AAVS1-PDi-Cas9 (TJC8iCas9; hereafter referred as TJC8) iPSCs were cultured on matrigel-coated plates in StemMACS™ iPS-Brew media (Miltenyi Biotech) supplemented with 0.5% of antibiotic-antimycotic solution (GIBCO). The iPSCs were passaged when they reached 70-80% confluency using TrypLE express (GIBCO). The cells were reseeded on matrigel-coated plates in the presence of RevitaCell supplement (GIBCO).

4.2.2 Designing and cloning of guide RNAs

The puromycin resistance gene was synthesized along with *Bam*HI restriction site at the 5' end and *Bsr*GI restriction site at the 3' end. The synthesized construct was cloned into the pKLV-2.2-dgRNA-BB vector. Restriction digestion analysis was performed to confirm the cloning. The gRNAs against *ZSCAN4* gene were designed using various online tools, namely CHOP-CHOP, Synthego, Integrated DNA technologies, and E-CRISP (Table S1). The gRNA cassettes with *Kfl*I restriction site at 5' end and *Bam*HI restriction site at 3' end was synthesized. Consequently, the gRNAs were cloned into the pKLV2.2 vector using the mentioned restriction enzymes. Restriction digestion analysis and Sanger sequencing was performed to confirm the cloning of the gRNAs. The lentiCas9-Blast plasmid (Addgene ID: 52962) was purchased from Addgene and used in our experiments for the expression of Cas9.

4.2.3 Lentivirus generation

Both pKLV2.2-dgRNA and lentiCas9-Blast lentiviruses were generated as previously described (Haridhasapavalan et al., 2022). Briefly, HEK293T cells were seeded at a density of $5x10^6$ cells in a 10 cm dish. The cells were replenished with fresh transfection media (2% FBS, 1% antibiotic-antimycotic in advanced DMEM; GIBCO) 2 h before transfection. Subsequently, the cells were transfected with 2nd generation packaging plasmid (psPAX2; Addgene ID: 12260), envelope plasmid (pMD2.G; Addgene ID: 12259), along with transfer plasmid (either pKLV2.2-dgRNA or lentiCas9-Blast plasmid). Post-transfection after 6 h and 18 h, the cells were replenished with fresh production media (5% FBS, 1% antibiotic-antimycotic in advanced DMEM; GIBCO). The generated viruses were collected at three time points, 42 h, 54 h, and 66 h. The collected viral supernatant was concentrated using a virus concentrator solution (40% PEG and 1.2 M sodium chloride in 1X PBS). The viral supernatant was mixed with the concentrator solution at a ratio of 1:3, incubated at 4 °C under rocking conditions for 4 h, and then centrifuged at 4 °C for 1 h at 1600xg. The viral pellet thus obtained was resuspended in DMEM without serum, aliquoted, and stored at -80 °C until further use.

4.2.4 Virus transduction

The cells were seeded at a density of $5x10^4$ cells per well of a 12-well plate for iPSCs and $8x10^4$ cells per well of a 12-well plate for AHDF cells. In case of iPSCs, cells were dissociated using TrypLE express (GIBCO) and the respective number of cells were seeded in the presence of RevitaCell supplement (GIBCO). After 24 h, the concentrated lentiviruses were added to cells at the respective transduction ratio in the presence of 5 µg/ml of polybrene. Subsequently, 24 h post-transduction, media was changed. Selection with the respective antibiotic or microscopic imaging was done using an inverted fluorescence microscope 48 h post-transduction. After selection with the respective antibiotic, the cells were expanded, and subsequently, genomic DNA or RNA was isolated for further analysis.

4.2.5 Real-Time quantitative PCR (RT-qPCR)

The RT-qPCR was performed as previously described in section 3.2.3. Briefly, RNA was isolated and cDNA was synthesized using cDNA synthesis kit (Bio-Rad) following the manufacturer's instructions. RT-qPCR was performed using 2X SYBR green supermix (Bio-Rad) in AriaMx quantitative PCR systems (Agilent). The results thus

obtained were visualized and analyzed using AriaMx software (Agilent). The relative expression levels were calculated using the standard $2^{-\Delta\Delta ct}$ method, and the graphs and the statistical analyses were done using GraphPad Prism 8 software. The results were plotted as mean±SEM and were statistically analyzed using student's t-test with 95% confidence interval.

4.2.6 Genomic DNA isolation

The genomic DNA was isolated using the protocol previously described in section 3.2.7. Briefly, one volume of PCI was added to the cells, the mixture was centrifuged at 16,000xg for 5 minutes, and the upper aqueous phase containing the genomic DNA was collected. To the aqueous phase, glycogen, ammonium acetate, ethanol, RNAse, and Proteinase K were added and incubated at -20 °C overnight to precipitate the genomic DNA. Subsequently, the precipitated genomic DNA was pelleted by centrifugation at 16,000xg for 30 minutes at 4 °C. Then, the pellet was washed with 70% ethanol, air dried, and resuspended in TE buffer. The isolated genomic DNA was stored at -20 °C until further use.

4.2.7 Semi-quantitative PCR

Semi-quantitative PCR was performed using the isolated genomic DNA as per the protocol previously described in section 3.2.8. Briefly, 100 ng of genomic DNA was used per reaction of PCR and nuclease-free water was used as a no-template control. The PCR was performed using specific primers (Table S2) with Taq DNA polymerase (New England Biolabs) in a Mastercycler Nexus X2 Gradient Thermal Cycler (Eppendorf). The PCR conditions were as follows: 95 °C for 5 min, 35 cycles of 95 °C for 30s, respective annealing temperature for 30s, 72 °C for 30s, and final extension at 72 °C for 10 min. The PCR products were resolved using 2% agarose gel and imaged using ChemiDoc™ XRS+ (Bio-Rad). The image was analyzed and processed using Image Lab™ software (Bio-Rad, California, USA).

4.2.8 Clonal expansion

To isolate a homozygous clonal population, clonal expansion was performed using two different methods, namely array dilution and single cell seeding. The heterogeneous population of knockout iPSCs was dissociated using TrypLE express and counted. Single cell expansion was performed on matrigel-coated 96-well plates, and the StemMACS™ iPS-Brew was supplemented with CloneR-2 supplement (STEMCELL Technologies) and

RevitaCell supplement (GIBCO). In case of array dilution method, 100 µL of StemMACS™ iPS-Brew supplemented with CloneR-2 supplement supplement (STEMCELL Technologies) was added to all the wells of a matrigel-coated 96-well plate. The first well was seeded with 2000 cells. Then the cells were serially diluted along the first column at a ratio of 1:1. Subsequently, the cells in first column were diluted serially along all the columns (column wise) using a multi-channel pipette at a ratio of 1:1. In case of single cell seeding method, the cells were diluted at a density of 80 cells per 10 mL of media, and 100 µL of cells were seeded in each well of the matrigel-coated 96-well plate. After 48 h post-seeding, media was changed to StemMACS™ iPS-Brew supplemented with CloneR-2 supplement. Subsequently, the media was changed every 48 h till single colonies were observed. These single clones were further expanded on a matrigel-coated 48-well plate and then on a matrigel-coated 24-well plate. From the 24-well plate, crude lysates from all the expanded single cell clones were collected using the heat shock method, and PCR was performed to identify the homozygous *ZSCAN4* knockout clones.

4.3 Results

4.3.1 Cloning of puromycin resistance gene in the desired vector

The plasmid pKLV2.2-h7SKgRNA5-hU6gRNA5-PGK-EGFP-W [(Raina et al., 2025); a kind gift from Prof. RV Shaji (CSCR-CMCV)] was used in our experiments. This vector has EGFP as a fluorescence selection marker. Further, to use puromycin as an antibiotic selection marker, the vector, pKLV2.2-h7SKgRNA5-hU6gRNA5-PGK-EGFP-W, was modified to express puromycin resistance gene. The coding sequence of the puromycin resistance gene was synthesized and cloned using *Bam*HI and *Bsr*GI restriction

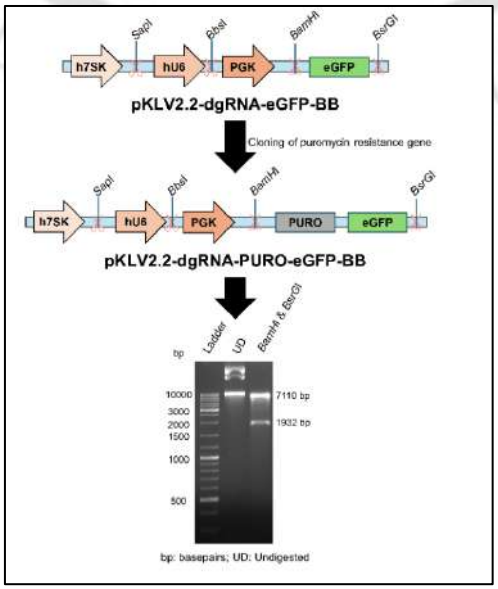


Fig. 4.1. Schematic and confirmation of cloning of puromycin resistance gene. Schematic depicting cloning of the puromycin resistance gene in pKLV2.2-dgRNA vector (*top*) and its confirmation using restriction digestion (*bottom*). bp: basepairs; UD: undigested.

enzymes, right before the coding sequence of EGFP, yielding the final plasmid pKLV2.2-h7SKgRNA5-hU6gRNA5-PGK-Puro-EGFP-W (hereafter referred as pKLV2.2-dgRNA-BB). The cloning was confirmed with restriction digestion, which showed the successful cloning of puromycin resistance gene in the pKLV2.2 vector (Fig. 4.1). Further, the integrity of the cloned gene was verified using DNA sequencing (data not shown).

4.3.2 Designing and cloning of dual gRNAs in the desired vector

In this study, we have utilized dual gRNA (dgRNA) based strategy, where two distinct gRNAs will target the same region in the genomic DNA approximately 50 to 100 bp apart (Fig 4.2). The Cas9 nuclease will induce double-strand breaks (DSB) at both of these sites, resulting in excision of a fragment of DNA from the target region (between two gRNA binding sites), rendering the target gene non-functional (Fig. 4.2). Subsequently, the knockout can be confirmed by performing PCR using primers flanking the target site (Fig. 4.2). The amplicon sizes before and after deletion are listed in Table 4.1. For the dgRNA setup, two gRNAs that bind to the target region within 100 bp were chosen. The dgRNAs pairs were chosen in such a way that they target either exon 3 or exon 5 of human *ZSCAN4* gene. This is due to the fact that two important domains of ZSCAN4 protein, the SCAN domain and zinc finger domain, are coded by exon 3 and exon 5, respectively. These gRNAs were designed using CHOP-CHOP online tool.

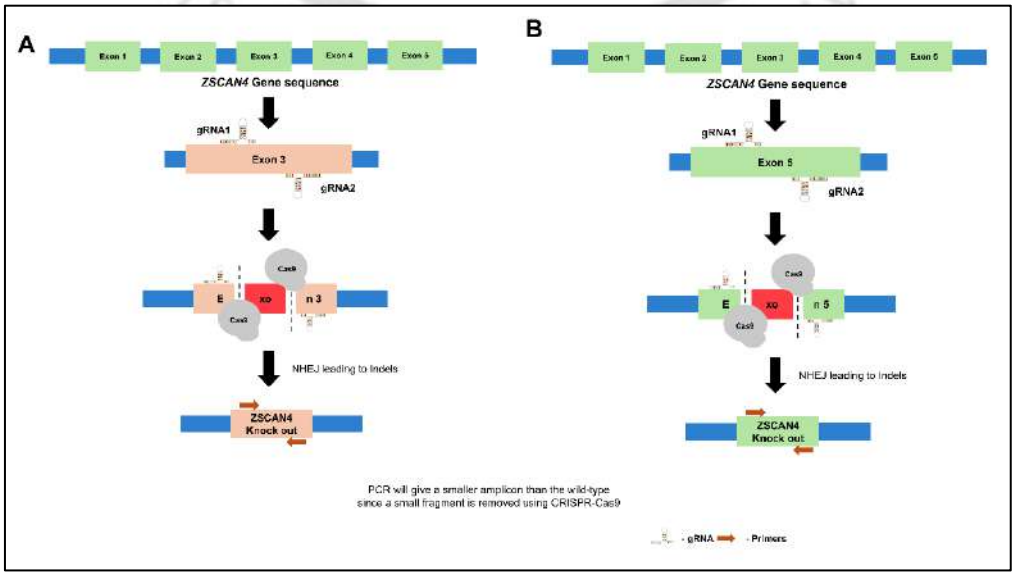


Fig. 4.2. Schematic representation of *ZSCAN4* knockout strategy. Schematic representation detailing *ZSCAN4* knockout strategy using dgRNAs against exon 3 **(A)** and exon 5 **(B)**.

Table 4.1. Expected band size for the amplicon for the wild-type and knockout cells.

Target exon	Wild-Type (WT) band size	Deletion size	Expected knockout (KO) band size
Exon 3	211 bp	62 bp	148 bp
Exon 5	286 bp	60 bp	226 bp

Subsequently, the designed gRNAs were synthesized and cloned into pKLV2.2-dgRNA-BB vector in a sequential manner, i.e., first gRNA1 was cloned initially using *Sap*I restriction enzyme, and cloning was verified using sequencing (Fig. 4.3). Subsequently, gRNA2 was cloned using *Bbs*I restriction enzyme, which was also further confirmed using Sanger sequencing (Fig. 4.3). This sequential cloning was performed for both the dgRNAs targeting exon 3 and exon 5 of human *ZSCAN4* gene. Finally, after the cloning, two different dgRNA plasmids were generated, pKLV2.2-dgRNA-Ex3 (dgRNA targeting exon 3 of *ZSCAN4* gene) and pKLV2.2-dgRNA-Ex5 (dgRNA targeting exon 5 of *ZSCAN4* gene).

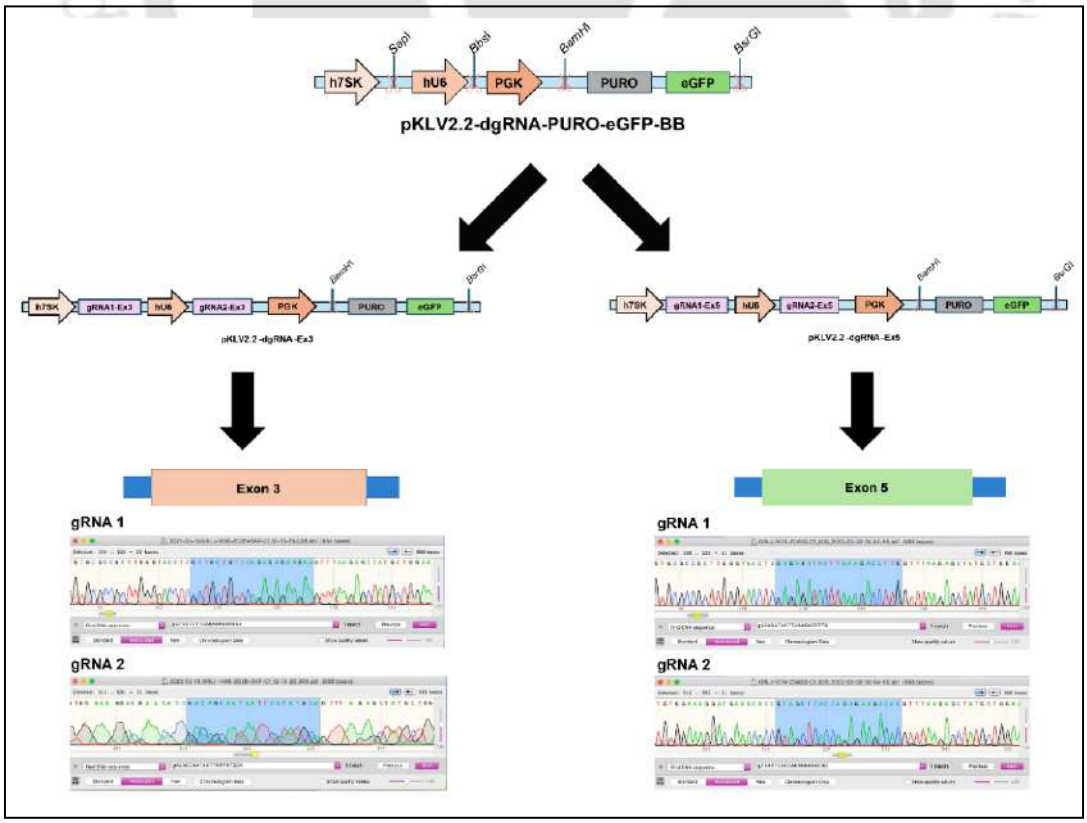


Fig. 4.3. Schematic and confirmation of cloning of dgRNAs. Schematic depicting cloning of dgRNAs against exon 3 (*left*) and exon 5 (*right*) in pKLV2.2-dgRNA vector and its confirmation using Sanger sequencing.

4.3.3 Functional validation of the generated lentiviruses

The lentiviruses were generated using HEK293T cells as described previously (Haridhasapavalan et al., 2022). The overall workflow of lentivirus generation is depicted in Fig. 4.4A. To validate the functionality of the dgRNA and Cas9 lentiviruses, the HEK293T and AHDF cells were transduced with either the dgRNA virus or the Cas9 virus. Consequently, expression of EGFP was analyzed to validate the functionality of pKLV2.2-dgRNA-BB virus. The fluorescence microscopy images of the cells transduced with pKLV2.2-dgRNA-BB virus showed the expression of EGFP (in case of both HEK293T cells and AHDF cells), confirming the successful transduction and expression of transgenes in the cells (Fig. 4.4B and 4.4C; *left*). Parallelly, RT-qPCR was performed using primers specific to the Cas9 transgene to validate the functionality of lentiCas9-Blast virus. The RT-qPCR results showed almost ~600-fold and ~1500-fold relative expression of Cas9 in lentiCas9-Blast transduced cells compared to non-transduced HEK293T and AHDF cells, respectively (Fig. 4.4B and 4.4C; *right*). These results proved that both pKLV2.2-dgRNA and lentiCas9-Blast lentiviruses generated have integrated into the target cells and are expressing the transgene.

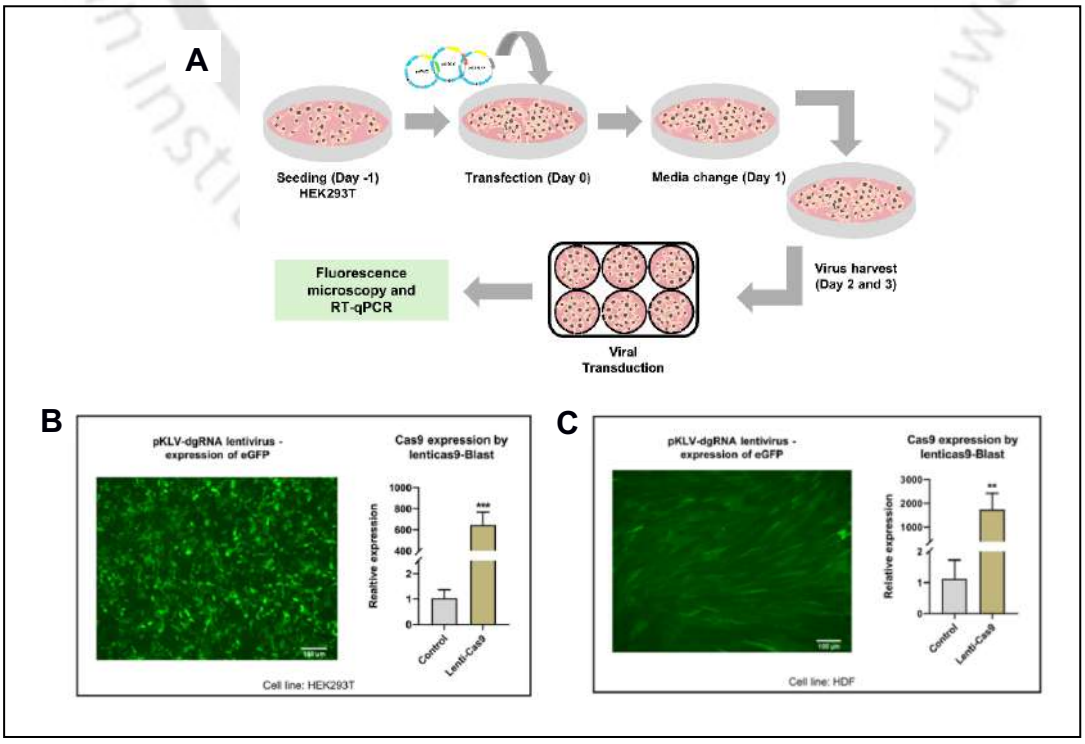


Fig. 4.4. Lentivirus generation, transduction, and functional validation. A) Workflow detailing lentivirus generation protocol using HEK293T cells and its subsequent transduction in HEK293T cells to verify the functionality of the generated viruses. Functional validation of pKLV2.2-dgRNA virus and lentiCas9-Blast virus by observing the expression of EGFP using an inverted fluorescence microscope and by performing RT-qPCR, respectively, in HEK293T cells (**B**) and AHDF cells (**C**). The expression levels were normalized using GAPDH, and the results were plotted as mean±SEM. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. Scale bar: 100 µm.

4.3.4 Optimization of transduction ratio of the lentivirus in AHDF cells

To determine the optimal transduction ratio of the lentiviruses, AHDF cells were transduced with pKLV2.2-dgRNA-BB in different ratios (virus to media ratio) ranging from 1:5 to 1:100. Subsequently, the expression of EGFP was observed and analyzed to finalize the optimal transduction ratio for the cells (Fig. 4.5). The optimal transduction ratio was chosen in such a way that around 60% to 70% of the cells express EGFP without compromising the health of the cells. As expected, the microscopy images showed an inverse relation between transduction ratio and EGFP expression. There is a gradual decrease in expression of EGFP as the transduction ratio increases (Fig. 4.5).

4.3.5 Influence of transduction ratio on the knockout efficiency

Firstly, AHDF cells were transduced with lentiCas9-Blast viruses at a transduction ratio of 1:10. The transduced cells were selected using Blasticidin, and subsequently, Cas9-expressing cells were transduced with either pKLV2.2-dgRNA-Ex3 or pKLV2.2-dgRNA-Ex5 viruses. The pKLV2.2-dgRNA-BB virus was used as a negative control in the experiment. Then, cells were selected using puromycin, genomic DNA was isolated, and the knockout efficiency was analyzed by PCR using locus-specific primers (Table S2). In case of knockout using dgRNA against exon 3, the expected band size for the wild-type cells is 211 bp. This set of dgRNA leads to a deletion of 62 bp, leading to a knockout band of 148 bp (Table 4.1). In case of knockout using dgRNA against exon 5, the expected band size for wild-type cells is 286 bp. The corresponding deletion is 60 bp, leading to a knockout band of 226 bp (Table 4.1). On performing PCR for cells transduced with dgRNA against exon 3 of *ZSCAN4*, two bands were observed in the agarose gel, corresponding to both wild-type (211 bp) and knockout (148 bp) (Fig. 4.6A). Analyzing the intensity of the bands using ImageJ revealed that the wild-type band contributed to ~84% and the knockout band contributed to ~16% (Fig. 4.6A). Similarly, cells transduced with dgRNA against exon 5

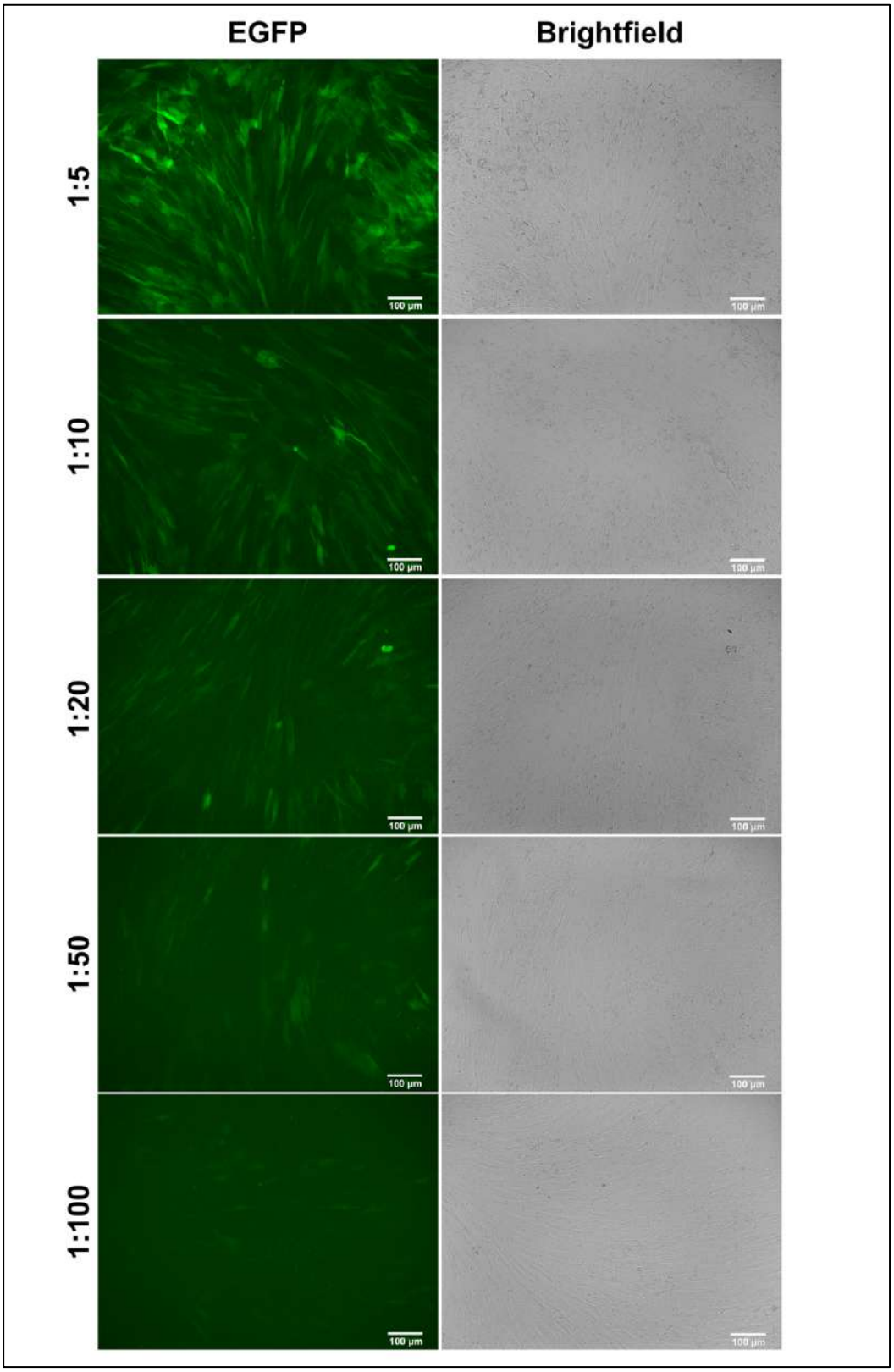


Fig. 4.5. Optimization of lentivirus transduction ratio. Screening of optimal transduction ratio for the lentiviruses by transducing different ratios of pKLV2.2-dgRNA-BB viruses in AHDF cells and observing the expression of EGFP using an inverted fluorescence microscope. Scale bar: 100 µm.

of *ZSCAN4* gene also led to partial knockout, leading to the formation of two different-sized amplicons on performing PCR, at 286 bp corresponding to wild-type and at 226 bp corresponding to knockout (Fig. 4.6A). Quantifying these bands using ImageJ revealed that the wild-type population contributed to ~55.5% and the knockout population contributed to ~44.5% (Fig. 4.6A). These results show that knockout efficiency is ~16% for dgRNA against exon 3 and ~44.5% for dgRNA against exon 5. Both the dgRNA against *ZSCAN4* gene led to partial knockout in cells. To enhance the knockout efficiency, this population of cells was subjected to another round of transduction with both lentiCas9-Blast and pKLV-dgRNA viruses. After the second round of transduction with the viruses, an increased knockout efficiency was observed, but it was not sufficient to generate a complete knockout of ZSCAN4 in cells. Quantifying the bands using ImageJ showed that in case of dgRNA against exon 3, the knockout efficiency increased from ~16% (single transduction) to ~35% (double transduction) (Fig. 4.6B). Similarly, in case of dgRNA against exon 5, the knockout efficiency increased from ~44.5% (single transduction) to ~60.5% (double transduction) (Fig. 4.6B). Even though, the second round of transduction with Cas9 and dgRNA viruses resulted in enhanced knockout efficiency, it was unable to generate complete *ZSCAN4* knockout cells.

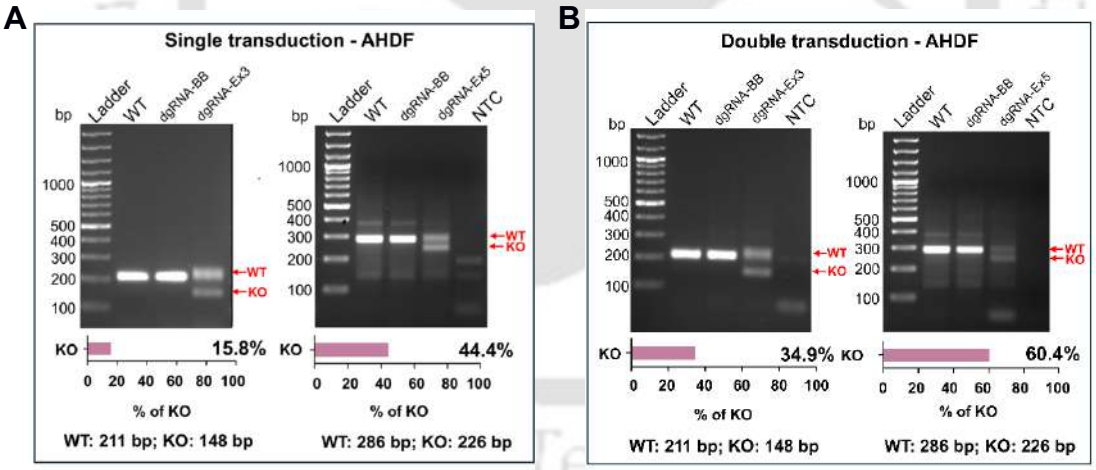


Fig. 4.6. Analysis of knockout of *ZSCAN4* gene in AHDF cells using single and double transduction. Analysis of knockout of *ZSCAN4* using dgRNA against exon 3 and exon 5 by performing semi-quantitative PCR in AHDF cells after single transduction **(A)** and in AHDF cells after double transduction **(B)**. bp: basepairs; WT: Wild-type; dgRNA-BB: dual gRNA vector backbone only; dgRNA-Ex3: dual gRNA against exon3 of ZSCAN4; dgRNA-Ex5: dual gRNA against exon5 of ZSCAN4; NTC: no-template control.

4.3.6 Influence of chromatin state on the knockout efficiency

To relax the chromatin in fibroblasts, we incorporated the use of a global DNA demethylating agent, 5-Aza-2'-deoxycytidine (5-Aza-dC), which can globally alter the chromatin. The cells were treated with 5-Aza-dC for 24 h prior to viral transduction, and the treatment was continued for the whole time period of the experiment till genomic DNA isolation. Performing PCR with the genomic DNA isolated from the cells that underwent 5-Aza-dC treatment revealed partial knockout of *ZSCAN4* gene in the fibroblasts with a modest enhancement in knockout efficiency compared to previous cases. In case of knockout using dgRNA against exon 3, the knockout efficiency is ~54% (Fig. 4.7A), which is higher than the efficiency using double transduction of viruses. In case of dgRNA against exon 5, quantitative analysis revealed a knockout efficiency of ~57% (Fig. 4.7A), which is comparable to the knockout efficiency using double transduction of viruses. Hence, the usage of a DNA demethylating agent, 5-Aza-dC, assisted in enhanced knockout efficiency, but still was not able to achieve complete knockout of *ZSCAN4* gene.

To further analyze the influence of chromatin state on the efficiency of CRISPR-Cas9 based knockout, we utilized cells having an euchromatin (open chromatin) structure to knockout *ZSCAN4* using our CRISPR-Cas9 toolbox. Knocking out *ZSCAN4* gene using the dgRNA toolbox in MDA-MB-231 cells resulted in a partial knockout, similar to

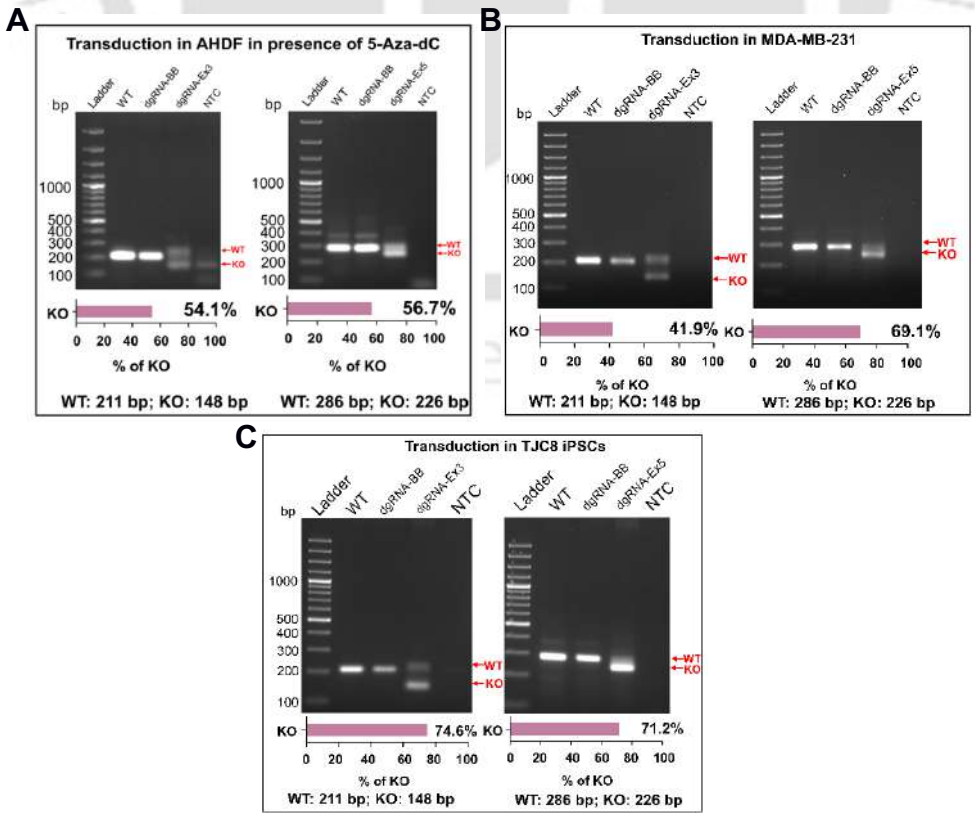


Fig. 4.7. Analysis of knockout of *ZSCAN4* gene in cells having different chromatin state. Analysis of knockout of *ZSCAN4* gene using dgRNA against exon 3 and exon 5 by performing semi-quantitative PCR, in AHDF cells after single transduction in AHDF cells after treatment with 5-Aza-dC **(A)**, in MDA-MB-231 cells **(B)**, and in iPSCs **(C)**. bp: basepairs; WT: Wild-type; dgRNA-BB: dual gRNA vector backbone only; dgRNA-Ex3: dual gRNA against exon3 of ZSCAN4; dgRNA-Ex5: dual gRNA against exon5 of ZSCAN4; NTC: no-template control.

previous cases. Further quantification revealed that dgRNA against exon 3 resulted in ~42% (Fig. 4.7B) knockout efficiency, and dgRNA against exon 5 resulted in ~69% (Fig. 4.7B), which is higher than the previous cases. Moreover, knockout in TJC8 iPSCs also resulted in partial knockout of *ZSCAN4*. Interestingly, the knockout efficiency was the highest in case of TJC8 iPSCs, yielding ~75% and ~71% (Fig. 4.7C) using dgRNA against exon 3 and exon 5, respectively. These results prove that chromatin state exerts a significant influence on the knockout efficiency of CRISPR-Cas9 toolbox.

4.3.7 Influence of gRNA potency on the knockout efficiency

Next, we evaluated the knockout efficiency of gRNAs with varying potencies. To achieve this, gRNAs were designed using four different computational online tools (Table 4.2), and the top ranked gRNAs were chosen for knocking out human *ZSCAN4* gene in AHDF cells. Notably, each tool yielded a unique set of candidate gRNAs for maximal knockout efficiency (Table 4.2). These dgRNAs were synthesized and cloned. The cloning was validated using restriction digestion analysis (Fig. 4.8) and further the fidelity of the sequences was verified using Sanger sequencing (data not shown).

Table 4.2. Expected band size for the amplicon for the wild-type and knockout cells and the tools used for designing the gRNAs.

dgRNA name	Wild-Type (WT) band size	Deletion size	Expected knockout (KO) band size	Online tool used for designing
dgRNA-set 1	281 bp	75 bp	206 bp	Synthego
dgRNA-set 2	484 bp	114 bp	370 bp	Integrated DNA technologies
dgRNA-set 3	309 bp	72 bp	237 bp	E-CRISP
dgRNA-set 4	309 bp	59 bp	250 bp	CHOP-CHOP
dgRNA-set 5	484 bp	47 bp	437 bp	CHOP-CHOP

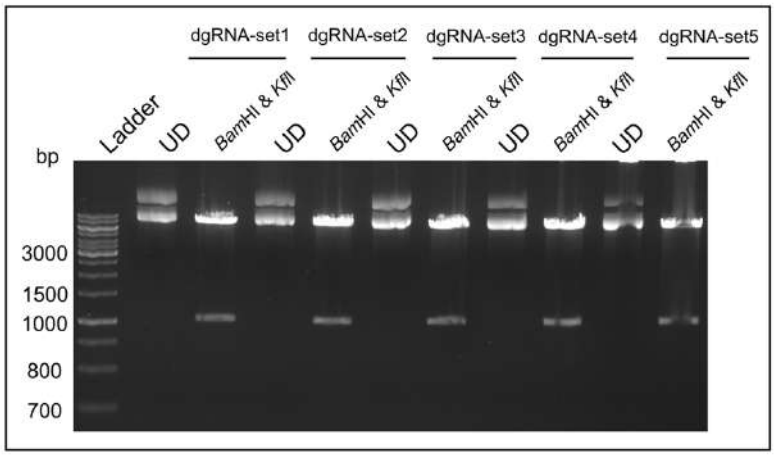
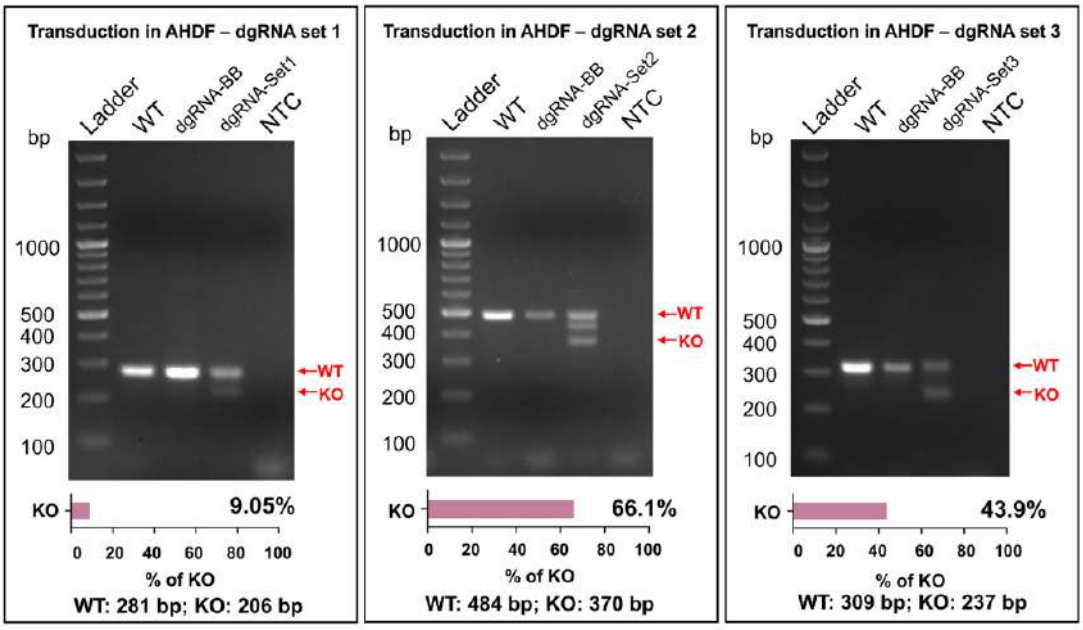


Fig 4.8. Confirmation of cloning of new sets of dgRNA cassettes. The newly synthesized dgRNA cassettes were cloned in pKLV2.2-dgRNA vector and the cloning was confirmed using restriction digestion. bp: basepairs; UD: undigested.

Further, the knockout of *ZSCAN4* was carried out in AHDF cells using these new sets of dgRNAs. The knockout efficiency was assessed via semi-quantitative PCR and quantifying the intensity of the amplicons using ImageJ software. The expected amplicon lengths for the wild-type and knockout are listed in Table 4.2. The results revealed all the sets of dgRNAs incorporated except for set 4, resulted in partial knockout of *ZSCAN4* gene with varying efficiencies (Fig. 4.9). In case of dgRNA set 1, two different sized amplicons were observed, one band corresponding to wild-type and other one corresponding to knockout (Fig. 4.9). Quantifying the band intensities using ImageJ revealed a knockout



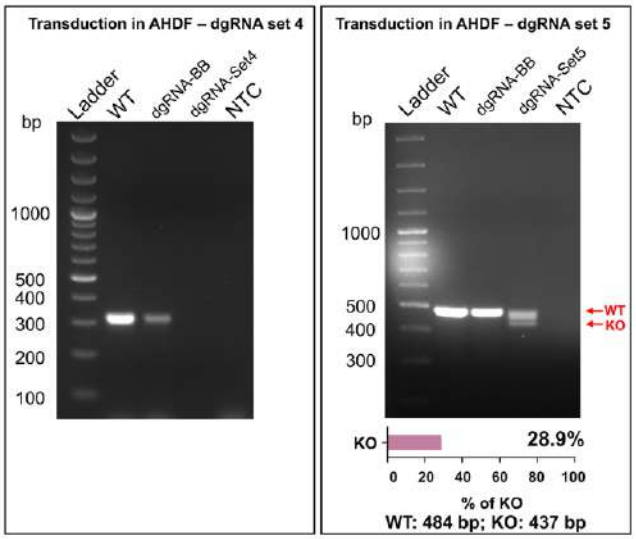


Fig. 4.9. Analysis of knockout of *ZSCAN4* gene using new sets of dgRNAs in AHDF cells. The AHDF cells were transduced with new sets of dgRNA lentivirus and Cas9 lentivirus, followed by respective antibiotic selection. The knockout was verified using semi-quantitative PCR using genomic DNA isolated from these cells. The band intensity was quantified using ImageJ software. bp: basepairs; WT: Wild-type; dgRNA-BB: dual gRNA vector backbone only; dgRNA-Set 1: 1st set of dual gRNA against *ZSCAN4*; dgRNA-Set 2: 2nd set of dual gRNA against *ZSCAN4*; dgRNA-Set 3: 3rd set of dual gRNA against *ZSCAN4*; dgRNA-Set 4: 4th set of dual gRNA against *ZSCAN4*; dgRNA-Set 5: 5th set of dual gRNA against *ZSCAN4*; NTC: no-template control.

efficiency of ~9%, which is much lesser than the previous dgRNAs that have been used. The knockout using dgRNA set 2, yielded a knockout efficiency of ~66% (Fig. 4.9), which is comparable to the efficiency that was obtained using dgRNA against exon 5. Interestingly, in case of dgRNA set 2, three distinct amplicons were observed. One band corresponding to wild-type, second band corresponding to knockout and the third band in between these amplicons (Fig. 4.9). The knockout efficiency of ~44% was observed using dgRNA set 3 in AHDF cells (Fig. 4.9). Similar to dgRNA set 1, in case of dgRNA set 3, two different sized amplicons were observed corresponding to wild-type and knockout (Fig. 4.9). In case of dgRNA set 4, most of the cells died after selection with respective antibiotics and therefore genomic DNA isolated from these cells was not sufficient to perform a PCR. So, there was no band observed in the PCR reaction using dgRNA set 4 (Fig. 4.9). Lastly, using dgRNA set 5, a knockout efficiency of ~29% was observed in AHDF cells (Fig. 4.9). Similar to other dgRNAs, two different sized amplicons were observed corresponding to wild-type and knockout (Fig. 4.9). All these results

demonstrated that gRNA potency plays a vital role in knockout efficiency. All the dgRNAs gave varying knockout efficiencies against the same target gene, *ZSCAN4* and in the same cell line, AHDF. Given that the dgRNA targeting exon 5 demonstrated one of the highest knockout efficiencies in both AHDF cells and iPSCs and targets a functionally important domain of ZSCAN4, it was selected for further experiments.

4.3.8 Clonal expansion led to homogeneous knockout population of iPSCs

Following transduction of iPSCs with dgRNA targeting exon 5 and induction of Cas9 expression using doxycycline, clonal expansion was performed. A total of nine clones were isolated, expanded, and subjected to genotypic analysis.

Among these, clones C2, C4, C7, and C9 exhibited a single PCR amplicon of ~226 bp, consistent with a ~60 bp deletion at the target site (Fig. 4.10; *top*). Sanger sequencing confirmed the deletion within exon 5 of *ZSCAN4* gene in all four clones (Fig. 4.10; *bottom*). The presence of a single deletion-specific band and absence of a wild-type amplicon indicate biallelic editing, consistent with a homozygous knockout in these clones.

Clone C1 exhibited two distinct PCR amplicons (~286 bp and ~226 bp; Fig. 4.10; *top*), indicating the presence of both a non-deleted and a deletion allele. Sequencing revealed that the wild-type-sized amplicon harbors three nucleotide substitutions resulting in two amino acid changes (Fig. 4.10; *bottom*). These findings indicate heterozygous editing, with one allele containing a deletion and the other harboring missense mutations, consistent with compound heterozygosity in this clone.

Clone C3 displayed a single PCR amplicon of ~250 bp, which was intermediate between the wild-type (~286 bp) and expected knockout (~226 bp) bands (Fig. 4.10; *top*). Sequencing analysis revealed indels at the target locus within exon 5 (Fig. 4.10; *bottom*), suggesting CRISPR-Cas9-mediated editing. The absence of a wild-type band and the presence of a single amplicon indicate biallelic editing.

Clone C5 also displayed two distinct PCR amplicons (~286 bp and >300 bp; Fig. 4.10; *top*), indicating the presence of wild-type–sized amplicon and insertional events at the target locus consistent with CRISPR-Cas9-mediated insertion. Sequencing analysis revealed multiple point mutations in the wild-type-sized amplicon (Fig. 4.10; *bottom*). These results indicate heterozygous editing, with one allele harboring an insertion and the other containing point mutations.

Clone C6 exhibited two PCR amplicons of ~226 bp and ~300 bp (Fig. 4.10; *top*), corresponding to deletion and insertion events, respectively. Sequencing confirmed the presence of multiple indels in the ~300 bp amplicon size within exon 5 of *ZSCAN4* gene (Fig. 4.10; *bottom*). The absence of a wild-type amplicon indicates biallelic editing, with one allele harboring a deletion and the other containing insertional mutations, likely resulting in functional disruption of *ZSCAN4* gene.

Clone C8 showed two PCR amplicons, one corresponding to the expected knockout band (~226 bp) and a second smaller band (<226 bp), with no detectable wild-type–sized amplicon (Fig. 4.10; *top*). Sequencing revealed deletions accompanied by minor nucleotide changes at the target locus (Fig. 4.10; *bottom*). These findings indicate biallelic editing with distinct deletion events, consistent with compound heterozygous disruption of *ZSCAN4* gene.

As a control, wild-type TJC8 cells and cells transduced with dgRNA-BB exhibited a single PCR amplicon corresponding to the wild-type band (286 bp; Fig. 4.10; *top*). Sequencing confirmed 100% identity with the reference sequence (Fig. 4.10; *bottom*), indicating the absence of nucleotide changes. These samples served as unedited controls, validating the specificity and accuracy of CRISPR-Cas9-mediated editing observed in the experimental clones (Clones C1 to C9).

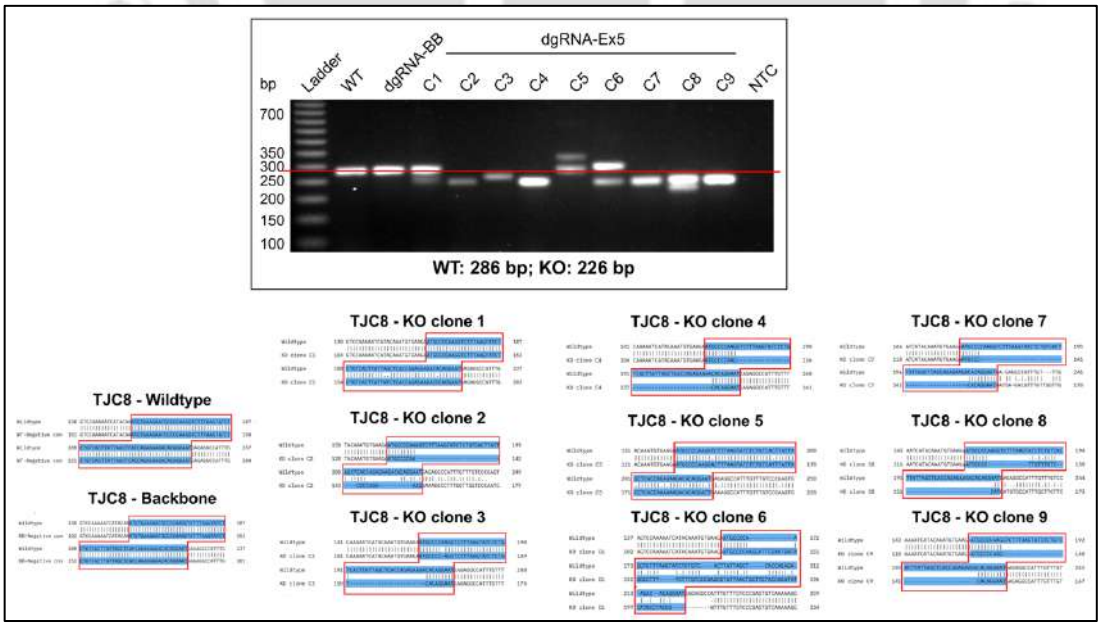


Fig. 4.10. Confirmation of *ZSCAN4* knockout in the isolated clones using semi-quantitative PCR and Sanger sequencing. The single cell clones isolated after *ZSCAN4* knockout were subjected to semi-quantitative to confirm the 60 bp deletion in exon 5 region (*top*) and further the amplicons were analyzed by Sanger sequencing (*bottom*). The red line

corresponds to wild-type amplicon size of 286 bp. bp: basepairs; WT: wild-type; dgRNA-BB: TJC8 cells transduced with dgRNA-BB; C1 to C9: *ZSCAN4* knockout TJC8 iPSC clones; NTC: no-template control; KO: knockout.

A summary of the genotype analysis of *ZSCAN4* knockout iPSC clones, based on semi-quantitative PCR and Sanger sequencing, is presented in Table 4.3.

Table 4.3. Summary of genotype analysis of *ZSCAN4* knockout iPSC clones derived from semi-quantitative PCR and Sanger sequencing.

Clone	PCR Bands	Genotype	Editing Type	Functional Interpretation
C1	~286 bp + ~226 bp	Heterozygous	Deletion + point mutations	Compound heterozygous mutant
C2	~226 bp	Biallelic	Deletion (both alleles)	Homozygous knockout
C3	~250 bp	Biallelic	Deletions (indel variant)	Homozygous knockout
C4	~226 bp	Biallelic	Deletion (both alleles)	Homozygous knockout
C5	~286 bp + >300 bp	Heterozygous	Insertion + point mutations	Compound heterozygous mutant
C6	~226 bp + ~300 bp	Biallelic	Deletion + insertion	Functional knockout (biallelic disruption)
C7	~226 bp	Biallelic	Deletion (both alleles)	Homozygous knockout
C8	~226 bp + <226 bp	Biallelic	Two deletions	Functional knockout (biallelic disruption)
C9	~226 bp	Biallelic	Deletion (both alleles)	Homozygous knockout
Control	286 bp	WT/WT	No editing	Wild-type

4.4 Discussion

The transcription factor ZSCAN4 has been reported to be vital for various cellular functions during embryonic development, as well as in ESCs, iPSCs, and cancer cells. However, the exact mechanistic role of ZSCAN4 in reprogramming of human somatic cells and in the maintenance of self-renewal and pluripotency property of iPSCs is still unclear. Therefore, understanding its role by complete deletion of *ZSCAN4* gene is of paramount importance, which will help us advance the therapeutic application of iPSCs. The CRISPR-Cas9 toolbox is a powerful and precise tool to edit target regions in the genome (Doudna and Charpentier, 2014). Certain biological hurdles need to be addressed to make the toolbox much more efficient. This chapter identified such hurdles and demonstrated its significant influence on the efficiency of CRISPR-Cas9 toolbox. This chapter employed lentivirus-based expression of dgRNA and Cas9 in the target cells. Lentivirus is proven to be one of the efficient tools to deliver the transgenes in the target cells with high efficiency (Denning et al., 2013; Durand and Cimarelli, 2011; Vigna and Naldini, 2000). Viral transduction ratio was identified as a critical factor to attain efficient delivery of the transgenes without compromising the health of the cells. Firstly, we demonstrated the influence of the transduction ratio on the knockout efficiency. Our results demonstrated a positive correlation between the transduction ratio and the knockout efficiency. Increasing the transduction load led to an increased amount of transgene expression (Denning et al., 2013; Zhang et al., 2004), which in turn helped in improving in knockout efficiency of CRISPR-Cas9 system. Although increasing the transduction ratio enhanced the knockout efficiency, it was unable to generate a homozygous knockout of *ZSCAN4* gene. The cells still displayed partial deletion of *ZSCAN4* gene.

To date, there are no studies reporting the expression of ZSCAN4 in human fibroblasts. It is hypothesized that the genomic region of *ZSCAN4* may exist in the heterochromatin region of the genome, making it inaccessible for the dgRNA and Cas9 nuclease. Previous reports suggest that eukaryotic chromatin state can significantly impede the activity and the efficiency of CRISPR-Cas9 (Verkuijl and Rots, 2019; Wu et al., 2014), which is due to the inaccessibility of the protospacer adjacent motif sequence (Hinz et al., 2015). Therefore, we speculated that the observed partial knockout of *ZSCAN4* gene in fibroblasts was a direct consequence of this inaccessibility. Similar results have been observed in one of our previous studies where the target gene, *UTF1*, not expressed in

human fibroblasts was deleted partially in fibroblasts using CRISPR-Cas9 toolbox (Raina et al., 2025). To test this hypothesis, we employed the use of 5-Aza-dC, a global DNA demethylating agent. This molecule is reported to relax the chromatin structure of cells, shifting them towards euchromatin state (Christman, 2002; Patra and Bettuzzi, 2009). Therefore, human fibroblast cells were treated with 5-Aza-dC before and during lentiviral transduction to enhance the accessibility of *ZSCAN4* locus by the dgRNA and Cas9. As anticipated, treatment with 5-Aza-dC resulted in higher knockout efficiency in both dgRNA against exon 3 and exon 5. However, even with chromatin modification, knockout of *ZSCAN4* gene was still partial as observed by the presence of two amplicons corresponding to wild-type and knockout sizes. This observation confirmed that the inherent eukaryotic chromatin state exerts a substantial influence on editing efficacy. Consequently, we made the strategic decision to employ the usage of cell lines that have relatively open chromatin profile to knockout *ZSCAN4* gene.

To accomplish this, we used MDA-MB-231 breast cancer cells and iPSCs, which are reported to have a relatively open chromatin structure compared to fibroblasts (Cai et al., 2020; Gaspar-Maia et al., 2011; Harris et al., 2023; Knaupp et al., 2017; Li et al., 2021). It is a well-known prerequisite that chromatin opens up and becomes more accessible during reprogramming to generate iPSCs (Knaupp et al., 2017; Li et al., 2021). This open chromatin is one of the crucial factors for the maintenance of pluripotency in iPSCs (Gaspar-Maia et al., 2011). The iPSCs used in this study, TJC8, is a well characterized iPSC line (Manian et al., 2018; Thamodaran et al., 2021). The TJC8 iPSC used in this study has a dox-inducible Cas9 cassette integrated at the safe harbor AAVS1 site (Thamodaran et al., 2021). Therefore, TJC8 iPSCs were transduced only with pKLV2.2-dgRNA viruses and induced with Dox for the expression of Cas9. As our results suggest, the cells with a relatively open chromatin, like MDA-MB-231 and iPSCs, led to higher knockout efficiency when compared to fibroblasts. Importantly, we were able to achieve highest knockout efficiency so far in this study, in case of iPSCs. Taken together, these results suggest that eukaryotic chromatin state has a very significant influence on the knockout efficiency while using CRISPR-Cas9 toolbox. However, despite this enhanced knockout efficiency in iPSCs, we were unable to generate homozygous *ZSCAN4* knockout cells. This is in contrast with one of our previous studies, where an incomplete knockout of the target gene, *UTF1*, was observed in fibroblasts, but homozygous knockout of *UTF1* was observed in case of iPSCs (Raina et al., 2025), suggesting a *ZSCAN4*-specific resistance to complete knockout.

Earlier studies revealed controlled expression of ZSCAN4, where it is reported to be expressed only in ~5% of cells in culture in mouse ESCs and iPSCs (Hirata et al., 2012; Zalzman et al., 2010). This tight control over its expression suggests that *ZSCAN4* locus might be present in the euchromatin region of iPSCs, leading to inaccessibility by the dgRNA and Cas9 nuclease as discussed earlier (Verkuijl and Rots, 2019; Wu et al., 2014). This, in turn, might be generating a heterozygous knockout of *ZSCAN4* gene in target cells.

Despite, optimizing transduction ratio and employing multiple cell lines with relatively open chromatin, CRISPR-Cas9-based knockout of *ZSCAN4* gene failed to achieve a complete, homozygous knockout. Previous study reported gRNA potency as one of the vital parameters for the success of CRISPR-Cas9-based knockout of target gene (Yuen et al., 2017). The study suggested that all the gRNAs have their own intrinsic potency, the efficiency with which the gRNA will bind to the target locus and induce a DSB at the locus (Yuen et al., 2017). Therefore, screening and choosing the best gRNA with maximum potency is crucial for the success of CRISPR-Cas9-based knockout of target gene. Hence, we performed a comparative screening of multiple pairs of gRNAs, which were designed using different online tools and subsequently, their potency to knockout *ZSCAN4* gene was assessed in AHDF cells. Although, all the dgRNA sets were ranked as most efficient ones by their respective designing tool, experimental validation revealed a significant variability in knockout efficiency, emphasizing the significance of *in vitro* experimental validation over computational modelling. All the chosen new dgRNAs resulted in partial knockout of *ZSCAN4* in AHDF cells with varying knockout efficiencies. Interestingly, in case of dgRNA set 2, a third band was observed with an amplicon size lying between corresponding wild-type and knockout amplicon sizes. This third band might be due to indels incorporated from the process of non-homologous end joining (NHEJ) at the target locus by the DNA repair machinery of the cells. These results signify the screening of gRNA that has best potency is crucial for successful knockout of the target gene. Among all the gRNAs that were screened in this study, the dgRNA against exon 5 demonstrated superior knockout efficiency across all evaluated cell lines. Moreover, dgRNA against exon 5 specifically targets the Zinc finger domain, a functionally indispensable domain of ZSCAN4 protein (Thool et al., 2022). Targeting this conserved domain will increase the probability of rendering the protein non-functional. Therefore, dgRNA against exon 5 was chosen for further experiments.

The persistent partial knockout across different cell lines and with gRNAs with varying potencies suggested the cell populations attained after selection were heterogeneous, indicated by the presence of two bands (WT and KO) in the PCR analysis. This heterogeneity can be attributed to two main scenarios. Firstly, DSBs have been induced at the target site, *ZSCAN4*, in one of the alleles, leaving the other allele intact. This scenario would have led to the formation of two amplicons of different sizes on performing PCR. Secondly, population mosaicism, which can be attributed to the presence of a heterogeneous population of knocked out and wild-type cells. We hypothesize that the second scenario may arise from the inherent varying potency of gRNAs as discussed earlier (Yuen et al., 2017). One of the previous studies showed that all the gRNAs possess an inherent potency with which they can bind to the target and eventually induce a DSB at the target locus (Yuen et al., 2017). The gRNAs that are being employed in this study have varying potencies to induce a DSB at the target site, *ZSCAN4*, as evident from the results, leading to a heterogeneous population of mixture of wild-type and knockout cells.

To resolve the issue of heterogeneity and attain a homozygous population of *ZSCAN4* knocked out cells, we performed clonal expansion (Bhargava et al., 2022; Chen and Pruett-Miller, 2018). In this method, single cells were seeded in a 96-well plate and eventually expanded. Since the population will have emerged from a single cell, it will generate a homogeneous, clonal population having uniform genotype, which in our case will be *ZSCAN4* knockout. Isolating clonal population is one of the major bottlenecks in CRISPR-Cas9 based genome modification (Bhargava et al., 2022; Chen and Pruett-Miller, 2018; DeWeirdt et al., 2020). Since iPSCs possess indefinite self-renewal capability and the ability to generate a clonal population, clonal expansion was performed with iPSCs (Bhargava et al., 2022; Chen and Pruett-Miller, 2018; Singh, 2019). The clonal populations attained were screened for EGFP expression. Those clones expressing EGFP were further expanded, genomic DNA was isolated, and the knockout was analyzed by performing PCR. Isolating clonal population is one of the major bottlenecks in CRISPR-Cas9 based genome modification.

Among the nine clones analyzed following clonal expansion, seven clones (77.8%; C2, C3, C4, C6, C7, C8, and C9) exhibited biallelic disruptive mutations with no detectable wild-type allele, indicating high-confidence functional knockout of *ZSCAN4* gene. These included clones harboring either precise deletions or compound indel events affecting both

alleles. In contrast, the remaining two clones (22.2%; C1 and C5) retained a wild-type-sized amplicon (although sequencing revealed nucleotide substitutions in these alleles) along with an additional smaller or larger band, consistent with heterozygous editing and incomplete disruption of *ZSCAN4* gene. Overall, the majority of clones demonstrated high-confidence biallelic disruption of *ZSCAN4* gene, highlighting the efficiency of CRISPR-Cas9-mediated gene editing in human iPSCs. Functional validation at the protein level will be required to confirm complete loss of ZSCAN4 activity in all these clones.

Notably, there were four clones (C1, C5, C6, and C8) out of the nine clones that exhibited heterozygous amplification profile, yielding two different-sized amplicons with indels or deletions (sequencing data). This suggests that both the alleles in that clonal population have repaired the DSB in an independent and distinct manner. This heterogeneity is a hallmark of the error-prone DNA repair mechanism of the cell via NHEJ pathway (Gaj et al., 2013), which repairs DSBs with unique indels in each cell, and thereby the clonal population expanded from them (Fischer et al., 2022; Wang et al., 2019). This led to the generation of two amplicons of different sizes corresponding to each allele. This phenomenon of obtaining heterozygous knockout in a clonal population is observed in previous studies in case of fibroblasts as well as in iPSCs (Fischer et al., 2022; Wang et al., 2015, 2019). The remaining five clones (C2, C3, C4, C7, and C9) exhibited homozygous amplification profile, yielding single amplicon upon semi-quantitative PCR. The sequencing results corroborated the same. This phenomenon underscores the necessity of rigorous clonal selection and sequencing to ensure that the functional genetic disruptions observed are tied to a clearly defined genomic state.

Despite initial heterogeneity in knockout, final optimization yielded a homogeneous, clonal population having complete knockout of *ZSCAN4* gene in iPSCs. These cells can be further characterized to get a deeper understanding on the role of *ZSCAN4* in the maintenance of self-renewal and pluripotency property of iPSCs. This is the first study to design a lentivirus-based, dgRNA CRISPR-Cas9 toolbox for robust ablation of human *ZSCAN4* gene in any target cell and to provide a validated strategy for the generation of homozygous *ZSCAN4* knockout population.

4.5 Conclusion

CRISPR-Cas9 is an effective tool in molecular biology for gene editing, which enables us to gain an in-depth insight on the mechanistic role of the target gene. However, there are few hurdles that needs to be addressed for an efficient application of the CRISPR-Cas9 toolbox. This study systematically investigated various critical factors, such as viral transduction, chromatin state of the cells, gRNA potency, and clonal selection, and the influence of these factors on the efficiency of knockout of human *ZSCAN4* gene using CRISPR-Cas9 toolbox. It is evident that factors like viral transduction ratio and chromatin state of the cells have a significant influence on the efficiency of gene knockout using CRISPR-Cas9 toolbox. We have demonstrated that using chemical molecules like 5-Aza-dC to relax the chromatin state enhanced the efficiency of the knockout. Moreover, the efficiency of knockout of *ZSCAN4* gene is considerably higher in cells having an open chromatin structure (like MDA-MB-231 and iPSCs) compared to those cells having a closed chromatin structure (like fibroblasts). Furthermore, this study demonstrated that the inherent potency of gRNA plays a crucial role in determining knockout efficiency and that experimental validation is absolutely necessary in spite of higher efficiencies calculated by computational tools. Despite optimizing transduction ratio, chromatin state, and gRNA potency, our efforts resulted in a heterogeneous population of cells having a mixture of wild-type and knockout cells. To overcome this, clonal expansion was performed to attain a homozygous population of *ZSCAN4* knockout cells. This is the first study to develop an efficient, dgRNA-based toolbox to knockout *ZSCAN4* gene in human cells. By elucidating and addressing the influence of factors such as viral delivery and chromatin accessibility, this work provides an optimized methodology that will facilitate the successful and efficient utilization of the CRISPR-Cas9 system in molecular biology research.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Chapter 5

Characterization of *ZSCAN4* knockout iPSCs

Overview

The knockout of human *ZSCAN4* gene in human induced Pluripotent Stem Cells (iPSCs) was carried out successfully in the previous Chapter 4. Homozygous *ZSCAN4* knockout clones were identified, expanded, and cryopreserved. Out of the nine *ZSCAN4* knockout clones, four of them were chosen for expansion and comprehensive characterization to understand the functional role of *ZSCAN4* in the maintenance of iPSCs. Morphological analysis revealed a divergent response to ZSCAN4 deficiency among the clones. Two of the knockout clones exhibited iPSC-like morphology with round, tightly packed cells, while the other two clones showed differentiated morphology that was characteristically different from iPSC-like morphology, indicating their spontaneous differentiation. Gene expression profiling of core pluripotency markers, namely OCT4, SOX2, NANOG, and REX1, revealed significant downregulation of the pluripotency genes analyzed in the clones that showed differentiated morphology. Further, immunofluorescence analysis of pluripotency surface (TRA-1-60, TRA-1-81, and SSEA4) and nuclear markers (OCT4, SOX2, and NANOG) revealed comparable expression of pluripotency markers in the clones that showed iPSC-like morphology relative to negative control. Interestingly, the clone exhibiting differentiated morphology showed complete loss of pluripotency marker expression, with the exception of TRA-1-60. Furthermore, studying the relative telomere length in *ZSCAN4* knockout clones revealed no significant differences, indicating the existence of compensatory or redundant mechanisms involved in telomere maintenance. Moreover, analysis of lineage-specific markers in the clone that showed differentiated morphology, revealed that the clone differentiated spontaneously towards all the three germ layers, resulting in a phenotype consistent with carcinoma-like cells. Finally, a semi-quantitative PCR was performed on RNA isolated from knockout clones to confirm the absence of *ZSCAN4* knockout at transcript level at passage 10, which showed the maintenance of stable knockout cells. Taken together, all these results show that knockout of *ZSCAN4* gene in human iPSCs led to heterogeneous behavior in the analyzed homozygous clones, and further studies are required to completely understand the vitality of ZSCAN4 in the maintenance of self-renewal and pluripotency characteristics of these cells.

5.1 Introduction

The CRISPR–Cas9 toolbox is widely employed to investigate and elucidate the functional roles of genes in cellular pathways that were previously not investigated or not well understood. Since its inception in 2012, the CRISPR-Cas9 genome editing toolbox has revolutionized the field of functional genomics (Jinek et al., 2012). This platform has been extensively utilized to generate targeted knockouts, facilitating a deeper understanding of gene involvement in complex metabolic and regulatory networks (Ben Jehuda et al., 2018; Chehelgerdi et al., 2024). Rigorous characterization of these gene knockout clones is an essential prerequisite for defining their biological significance. This approach has been widely applied to iPSCs to study gene functions across various applications (Geng et al., 2020; Harper et al., 2023; Wu et al., 2018; Xu et al., 2019). Through various molecular and cellular assays, we can observe the phenotypic, genotypic, and metabolic changes resulting from a knockout, providing a direct insight into the functionality of the target gene (Geng et al., 2020; Harper et al., 2023; Wu et al., 2018; Xu et al., 2019). The broader applications of CRISPR-Cas9 toolbox in disease modelling and functional analysis are discussed in brief elsewhere (Li et al., 2023; Sundaravadivelu et al., 2025). This chapter details the extensive phenotypic and molecular characterization of *ZSCAN4* knockout clones to determine the functional requirement of this gene in the long-term maintenance of human iPSCs.

This chapter employs a multifaceted experimental framework comprising RT-qPCR, immunofluorescence and telomere qPCR to systematically evaluate the consequences of the absence of *ZSCAN4* in human iPSCs. Ensuring the permanent stability of the genetic disruption is of paramount importance for accurate interpretation of long-term knockout effects. Consequently, the stability of homozygous knockout of *ZSCAN4* gene in the clones at a transcriptome level was studied by performing a semi-quantitative PCR with specific primers. Notably, this research represents inaugural study to achieve and characterize homozygous knock out of *ZSCAN4* in human iPSCs, providing a foundational baseline for understanding its essentiality in the maintenance of iPSCs.

5.2 Materials and Methods

5.2.1 Semi-quantitative PCR

Semi-quantitative PCR was performed in accordance with the protocol detailed in section 3.2.8. Briefly, either genomic DNA or cDNA was used as the template for

amplification. The PCR was performed using target-specific primers at their optimized annealing temperatures (Table S2). The PCR products were resolved using 2% agarose gel and imaged using the ChemiDoc XRS+ system (Bio-Rad).

5.2.2 Immunofluorescence

Immunofluorescence was performed following the protocol previously mentioned in section 2.2.8. Briefly, the cells were fixed, permeabilized, blocked, incubated with respective primary and secondary antibodies (Table S3), counterstained with Hoechst nuclear stain, and imaged using an inverted fluorescence microscope (ZOE™ Fluorescent Cell Imager, Bio-Rad). The images were further processed using ImageJ software.

5.2.3 Real-Time quantitative PCR (RT-qPCR)

The total RNA was isolated from the cells and RT-qPCR was performed following the protocol previously mentioned in section 3.2.3. Briefly, RNA was isolated, cDNA was synthesized, and RT-qPCR was performed. The results were normalized using GAPDH as the reference gene. The final results were plotted as mean±SEM, and were statistically analyzed using student's t-test or one-way ANOVA with 95% confidence interval.

5.2.4 Telomere qPCR

The relative telomere length was quantified by performing qPCR as previously described (Cawthon, 2009; Joglekar et al., 2020). Briefly, genomic DNA was isolated following the protocol previously mentioned in section 4.2.6. For performing telomere qPCR, 60 ng of genomic DNA was used as a template per PCR reaction. The qPCR was performed by using 100 nM of telomere-specific primers (Cawthon, 2009) and 180 nM of primers specific to human β-globin (Joglekar et al., 2020). The human β-globin is a single copy gene and was used as a reference gene for normalization. No template control was used as a negative control. The relative telomere length was calculated using $2^{-\Delta\Delta ct}$ method as previously described (Joglekar et al., 2020). The relative telomere length values were plotted using GraphPad Prism software and were statistically analyzed using one-way ANOVA with 95% confidence interval.

5.3 Results

5.3.1 Phenotypic analysis of ZSCAN4 knockout iPSC clones

From the semi-quantitative PCR results from previous chapter, it was revealed that four of the nine clones (C2, C4, C7, and C9) exhibited homozygous *ZSCAN4* knockout yielding the expected single amplicon at ~226 bp. For this chapter, these four clones were chosen for further expansion and characterization to understand the functional role of ZSCAN4 in the maintenance of iPSCs.

Further, the phenotypic consequences of *ZSCAN4* knockout in iPSCs, were observed using brightfield microscopy. Wild-type iPSCs and iPSCs transduced with dgRNA-BB maintained characteristic pluripotent morphology defined by small, rounded cells organized in tightly packed colonies (Fig. 5.1; *left panel*). In contrast, the four confirmed *ZSCAN4* knockout clones exhibited a heterogeneous phenotypic spectrum. Clones C4 and C7 closely resembled wild-type iPSCs, displaying compact colonies composed of small, rounded cells consistent with typical iPSC morphology (Fig. 5.1, *middle panel*). Conversely, clones C2 and C9 exhibited differentiated morphologies distinct from the pluripotent state. Clone C2 displayed an elongated cellular morphology, whereas clone C9 consisted of comparatively larger, rounded cells (Fig. 5.1, *right panel*). Upon subsequent passaging, both C2 and C9 progressively lost their characteristic iPSC morphology and underwent complete differentiation. In contrast, clones C4 and C7 retained typical iPSC morphology and showed no signs of spontaneous differentiation up to passage 15.

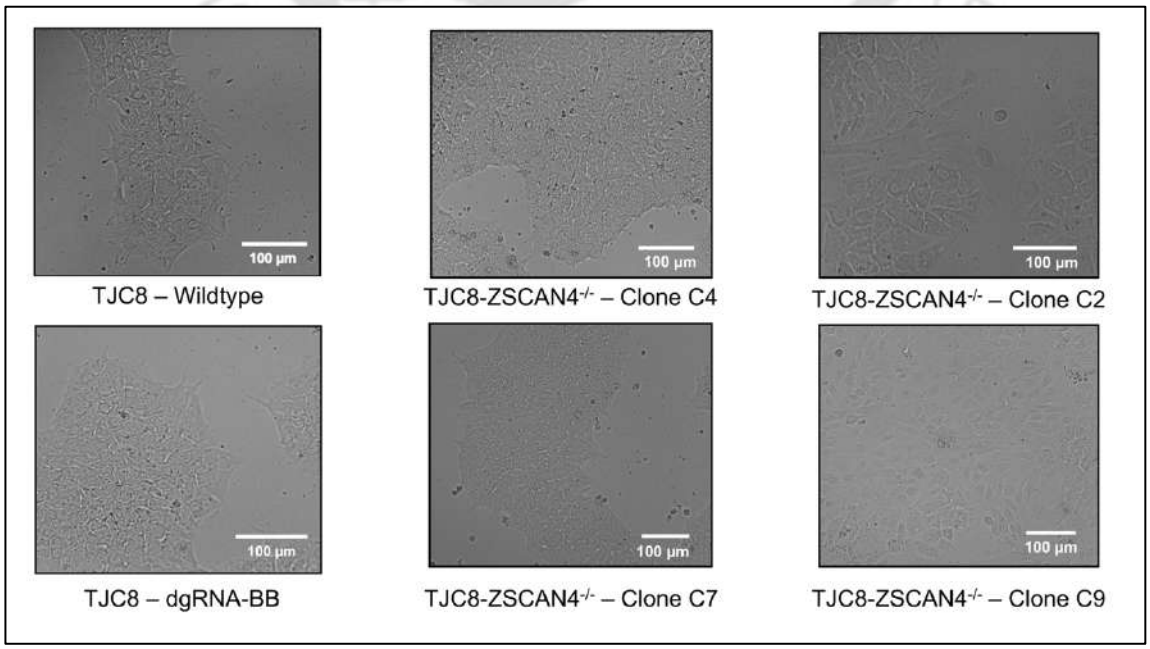


Fig. 5.1. Phenotypic analysis of *ZSCAN4* knockout TJC8 iPSC clones. The four chosen *ZSCAN4* knockout clones were analyzed using brightfield microscopy for any morphological changes in their phenotype. The wild-type TJC8 cells and TJC8 cells transduced with dgRNA-BB served as negative controls. Scale bar: 100 µm.

Collectively, these observations indicate that, despite all the four clones being confirmed as homozygous *ZSCAN4* knockouts at the genomic level, loss of ZSCAN4 resulted in a heterogeneous phenotypic response. This suggests that ZSCAN4 plays a context-dependent role in maintaining pluripotency and cellular stability in some clones, potentially mediated by compensatory or parallel regulatory pathways.

5.3.2 Assessment of relative expression of pluripotency markers in ZSCAN4 knockout iPSC clones

The relative expression levels of pluripotency markers were analyzed using RT-qPCR to evaluate the molecular impact of *ZSCAN4* knockout in iPSCs. The expression of core pluripotency markers, namely OCT4, SOX2, NANOG, and REX1 was analyzed. Although clone C2 exhibited rapid phenotypic collapse and did not survive beyond passage 3, total RNA was isolated successfully at passage 2 to enable this analysis. The results revealed differential response among *ZSCAN4* knockout clones compared to the cells transduced with dgRNA-BB virus, which served as the negative control in this experiment. The expression level of OCT4 exhibited a marked increase in clone C4 (~2.2-fold), while its expression was significantly downregulated in the other three clones relative to control, C2 (~1000-fold), C7 (~2-fold), and C9 (~50-fold) (Fig. 5.2). A similar pattern was observed for the expression of SOX2, where in clone C4 showed significant upregulation (~1.8-fold) whereas significant downregulation was observed in the clones C2 (~500-fold), C7 (~5-fold), and C9 (~11-fold) relative to control cells (Fig. 5.2).

Next, relative expression levels of NANOG was assessed in *ZSCAN4* knockout iPSC clones. Unlike the divergent patterns observed for OCT4 and SOX2, NANOG expression was universally and significantly downregulated in all the four clones relative to control cells, C2 (~1000-fold), C4 (~2-fold), C7 (~2.5-fold), and C9 (~100-fold) (Fig. 5.2). Expression of REX1, a marker of naïve pluripotency, was significantly increased in C4 (~2.8-fold), while being significantly reduced in the remaining clones compared to the control (C2: ~50-fold; C7: ~50-fold; C9: ~25-fold) (Fig. 5.2). Collectively, the expression levels of OCT4, SOX2, and REX1 followed a similar pattern

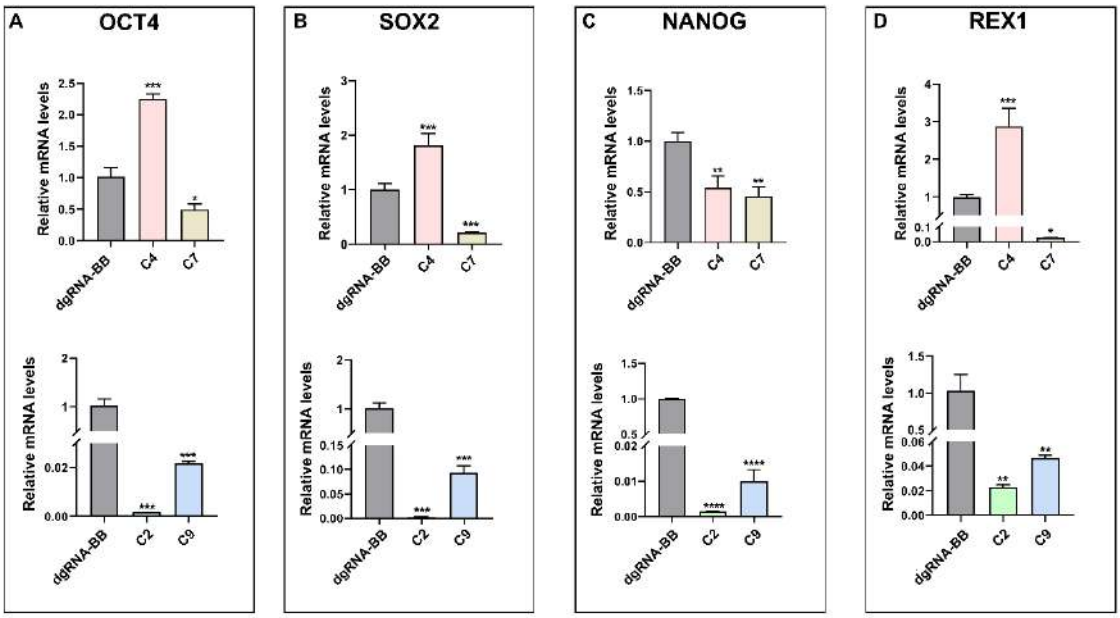


Fig. 5.2. Relative expression analysis of pluripotency genes in *ZSCAN4* knockout TJC8 iPSC clones. RT-qPCR was performed to analyze the relative expression of pluripotency markers, namely OTC4, SOX2, NANOG, and REX1 in *ZSCAN4* knockout TJC8 iPSC clones. The expression levels were normalized using GAPDH, and the results were plotted as mean±SEM. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. dgRNA-BB: TJC8 cells transduced with dgRNA-BB; C2, C4, C7, C9: *ZSCAN4* knockout homozygous TJC8 iPSC clones.

in *ZSCAN4* knockout clones, whereas their expression was significantly upregulated in C4 but suppressed in other three clones (C2, C7, and C9). NANOG stood out as the only factor that was downregulated in all the four knockout clones regardless of their initial morphology. These data suggest that while *ZSCAN4* deficiency consistently destabilizes NANOG expression, it may trigger compensatory upregulation of other core factors like OCT4, SOX2 or REX1 in specific cellular contexts like clone C4.

5.3.3 Immunofluorescence analysis of pluripotent markers in ZSCAN4 knockout iPSC clones

The expression of surface and nuclear pluripotent markers was analyzed in *ZSCAN4* knockout iPSC clones using immunofluorescence. Since the clone C2 did not survive beyond passage 3, the immunofluorescence analysis was not carried out for this clone and was performed for the other three clones (C4, C7, and C9). The analysis revealed robust expression of the surface markers, TRA-1-60, TRA-1-81, and SSEA4, in the clones C4 and C7 (Fig. 5.3). In case of C9, no expression was observed when the cells were stained for

TRA-1-81 and SSEA4 (Fig. 5.3). Notably, a faint signal for TRA-1-60 was detected in clone C9, despite its differentiated morphology (Fig. 5.3).

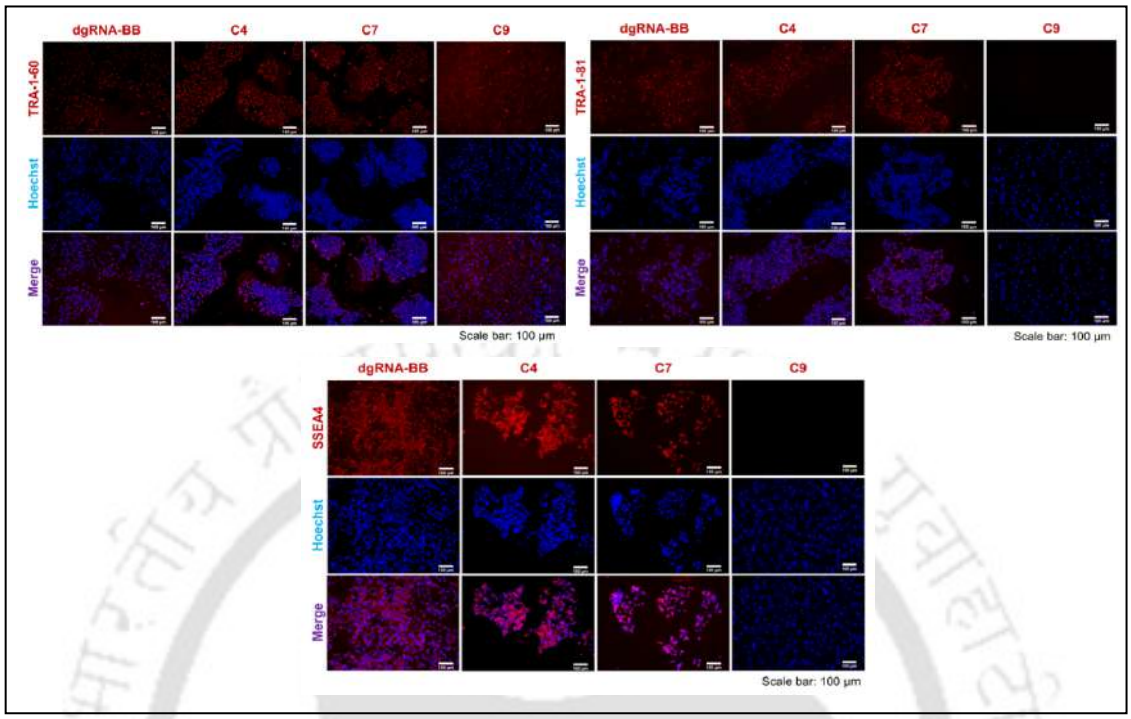


Fig. 5.3. Immunofluorescence analysis of pluripotency surface markers in *ZSCAN4* knockout TJC8 iPSC clones. *ZSCAN4* knockout clones, C4, C7, and C9, were analyzed for the expression of pluripotency surface markers, namely TRA-1-60, TRA-1-81, and SSEA4 using immunofluorescence. Scale bar: 100 µm. dgRNA-BB: TJC8 cells transduced with dgRNA-BB; C4, C7, C9: *ZSCAN4* knockout homozygous TJC8 iPSC clones.

Subsequently, the expression of nuclear pluripotency markers was also analyzed. The analysis revealed robust expression of the core nuclear markers, namely OCT4, SOX2, and NANOG, in the clones C4 and C7 (Fig. 5.4). In contrast, no detectable expression of the nuclear markers was observed in the clone C9 (Fig. 5.4). The complete lack of these master regulators confirmed that the clone C9 has exited pluripotent state and undergone spontaneous differentiation.

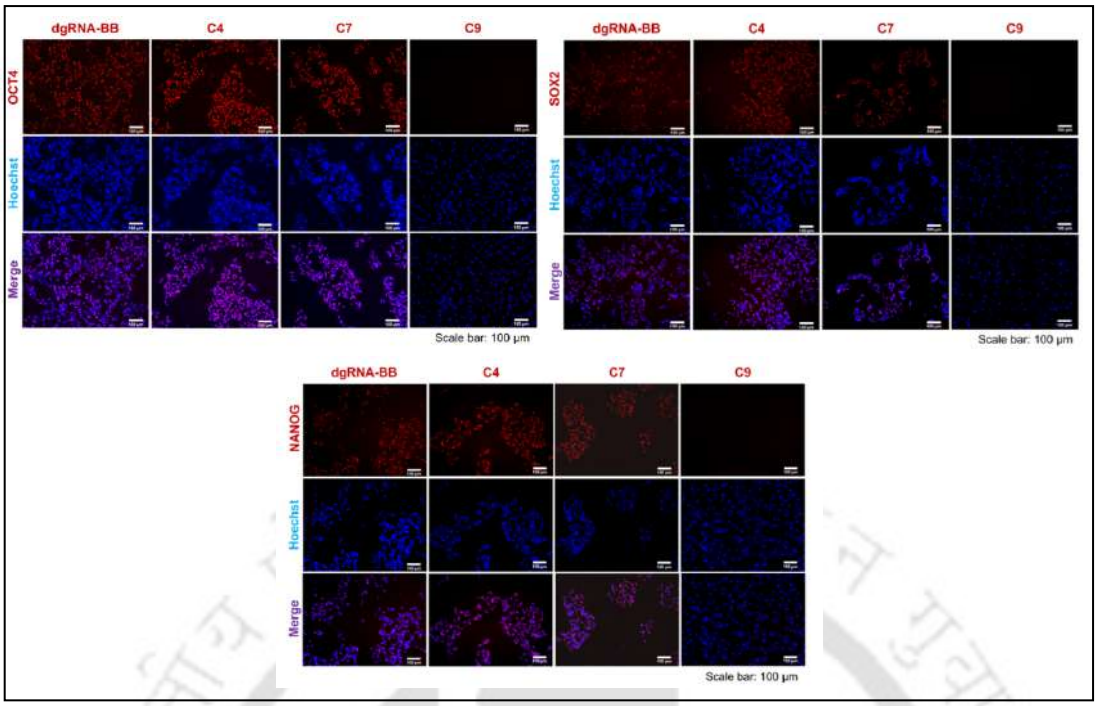


Fig. 5.4. Immunofluorescence analysis of pluripotency nuclear markers in *ZSCAN4* knockout TJC8 iPSC clones. *ZSCAN4* knockout clones C4, C7, and C9 were analyzed for the expression of pluripotency nuclear markers, namely OCT4, SOX2, and NANOG using immunofluorescence. Scale bar: 100 µm. dgRNA-BB: TJC8 cells transduced with dgRNA-BB; C4, C7, C9: *ZSCAN4* knockout homozygous TJC8 iPSC clones.

5.3.4 Spontaneous differentiation in ZSCAN4 knockout iPSC clone C9

As mentioned earlier, one of *ZSCAN4* knockout clones C9, underwent rapid differentiation in early passage and maintained its differentiated state at least until passage 15. To determine whether this phenotypic change in the absence of *ZSCAN4* in clone C9 is random or is a directed transition towards a particular lineage, relative expression of lineage-specific markers was observed. The RT-qPCR results revealed that differentiation is random (towards all three lineages) with the cells leaning towards mesoderm lineage (Fig. 5.5). In case of mesodermal markers, maximum upregulation was observed for CDX2 (~22,000-fold), and HAND1 (~3500-fold) (Fig. 5.5). Similarly, a significant induction of ectodermal transcripts was observed, with OTX2 and DLK1 showing approximately ~2.5-fold and ~40.5-fold upregulation, respectively, relative to negative control cells (Fig. 5.5). In the case of endodermal markers, a marked upregulation was also detected, with EOMES and FOXA2 exhibiting an increase of ~360-fold and ~3-fold, respectively, compared to its controls.

Collectively, these results demonstrate that the loss of ZSCAN4 in clone C9 resulted in a spontaneous collapse of the pluripotent state. While the resulting differentiation was heterogeneous, containing signatures of all three germ layers, the molecular profile was heavily skewed toward mesoderm lineage.

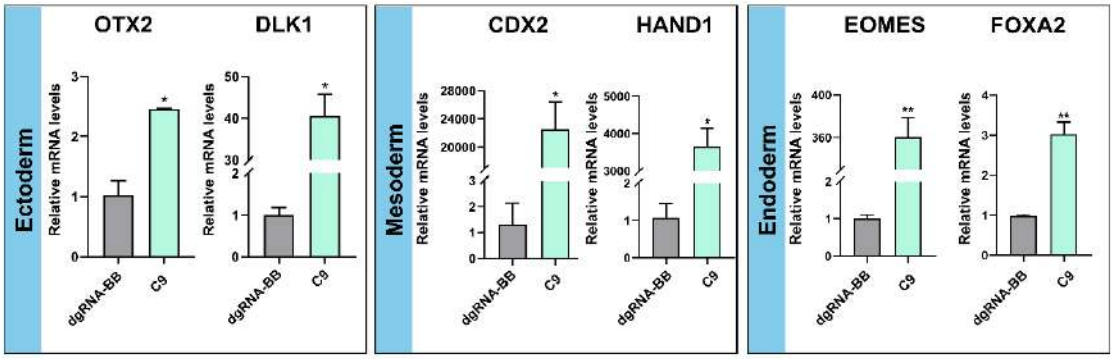


Fig. 5.5. Relative expression analysis of lineage-specific markers in *ZSCAN4* knockout TJC8 iPSC clone C9. RT-qPCR was performed to analyze the relative expression of lineage-specific markers in *ZSCAN4* knockout clone C9. Ectodermal markers (OTX2, DLK1), mesodermal markers (CDX2, HAND1), and endodermal markers (EOMES, FOXA2) were evaluated. Gene expression levels were normalized to GAPDH, and the results were plotted as mean±SEM. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. dgRNA-BB: TJC8 cells transduced with dgRNA-BB; C9: *ZSCAN4* knockout homozygous TJC8 iPSC clones.

5.3.5 Assessment of relative telomere length of ZSCAN4 knockout iPSC clones

ZSCAN4 is a critical transcription factor having a major role in regulating telomere length in iPSCs via ALT pathway. Therefore, to assess the effect of loss of *ZSCAN4* on telomere length, telomere qPCR was performed using the genomic DNA isolated from *ZSCAN4* knockout iPSC clones. The telomere qPCR revealed no change in relative telomere length compared to the iPSCs transduced with dgRNA-BB (Fig. 5.6). All the four *ZSCAN4* knockout clones (C2, C4, C7, and C9) did not have significant difference in relative telomere length (Fig. 5.6). Notably, the maintenance of telomere length was observed regardless of the morphological state of the clone. Even in clone C2, which was not viable beyond passage 3, the telomeres remained at length statistically insignificant from control cells. Similarly, the differentiated clone C9 maintained its telomeric length comparable to control cells.

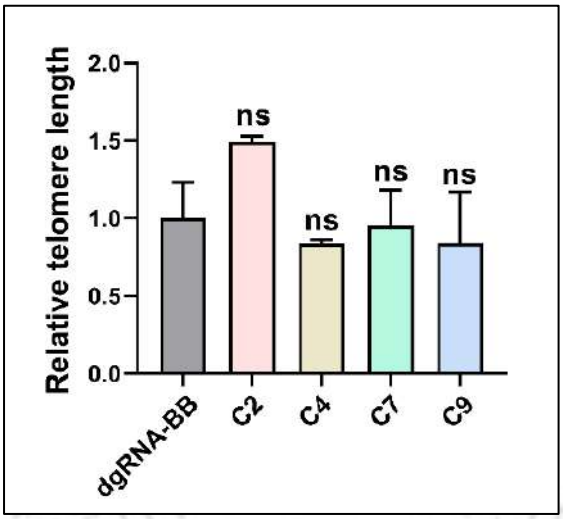


Fig. 5.6. Relative telomere length analysis of *ZSCAN4* knockout TJC8 iPSC clones. Telomere qPCR was performed to analyze the relative telomere length in *ZSCAN4* knockout TJC8 iPSC clones. The results were normalized using human β-globin, a single copy gene, and the results were plotted as mean±SEM. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. dgRNA-BB: TJC8 cells transduced with dgRNA-BB; C2, C4, C7, C9: *ZSCAN4* knockout homozygous TJC8 iPSC clones.

5.3.6 Confirmation of stability of ZSCAN4 knockout in iPSCs at transcriptome level

The knockout of *ZSCAN4* was previously confirmed at the genomic DNA level using semi-quantitative PCR and Sanger sequencing. To assess the long-term stability of the genetic disruption, total RNA was isolated from ZSCAN4 knockout iPSC clones at passage 10 and analyzed by semi-quantitative PCR to confirm maintenance of the knockout during extended culture. The cDNA synthesized from the isolated RNA was used as the template for the PCR. Firstly, the PCR was performed using the dgRNA-Ex5 primers (Table S2), which were designed to flank the DSBs induced by the Cas9 nuclease (Fig. 5.7A). Therefore, the cells transduced with dgRNA-BB, will give an amplicon of size 286 bp, whereas knockout clones will give an amplicon of size 226 bp. As expected, the cells transduced with dgRNA-BB gave a single amplicon of size ~286 bp (Fig. 5.7B; *second lane*), whereas *ZSCAN4* knockout clones, namely C4, C7, and C9 produced a single amplicon of ~226 bp (Fig. 5.7B). This correlated to a deletion of 60 bp fragment in the target region of exon 5. This data confirmed that *ZSCAN4* knockout remained stable at the mRNA level across 10 passages.

To provide an additional layer of validation, another set of primer was designed in such a way that the forward primer will bind exactly in the 60 bp deletion region as depicted in fig. 5.7C and the reverse primer will bind in the region downstream of DSB induced by

Cas9 nuclease (Fig. 5.7C). This primer set will result in a formation of an amplicon of 242 bp in cells transduced with dgRNA-BB, where the target locus will remain intact, resulting in the production of an amplicon. In ZSCAN4 knockout clones, the deletion disrupted the forward primer binding site, thereby preventing primer annealing and resulting in the formation of no PCR product. The PCR results revealed formation of amplicon in cells transduced with dgRNA-BB corresponding to ~242 bp (Fig. 5.7D). As expected, *ZSCAN4* knockout clones did not produce any PCR products (Fig. 5.7D), confirming the deletion of 60 bp fragment in the target region. The consistency of these results at passage 10 in all the three homozygous clones C4, C7, and C9 indicated that *ZSCAN4* knockout is transcriptomically stable over long-term culture.

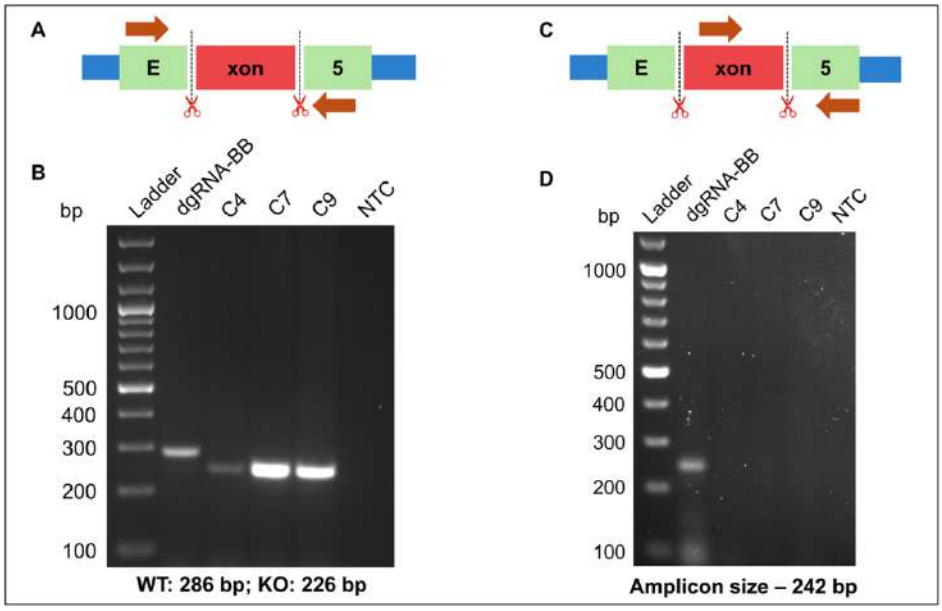


Fig. 5.7. Confirmation of long-term stability of *ZSCAN4* knockout at transcript level in *ZSCAN4* knockout TJC8 iPSC clones. To confirm the stability knockout of *ZSCAN4* at RNA level, the knockout clones from passage 10 were analyzed using semi-quantitative PCR. The PCR was performed with two different sets of primers as depicted in (**A**) and (**B**). The PCR products were resolved in 2% agarose gels (**C**) and (**D**). bp: basepairs; dgRNA-BB: TJC8 cells transduced with dgRNA-BB; C4, C7, C9: *ZSCAN4* knockout homozygous TJC8 iPSC clones; NTC: no template control; WT: wild-type; KO: knockout.

5.4 Discussion

ZSCAN4 is one of the crucial transcription factors for the generation of mouse iPSCs and their maintenance (Hirata et al., 2012; Jiang et al., 2013). While the role of ZSCAN4 has been established in murine system, its role in human iPSC reprogramming

and maintenance is greatly unexplored. Therefore, understanding the role of ZSCAN4 from the perspective of human iPSC reprogramming and maintenance is of utmost importance. This chapter addresses that, by employing CRISPR-Cas9 toolbox to knockout *ZSCAN4* in human iPSCs and studying the effect of loss of *ZSCAN4* on the maintenance of the pluripotent and self-renewal property of iPSCs.

The morphological analysis of *ZSCAN4* knockout clones revealed a distinct phenotypic divergence immediately following clonal selection. The morphological analysis revealed that two of the clones exhibited signs of differentiation right from passage 1 (C2 and C9), of which clone C2 was not viable beyond passage 3. In contrast, the other two clones (C4 and C7) maintained their iPSC-like morphology at least till passage 15. The ablation of *ZSCAN4* directly led to the loss of pluripotency in 50% of the characterized clones with one clone failing to survive beyond early passage, highlighting the vital role of ZSCAN4 in maintaining the self-renewal property of human iPSCs.

The molecular profile of *ZSCAN4* knockout clones was further elucidated through quantification of expression of core pluripotency factors. The analysis revealed that clones exhibiting differentiated morphology (C2 and C9) showed a pronounced and statistically significant downregulation of the pluripotency network. The expression level of the pluripotency genes analyzed, namely OCT4, SOX2, NANOG and REX1 was significantly downregulated (less than ~0.1-fold relative to control) in both the differentiated *ZSCAN4* knockout iPSC clones C2 and C9. This marked suppression indicates a coordinated collapse of the gene regulatory network required to maintain the undifferentiated state, confirming that the observed morphological changes in these clones are driven by loss of the core pluripotency program. The expression of these pluripotent factors in the other two iPSC clones that maintained the pluripotent morphology (C4 and C7) was either upregulated or comparable to the expression levels in iPSCs transduced with dgRNA-BB. The observed upregulation in certain factors suggests a potential compensatory mechanism where the cell attempts to stabilize the pluripotent network in the absence of *ZSCAN4*.

The expression of pluripotent markers was further corroborated at protein level by performing immunofluorescence. The results revealed robust and comparable expression of pluripotency surface and nuclear markers in *ZSCAN4* knockout clones C4 and C7, which maintained their pluripotent morphology. These results indicate that in these two *ZSCAN4* knockout clones (C4 and C7), the core pluripotent protein network remains functionally

intact. Conversely, clone C9, which exhibited a differentiated morphology, showed an absence of both surface and nuclear pluripotency markers, with the exception of residual TRA-1-60 expression.

Further, to definitively validate that the differentiated state of clone and characterize its development trajectory, the relative expression levels of lineage-specific markers was analyzed. The analysis revealed that the clone C9 has showed a random and rapid exit from pluripotency and has differentiated spontaneously towards all three lineages, namely ectoderm, mesoderm, and endoderm. While the population is heterogeneous, a distinct lineage bias was observed. The expression profile suggested that the cells are heavily skewed towards a mesodermal fate. Taken together, all these results suggests that the clone C9 has lost its pluripotent property and spontaneously differentiated towards all three lineages with a biasness towards mesoderm.

Even though TRA-1-60 is considered as one of the characteristic surface markers for identification of stem cells, its expression is not exclusive to pluripotent stem cells. Previous reports suggested that expression of TRA-1-60 is found in other cells as well, namely embryonal carcinomas of testes (Schopperle and DeWolf, 2007), embryonal carcinomas of ovaries (Malecki et al., 2012), follicular lymphomas (Takata et al., 2019), blood cells of patients with metastatic prostate cancer (Schäfer et al., 2020), and in differentiated cells like foetal and adult human kidney (Fesenko et al., 2010). Notably, TRA-1-60 is highly expressed in embryonal carcinoma. Embryonal carcinoma is a highly malignant tumor arising in germ cells and is considered as the malignant counterparts to ESCs (Schopperle and DeWolf, 2007), since they possess self-renewal and pluripotent characteristics. The phenotypic and molecular profile of *ZSCAN4* knockout clone C9 aligns closely with these characteristics of embryonal carcinoma cells. The clone C9 expresses only TRA-1-60 out of all the pluripotency markers analyzed, exhibited differentiated morphology, was viable for up to 15 passages despite its differentiated state and simultaneously expressed markers of all three lineages, which suggested that the clone C9 has transformed into carcinoma-like cells. However, further functional assays are required to confirm this malignant transformation.

As discussed previously, pluripotent stem cells regulate their telomere length using two different independent pathways, namely telomerase-dependent pathway and telomerase-independent pathway also called as ALT pathway. Existing literature has

robustly characterized ZSCAN4 as one of the master regulators of ALT pathway (Thool et al., 2022; Wang et al., 2012; Zalzman et al., 2010). Also, inactivation of ZSCAN4 led to telomere shortening in mouse ESCs and in specific ALT-positive cancer cell lines as reported by previous studies (Dan et al., 2022; Zalzman et al., 2010), which underscores the importance of ZSCAN4 in telomere length regulation. Since telomere length was shortened in the absence of ZSCAN4 in mouse ECSs (Zalzman et al., 2010), we were curious if ZSCAN4 has same functionality in human iPSCs as well. However, our investigation in human iPSCs revealed a divergent outcome. Despite the successful homozygous knockout of *ZSCAN4*, no significant difference in relative telomere length was observed in the characterized clones. We hypothesize that this might be due to the fact that stem cells have two independent pathways to regulate telomere length and since ZSCAN4 is vital for one of the pathways, absence of which might have triggered the upregulation of the telomerase-dependent pathway to compensate for the absence of ZSCAN4. This metabolic shift in telomere maintenance could preserve the telomere length masking the deficit caused by the inactivation of ZSCAN4. While our data confirmed telomeric length regulation, further assays such as TRAP assays and expression of profiling of TERT and TERC are required to definitively confirm whether telomerase-dependent pathway is indeed compensating for the loss of ZSCAN4 in these knock out clones.

Despite the definitive validation of *ZSCAN4* deletion at genomic level in all the four knockout clones, each line exhibited a distinct and unique characteristics. This necessitated a rigorous investigation into the potential for clonal stability. A primary hypothesis for the differential behavior of the knockout clones was the possibility of an incomplete or unstable knockout. Specifically, we considered whether a minor population of wild-type cells could have persisted during initial stage due to technical limitations. In such a scenario, even a negligible number of wild-type cells could exert a selective growth advantage over *ZSCAN4* knockout population. Over the course of multiple passages, this wild-type fraction could potentially outcompete the knockout cells, eventually dominating the culture and restoring a wild-type like phenotype and molecular signature. This speculation is further supported by the observed heterogeneity in the early stages of *ZSCAN4* knockout process.

Therefore, a semi-quantitative PCR was performed from the RNA isolated from clones at passage 10 to validate the stability of knockout of *ZSCAN4* at transcriptome level,. The PCR results revealed that the knockout of *ZSCAN4* remained stable after 10 passages

at transcriptome level in the knockout clones C4, C7, and C9. The absence of the wild-type amplicon in knockout clones, together with consistent detection of the truncated knockout transcript, provides strong evidence that the isolated clones are homozygous.

The transcriptomic stability of the knockout across all clones suggested that the observed phenotypic heterogeneity may be driven by extrinsic or process-related factors rather than genomic instability. We propose two hypotheses as to why *ZSCAN4* knockout clones exhibit this differential behavior. Firstly, iPSCs are inherently epithelial and rely on growing as tightly packed colonies to maintain their pluripotent state. This growth architecture is essential as this arrangement provides a microenvironment niche that facilitate cell-to-cell contact and aid in localized concentration of create certain growth factors that assist in the maintenance of their pluripotent state (Hicks and Pyle, 2023). The process of clonal isolation mandates seeding these cells as individual cells, which intrinsically disrupts this microenvironment niche. This might activate some non-specific differentiation pathways leading to spontaneous differentiation, which in turn leading to formation of carcinoma-like cells. Because this lineage commitment was not uniform across all *ZSCAN4* knockout clones, it may represent a technical artifact of the isolation stage rather than direct consequence of *ZSCAN4* deletion. Further investigation into the correlation between seeding density and differentiation kinetics in *ZSCAN4* knockout line is required to validate this hypothesis.

An alternative hypothesis for the observed phenotypic divergence is the inherent potential of off-target effects associated with the CRISPR-Cas9 system (Gaj et al., 2013; Jinek et al., 2012; Khan, 2019). Despite high specificity, the Cas9 nuclease can occasionally induce DSBs at genomic loci that possess partial sequence complementarity to the designed gRNAs (Gaj et al., 2013; Jinek et al., 2012; Khan, 2019). If these off-target sites are functionally significant genes that has role in maintaining pluripotency or survival of the cells, induction of DSBs in those sites might lead to deleterious effects. In our study, such off-target effects could provide a mechanistic explanation for the differential behavior observed in the knockout clones. Because CRISPR-mediated off-target effects can vary between individual cells, they can result in highly heterogeneous clonal populations. To definitively distinguish primary effects of ZSCAN4 deletion from secondary technical artifacts, whole-genome sequencing and transcriptome analysis by RNA sequencing are required. These approaches will enable assessment of potential off-target effects associated

with the CRISPR–Cas9 system used in this study, thereby providing a possible explanation for the differential behavior observed among knockout clones.

5.5 Conclusion

ZSCAN4 is recognized as one of the important transcription factors that has multifaceted roles in the maintenance of mouse ESCs and iPSCs. While its inclusion in reprogramming generated genomically stable mouse iPSCs, its functional significance in human iPSCs remains largely unexplored. Therefore, this study generated stable, homozygous *ZSCAN4* knockout iPSC clones and characterized them extensively to gain an in-depth knowledge on its role. The isolated homozygous clones exhibited differential behavior. Out of the four homozygous *ZSCAN4* knockout clones isolated, C2 did not survive beyond passage 3, C9 displayed differentiated morphology maintained up to at least passage 15, and C4 and C7 retained pluripotent-like morphology. Quantifying the relative expression of pluripotency markers revealed profound transcriptional collapse in the clones C2 and C9 in all the markers analyzed. The clones C4 and C7 exhibited a different expression pattern compared to control cells, where OCT4, SOX2 and REX1 are significantly upregulated in clone C4, suggesting a compensatory response. In contrast clone C7 demonstrated significant downregulation of all the three markers. The expression of NANOG was significantly downregulated in both C4 and C7 clones. The immunofluorescence of surface and nuclear markers corroborated these findings where robust expression of the markers was observed in clones C4 and C7, whereas the clone C9 showed loss of pluripotency machinery, except for TRA-1-60. Analysis of relative expression lineage-specific markers in clone C9 revealed significant expression of markers for all three lineages with predominant expression mesodermal markers. This multi-lineage signature, combined with the persistence of TRA-1-60 suggested that ZSCAN4 deficiency led to carcinoma-like transition. Interestingly, analysis of relative telomere lengths of *ZSCAN4* knockout clones revealed no significant difference. This suggests that in human iPSCs, ZSCAN4 may not be the primary determinant of telomere length or that other maintenance pathways are activated to compensate for the loss of *ZSCAN4*. Finally, to verify the stability of *ZSCAN4* knockout in these clones, it was confirmed at transcriptome level at passage 10. The results showed that the knockout of *ZSCAN4* was stable long term, ruling out the possibility of wild-type reversion as a cause of differential behavior of knockout clones. The diverse outcomes following ZSCAN4 deletion underscore the

complexity of its regulatory role. Consequently, high resolution analyses such as whole genome sequencing and RNA-sequencing are essential to definitively elucidate the mechanistic pathways through which ZSCAN4 governs pluripotency and self-renewal in iPSCs.

Chapter 6

Conclusions and Future Perspectives

6.1 Overall conclusions

The transcription factor ZSCAN4 is recognized for its critical role in maintaining genomic stability and generating quality, *bonafide* iPSCs in mouse system. Given the high degree of functional conservation between mouse and human orthologs, elucidating the role of human ZSCAN4 is pivotal for enabling the effective translation of induced pluripotent stem cells (iPSCs) into clinical therapeutics and advancing personalized medicine. Consequently, defining the regulatory mechanisms of human ZSCAN4 in iPSC homeostasis is of paramount importance. Hence, the primary focus of this thesis was to develop an cell-permeant recombinant protein and knockout toolbox for ZSCAN4 and utilize them to gain a resourceful insight on the functional role of ZSCAN4 in the maintenance of self-renewal and pluripotent property of iPSCs. To achieve this, we have systematically designed four main objectives in this thesis, which are segregated as four main chapters.

In the first chapter, we established a meticulously designed study for the heterologous expression and purification of the cell-permeant recombinant ZSCAN4 fusion protein from *E. coli* under native conditions. To supplement this molecular toolbox, we have also generated domain deleted variants of recombinant ZSCAN4 protein to serve as experimental controls. Further, in this chapter we demonstrated the influence of various factors that influence recombinant protein expression, purification and stability. The purified fusion proteins exhibited cell transduction ability with nuclear localization.

The second chapter delineates the generation of human iPSCs by reprogramming adult human dermal fibroblasts, by electroporation of Y4 episomal plasmids containing reprogramming factors. A representative clone, designated as IITGi001-B, was picked, expanded, and characterized extensively for the expression of pluripotency markers, *in vitro* trilineage differentiation capability, absence of integration of episomal plasmids, genomic stability, identity, and absence of mycoplasma.

The third chapter details the designing and validation of CRISPR-Cas9 toolbox for targeted knockout of *ZSCAN4* gene in human cells. This was achieved by systematically

investigating the influence of various critical factors, such as viral transduction, chromatin structure, gRNA potency, and the rigor of clonal selection on the efficiency of knockout of *ZSCAN4* gene. Our findings demonstrated that all the screened factors have varied effects on the knockout efficiency of *ZSCAN4* in AHDF cells. Additionally, the data suggests that clonal expansion is an essential requirement for the derivation of homozygous *ZSCAN4* knockout iPSCs. We isolated nine *ZSCAN4* knockout clones, out of which all nine clones exhibited *ZSCAN4* knockout genotype, resulting the knockout efficiency to be 100%.

The fourth and final chapter provides the extensive characterization of *ZSCAN4* knockout iPSC clones. Among the nine clones that exhibited *ZSCAN4* knockout genotype, four of the clones (C2, C4, C7, C9) that exhibited the target homozygous *ZSCAN4* knockout was chosen for further experiments. Out of those four homozygous *ZSCAN4* knockout clones chosen, one of them failed to survive beyond passage 3 (C2), the other showed differentiated morphology (C9) that was maintained at least till passage 15, and two of the clones (C4 and C7) maintained pluripotent-like morphology. The chosen homozygous clones exhibited differential behavior in terms of morphology and expression profile of pluripotency markers. Furthermore, a semi-quantitative PCR confirmed the long-term stability of ZSCAN4 ablation at the transcript level. The diverse outcomes observed following *ZSCAN4* deletion, underscore the intricate and multifaceted regulatory role this transcription factor plays in maintaining iPSC homeostasis.

6.2 Future perspectives

The recombinant protein toolbox developed in this study enables modulation of the application of ZSCAN4 in a time- and dose-dependent manner in wild-type and knockout iPSC lines, paving the way to gain deeper insights. Moreover, the non-integrative nature of recombinant proteins offers a significant advantage for clinical and basic research applications bypassing the risks associated with genomic alterations. As depicted in Fig. 6.1, this toolbox can be utilized to investigate the stage-specific role of ZSCAN4 in reprogramming.

The iPSC line, IITGi001-B, established in this study proved to be a *bonafide*, integration-free cell line, which will provide a platform for various biomedical applications like disease modelling, drug discovery and toxicity screening, functional genomic studies, personalized regenerative medicine, and understanding human developmental biology.

Further, this study designed an optimized CRISPR-Cas9 toolbox to generate *ZSCAN4* knockout cells. This toolbox can be utilized to generate *ZSCAN4*-deficient cells to study the importance of ZSCAN4 in various cellular and metabolic processes. Given that the mechanistic role of ZSCAN4 in the generation of human iPSCs remains unexplored to date, the CRISPR-Cas9 toolbox generated and optimized in this thesis can be utilized to knockout *ZSCAN4* in human fibroblasts. By subjecting these cells to reprogramming alongside wild-type fibroblasts, definitive insights can be gained into the vitality of ZSCAN4 in reprogramming as depicted in Fig. 6.1. Furthermore, the recombinant protein generated in this thesis can be utilized in rescue experiments to determine if the changes observed in ZSCAN4-deficient cells are reversible. Thereby we can study the direct functional role of the transcription factor as depicted in Fig. 6.1.

Moreover, in this study, we successfully generated *ZSCAN4*-knockout iPSCs providing a critical model for investigating its regulatory functions. Preliminary characterization of these *ZSCAN4* knockout iPSC clones revealed the complex role of ZSCAN4 in the maintenance of self-renewal and pluripotency characteristics of iPSCs. To further understand this, high resolution analyses such as whole genome sequencing and RNA-sequencing in *ZSCAN4* knockout iPSCs are essential to definitively elucidate integrity of the genome and mechanistic pathways, respectively, through which ZSCAN4 governs pluripotency and self-renewal in iPSCs.

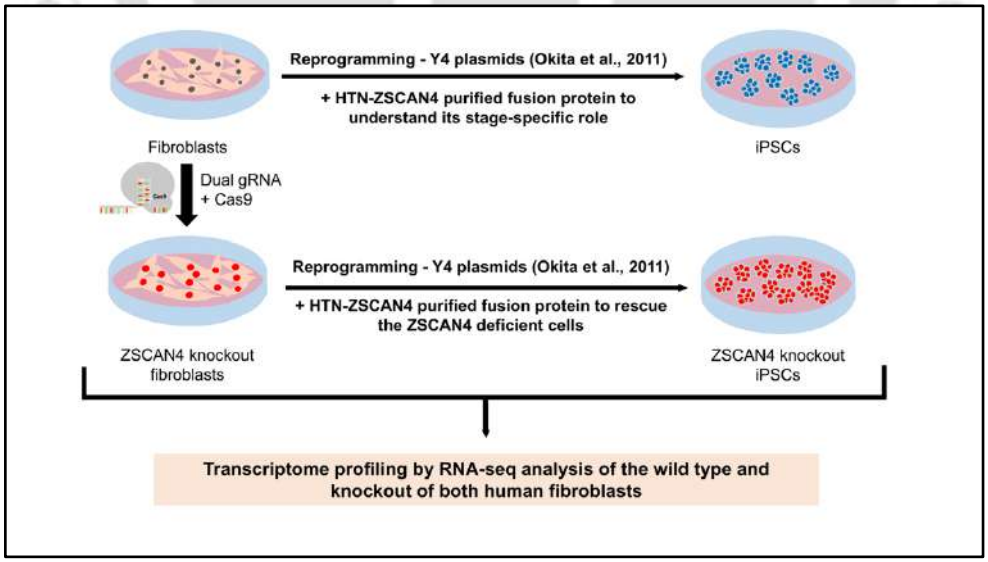


Fig. 6.1. Schematic depiction of future perspectives. The CRISPR-Cas9 toolbox can be utilized to knockout *ZSCAN4* in human fibroblasts and these fibroblasts can be reprogrammed to generate iPSCs to understand the role of ZSCAN4. The HTN-ZSCAN4 purified protein can be used in stage-specific manner to deduce its role in reprogramming and also in rescue experiments.

References

Achouri, A., Nail, V., Boye, J.I., 2012. Sesame protein isolate: Fractionation, secondary structure and functional properties. Food Research International 46, 360–369. https://doi.org/10.1016/j.foodres.2012.01.001

Aguirre, M., Escobar, M., Forero Amézquita, S., Cubillos, D., Rincón, C., Vanegas, P., Tarazona, M.P., Atuesta Escobar, S., Blanco, J.C., Celis, L.G., 2023. Application of the Yamanaka Transcription Factors Oct4, Sox2, Klf4, and c-Myc from the Laboratory to the Clinic. Genes (Basel). 14, 1697. https://doi.org/10.3390/genes14091697

Akiyama, T., Ishiguro, K.I., Chikazawa, N., Ko, S.B.H., Yukawa, M., Ko, M.S.H., 2024. ZSCAN4-binding motif - TGCACAC is conserved and enriched in CA/TG microsatellites in both mouse and human genomes. DNA Research 31. https://doi.org/10.1093/dnares/dsad029

Akiyama, T., Xin, L., Oda, M., Sharov, A.A., Amano, M., Piao, Y., Cadet, J.S., Dudekula, D.B., Qian, Y., Wang, W., Ko, S.B.H., Ko, M.S.H., 2015. Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. DNA Research 22, 307–318. https://doi.org/10.1093/dnares/dsv013

Allsopp, R., 2012. Telomere length and iPSC re-programming: Survival of the longest. Cell Res. 22, 614–615. https://doi.org/10.1038/cr.2012.6

Amano, T., Hirata, T., Falco, G., Monti, M., Sharova, L. V., Amano, M., Sheer, S., Hoang, H.G., Piao, Y., Stagg, C.A., Yamamizu, K., Akiyama, T., Ko, M.S.H., 2013. Zscan4 restores the developmental potency of embryonic stem cells. Nat. Commun. 4, 1–10. https://doi.org/10.1038/ncomms2966

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D.R., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N.J., Oinn, T.M., Pagni, M., Servant, F., Sigrist, C.J.A., Zdobnov, E.M., 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. 29, 37–40. https://doi.org/10.1093/nar/29.1.37

Balasubramanian, P., Zhao, L.-J., Chinnadurai, G., 2003. Nicotinamide adenine dinucleotide stimulates oligomerization, interaction with adenovirus E1A and an intrinsic dehydrogenase activity of CtBP. FEBS Lett. 537, 157–160. https://doi.org/10.1016/S0014-5793(03)00119-4

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science. 315, 1709–1712. https://doi.org/10.1126/science.1138140

Barrow, C.J., Yasuda, A., Kenny, P.T.M., Zagorski, M.G., 1992. Solution conformations and aggregational properties of synthetic amyloid β-peptides of Alzheimer's disease. J. Mol. Biol. 225, 1075–1093. https://doi.org/10.1016/0022-2836(92)90106-T

Ben Jehuda, R., Shemer, Y., Binah, O., 2018. Genome Editing in Induced Pluripotent Stem Cells using CRISPR/Cas9. Stem Cell Rev. Rep. 14, 323-336. https://doi.org/10.1007/s12015-018-9811-3

Berg, J.M., H-y, F., 1995. 1994, 1, 345. (10) Hanas, Curr. Opin. Struct. Biol. https://doi.org/0001-4842/95/0128-0014$09.00/0

Bhargava, N., Thakur, P., Muruganandam, T.P., Jaitly, S., Gupta, P., Lohani, N., Goswami, S.G., Saravanakumar, V., Bhattacharya, S.K., Jain, S., Ramalingam, S., 2022. Development of an efficient single-cell cloning and expansion strategy for genome edited induced pluripotent stem cells. Mol. Biol. Rep. 49, 7887–7898. https://doi.org/10.1007/s11033-022-07621-9

Bhat, E., Abdalla, M., Rather, I., 2018. Key Factors for Successful Protein Purification and Crystallization. Global Journal of Biotechnology and Biomaterial Science. 4, 001–007. https://doi.org/10.17352/gjbbs.000010

Blasco, M.A., 2005. Telomeres and human disease: ageing, cancer and beyond. Nat. Rev. Genet. 6, 611–622. https://doi.org/10.1038/nrg1656

Blum, M., Chang, H.-Y.Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E.E., Wu, C.H., Bateman, A., Finn, R.D., 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49, D344–D354. https://doi.org/10.1093/nar/gkaa977

References

Boivin, S., Kozak, S., Meijers, R., 2013. Optimization of protein purification and characterization using Thermofluor screens. Protein Expr. Purif. 91, 192-206. https://doi.org/10.1016/j.pep.2013.08.002

Bone, H.K., Damiano, T., Bartlett, S., Perry, A., Letchford, J., Ripoll, Y.S., Nelson, A.S., Welham, M.J., 2009. Involvement of GSK-3 in Regulation of Murine Embryonic Stem Cell Self-Renewal Revealed by a Series of Bisindolylmaleimides. Chem. Biol. 16, 15–27. https://doi.org/10.1016/j.chembiol.2008.11.003

Borgohain, M.P., Haridhasapavalan, K.K., Dey, C., Adhikari, P., Thummer, R.P., 2019. An Insight into DNA-free Reprogramming Approaches to Generate Integration-free Induced Pluripotent Stem Cells for Prospective Biomedical Applications. Stem Cell Rev. Rep. 15, 286–313. https://doi.org/10.1007/s12015-018-9861-6

Borgohain, M.P., Narayan, G., Haridhasapavalan, K.K., Dey, C., Thummer, R.P., 2018. Maximizing Expression and Yield of Human Recombinant Proteins from Bacterial Cell Factories for Biomedical Applications, in: Kumar, P., Patra, J.K., Chandra, P. (Eds.), Advances in Microbial Biotechnology. Apple Academic Press, New York, pp. 447–486. https://doi.org/10.1201/9781351248914

Bos, T.A., Polyakova, E., DeRuiter, M.C., Nühn, E., Jongbloed, M.R.M., 2026. Generation of three induced pluripotent stem cell clones from a functional single ventricle patient carrying the BRAF c.1897 T > C variant. Stem Cell Res. 90, 103877. https://doi.org/10.1016/j.scr.2025.103877

Briggs, R., King, T.J., 1952. Transplantation of living nuclei from blastula cells into enucleated frogs' eggs, J. Cell. & Comp. Physiol. 38, 455-463. https://doi.org/10.1073/pnas.38.5.455

Bryan, T.M., Englezou, A., Dalla-Pozza, L., Dunham, M.A., Reddel, R.R., 1997. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. Nat. Med. 3, 1271–1274. https://doi.org/10.1038/nm1197-1271

Bryan, T.M., Englezou, A., Gupta, J., Bacchetti, S., Reddel, R.R., 1995. Telomere elongation in immortal human cells without detectable telomerase activity. EMBO J. 14, 4240–4248. https://doi.org/10.1002/j.1460-2075.1995.tb00098.x

Buchan, D.W.A., Jones, D.T., 2019. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. 47, W402–W407. https://doi.org/10.1093/nar/gkz297

Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., Gileadi, O., 2008. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. Protein Expr. Purif. 59, 94–102. https://doi.org/10.1016/j.pep.2008.01.008

Cai, W.L., Greer, C.B., Chen, J.F., Arnal-Estapé, A., Cao, J., Yan, Q., Nguyen, D.X., 2020. Specific chromatin landscapes and transcription factors couple breast cancer subtype with metastatic relapse to lung or brain. BMC Med. Genomics 13. https://doi.org/10.1186/s12920-020-0695-0

Carter, M.G., Stagg, C.A., Falco, G., Yoshikawa, T., Bassey, U.C., Aiba, K., Sharova, L. V., Shaik, N., Ko, M.S.H., 2008. An in situ hybridization-based screen for heterogeneously expressed genes in mouse ES cells. Gene Expression Patterns 8, 181–198. https://doi.org/10.1016/j.gep.2007.10.009

Cawthon, R.M., 2009. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic Acids Res. 37, 1–7. https://doi.org/10.1093/nar/gkn1027

Cesare, A.J., Reddel, R.R., 2010. Alternative lengthening of telomeres: models, mechanisms and implications. Nat. Rev. Genet. 11, 319–330. https://doi.org/10.1038/nrg2763

Chehelgerdi, Mohammad, Chehelgerdi, Matin, Khorramian-Ghahfarokhi, M., Shafieizadeh, M., Mahmoudi, E., Eskandari, F., Rashidi, M., Arshi, A., Mokhtari-Farsani, A., 2024. Comprehensive review of CRISPR-based gene editing: mechanisms, challenges, and applications in cancer therapy. Mol. Cancer. 23, 9. https://doi.org/10.1186/s12943-023-01925-5

Chen, L., Khillan, J.S., 2010. A Novel Signaling by Vitamin A/Retinol Promotes Self Renewal of Mouse Embryonic Stem Cells by Activating PI3K/Akt Signaling Pathway via Insulin-Like Growth Factor-1 Receptor . Stem Cells 28, 57–63. https://doi.org/10.1002/stem.251

Chen, L., Khillan, J.S., 2008. Promotion of Feeder-Independent Self-Renewal of Embryonic Stem Cells by Retinol (Vitamin A). Stem Cells 26, 1858–1864. https://doi.org/10.1634/stemcells.2008-0050

Chen, Y.H., Pruett-Miller, S.M., 2018. Improving single-cell cloning workflow for gene editing in human pluripotent stem cells. Stem Cell Res. 31, 186–192. https://doi.org/10.1016/j.scr.2018.08.003

Cheng, Z.L., Zhang, M.L., Lin, H.P., Gao, C., Song, J. Bin, Zheng, Z., Li, L., Zhang, Y., Shen, X., Zhang, H., Huang, Z., Zhan, W., Zhang, C., Hu, X., Sun, Y.P., Jiang, L., Sun, L., Xu, Y., Yang, C., Ge, Y., Zhao, Y., Liu, X., Yang, H., Liu, P., Guo, X., Guan, K.L., Xiong, Y., Zhang, M., Ye, D., 2020. The Zscan4-Tet2 Transcription Nexus Regulates Metabolic Rewiring and Enhances Proteostasis to Promote Reprogramming. Cell Rep. 32. https://doi.org/10.1016/j.celrep.2020.107877

Choi, D.Y., Leem, J., Lee, C., Oh, J.S., 2025. ZSCAN4 functions as a safeguard to maintain centromere integrity during oocyte meiosis. Genome Biol. 26. https://doi.org/10.1186/s13059-025-03687-3

Choi, S.M., Ma, C.Y., 2005. Conformational study of globulin from common buckwheat (Fagopyrum esculentum Moench) by fourier transform infrared spectroscopy and differential scanning calorimetry. J. Agric. Food Chem. 53, 8046–8053. https://doi.org/10.1021/jf051040v

Christman, J.K., 2002. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21, 5483–5495. https://doi.org/10.1038/sj.onc

Cowan, C.A., Atienza, J., Melton, D.A., Eggan, K., 2005. Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem cells. Science. 309, 1369–1373. https://doi.org/10.1126/science.111644

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: A sequence logo generator. Genome Res. 14, 1188–1190. https://doi.org/10.1101/gr.849004

Damodaran, S., Kinsella, J.E., 1982. Effects of Ions on Protein Conformation and Functionality. pp. 327–357. https://doi.org/10.1021/bk-1982-0206.ch013

Dan, J., Li, M., Yang, J., Li, J., Okuka, M., Ye, X., Liu, L., 2013. Roles for Tbx3 in regulation of two-cell state and telomere elongation in mouse ES cells. Sci. Rep. 3, 1–9. https://doi.org/10.1038/srep03492

Dan, J., Liu, Y., Liu, N., Chiourea, M., Okuka, M., Wu, T., Ye, X., Mou, C., Wang, Lei, Wang, Lingling, Yin, Y., Yuan, J., Zuo, B., Wang, F., Li, Z., Pan, X., Yin, Z., Chen, L., Keefe, D.L., Gagos, S., Xiao, A., Liu, L., 2014. Rif1 Maintains Telomere Length Homeostasis of ESCs by Mediating Heterochromatin Silencing. Dev. Cell 29, 7–19. https://doi.org/10.1016/j.devcel.2014.03.004

Dan, J., Rousseau, P., Hardikar, S., Veland, N., Wong, J., Autexier, C., Chen, T., 2017. Zscan4 Inhibits Maintenance DNA Methylation to Facilitate Telomere Elongation in Mouse Embryonic Stem Cells. Cell Rep. 20, 1936–1949. https://doi.org/10.1016/j.celrep.2017.07.070

Dan, J., Zhou, Z., Wang, F., Wang, H., Guo, R., Keefe, D.L., Liu, L., 2022. Zscan4 Contributes to Telomere Maintenance in Telomerase-Deficient Late Generation Mouse ESCs and Human ALT Cancer Cells. Cells 11, 456. https://doi.org/10.3390/cells11030456

De Lange, T., 2009. How Telomeres Solve the End-Protection Problem. Science. 326, 948–952. https://doi.org/10.1126/science.1170633

Denning, W., Das, S., Guo, S., Xu, J., Kappes, J.C., Hel, Z., 2013. Optimization of the transductional efficiency of lentiviral vectors: Effect of sera and polycations. Mol. Biotechnol. 53, 308–314. https://doi.org/10.1007/s12033-012-9528-5

DeWeirdt, P.C., Sangree, A.K., Hanna, R.E., Sanson, K.R., Hegde, M., Strand, C., Persky, N.S., Doench, J.G., 2020. Genetic screens in isogenic mammalian cell lines without single cell cloning. Nat. Commun. 11, 752. https://doi.org/10.1038/s41467-020-14620-6

Dey, C., Raina, K., Haridhasapavalan, K.K., Thool, M., Sundaravadivelu, P.K., Adhikari, P., Gogoi, R., Thummer, R.P., 2021a. An overview of reprogramming approaches to derive integration-free induced pluripotent stem cells for prospective biomedical applications, in: Recent Advances in IPSC Technology. Elsevier, pp. 231–287. https://doi.org/10.1016/B978-0-12-822231-7.00011-4

Dey, C., Raina, K., Thummer, R.P., 2023. Production of a Bioactive Recombinant Human UTF1 Protein from E. coli, in: Healthcare Research and Related Technologies. Springer Nature Singapore, Singapore, pp. 63–83. https://doi.org/10.1007/978-981-99-4056-1_4

Dey, C., Thool, M., Bhattacharyya, S., Sudhagar, S., Thummer, R.P., 2021b. Generation of biologically active recombinant human OCT4 protein from *E. coli*. 3 Biotech 11, 1–16. https://doi.org/10.1007/s13205-021-02758-z

Dey, C., Venkatesan, V., Thummer, R.P., 2022. Identification of Optimal Expression Parameters and Purification of a Codon-Optimized Human GLIS1 Transcription Factor from *Escherichia coli*. Mol. Biotechnol. 64, 42–56. https://doi.org/10.1007/s12033-021-00390-z

Dinsdale, C., Mirza, F., Wiebe, J., 1992. Glycerol alters cytoskeleton and cell adhesion while inhibiting cell proliferation. Cell Biol. Int. Rep. 16, 591–602. https://doi.org/10.1016/S0309-1651(06)80001-9

Doudna, J.A., Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. Science. 346. https://doi.org/10.1126/science.1258096

Durand, S., Cimarelli, A., 2011. The Inside Out of Lentiviral Vectors. Viruses 3, 132–159. https://doi.org/10.3390/v3020132

Eckersley-Maslin, M.A., Svensson, V., Krueger, C., Stubbs, T.M., Giehr, P., Krueger, F., Miragaia, R.J., Kyriakopoulos, C., Berrens, R. V., Milagre, I., Walter, J., Teichmann, S.A., Reik, W., 2016. MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. Cell Rep. 17, 179–192. https://doi.org/10.1016/j.celrep.2016.08.087

Edelstein, L.C., Collins, T., 2005. The SCAN domain family of zinc finger transcription factors. Gene 359, 1–17. https://doi.org/10.1016/j.gene.2005.06.022

Falco, G., Lee, S.L., Stanghellini, I., Bassey, U.C., Hamatani, T., Ko, M.S.H., 2007. Zscan4: A novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. Dev. Biol. 307, 539–550. https://doi.org/10.1016/j.ydbio.2007.05.003

Feng, S., Yan, Y., 2008. Effects of glycerol on the compaction and stability of the wild type and mutated rabbit muscle creatine kinase. Proteins: Structure, Function, and Bioinformatics 71, 844–854. https://doi.org/10.1002/prot.21744

Ferreboeuf, M., Mariot, V., Bessières, B., Vasiljevic, A., Attié-Bitach, T., Collardeau, S., Morere, J., Roche, S., Magdinier, F., Robin-Ducellier, J., Rameau, P., Whalen, S., Desnuelle, C., Sacconi, S., Mouly, V., Butler-Browne, G., Dumonceaux, J., 2014. DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. Hum. Mol. Genet. 23, 171–181. https://doi.org/10.1093/hmg/ddt409

Fesenko, I., Franklin, D., Garnett, P., Bass, P., Campbell, S., Hardyman, M., Wilson, D., Hanley, N., Collins, J., 2010. Stem cell marker TRA-1-60 is expressed in foetal and adult kidney and upregulated in tubulo-interstitial disease. Histochem. Cell Biol. 134, 355–369. https://doi.org/10.1007/s00418-010-0741-7

Fischer, B., Schmidt, V., Ly, T.D., Kleine, A., Knabbe, C., Faust-Hinse, I., 2022. First Characterization of Human Dermal Fibroblasts Showing a Decreased Xylosyltransferase-I Expression Induced by the CRISPR/Cas9 System. Int. J. Mol. Sci. 23. https://doi.org/10.3390/ijms23095045

Frankel, A.D., Berg, J.M., Pabo, C.O., 1987. Metal-dependent folding of a single zinc finger from transcription factor IIIA. Proceedings of the National Academy of Sciences 84, 4841–4845. https://doi.org/10.1073/pnas.84.14.4841

Fu, H., Tian, C. lei, Ye, X., Sheng, X., Wang, H., Liu, Y., Liu, L., 2018. Dynamics of Telomere Rejuvenation during Chemical Induction to Pluripotent Stem Cells. Stem Cell Reports 11, 70–87. https://doi.org/10.1016/j.stemcr.2018.05.003

Gaj, T., Gersbach, C.A., Barbas, C.F., 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31, 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004

Galloway, C.A., Sowden, M.P., Smith, H.C., 2003. Increasing the yield of soluble recombinant protein expressed in E. coli by Induction during late log phase. Biotechniques 34, 524–530. https://doi.org/10.2144/03343st04

García-Fraga, B., da Silva, A.F., López-Seijas, J., Sieiro, C., 2015. Optimized expression conditions for enhancing production of two recombinant chitinolytic enzymes from different prokaryote domains. Bioprocess Biosyst. Eng. 38, 2477–2486. https://doi.org/10.1007/s00449-015-1485-5

Garnier, J., Gibrat, J.-F., Robson, B., 1996. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol. 266, 540–553. https://doi.org/10.1016/S0076-6879(96)66034-0

Garza, A.S., Ahmad, N., Kumar, R., 2009. Role of intrinsically disordered protein regions/domains in transcriptional regulation. Life Sci. 84, 189–193. https://doi.org/10.1016/j.lfs.2008.12.002

Gaspar-Maia, A., Alajem, A., Meshorer, E., Ramalho-Santos, M., 2011. Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. https://doi.org/10.1038/nrm3036

Geng, B. chuan, Choi, K.H., Wang, S. zhi, Chen, P., Pan, X. di, Dong, N. guo, Ko, J.K., Zhu, H., 2020. A simple, quick, and efficient CRISPR/Cas9 genome editing method for human induced pluripotent stem cells. Acta Pharmacol. Sin. 41, 1427–1432. https://doi.org/10.1038/s41401-020-0452-0

Geourjon, C., Deléage, G., 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics 11, 681–684. https://doi.org/10.1093/bioinformatics/11.6.681

Ghahghaei, A., Bathaie, S., Shahraki, A., Asgarabad, F.R., 2011. Comparison of the chaperoning action of glycerol and β-casein on aggregation of proteins in the presence of crowding Zahra agent. Int. J. Pept. Res. Ther. 17, 101–111. https://doi.org/10.1007/s10989-011-9247-y

Greenfield, N.J., 2007. Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. 1, 2876–2890. https://doi.org/10.1038/nprot.2006.202

Gu, X.N., Li, N., Zheng, Y.F., Ruan, L., 2011. In vitro degradation performance and biological response of a Mg-Zn-Zr alloy, in: Materials Science and Engineering, pp. 1778–1784. https://doi.org/10.1016/j.mseb.2011.05.032

Gu, X.N., Zheng, Y.F., Chen, L.J., 2009. Influence of artificial biological fluid composition on the biocorrosion of potential orthopedic Mg-Ca, AZ31, AZ91 alloys. Biomedical Materials 4. https://doi.org/10.1088/1748-6041/4/6/065011

Guo, R., Ye, X., Yang, Jiao, Zhou, Z., Tian, C., Wang, Hua, Wang, Haiying, Fu, H., Liu, C., Zeng, M., Yang, Jun, Liu, L., 2018. Feeders facilitate telomere maintenance and chromosomal stability of embryonic stem cells. Nat. Commun. 9. https://doi.org/10.1038/s41467-018-05038-2

Gurdon, J.B., 1962. The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. 10, 622-640. https://doi.org/10.1242/dev.10.4.622

Hamatani, T., Carter, M.G., Sharov, A.A., Ko, M.S.H., 2004. Dynamics of Global Gene Expression Changes during Mouse Preimplantation Development. Dev. Cell 6, 117–131. https://doi.org/10.1016/S1534-5807(03)00373-3

Haridhasapavalan, K.K., Ranjan, S.H., Bhattacharyya, S., Thummer, R.P., 2021a. Soluble expression, purification, and secondary structure determination of human MESP1 transcription factor. Appl. Microbiol. Biotechnol. 105, 2363–2376. https://doi.org/10.1007/s00253-021-11194-1

Haridhasapavalan, K.K., Sundaravadivelu, P.K., Bhattacharyya, S., Ranjan, S.H., Raina, K., Thummer, R.P., 2021b. Generation of cell-permeant recombinant human transcription factor GATA4 from *E. coli*. Bioprocess Biosyst. Eng. 44, 1131–1146. https://doi.org/10.1007/s00449-021-02516-8

Haridhasapavalan, K.K., Sundaravadivelu, P.K., Joshi, N., Das, N.J., Mohapatra, A., Voorkara, U., Kaveeshwar, V., Thummer, R.P., 2022. Generation of a recombinant version of a biologically active cell-permeant human HAND2 transcription factor from *E. coli*. Sci. Rep. 12, 16129. https://doi.org/10.1038/s41598-022-19745-w

Haridhasapavalan, K.K., Sundaravadivelu, P.K., Voorkara, U., Kaveeshwar, V., Thummer, R.P., 2023. Generation of Recombinant Version of a Bioactive Human MEF2C Transcription Factor from *E. coli*, in: Healthcare Research and Related Technologies. Springer Nature Singapore, Singapore, pp. 3–27. https://doi.org/10.1007/978-981-99-4056-1_1

Haridhasapavalan, K.K., Borgohain, M.P., Dey, C., Saha, B., Narayan, G., Kumar, S., Thummer, R.P., 2019. An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells. Gene 686, 146–159. https://doi.org/10.1016/j.gene.2018.11.069

Harper, T.C., Oberlick, E.M., Smith, T.J., Nunes, D.E., Bray, M.A., Park, S., Driscoll, C.D., Mowbray, S.F., Antczak, C., 2023. GATA1 deletion in human pluripotent stem cells increases differentiation yield and maturity of neutrophils. iScience 26. https://doi.org/10.1016/j.isci.2023.107804

Harris, A.R., Panigrahi, G., Liu, H., Koparde, V.N., Bailey-Whyte, M., Dorsey, T.H., Yates, C.C., Ambs, S., 2023. Chromatin Accessibility Landscape of Human Triple-negative Breast Cancer Cell Lines Reveals Variation by Patient Donor Ancestry. Cancer Research Communications 3, 2014–2029. https://doi.org/10.1158/2767-9764.CRC-23-0236

Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.-W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., Nix, D.A., Peterson, C.M., Tapscott, S.J., Carrell, D.T., Cairns, B.R., 2017. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet. 49, 925–934. https://doi.org/10.1038/ng.3844

Hicks, M.R., Pyle, A.D., 2023. The emergence of the stem cell niche. Trends Cell Biol. 33, 112-123. https://doi.org/10.1016/j.tcb.2022.07.003

Hinz, J.M., Laughery, M.F., Wyrick, J.J., 2015. Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. Biochemistry 54, 7063–7066. https://doi.org/10.1021/acs.biochem.5b01108

Hirata, T., Amano, T., Nakatake, Y., Amano, M., Piao, Y., Hoang, H.G., Ko, M.S.H., 2012. Zscan4 transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells. Sci. Rep. 2, 208. https://doi.org/10.1038/srep00208

Hiyama, E., Hiyama, K., 2007. Telomere and telomerase in stem cells. Br. J. Cancer 96, 1020–1024. https://doi.org/10.1038/sj.bjc.6603671

Hung, S.S.C., Wong, R.C.B., Sharov, A.A., Nakatake, Y., Yu, H., Ko, M.S.H., 2013. Repression of global protein synthesis by Eif1a-like genes that are expressed specifically in the two-cell embryos and the transient Zscan4-positive state of embryonic stem cells. DNA Research 20, 391–401. https://doi.org/10.1093/dnares/dst018

Ishiguro, K. ichiro, Monti, M., Akiyama, T., Kimura, H., Chikazawa-Nohtomi, N., Sakota, M., Sato, S., Redi, C.A., Ko, S.B.H., Ko, M.S.H., 2017a. Zscan4 is expressed specifically during late meiotic prophase in both spermatogenesis and oogenesis. In Vitro Cell. Dev. Biol. Anim. 53, 167–178. https://doi.org/10.1007/s11626-016-0096-z

Ishiguro, K. ichiro, Nakatake, Y., Chikazawa-Nohtomi, N., Kimura, H., Akiyama, T., Oda, M., Ko, S.B.H., Ko, M.S.H., 2017b. Expression analysis of the endogenous Zscan4 locus and its coding proteins in mouse ES cells and preimplantation embryos. In Vitro Cell. Dev. Biol. Anim. 53, 179–190. https://doi.org/10.1007/s11626-016-0097-y

Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J.M., Torres-Padilla, M.-E., 2015. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat. Struct. Mol. Biol. 22, 662–671. https://doi.org/10.1038/nsmb.3066

Jiang, J., Lv, W., Ye, X., Wang, L., Zhang, M., Yang, H., Okuka, M., Zhou, C., Zhang, X., Liu, L., Li, J., 2013. Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPS cells as demonstrated by tetraploid complementation. Cell Res. 23, 92–106. https://doi.org/10.1038/cr.2012.157

References

Jiang, W., Chen, C., Zhang, H., Zhu, S., 2026. Establishment and characterization of a human induced pluripotent stem cell line from a child with epilepsy. Stem Cell Res. 90, 103900. https://doi.org/10.1016/j.scr.2025.103900

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 337, 816–821. https://doi.org/10.1126/science.1225829

Joglekar, M. V., Satoor, S.N., Wong, W.K.M., Cheng, F., Ma, R.C.W., Hardikar, A.A., 2020. An optimised step-by-step protocol for measuring relative telomere length. Methods Protoc. 3, 1–12. https://doi.org/10.3390/mps3020027

Johnson, I., Jiang, W., Liu, H., 2017. The Effects of Serum Proteins on Magnesium Alloy Degradation in Vitro. Sci. Rep. 7. https://doi.org/10.1038/s41598-017-14479-6

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202. https://doi.org/10.1006/jmbi.1999.3091

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Keeney, S., Giroux, C.N., Kleckner, N., 1997. Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family. Cell 88, 375–384. https://doi.org/10.1016/S0092-8674(00)81876-0

Khan, S.H., 2019. Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. Mol. Ther. Nucleic Acids 16, 326–334. https://doi.org/10.1016/j.omtn.2019.02.027

Khow, O., Suntrarachun, S., 2012. Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pac. J. Trop. Biomed. 2, 159–162. https://doi.org/10.1016/S2221-1691(11)60213-X

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W., Kim, J.-S., 2009. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res. 19, 1279–1288. https://doi.org/10.1101/gr.089417.108

Knaupp, A.S., Buckberry, S., Pflueger, J., Lim, S.M., Ford, E., Larcombe, M.R., Rossello, F.J., de Mendoza, A., Alaei, S., Firas, J., Holmes, M.L., Nair, S.S., Clark, S.J., Nefzger, C.M., Lister, R., Polo, J.M., 2017. Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. Cell Stem Cell 21, 834-845.e6. https://doi.org/10.1016/j.stem.2017.11.007

Ko, M.S., Kitchen, J.R., Wang, X., Threat, T.A., Wang, X., Hasegawa, A., Sun, T., Grahovac, M.J., Kargul, G.J., Lim, M.K., Cui, Y., Sano, Y., Tanaka, T., Liang, Y., Mason, S., Paonessa, P.D., Sauls, A.D., DePalma, G.E., Sharara, R., Rowe, L.B., Eppig, J., Morrell, C., Doi, H., 2000. Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. Development 127, 1737–1749. https://doi.org/10.1242/dev.127.8.1737

Ko, M.S.H., 2016. Zygotic Genome Activation Revisited: Looking Through the Expression and Function of Zscan4. Curr. Top. Dev. Biol. 120, 103–124. https://doi.org/10.1016/bs.ctdb.2016.04.004

Ko, S.B.H., Azuma, S., Yokoyama, Y., Yamamoto, A., Kyokane, K., Niida, S., Ishiguro, H., Ko, M.S.H., 2013. Inflammation increases cells expressing ZSCAN4 and progenitor cell markers in the adult pancreas. Am. J. Physiol. Gastrointest. Liver Physiol. 304, G1103–G1116. https://doi.org/10.1152/ajpgi.00299.2012

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Kumpfmueller, B., 2011. Control of embryonic stem cell fate : the role of phosphoinositide 3-kinase signalling and Zscan4.

Kwon, Y.W., Paek, J.S., Cho, H.J., Lee, C.S., Lee, H.J., Park, I.H., Roh, T.Y., Kang, C.M., Yang, H.M., Park, Y.B., Kim, H.S., 2015. Role of Zscan4 in secondary murine iPSC derivation mediated by protein extracts of ESC or iPSC. Biomaterials 59, 102–115. https://doi.org/10.1016/j.biomaterials.2015.03.031

Le, R., Huang, Y., Zhang, Y., Wang, H.H., Lin, J., Dong, Y., Li, Z., Guo, M., Kou, X., Zhao, Y., Chen, M., Zhu, Q., Zhao, A., Yin, J., Sun, J., Su, Z., Shi, K., Gao, Y., Chen, J., Liu, W., Kang, L., Wang, Y., Li, C., Liu, X., Gao, R., Wang, H.H., Ju, Z., Gao, S., 2021. Dcaf11 activates Zscan4-mediated alternative telomere lengthening in early embryos and embryonic stem cells. Cell Stem Cell 28, 732-747.e9. https://doi.org/10.1016/j.stem.2020.11.018

Lee, K., Gollahon, L.S., 2015. ZSCAN4 and TRF1: A functionally indirect interaction in cancer cells independent of telomerase activity. Biochem. Biophys. Res. Commun. 466, 644–649. https://doi.org/10.1016/j.bbrc.2015.09.107

Lee, K., Gollahon, L.S., 2014. Zscan4 interacts directly with human Rap1 in cancer cells regardless of telomerase status. Cancer Biol. Ther. 15, 1094–1105. https://doi.org/10.4161/cbt.29220

Li, D., Shu, X., Zhu, P., Pei, D., 2021. Chromatin accessibility dynamics during cell fate reprogramming. EMBO Rep. 22. https://doi.org/10.15252/embr.202051644

Li, K., Ouyang, M., Zhan, J., Tian, R., 2023. CRISPR-based functional genomics screening in human-pluripotent-stem-cell-derived cell types. Cell Genomics. 3. https://doi.org/10.1016/j.xgen.2023.100300

Li, P., Schille, C., Schweizer, E., Kimmerle-Müller, E., Rupp, F., Heiss, A., Legner, C., Klotz, U.E., Geis-Gerstorfer, J., Scheideler, L., 2019. Selection of extraction medium influences cytotoxicity of zinc and its alloys. Acta Biomater. 98, 235–245. https://doi.org/10.1016/j.actbio.2019.03.013

Li, X.H., Sun, M.H., Jiang, W.J., Zhou, D., Lee, S.H., Heo, G., Chen, Z., Cui, X.S., 2023. ZSCAN4 Regulates Zygotic Genome Activation and Telomere Elongation in Porcine Parthenogenetic Embryos. Int. J. Mol. Sci. 24. https://doi.org/10.3390/ijms241512121

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., Cech, T.R., 1997. Reverse Transcriptase Motifs in the Catalytic Subunit of Telomerase. Science. 276, 561–567. https://doi.org/10.1126/science.276.5312.561

Liu, X., Yang, H., Liu, Y., Xiong, P., Guo, H., Huang, H.-H., Zheng, Y., 2019. Comparative Studies on Degradation Behavior of Pure Zinc in Various Simulated Body Fluids. JOM 71, 1414–1425. https://doi.org/10.1007/s11837-019-03357-3

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., Zhang, Y., 2014. Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev. 28, 2103–2119. https://doi.org/10.1101/gad.248005.114

Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B., Pfaff, S.L., 2011. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev. 25, 594–607. https://doi.org/10.1101/gad.2008511

Maekawa, M., Yamaguchi, K., Nakamura, T., Shibukawa, R., Kodanaka, I., Ichisaka, T., Kawamura, Y., Mochizuki, H., Goshima, N., Yamanaka, S., 2011. Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. Nature 474, 225–228. https://doi.org/10.1038/nature10106

Maertens, B., Spriestersbach, A., Von Groll, U., Roth, U., Kubicek, J., Gerrits, M., Graf, M., Liss, M., Daubert, D., Wagner, R., Schäfer, F., 2010. Gene optimization mechanisms: A multi-gene study reveals a high success rate of full-length human proteins expressed in *Escherichia coli*. Protein Science 19, 1312–1326. https://doi.org/10.1002/pro.408

Mahadevaiah, S.K., Turner, J.M.A., Baudat, F., Rogakou, E.P., de Boer, P., Blanco-Rodríguez, J., Jasin, M., Keeney, S., Bonner, W.M., Burgoyne, P.S., 2001. Recombinational DNA double-strand breaks in mice precede synapsis. Nat. Genet. 27, 271–276. https://doi.org/10.1038/85830

Malecki, M., Anderson, M., Beauchaine, M., Seo, S., Tombokan, X., Malecki, R., 2012. TRA-1-60+, SSEA-4+, Oct4A+, Nanog+ Clones of Pluripotent Stem Cells in the Embryonal Carcinomas of the Ovaries. J. Stem Cell Res. Ther. 02, 130. https://doi.org/10.4172/2157-7633.1000130

Manian, K. V., Bharathan, S.P., Maddali, M., Srivastava, V.M., Srivastava, A., Velayudhan, S.R., 2018. Generation of an integration-free iPSC line (CSCRi005-A) from erythroid progenitor cells of a healthy Indian male individual. Stem Cell Res. 29, 148–151. https://doi.org/10.1016/j.scr.2018.03.020

Meng, X., Neises, A., Su, R.J., Payne, K.J., Ritter, L., Gridley, D.S., Wang, J., Sheng, M., Lau, K.H.W., Baylink, D.J., Zhang, X.B., 2012. Efficient reprogramming of human cord blood CD34+ cells into induced pluripotent stem cells with OCT4 and SOX2 alone. Molecular Therapy 20, 408–416. https://doi.org/10.1038/mt.2011.258

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., Gregory, P.D., Pabo, C.O., Rebar, E.J., 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat. Biotechnol. 25, 778–785. https://doi.org/10.1038/nbt1319

Nandakumar, J., Cech, T.R., 2013. Finding the end: recruitment of telomerase to telomeres. Nat. Rev. Mol. Cell Biol. 14, 69–82. https://doi.org/10.1038/nrm3505

Niwa, H., 2007. How is pluripotency determined and maintained? Development 134, 635–646. https://doi.org/10.1242/dev.02787

Ochiai, H., Yamamoto, T., 2017. Construction and Evaluation of Zinc Finger Nucleases. pp. 1–24. https://doi.org/10.1007/978-1-4939-7128-2_1

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K.I., Shibata, T., Kunisada, T., Takahashi, M., Takahashi, J., Saji, H., Yamanaka, S., 2011.

A more efficient method to generate integration-free human iPS cells. Nat. Methods 8, 409–412. https://doi.org/10.1038/nmeth.1591

Oliva, J., Galasinski, S., Richey, A., Campbell, A.E., Meyers, M.J., Modi, N., Zhong, J.W., Tawil, R., Tapscott, S.J., Sverdrup, F.M., 2019. Clinically advanced p38 inhibitors suppress DUX4 expression in cellular and animal models of facioscapulohumeral muscular dystrophys. Journal of Pharmacology and Experimental Therapeutics 370, 219–230. https://doi.org/10.1124/jpet.119.259663

O'Malley, J., Woltjen, K., Kaji, K., 2009. New strategies to generate induced pluripotent stem cells. Curr. Opin. Biotechnol. 20, 516–521. https://doi.org/10.1016/j.copbio.2009.09.005

Ou, J., Wang, L., Ding, X., Du, J., Zhang, Y., Chen, H., Xu, A., 2004. Stationary phase protein overproduction is a fundamental capability of *Escherichia coli*. Biochem. Biophys. Res. Commun. 314, 174–180. https://doi.org/10.1016/j.bbrc.2003.12.077

Paduano, V., Tagliaferri, D., Falco, G., Ceccarelli, M., 2013. Automated identification and location analysis of marked stem cells colonies in optical microscopy images. PLoS One 8, 1–11. https://doi.org/10.1371/journal.pone.0080776

Paling, N.R.D., Wheadon, H., Bone, H.K., Welham, M.J., 2004. Regulation of Embryonic Stem Cell Self-renewal by Phosphoinositide 3-Kinase-dependent Signaling. Journal of Biological Chemistry 279, 48063–48070. https://doi.org/10.1074/jbc.M406467200

Palm, W., De Lange, T., 2008. How shelterin protects mammalian telomeres. Annu. Rev. Genet. 42, 301–334. https://doi.org/10.1146/annurev.genet.41.110306.130350

Panopoulos, A.D., Yanes, O., Ruiz, S., Kida, Y.S., Diep, D., Tautenhahn, R., Herrerías, A., Batchelder, E.M., Plongthongkum, N., Lutz, M., Berggren, W.T., Zhang, K., Evans, R.M., Siuzdak, G., Belmonte, J.C.I., 2012. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. Cell Res. 22, 168–177. https://doi.org/10.1038/cr.2011.177

Papadopoulos, J.S., Agarwala, R., 2007. COBALT: Constraint-based alignment tool for multiple protein sequences. Bioinformatics 23, 1073–1079. https://doi.org/10.1093/bioinformatics/btm076

Park, H.S., Hwang, I., Choi, K.A., Jeong, H., Lee, J.Y., Hong, S., 2015. Generation of induced pluripotent stem cells without genetic defects by small molecules. Biomaterials 39, 47–58. https://doi.org/10.1016/j.biomaterials.2014.10.055

Patra, S.K., Bettuzzi, S., 2009. Epigenetic DNA-(cytosine-5-carbon) modifications: 5-aza-2′- deoxycytidine and DNA-demethylation. Biochemistry (Moscow) 74, 613–619. https://doi.org/10.1134/S0006297909060042

Pavletich, N.P., Pabo, C.O., 1991. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. Science. 252, 809–817. https://doi.org/10.1126/science.2028256

Pengue, G., Calabró, V., Bartoli, P.C., Pagliuca, A., Lania, L., 1994. Repression of transcriptional activity at a distance by the evolutionary conserved KRAB domain present in a subfamily of zinc finger proteins. Nucleic Acids Res. 22, 2908–2914. https://doi.org/10.1093/nar/22.15.2908

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V, Lee, Y.-L., Guschin, D.Y., Rupniewski, I., Waite, A.J., Carpenito, C., Carroll, R.G., S Orange, J., Urnov, F.D., Rebar, E.J., Ando, D., Gregory, P.D., Riley, J.L., Holmes, M.C., June, C.H., 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat. Biotechnol. 26, 808–816. https://doi.org/10.1038/nbt1410

Portney, B.A., Arad, M., Gupta, A., Brown, R.A., Khatri, R., Lin, P.N., Hebert, A.M., Angster, K.H., Silipino, L.E., Meltzer, W.A., Taylor, R.J., Zalzman, M., 2020. ZSCAN4 facilitates chromatin remodeling and promotes the cancer stem cell phenotype. Oncogene 39, 4970–4982. https://doi.org/10.1038/s41388-020-1333-1

Portney, B.A., Khatri, R., Meltzer, W.A., Mariano, J.M., Zalzman, M., 2018. ZSCAN4 is negatively regulated by the ubiquitin-proteasome system and the E3 ubiquitin ligase RNF20. Biochem. Biophys. Res. Commun. 498, 72–78. https://doi.org/10.1016/j.bbrc.2018.02.155

Qi, R., Guo, J., Liu, Y., Zhang, R., Gan, Z., 2019. Effects of salt content on secondary structure of protein in sodium alginate/antarctic krill protein composite system and characterization of fiber properties. Dyes and Pigments 171. https://doi.org/10.1016/j.dyepig.2019.107686

Raab, S., Klingenstein, M., Liebau, S., Linta, L., 2014. A Comparative View on Human Somatic Cell Sources for iPSC Generation. Stem Cells Int. 2014. https://doi.org/10.1155/2014/768391

Raina, K., Joshi, G., Modak, K., Premkumar, C., Priyanka, S., Rajesh, P., Velayudhan, S.R., Thummer, R.P., 2023. Generation and characterization of induced pluripotent stem cell line IITGi001-A derived from adult human primary dermal fibroblasts. Stem Cell Res. 71. https://doi.org/10.1016/j.scr.2023.103159

Raina, K., Modak, K., Premkumar, C., Joshi, G., Palani, D., Nandy, K., Sivamani, Y., Velayudhan, S.R., Thummer, R.P., 2025. UTF1 Expression is Important for the Generation and Maintenance of Human iPSCs. Stem Cell Rev. Rep. 21, 859–871. https://doi.org/10.1007/s12015-024-10836-x

Ray, A., Joshi, J.M., Sundaravadivelu, P.K., Raina, K., Lenka, N., Kaveeshwar, V., Thummer, R.P., 2021. An Overview on Promising Somatic Cell Sources Utilized for the Efficient Generation of Induced Pluripotent Stem Cells. Stem Cell Rev. Rep. 17, 1954-1974. https://doi.org/10.1007/s12015-021-10200-3

Reinholdt, L.G., Schimenti, J.C., 2005. Mei1 is epistatic to Dmc1 during mouse meiosis. Chromosoma 114, 127–134. https://doi.org/10.1007/s00412-005-0346-4

San-Miguel, T., Pérez-Bermúdez, P., Gavidia, I., 2013. Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. Springerplus 3, 1–4. https://doi.org/10.1186/2193-1801-2-89

Schäfer, C., Ju, Y., Tak, Y., Vazquez, C., Han, S.J., Tan, E., Shay, J.W., Holmqvist, M., Danuser, G., Schopperle, W.M., Bubley, G., 2020. TRA-1-60-positive/CD45low cells found in the peripheral blood of prostate cancer patients with metastatic disease – A proof-of-concept study. Heliyon 6. https://doi.org/10.1016/j.heliyon.2020.e03263

Scheideler, L., Füger, C., Schille, C., Rupp, F., Wendel, H.P., Hort, N., Reichel, H.P., Geis-Gerstorfer, J., 2013. Comparison of different in vitro tests for biocompatibility screening of Mg alloys. Acta Biomater. 9, 8740–8745. https://doi.org/10.1016/j.actbio.2013.02.020

Schmitt, C.E., Morales, B.M., Schmitz, E.M.H., Hawkins, J.S., Lizama, C.O., Zape, J.P., Hsiao, E.C., Zovein, A.C., 2017. Fluorescent tagged episomals for stoichiometric induced pluripotent stem cell reprogramming. Stem Cell Res. Ther. 8. https://doi.org/10.1186/s13287-017-0581-7

Schneider, T.D., Stephens, R.M., 1990. Sequence logos: A new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100. https://doi.org/10.1093/nar/18.20.6097

Schopperle, W.M., DeWolf, W.C., 2007. The TRA-1-60 and TRA-1-81 Human Pluripotent Stem Cell Markers Are Expressed on Podocalyxin in Embryonal Carcinoma. Stem Cells 25, 723–730. https://doi.org/10.1634/stemcells.2005-0597

Sharova, L. V., Sharov, A.A., Piao, Y., Stagg, C.A., Amano, T., Qian, Y., Dudekula, D., Schlessinger, D., Ko, M.S.H., 2016. Emergence of undifferentiated colonies from mouse embryonic stem cells undergoing differentiation by retinoic acid treatment. In Vitro Cell. Dev. Biol. Anim. 52, 616–624. https://doi.org/10.1007/s11626-016-0013-5

Siegl-Cachedenier, I., Flores, I., Klatt, P., Blasco, M.A., 2007. Telomerase reverses epidermal hair follicle stem cell defects and loss of long-term survival associated with critically short telomeres. Journal of Cell Biology 179, 277–290. https://doi.org/10.1083/jcb.200704141

Singh, A.M., 2019. An Efficient Protocol for Single-Cell Cloning Human Pluripotent Stem Cells. Front. Cell Dev. Biol. 7. https://doi.org/10.3389/fcell.2019.00011

Smith, L.J., Fiebig, K.M., Schwalbe, H., Dobson, C.M., 1996. The concept of a random coil: Residual structure in peptides and denatured proteins. Fold. Des. 1, R95–R106. https://doi.org/10.1016/S1359-0278(96)00046-6

Sommer, C.A., Mostoslavsky, G., 2013. The evolving field of induced pluripotency: Recent progress and future challenges. J. Cell. Physiol. 228, 267–275. https://doi.org/10.1002/jcp.24155

Sørensen, H.P., Mortensen, K.K., 2005. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. Microb. Cell Fact. 4, 1–8. https://doi.org/10.1186/1475-2859-4-1

Srinivasan, R., Nady, N., Arora, N., Hsieh, L.J., Swigut, T., Narlikar, G.J., Wossidlo, M., Wysocka, J., 2020. Zscan4 binds nucleosomal microsatellite DNA and protects mouse two-cell embryos from DNA damage. Sci. Adv. 6. https://doi.org/10.1126/sciadv.aaz9115

Storm, M.P., Kumpfmueller, B., Bone, H.K., Buchholz, M., Sanchez Ripoll, Y., Chaudhuri, J.B., Niwa, H., Tosh, D., Welham, M.J., 2014. Zscan4 is regulated by PI3-kinase and DNA-damaging agents and directly interacts with the transcriptional repressors LSD1 and CtBP2 in mouse embryonic stem cells. PLoS One 9. https://doi.org/10.1371/journal.pone.0089821

Storm, M.P., Kumpfmueller, B., Thompson, B., Kolde, R., Vilo, J., Hummel, O., Schulz, H., Welham, M.J., 2009. Characterization of the phosphoinositide 3-kinase-dependent transcriptome in murine embryonic stem cells: Identification of novel regulators of pluripotency. Stem Cells 27, 764–775. https://doi.org/10.1002/stem.3

Su, R.J., Yang, Y., Neises, A., Payne, K.J., Wang, J., Viswanathan, K., Wakeland, E.K., Fang, X., Zhang, X.B., 2013. Few Single Nucleotide Variations in Exomes of Human Cord Blood Induced Pluripotent Stem Cells. PLoS One 8. https://doi.org/10.1371/journal.pone.0059908

Su, R.-J.J., Baylink, D.J., Neises, A., Kiroyan, J.B., Meng, X., Payne, K.J., Tschudy-Seney, B., Duan, Y., Appleby, N., Kearns-Jonker, M., Gridley, D.S., Wang, J., Lau, K.-H.W.H.W., Zhang, X.-B.B., 2013. Efficient Generation of Integration-Free iPS Cells from Human Adult Peripheral Blood Using BCL-XL Together with Yamanaka Factors. PLoS One 8, 1–12. https://doi.org/10.1371/journal.pone.0064496

Sun, H., Kim, P., Jia, P., Park, A.K., Liang, H., Zhao, Z., 2019. Distinct telomere length and molecular signatures in seminoma and non-seminoma of testicular germ cell tumor. Brief. Bioinform. 20, 1502–1512. https://doi.org/10.1093/bib/bby020

Sundaravadivelu, P.K., Borthakur, A., Agrawal, A., Sen, T., Thummer, R.P., 2025. An Overview of the Generation of Integration-free Induced Pluripotent Stem Cells and Their Subsequent Differentiation into Cardiomyocytes and β-cells, in: Stem Cell Biology. World Scientific, pp. 95–149. https://doi.org/10.1142/9789811294891_0005

Sundaravadivelu, P.K., Raina, K., Thool, M., Ray, A., Joshi, J.M., Kaveeshwar, V., Sudhagar, S., Lenka, N., Thummer, R.P., 2021. Tissue-Restricted Stem Cells as Starting Cell Source for Efficient Generation of Pluripotent Stem Cells: An Overview. Adv Exp Med Biol - Cell Biology and Translational Medicine. pp- 151-180. https://doi.org/https://doi.org/10.1007/5584_2021_660

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., Von Mering, C., 2019. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47, D607–D613. https://doi.org/10.1093/nar/gky1131

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., Tada, T., 2001. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr. Biol. 11, 1553-1558. 10.1016/S0960-9822(01)00459-6

Tagliaferri, D., De Angelis, M.T., Russo, N.A., Marotta, M., Ceccarelli, M., Del Vecchio, L., De Felice, M., Falco, G., 2016. Retinoic acid specifically enhances embryonic stem cell metastate marked by Zscan4. PLoS One 11, 1–15. https://doi.org/10.1371/journal.pone.0147683

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., Yamanaka, S., 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131, 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Takahashi, K., Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126, 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Takata, K., Saito, K., Maruyama, S., Miyata-Takata, T., Iioka, H., Okuda, S., Ling, Y., Karube, K., Miki, Y., Maeda, Y., Yoshino, T., Steidl, C., Kondo, E., 2019. Identification of TRA-1-60-positive cells as a potent refractory population in follicular lymphomas. Cancer Sci. 110, 443–457. https://doi.org/10.1111/cas.13870

Thamodaran, V., Rani, S., Velayudhan, S.R., 2021. Gene Editing in Human Induced Pluripotent Stem Cells Using Doxycycline-Inducible CRISPR-Cas9 System. pp. 755–773. https://doi.org/10.1007/7651_2021_348

Thool, M., Dey, C., Bhattacharyya, S., Sudhagar, S., Thummer, R.P., 2021. Generation of a Recombinant Stem Cell-Specific Human SOX2 Protein from *Escherichia coli* Under Native Conditions. Mol. Biotechnol. 63, 327–338. https://doi.org/10.1007/s12033-021-00305-y

Thool, M., Sudhagar, S., Thummer, R.P., 2023. Soluble Expression and Purification of Biologically Active Human NANOG from *Escherichia coli*, in: Healthcare Research and Related Technologies. Springer Nature Singapore, Singapore, pp. 99–118. https://doi.org/10.1007/978-981-99-4056-1_6

Thool, M., Sundaravadivelu, P.K., Sudhagar, S., Thummer, R.P., 2022. A Comprehensive Review on the Role of ZSCAN4 in Embryonic Development, Stem Cells, and Cancer. Stem Cell Rev. Rep. 18, 2740–2756. https://doi.org/10.1007/s12015-022-10412-1

Tsai, L.K., Peng, M., Chang, C.C., Wen, L., Liu, L., Liang, X., Chen, Y.E., Xu, J., Sung, L.Y., 2023. ZSCAN4 interacts with PARP1 to promote DNA repair in mouse embryonic stem cells. Cell Biosci. 13. https://doi.org/10.1186/s13578-023-01140-1

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., Holmes, M.C., 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435, 646–651. https://doi.org/10.1038/nature03556

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., Gregory, P.D., 2010. Genome editing with engineered zinc finger nucleases. Nat. Rev. Genet. 11, 636–646. https://doi.org/10.1038/nrg2842

Vagenende, V., Yap, M.G.S., Trout, B.L., 2009. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. Biochemistry 48, 11084–11096. https://doi.org/10.1021/bi900649t

Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Žídek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., Birney, E., Steinegger, M., Hassabis, D., Velankar, S., 2024. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Res. 52, D368–D375. https://doi.org/10.1093/nar/gkad1011

Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., Belmonte, J.C.I., 2011. Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. Development 138, 3699–3709. https://doi.org/10.1242/dev.064741

Verkuijl, S.A., Rots, M.G., 2019. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. Curr. Opin. Biotechnol. 55, 68-73. https://doi.org/10.1016/j.copbio.2018.07.005

Vigna, E., Naldini, L., 2000. Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy. J. Gene Med. 2, 308–316. https://doi.org/10.1002/1521-2254(200009/10)2:5<308::AID-JGM131>3.0.CO;2-3

Wang, F., Yin, Y., Ye, X., Liu, K., Zhu, H., Wang, L., Chiourea, M., Okuka, M., Ji, G., Dan, J., Zuo, B., Li, M., Zhang, Q., Liu, N., Chen, L., Pan, X., Gagos, S., Keefe, D.L., Liu, L., 2012. Molecular insights into the heterogeneity of telomere reprogramming in induced pluripotent stem cells. Cell Res. 22, 757–768. https://doi.org/10.1038/cr.2011.201

Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G.G., Krones, A., Ohgi, K.A., Zhu, P., Garcia-Bassets, I., Liu, F., Taylor, H., Lozach, J., Jayes, F.L., Korach, K.S., Glass, C.K., Fu, X.-D., Rosenfeld, M.G., 2007. Opposing LSD1 complexes function in developmental gene activation and repression programmes. Nature 446, 882–887. https://doi.org/10.1038/nature05671

Wang, P., Lin, M., Pedrosa, E., Hrabovsky, A., Zhang, Z., Guo, W., Lachman, H.M., Zheng, D., 2015. CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in neurodevelopment. Mol. Autism 6. https://doi.org/10.1186/s13229-015-0048-6

Wang, Q., Latham, K.E., 1997. Requirement for protein synthesis during embryonic genome activation in mice. Mol. Reprod. Dev. 47, 265–270. https://doi.org/10.1002/(SICI)1098-2795(199707)47:3<265::AID-MRD5>3.0.CO;2-J

Wang, Q.T., Piotrowska, K., Ciemerych, M.A., Milenkovic, L., Scott, M.P., Davis, R.W., Zernicka-Goetz, M., 2004. A Genome-Wide Study of Gene Activity Reveals Developmental Signaling Pathways in the Preimplantation Mouse Embryo. Dev. Cell 6, 133–144. https://doi.org/10.1016/S1534-5807(03)00404-0

Wang, R., Zhang, J.Y., Lu, K.H., Lu, S.S., Zhu, X.X., 2019. Efficient generation of GHR knockout Bama minipig fibroblast cells using CRISPR/Cas9-mediated gene editing. In Vitro Cell. Dev. Biol. Anim. 55, 784–792. https://doi.org/10.1007/s11626-019-00397-6

Wen, W., Zhang, J.P., Xu, J., Su, R.J., Neises, A., Ji, G.Z., Yuan, W., Cheng, T., Zhang, X.B., 2016. Enhanced generation of integration-free iPSCs from human adult peripheral blood mononuclear cells with an optimal combination of episomal vectors. Stem Cell Reports 6, 873–884. https://doi.org/10.1016/j.stemcr.2016.04.005

Whitmore, L., Wallace, B.A., 2008. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. Biopolymers 89, 392–400. https://doi.org/10.1002/bip.20853

Wiebe, J.P., Dinsdale, C.J., 1991. Inhibition of cell proliferation by glycerol. Life Sci. 48, 1511–1517. https://doi.org/10.1016/0024-3205(91)90275-G

Williams, A.J., Khachigian, L.M., Shows, T., Collins, T., 1995. Isolation and Characterization of a Novel Zinc-finger Protein with Transcriptional Repressor Activity. Journal of Biological Chemistry 270, 22143–22152. https://doi.org/10.1074/jbc.270.38.22143

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., Campbell, K.H.S., 1997. Viable offspring derived from fetal and adult mammalian cells. Nature 385, 810–813. https://doi.org/10.1038/385810a0

Wu, M., Liu, S., Gao, Y., Bai, H., Machairaki, V., Li, G., Chen, T., Cheng, L., 2018. Conditional gene knockout and reconstitution in human iPSCs with an inducible Cas9 system. Stem Cell Res. 29, 6–14. https://doi.org/10.1016/j.scr.2018.03.003

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., Jaenisch, R., Zhang, F., Sharp, P.A., 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. 32, 670–676. https://doi.org/10.1038/nbt.2889

Xu, H., Liang, H., 2022. The regulation of totipotency transcription: Perspective from in vitro and in vivo totipotency. Front. Cell Dev. Biol. 10. https://doi.org/10.3389/fcell.2022.1024093

Xu, H., Wang, B., Ono, M., Kagita, A., Fujii, K., Sasakawa, N., Ueda, T., Gee, P., Nishikawa, M., Nomura, M., Kitaoka, F., Takahashi, T., Okita, K., Yoshida, Y., Kaneko, S., Hotta, A., 2019. Targeted Disruption of HLA Genes via CRISPR-Cas9 Generates iPSCs with Enhanced Immune Compatibility. Cell Stem Cell 24, 566-578.e7. https://doi.org/10.1016/j.stem.2019.02.005

Xu, Yan, Xu, Yichen, Zhang, W., Li, M., Wendel, H.P., Geis-Gerstorfer, J., Li, P., Wan, G., Xu, S., Hu, T., 2022. Biodegradable Zn-Cu-Fe Alloy as a Promising Material for Craniomaxillofacial Implants: An in vitro Investigation into Degradation Behavior, Cytotoxicity, and Hemocompatibility. Front. Chem. 10. https://doi.org/10.3389/fchem.2022.860040

Yamamoto, A., Hiromoto, S., 2009. Effect of inorganic salts, amino acids and proteins on the degradation of pure magnesium in vitro. Materials Science and Engineering C 29, 1559–1568. https://doi.org/10.1016/j.msec.2008.12.015

Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J., Xia, N., 2020. Methods favoring homology-directed repair choice in response to crispr/cas9 induced-double strand breaks. Int. J. Mol. Sci. 21, 6461. https://doi.org/10.3390/ijms21186461

Yin, Y., Liu, N., Ye, X., Guo, R., Hao, J., Wang, F., Liu, L., 2014. Telomere elongation in parthenogenetic stem cells. Protein Cell 5, 8–11. https://doi.org/10.1007/s13238-013-0006-z

Yu, H., Kunarso, G., Hong, F.H., Stanton, L.W., 2009. Zfp206, Oct4, and Sox2 Are Integrated Components of a Transcriptional Regulatory Network in Embryonic Stem Cells. Journal of Biological Chemistry 284, 31327–31335. https://doi.org/10.1074/jbc.M109.016162

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., Slukvin, I.I., Thomson, J.A., 2007. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science. 318, 1917–1920. https://doi.org/10.1126/science.1151526

Yuen, G., Khan, F.J., Gao, S., Stommel, J.M., Batchelor, E., Wu, X., Luo, J., 2017. CRISPR/Cas9-mediated gene knockout is insensitive to target copy number but is dependent on guide RNA potency and Cas9/sgRNA threshold expression level. Nucleic Acids Res. 45, 12039–12053. https://doi.org/10.1093/nar/gkx843

Zalzman, M., Falco, G., Sharova, L. V., Nishiyama, A., Thomas, M., Lee, S.L., Stagg, C.A., Hoang, H.G., Yang, H. Te, Indig, F.E., Wersto, R.P., Ko, M.S.H., 2010. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464, 858–863. https://doi.org/10.1038/nature08882

Zeng, F., Baldwin, D.A., Schultz, R.M., 2004. Transcript profiling during preimplantation mouse development. Dev. Biol. 272, 483–496. https://doi.org/10.1016/j.ydbio.2004.05.018

Zhang, B., Fu, D., Xu, Q., Cong, X., Wu, C., Zhong, X., Ma, Y., Lv, Z., Chen, F., Han, L., Qian, M., Chin, Y.E., Lam, E.W.F., Chiao, P., Sun, Y., 2018. The senescence-associated secretory phenotype is potentiated by feedforward regulatory mechanisms involving Zscan4 and TAK1. Nat. Commun. 9. https://doi.org/10.1038/s41467-018-04010-4

Zhang, B., Metharom, P., Jullie, H., Ellem, K.A., Cleghorn, G., West, M.J., Wei, M.Q., 2004. The significance of controlled conditions in lentiviral vector titration and in the use of multiplicity of infection (MOI) for predicting gene transfer events. Genet. Vaccines Ther. 2, 6. https://doi.org/10.1186/1479-0556-2-6

Zhang, W., Walker, E., Tamplin, O.J., Rossant, J., Stanford, W.L., Hughes, T.R., 2006. Zfp206 regulates ES cell gene expression and differentiation. Nucleic Acids Res. 34, 4780–4790. https://doi.org/10.1093/nar/gkl631

Zhao, Yang, Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., Zhang, Q., Xiang, C., Hou, P., Song, Z., Liu, Y., Yong, J., Zhang, P., Cai, J., Liu, M., Li, H., Li, Y., Qu, X., Cui, K., Zhang, W., Xiang, T., Wu, Y., Zhao, Yiding, Liu, C., Yu, C., Yuan, K., Lou, J., Ding, M., Deng, H., 2008. Two Supporting Factors Greatly Improve the Efficiency of Human iPSC Generation. Cell Stem Cell. 3, 475-479. https://doi.org/10.1016/j.stem.2008.10.002

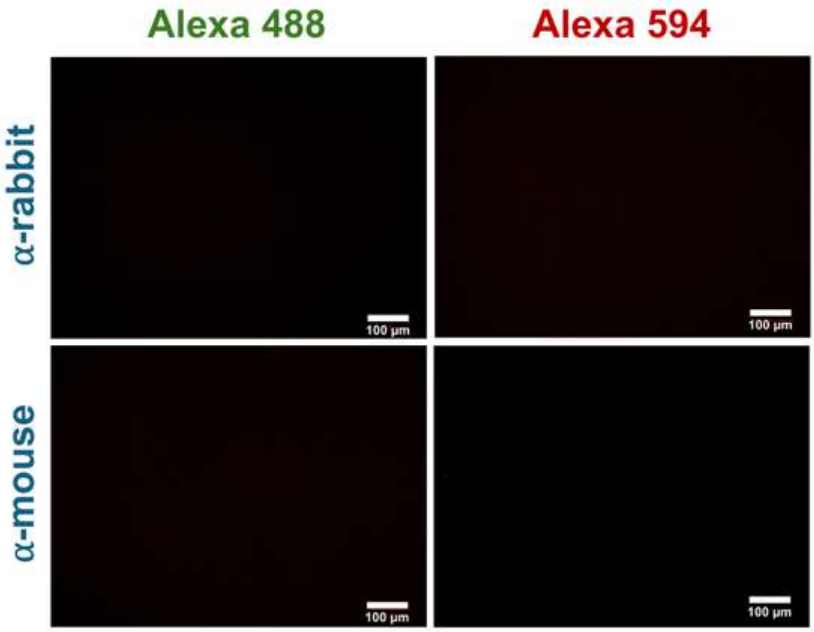


Fig. S1. The secondary antibody (control only) used in immunofluorescence experiments in this thesis. Scale bar: 100 µm.

```
Wildtype          1 MALDLRTTFQCEPSENNLGSENSAFQQSQGPAVQREEGISEFSRMVLNSF    50
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
KO clone 1        1 MALDLRTTFQCEPSENNLGSENSAFQQSQGPAVQREEGISEFSRMVLNSF    50

Wildtype         51 QDSNNSYARQELQRLYRIFHSWLQPEKHSKDEIISLLVLEQFMIGGHCND   100
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
KO clone 1       51 QDSNNSYARQELQRLYRIFHSWLQPEKHSKDEIISLLVLEQFMIGGHCND   100

Wildtype        101 KASVKEKWKSSGKNLERFIEDLTDDSINPPALVHVHMQGQEALFSEDMPL   150
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
KO clone 1      101 KASVKEKWKSSGKNLERFIEDLTDDSINPPALVHVHMQGQEALFSEDMPL   150

Wildtype        151 RDVIVHLTKQVNAQTTREANMGTPSQTSQDTSLETGQGYEDEQDGWNSSS   200
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
KO clone 1      151 RDVIVHLTKQVNAQTTREANMGTPSQTSQDTSLETGQGYEDEQDGWNSSS   200

Wildtype        201 KTTRVNENITNQGNQIVSLIIIQEENGPRPEEGGVSSDNPYNSKRAELVT   250
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
KO clone 1      201 KTTRVNENITNQGNQIVSLIIIQEENGPRPEEGGVSSDNPYNSKRAELVT   250

Wildtype        251 ARSQEGSINGITFQGVPMVMGAGCISQPEQSSPESALTHQSNEGNSTCEV   300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
KO clone 1      251 ARSQEGSINGITFQGVPMVMGAGCISQPEQSSPESALTHQSNEGNSTCEV   300

Wildtype        301 HQKGSHGVQKSYKCEECPKVFKYLCHLLAHQRRHRNERPFVCPECQKGFF   350
                    ||||||||||||||||||||||||||||:|||.|||||||||||||||||
KO clone 1      301 HQKGSHGVQKSYKCEECPKVFKYLCHLLSHQRRRRNERPFVCPECQKGFF   350

Wildtype        351 QISDLRVHQIIHTGKKPFTCSMCKKSFSHKTNLRSHERIHTGEKPYTCPF   400
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
KO clone 1      351 QISDLRVHQIIHTGKKPFTCSMCKKSFSHKTNLRSHERIHTGEKPYTCPF   400

Wildtype        401 CKTSYRQSSTYHRHMRTHEKITLPSVPSTPEAS    433
                    |||||||||||||||||||||||||||||||
KO clone 1      401 CKTSYRQSSTYHRHMRTHEKITLPSVPSTPEAS    433
```

Fig. S2. The pairwise alignment of ZSCAN4 wild-type protein sequence and the translated sequencing data of *ZSCAN4* knockout clone C1 showing two substitution mutations.

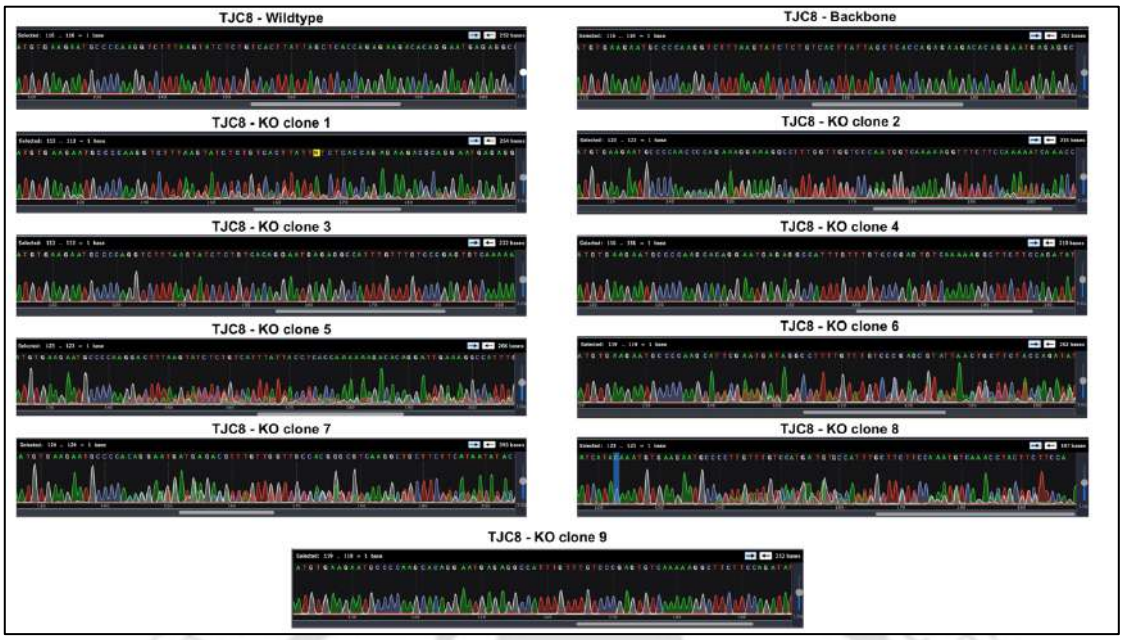


Fig. S3. The chromatogram corresponding to the sequencing of the TJC8 ZSCAN4 knockout clones.

Table S1. List of gRNAs used in this thesis.

Name of gRNA	gRNA	Sequence (5'-3')	Online tool used for designing	Deletion size (bp)
dgRNA-Exon 3	gRNA1	ACAGCAATAATTCATATGCA	CHOP-CHOP	63
	gRNA2	CTGCTGTTCAGAGAGAAGAA		
dgRNA-Exon 5	gRNA1	TAGCTCACCAGAGAAGACAC	CHOP-CHOP	60
	gRNA2	AGAGATACTTAAAGACCTTG		
dgRNA-set1	gRNA1	TCAGCGTTTCAACAAAGCCA	Synthego	75
	gRNA2	GGATGGTTCACACTGAAATA		
dgRNA-set2	gRNA1	GAACGGTCCTAGGCCTGAAG	Integrated DNA Technologies	114
	gRNA2	CCAAGGTGTCCCTATGGTGA		
dgRNA-set3	gRNA1	GGATTCTCTCATGAGACCGC	E-CRISP	72
	gRNA2	GGCGGTGGTATGTGGATGAC		
dgRNA-set4	gRNA1	CCACATACCACCGCCATATG	CHOP-CHOP	59
	gRNA2	CGGTAGCTTGTCTTACAAAA		
dgRNA-set5	gRNA1	CCACAAGAGAAGCAAACATG	CHOP-CHOP	47
	gRNA2	ACTTCCTTAGAAACAGGACA		

Table S2. List of primers used in this thesis.

Purpose	Name of Primer	Forward/Reverse primer (5'-3')	Amplicon size (bp)	Annealing temperature (°C)
Analysis of episome integration in genome	Oct4-epi-Tg	CATTCAAACTGAGGTAAGGG/ TAGCGTAAAAGGAGCAACATAG	124	54
	Klf4-epi-Tg	CCACCTCGCCTTACACATGAAGA / TAGCGTAAAAGGAGCAACATAG	156	62
	L-Myc-epi-Tg	GGCTGAGAAGAGGATGGCTAC/ TAGCGTAAAAGGAGCAACATAG	954	54
Pluripotency markers	OCT4	GTGGAGGAAGCTGACAACAATG/ TCTCACTCGGTTCTCGATACTG	105	51
	SOX2	GCACAACTCGGAGATCAG/ TAATCCGGGTGCTCCTTC	126	60
	NANOG	AGCCTAATCAGCGAGGTTTC/ ACAGAGCAAGACTCCGTTTC	101	57.7
	REX1	TTGCATACGCCTGTGTTCC/ GGCTCTTGCTGTTATCCAGTC	127	55
	GAPDH	GTCTCCTCTGACTTCAACAGCG/ ACCACCCTGTTGCTGTAGCCAA	131	58
Ectoderm-specific markers	OTX1	TACGCCCTCCTCTTCCTACT/ GCATGTGGGTGGTGATGATG	138	55
	OTX2	GGAAGCACTGTTTGCCAAGACC/ CTGTTGTTGGCGGCACTTAGCT	133	60
	DLK1	CCCCAAAATGGATTCTGCGAGG/ GGTTCTCCACAGAGTCCGTGAA	119	60
Mesoderm-specific markers	CDX2	ACAGTCGCTACATCACCATCCG/ CCTCTCCTTTGCTCTGCGGTTC	107	60
	TBX6	TCATCTCCGTGACAGCCTACCA/ CCGCAGTTTCCTCTTCACACGG	137	60
	CXCR4	CTCCTCTTTGTCATCACGCTTCC/ GGATGAGGACACTGCTGTAGAG	127	58
	HAND1	TCAAGGCTGAACTCAAGAAGG/ TGCGTCCTTTAATCCTCTTCTC	120	55
Endoderm-specific markers	CER1	CAGGACAGTGCCCTTCAGCCA/ ACAGTGAGAGCAGGAGGTATGG	142	60
	EOMES	AAATGGGTGACCTGTGGCAAAGC/ CTCCTGTCTCATCCAGTGGGAA	102	60
	FOXA2	GGAACACCACTACGCCTTCAAC/ AGTGCATCACCTGTTCGTAGGC	134	58
Confirmation of knockout	dgRNA-Ex3	CAGTGTGAACCATCCGAGA/ GCTGTGCTTTTCTGGTTGCA	WT: 211 KO: 148	55
	dgRNA-Ex5	CTTTCCAAGGTGTCCCTATG/ ACCCGTAGGTCTGATATCTG	WT: 286 KO: 226	55
	dgRNA-set1	AGAGACACAAGGCAAGAGAC/ CAGCCATGAGTGAAAGATCC	WT: 281 KO: 206	63
	dgRNA-set2	GTGAATGCCCAAACCACAAG/ TCATTGCTCTGGTGGGTAAG	WT: 484 KO: 370	63

	dgRNA-set3	GAGAGGCCATTTGTTTGTCC/ TATCAGACCAGCAGCTTAGG	WT: 309 KO: 237	63
	dgRNA-set4	GAGAGGCCATTTGTTTGTCC/ TATCAGACCAGCAGCTTAGG	WT: 309 KO: 250	63
	dgRNA-set5	ACGTCCACATGCAGGGACAG/ CTTCAGGCCTAGGACCGTTC	WT: 380 KO: 333	63
	dgRNA-Ex5-Verify at RNA level	CCAGAGAAGACACAGGAATG/ ATGTGGATGACTGGCGGTAG	WT: 242 KO: No band	55
	LentiCas9	AAGCAGTCCGGCAAGACAATCC/ AATGTGCTCGTGCAGGCTATCG	156	58
Telomere qPCR	Telomere	ACACTAAGGTTTGGGTTTGGGTT TGGGTTTGGGTTAGTGT/ TGTTAGGTATCCCTATCCCTATC CCTATCCCTATCCCTAACA	79	62
	Human β-globin	GCTTCTGACACAACTGTGTTCAC TAGC/ CACCAACTTCATCCACGTTCACC	120	58

Table S3. List of antibodies used in this thesis.

Name of the Antibody	Dilution for Western blot	Dilution for ICC	Dilution for flow cytometry	Company and Cat. No.
Histidine	1:5000	NA	NA	Biobharati Lifescience Cat. No.: BB-AB0010 RRID: AB_3741480
ZSCAN4	1:1000	1:250	1:250	Novus Biologicals Cat. No.: NBP1-77120 RRID: AB_10559205
TRA-1-60	NA	1:500	1:1600	Cell Signaling Technologies Cat. No.: 4746 RRID: AB_2119059
TRA-1-81	NA	1:500	1:1600	Cell Signaling Technologies Cat. No.: 4745 RRID: AB_2119060
SSEA4	NA	1:200	1:800	Cell Signaling Technologies Cat. No.: 4755 RRID: AB_1264259
OCT3/4	NA	1:100	1:400	SantaCruz Biotechnology Cat. No.: sc-5279 RRID: AB_628051

SOX2	NA	1:400	1:300	Cell Signaling Technologies Cat. No.: 3579 RRID: AB_2195767
NANOG	NA	1:200	1:100	Cell Signaling Technologies Cat. No.: 4903 RRID: AB_10559205
PAX6	NA	1:400	NA	Cell Signaling Technologies Cat. No.: 60433 RRID: AB_2797599
Brachyury	NA	1:1600	NA	Cell Signaling Technologies Cat. No.: 81694S RRID: AB_2799983
SOX17	NA	1:400	NA	Cell Signaling Technologies Cat. No.: 3579 RRID: AB_2650582
HRP-conjugated anti-rabbit secondary antibody	1:5000	NA	NA	Invitrogen Cat. No.: 31460 RRID: AB_228341
Goat anti-rabbit Alexa Fluor 488	NA	1:2000	1:400	Invitrogen Cat. No.: A-11034 RRID: AB_2576217
Goat anti-rabbit Alexa Fluor 594	NA	1:500	1:400	Invitrogen Cat. No.: A-11037 RRID: AB_2534095
Goat anti-mouse Alexa Fluor 488	NA	1:1000	1:400	Invitrogen Cat. No.: A-11029 RRID: AB_2534088
Goat anti-mouse Alexa Fluor 594	NA	1:500	1:400	Invitrogen Cat. No.: A-11032 RRID: AB_2534091

List of abbreviations

iPSCs	induced pluripotent stem cells
ZSCAN4	Zinc and SCAN domain containing 4
mESCs	murine embryonic stem cells
CRISPR-Cas9	Clustered regularly interspaced short palindromic repeats - CRISPR associated protein 9
HSCs	Hematopoietic stem cells
MSCs	Mesenchymal stem cells
SSCs	Spermatogonial stem cells
SCNT	Somatic cell nuclear transfer
OSKM	Oct3/4, Sox2, Klf4, c-Myc
OSNL	OCT4, SOX2, NANOG, LIN28
UTF1	Undifferentiated embryonic cell transcription factor 1
GLIS1	GLI-similar 1
ORF	Open reading frame
CD34	Cluster of differentiation 34
LDOC1	Leucine zipper downregulated in cancer 1
GOR4	Garnier-Osguthorpe-Robson version 4
SOPMA	Self-optimized prediction method with alignment
PSIPRED	PSI-BLAST based secondary structure prediction
COBALT	Constraint-based multiple alignment tool
MEGA	Molecular evolutionary genetics analysis
ICM	Inner cell mass
ZGA	Zygotic genome activation
Eif1a	Eukaryotic translation initiation factor 1A
MERVL	murine endogenous retroviruses with leucine tRNA primer
Dmnts	DNA methytransferases
PI3K	Phosphatidylinositol 3-kinase
GSK	Glycogen synthase kinase
SCE	Sister chromatid exchange
ALT	Alternate lengthening of telomeres
MEFs	Mouse embryonic fibroblasts

LIF	Leukemia inhibitory factor
LSD1	Lysine-specific histone demethylase 1A
Ctbp2	C-terminal binding protein 2
HTNT1	Histidine triad nucleotide-binding protein 1
terc	telomerase reverse transcriptase
terc	telomerase RNA component
T-SCE	Telomere sister chromatid exchange
TCA	Tetraploid complementation assay
CSC	Cancer stem cells
DSBs	Double stranded breaks
DDR	DNA damage response
HDR	Homology-directed repair
NHEJ	Non-homologous end joining
ZFNs	Zinc finger nucleases
TALENs	Transcription activator-like effector nucleases
RVDs	Repeat-variable di-residues
Pre-crRNA	Precursor crispr RNA
crRNA	CRISPR RNA
tracrRNA	trans-activating crRNA
PAM	Protospacer adjacent motif
CD	Circular dichroism
SNPs	Single nucleotide polymorphisms
OD_{600}	Optical density at 600 nm
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween-20
HRP	Horseradish peroxidase
PB	Phosphate buffer
DMEM	Dulbecco's modified eagle media
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
DMSO	Dimethyl sulfoxide
BSA	Bovine serum albumin
PBS	Phosphate buffered saline
GRCA	GenScript rare codon analysis

GCUA	Graphical codon usage analyzer 2.0
TAT	Transactivator of transcription
NLS	Nuclear localization sequence
E. coli	*Escherichia coli*
IPTG	Isopropyl β-D-1-thiogalactopyranoside
FBS	Fetal bovine serum
KOSR	Knockout serum replacement
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
PCI	Phenol chloroform isoamyl alcohol mixture
EBNA1	Epstein-barr virus nuclear antigen 1
OriP	Origin of replication
HDF	Human dermal fibroblasts
AHDF	Adult human dermal fibroblasts
FGM	Fibroblast growth media
STR	Short tandem repeats
PCR	Polymerase chain reaction
MET	Mesenchymal to epithelial transition
gRNA	guide RNA
EGFP	Enhanced green fluorescence protein
dgRNA	dual guide RNA
5-Aza-dC	5-Aza-2'-deoxycytidine
WT	Wild-type
KO	Knockout

Publications from PhD Thesis

Research articles

- **Sundaravadivelu PK**, Augustine ABHR, Thummer RP (**2026**) Generation of cell permeant recombinant ZSCAN4 protein for various biological applications, In Recent Developments in Biotechnology, Proceedings of Research and Industrial Conclave - Synergy 2025, Springer Nature, Singapore (Accepted).
- **Sundaravadivelu PK**, Raina K, Borthakur A, Mahto P, Shetty D, Dhamne C, Kaveeshwar V, Martin CA, Radhakrishnan S, Rela M, Thummer RP (**2026**) Derivation and characterization of the iPSC line IITGi001-B from primary human adult dermal fibroblasts. Stem Cell Research. Feb 13:103931. https://doi.org/10.1016/j.scr.2026.103931 (Elsevier).
- **Sundaravadivelu PK**, Thummer RP (**2026**) Design and validation of CRISPR-Cas9-mediated knockout of ZSCAN4 gene in mammalian cells, In Recent Developments in Biotechnology, Proceedings of Research and Industrial Conclave - Synergy 2025, Springer Nature, Singapore (Accepted).

Review articles

1. **Sundaravadivelu PK**, Borthakur A, Agrawal A, Sen T, Thummer RP (**2024**). An overview on the generation of integration-free induced pluripotent stem cells and their subsequent differentiation into cardiomyocytes and β-cells. In Advances in Stem Cell Research: An Indian Perspective. Eds Prof. Rama Shanker Verma. https://doi.org/10.1142/9789811294891_0005 (World Scientific).
2. Thool M*, **Sundaravadivelu PK***, Sudhagar S, Thummer RP (**2022**) A Comprehensive Review on the Role of ZSCAN4 in Embryonic Development, Stem Cells, and Cancer. Stem Cell Reviews and Reports. Jun 23:1-7. https://doi.org/10.1007/s12015-022-10412-1 (Springer Nature) *Equal contribution.
3. **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. In: Turksen K. (eds) Cell Biology and Translational Medicine. Advances in Experimental Medicine and Biology. https://doi.org/10.1007/5584_2021_660 (Springer Nature).
4. 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. https://doi.org/10.1007/s12015-021-10200-3 (Springer Nature). *Equal contribution.
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. https://doi.org/10.1016/B978-0-323-90059-1.00007-5 (Elsevier) *Equal contribution.
6. Dey C*, Raina K*, Haridhasapavalan KK*, Thool M*, **Sundaravadivelu PK**, Adhikari P, Gogoi R, Thummer RP (**2021**) An Overview of Reprogramming Approaches 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. https://doi.org/10.1016/B978-0-12-822231-7.00011-4 (Elsevier) *Equal contribution.

Publications from other collaborative research

1. Chaudhari H, **Sundaravadivelu PK**, Kulkarni R, Bhuyan MK, Thummer RP (**2025**). Cumulative learning-based segmentation aided cell mixtures classification in digital holographic microscopy. Optics & Laser Technology. Feb 1;181:112029 https://doi.org/10.1016/j.optlastec.2024.112029.
2. Chatterjee S, Sarkar T, **Sundaravadivelu PK**, Thummer RP, Satpati P (**2024**) De novo Design of Tryptophan containing Broad-Spectrum Cationic Antimicrobial Octapeptides. ChemMedChem.:e202400566. https://doi.org/10.1002/cmdc.202400566.
3. Chaudhari H, Kulkarni R, Bhuyan MK, **Sundaravadivelu PK**, Thummer R. (**2024**) Machine-learning-based classification of co-cultured cells from the phase images in digital holographic microscopy. In Unconventional Optical Imaging IV, Jun 18 (Vol. 12996, pp. 87-96). SPIE https://doi.org/10.1117/12.3017346.
4. Sarkar T, Ghosh S, **Sundaravadivelu PK**, Pandit G, Debnath S, Thummer RP, Satpati P, Chatterjee S (**2024**). Mechanism of Protease Resistance of D-Amino Acid Residue Containing Cationic Antimicrobial Heptapeptides. ACS Infectious Diseases. 10.1021/acsinfecdis.3c00491. (ACS publications).
5. Chaudhari H, Kulkarni R, **Sundaravadivelu PK**, Thummer RP, Bhuyan MK (**2024**). Digital hologram reconstruction algorithm based on the Fractional Fourier transform in non-telecentric digital holographic microscopy. Optics Letters. 49(2):182-5. https://doi.org/10.1364/OL.504723. (Optica publishing group).
6. Chetia M, Sarkar T, Yadav M, Dey C, **Sundaravadivelu PK**, Thummer RP, Chatterjee S (**2024**). Tuning of hydrophobic–hydrophilic balance for the development of a salt-tolerant and protease-resistant lipopeptide AMP. New Journal of Chemistry. 48(6):2764-77. 10.1039/D3NJ04815B. (Royal Society of Chemistry).
7. Chaudhari H, Kulkarni R, **Sundaravadivelu PK**, Thummer RP, Bhuyan MK (**2024**). Dimensionality reduction technique-based phase aberration compensation and spurious fringe removal in off-axis digital holographic microscopy. Optics and Lasers in Engineering. https://doi.org/10.1016/j.optlaseng.2023.107853. (Elsevier).

Conferences/ Workshops

1. Oral and Poster presentation in Research and Industrial Conclave, 2025 organized by IIT Guwahati, Guwahati, Assam, India.
2. Presented a poster in “ISSCR 2024 Annual meet: The Global Stem Cell Event” organised by the International Society for Stem Cell Research (ISSCR) held from 10th July to 13th July, 2024 in Hamburg, Germany.
3. Attended a Hands-on workshop on the Applications of Fluorescence-Activated Cell Sorting (FACS) organized by Department of Chemistry, Indian Institute of Technology Guwahati in association with Becton Dickinson India funded by DBT, Govt. of India from 4th March to 7th March, 2024.
4. Presented a poster entitled ”Optimizing expression and purification of biologically active recombinant human ZSCAN4 protein” in International conference on Biomaterials, Regenerative medicine and Devices held on Dec 14th-18th, 2022, at Indian Institute of Technology Guwahati, Guwahati, Assam, India.

Awards

1. Recipient of the Prime Minister Research Fellowship (PMRF) for a period of four years from 2021 to 2024 with a research contingency of INR 8 lakhs.
2. Recipient of the travel award by the International Society for Stem Cell Research (ISSCR), an amount of 1500 USD, for attending the ISSCR 2024 Annual meeting in Hamburg, Germany.