Establishment of a bioactive recombinant protein toolbox for the prospective generation of integration-free induced pluripotent stem cells and other biological applications

A Thesis

Submitted in Partial Fulfilment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Chandrima Dey

Under the supervision of

Dr. Rajkumar P. Thummer

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

October, 2022

Dedicated to

My Baba and Ma

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

DECLARATION

I do hereby declare that the content embodied in this thesis entitled "**Establishment of a bioactive recombinant protein toolbox for the prospective generation of integration-free induced pluripotent stem cells and other biological applications**" 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: 30th October 2022

Chandrima Dey

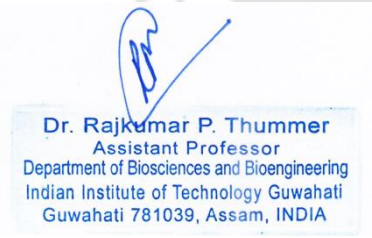
Chandrima Dey

(166106013)

Indian Institute of Technology Guwahati
Department of Biosciences and Bioengineering

CERTIFICATE

It is certified that the work described in this thesis entitled "**Establishment of a bioactive recombinant protein toolbox for the prospective generation of integration-free induced pluripotent stem cells and other biological applications**", by **Ms. Chandrima Dey** (Roll No. 166106013) 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: 30th October 2022

Acknowledgement

I want to extend my immense gratitude towards all the people who have been part of this long journey and who guided and kept their faith in me through thick and thin. I feel privileged and want to acknowledge their presence and immense support in achieving this milestone in my life.

My heartfelt gratitude towards my thesis supervisor **Dr. Rajkumar P. Thummer**, for giving me the opportunity to work under him and become a part of this esteemed institution. He showed confidence in me even in the most difficult times and helped me understand how science is not about one outcome and one perspective. His constant support was not only as my mentor but as a friend and a father-like figure with whom I got the confidence to share random scientific ideas and most importantly learn how to deal with life. His ability to be calm in the most tumultuous times and to look for plausible options encouraged me to evolve as a student of science as well as a person. Through exciting scientific discussions, non-agreements and agreements, disappointments and achievements, I learned the true meaning of working as a team and what it takes to be a leader. Forever, I will be indebted to his love, affection, and protection toward all of us and his constant guidance. The mindset that even if you are not the best but you can strive to be better than what you were before will stay with me forever.

I want to acknowledge **Dr. Shirisha Nagotu** Ma'am for providing high-end instrument facilities and for her immense support and encouragement and unbiased affection in every wake of this journey.

I want to extend my gratitude to the members of my Doctoral committee, **Prof. Anil Mukund Limaye** and **Prof. Sachin Kumar** of the Department of Biosciences and Bioengineering and **Dr. Sunanda Chatterjee** of the Department of Chemistry for their valuable suggestions, constant support, and advice which enabled me to improve my work.

I am grateful to current Head of the Department of Biosciences and Bioengineering, IIT Guwahati, **Prof. Rakhi Chaturvedi** and other former HODs – **Prof. Latha Rangan, Prof. Kannan Pakshiranjan** and **Prof. V Venkata Dasu** for providing me with the departmental facilities for carrying out my research work. I would thank all the technical

staff of the department for their help and assistance. I am also grateful to **Prof. Sidharth S Ghosh**, IIT Guwahati, for providing high-end instrument facilities. I would also like to thank the Ministry of Education, Government of India, for the financial support during my tenure.

"The pursuit of Ph.D. is enduring daring adventure."-Lailah Gifty Akita.
In this long and adventurous journey, I was privileged to have all these incredible humans supporting me through some of the toughest phases of my life.

Since the beginning of my journey here at IIT Guwahati and before we all could be part of our fully functional (SCERM) lab, my labmates cum friends-like family **Krishna** and **Gloria** stood with me through the test of time. They supported, cared and loved me at my lowest and cheered louder at my highest. This journey wouldn't be complete without their help and suggestions starting from setting up the lab to designing experiments, analyzing data, and every possible problem I have encountered at work or in life. I want to acknowledge **Khyati** for helping me with experiments whenever required and mostly bringing a sense of calm during my difficult work phases. I feel happy that I got the chance to be part of a team that uplifts each other and sets an example for a unit. This team and my thesis would not be complete without my incredibly hardworking and talented juniors **Vishalini, Pradeep, Sujal, Madhuri, Atreyee, Ronima, Rishabh, Badu, Bitan, Nayan Disha,** and **Bishakha**. They have played a crucial part in helping me through optimizing protocols, papers, experimental troubleshooting and essentially **Krishna, Pradeep and Ronima** for moral support and as an active helping hands during my thesis submission.

I want to thank all the OBCAL lab members (**Riddhi, Sahay, Rachayeeta di, Nayan bhaiyya, Terrence, Suchetana, Esha,** and **Tanveera** for being supportive throughout this journey.

This journey is incomplete without my friends outside my work as they form one of the most integral part of my life. I want to thank my friend **Satakshi** for all her support, incredible food, and love. I also want to thank **Muthuvel, Nivedita, Anil, Kamlesh, Bharathi**, **Jenny**, **Debojit**, **Shinchini** and the entire **Tamil Sangam group** for being the most fun-filled friends, and also I want to acknowledge my **Subansiri animal care group** for helping me learn compassion, courage and perseverance.

My journey over these years with music, creativity, laughter and unbiased affection was possible because of these incredible people **Gundappa Saha**, **Siddhanta Roy,** and **Pradip Das**. They held me close and guided me through some of the most difficult phases. They pushed me to be the best version of myself and gave me some of the most precious memories to cherish. I also want to acknowledge **Anjishnu**, **Rajesh**, **Surojit and Roop** for making my last 6 months on campus musically exciting.

I want to thank my father, **Late Shib Sanker Dey**, my mother **Kakali Dey** for all the love and support, and my aunt **Late Purnima Datta** who introduced me to this subject. I want to thank my uncles, **Late Indrajit Datta**, **Late Uday Sanker Dey**, and **Late Ravi Sanker Dey** for encouraging and nurturing my inclination towards art and music. I want to thank **my grandparents** and **my entire family** (Maternal) for loving and supporting me always. Nothing can be fun and fulfilling without siblings, and I want to thank **Debarun Dey**, **Ipshita Datta**, **Sangeeta Dey**, **Raktim Dey,** and **Gaurang Bhatt** for being my biggest support system. I want to thank my friends like my family (**Rajshree**, **Sumedh**, **Anandita**, **Sudeepta da** and **Rene**) for constantly supporting and taking care of me.

Signing off

Chandrima Dey

Contents

ABSTRACT

The derivation of induced pluripotent stem cells (iPSCs) over a decade ago ushered a new era in the cellular reprogramming paradigm. The very concept of reprogramming somatic cells into a pluripotent cell stage brought great enthusiasm in the scientific community as it opened new and promising avenues in the field of disease modeling, drug designing, understanding the intricacies of developmental biology, and wide use of autologous cell-based therapies. The generation of iPSCs enhances the prospects of pluripotent cells from bench to bedside, providing an opportunity to bring patient-specific therapies. However, the major limitation was the use of integrative approaches for the generation of iPSCs that drastically crippled their use in clinical applications due to random integrations, leading to mutations. Alternative to the integrative approaches, the non-integrative approaches provided a fruitful solution to overcome the problem of genomic integration. Over the years, several non-integrative approaches have emerged to generate integration-free iPSCs; among them, the recombinant protein-based approach is deemed to be the safest one. There are several challenges associated with protein-based cellular reprogramming. Hence, we aim to establish a recombinant protein toolbox (OCT4, SOX2, UTF1 and GLIS1) for the generation of integration-free human iPSCs by addressing the roadblocks. In this study, we have laid down simple and methodical strategies to generate recombinant proteins withholding native-like secondary structure conformations, thereby retaining their functionality. To achieve this, we codon-optimized the protein-coding nucleotide sequences for each of the transcription factors and then fused them with the nuclear localization signal, cell-penetrating peptide and poly-Histidine tag sequences. The fusion genes were then cloned into the expression vector and expressed in *E. coli* BL21(DE3). To obtain maximum soluble expression, we screened various expression parameters and analyzed the effect of the position of fusion tags at two terminals. The results demonstrated the importance of identifying suitable genetic constructs and optimal expression parameters for the soluble expression of these recombinant proteins concerning their quality and quantity. Notably, irrespective of the position of fusion tags, OCT4 and SOX2 showed significant soluble expression, whereas only C-terminally tagged UTF1 and GLIS1 showed soluble expression. Additionally, GLIS1 showed bacterial strain and culture media-specific dependency during expression in *E. coli*. The effect of fusion tag was also observed in the case of purified N-terminal OCT4 protein, where N-terminal tagged OCT4 showed a disordered secondary structure. However, no such influence of the position of

fusion tags was observed on the secondary structure of purified SOX2. Interestingly, UTF1 protein showed salt-dependent aggregation during purification and was found to be an ion-sensitive protein. GLIS1 showed compromised protein yield due to the presence of glycerol. Thus, we established a successful methodology to utilize one-step homogeneous purification for obtaining these recombinant fusion proteins under native and suitable buffer conditions for retaining their secondary structure conformation. Our fusion strategy of tagging the nuclear localizing signal and trans-activator of transcription (a cell-penetrating peptide sequence) facilitated cellular and nuclear delivery of these recombinant proteins into the mammalian cells, without the need of any transduction reagent. We showed that the generated recombinant fusion proteins are biologically active, as OCT4 showed reduction in cell migration and cell proliferation in human fibroblasts. Additionally, we also confirmed the biological activity of the purified OCT4 protein using a reporter system, where this protein binds to its own promoter, thereby expressing GFP. SOX2 and UTF1 showed tumorigenic and tumor-suppressive roles in cervical cancer cells, respectively. GLIS1 showed tumorigenic potential in breast cancer cells, with no significant effect on normal human fibroblasts. The established functionally active recombinant protein toolbox provides a path to generate integration-free iPSCs circumventing genetic manipulation. These proteins also open various opportunities to unravel their functions in different cancers and their potential as promising therapeutic targets and other biological applications in the near future.

List of Figures

List of Tables

Abbreviations

SCNT	Somatic cell nuclear transfer
ESCs	Embryonic stem cells
MEF	Mouse embryonic fibroblasts
ETS2	Avian erythroblastosis virus E26 oncogene homolog-2
iPSC	Induced pluripotent stem cell
hiPSC	Human induced pluripotent stem cell
mRNA	Messenger RNA
miRNA	Micro RNA
SeV	Sendai virus
AAV	Adeno-associated viruses
CMV	Cytomegalovirus
EBNA1	Epstein-Barr Virus Nuclear Antigen 1
oriP	origin of replication
SV40 LT	Simian Virus 40 Large T antigen
PB	PiggyBac
SB	Sleeping Beauty
IVT	In vitro transcription
psi	Pseudouridine
5mC	5-methylcytidine
VPA	Valproic Acid
5-AZA	5-Azacitidine
DNMTi	DNA methyltransferase inhibitor
CPP	Cell-penetrating peptides
NLS	Nuclear localisation signal/sequence
TAT	HIV-Transactivator of transcription
HFF	Human foreskin fibroblast
NTP	Nuclear trafficking peptide
TLR3	Toll-like receptor 3
Poly I:C	Polyinosinicpolycytidylic acid
UTF1	Undifferentiated embryonic cell transcription factor 1
RefSeq	Reference sequence
His	Histidine

GOI	Gene of interest
LB	Luria-Bertani broth
TB	Terrific broth
IPTG	Isopropyl β-D-1-thiogalactopyranoside
PB	Sodium phosphate buffer
NaCl	Sodium chloride
RT	Room temperature
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TBST	Tris-buffered saline tween-20
CAI	Codon adaptation index
IMAC	Immobilized metal ion affinity chromatography
Ni-NTA	Nickel-nitrilotriacetic acid
LC-MS/MS	Liquid chromatography-tandem mass spectroscopy
CD	Circular dichroism
UV	Ultra violet
BeStSel	Beta Structure Selection
DMEM	Dulbecco's modified eagle medium
FBS	Fetal bovine serum
P/S	Penicillin-streptomycin
CO_2	Carbon dioxide
NEAA	Non-essential amino acids
HDF	Human dermal fibroblasts
PBS	Phosphate buffered saline
DAPI	4′,6-diamidino-2-phenylindole
MTT	3- [4, 5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide; thiazolyl blue
DMSO	Dimethyl sulfoxide
RT-qPCR	Reverse transcription – quantitative polymerase chain reaction
CSCs	cancer stem cells
ESCC	esophageal squamous cell carcinoma
cDNA	complementary deoxyribonucleic acid
GAPDH	glyceraldehyde-3-phosphate dehydrogenase

CHAPTER 1

Introduction and Review of literature

Brief summary of the chapter

This chapter details the current research on reprogramming approaches used in the generation of iPSCs and their respective advantages and limitations in bringing this technology from bench to bedside. Previous studies have attested that iPSCs have great potential in patient-specific cell therapies, drug screening, and disease modeling. Thus, it is critical to generate clinical-grade iPSCs using a non-integrative based reprogramming approach. This chapter provides an overview of recombinant proteins (a promising and safer approach) in cellular reprogramming and their applications in various other fields. It also discusses the critical prerequisites for the expression and purification of human recombinant proteins in different expression systems, their advantages, and their disadvantages. It also details the challenges associated with heterologous recombinant protein expression and purification in a bacterial system. We discuss the importance of generating the recombinant protein toolbox and the strategies, which are involved in successfully delivering these recombinant proteins into mammalian cells. Further, we have discussed the rationale behind the selection of the four key transcription factors, OCT4, SOX2, UTF1 and GLIS1, and provided a summary of their role in the generation of iPSCs. OCT4 and SOX2 are the master regulators of reprogramming and UTF1 and GLIS1 have been reported to enhance the quality of the iPSCs generated. The chapter briefly summarizes the various roles of these proteins in different cancers. Finally, the chapter concludes with the motivation of the work carried out in this thesis work, addressing the limitations of recombinant protein production and their use in iPSC generation.

1.1 Review of Literature

1.1.1 Emergence of induced pluripotent stem cells

The idea of cellular reprogramming stems from the study performed on the transplantation of living cell nuclei into animal eggs by Briggs and Kings in 1951 (Briggs and King, 1951). A decade later in the year 1962, John Gurdon transferred the nucleus from a single cell procured from the tadpole intestine to the unfertilized non-nucleated cell, giving rise to a new offspring (Gurdon, 1962). The conclusions of this study were, firstly epigenetic changes are not permanent during cell differentiation and secondly, terminally differentiated adult mammalian cells are genetically totipotent (Stadtfeld and Hochedlinger, 2010). Based on this concept of somatic cell nuclear transfer (SCNT), Ian Wilmut in the year 1997 cloned a sheep (Dolly) (Wilmut et al., 1997). In 1981, Gail R Martin and Martin Evans were successful in the isolation of mouse embryonic stem cells (ESCs) (Martin, 1981). Seven years later, Thomson and group derived ESCs from human blastocysts that had the ability to differentiate into three germ layers, having a normal karyotype, and high telomerase activity. They stated that these pluripotent cells could be useful in understanding developmental biology, drug screening and transplantation medicine (Thomson et al. 1998). Further, in 1988, Smith and colleagues provided insight into the basics of culturing ESCs and the role of essential pluripotency factors (Smith et al., 1988). In the pioneering study, Takahashi and Yamanaka addressed the limitations encircling the ethical issues associated with ESCs and the technical challenges of SCNT technology (Takahashi and Yamanaka, 2006). It was accomplished by overexpressing stem cell-specific transcription factors (OCT4, SOX2, KLF4, c-MYC (OSKM)) in a terminally differentiated cell type, mouse embryonic fibroblasts (MEFs), generating stem cell-like cells, which they coined as induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2006).

Subsequently, the same group generated human iPSCs (hiPSCs) by reprogramming human fibroblasts using the same cocktail of transcription factors (Takahashi et al., 2007). Simultaneously, Thomson and coworkers used a different set of transcription factors, namely OCT4, SOX2, NANOG and LIN28 (OSNL) for generating hiPSCs (Yu et al., 2007). Timeline of cellular reprogramming towards the generation of iPSCs has been depicted in Figure 1.1.

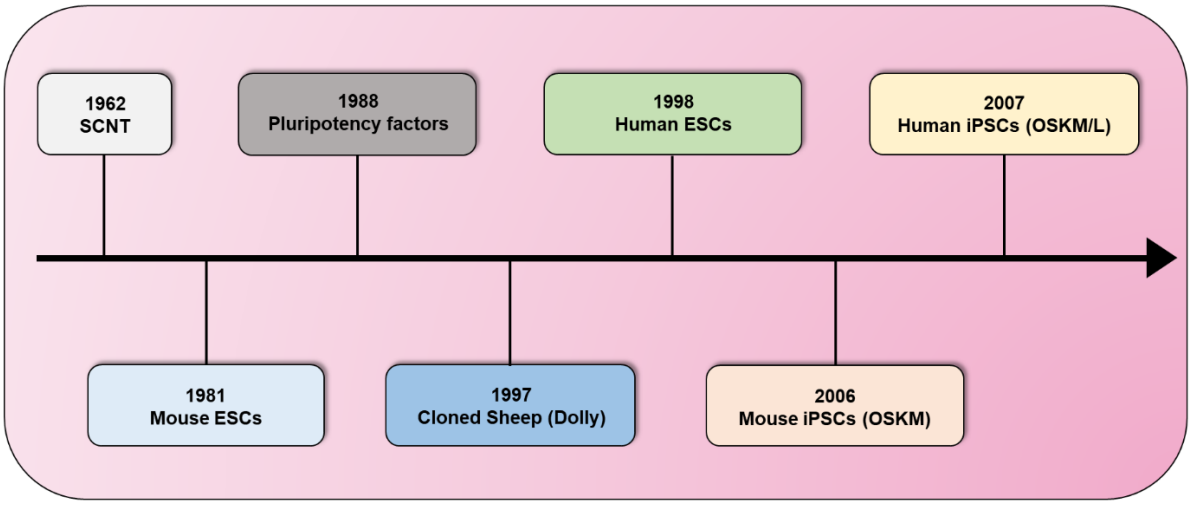


Figure 1.1 Timeline of cellular reprogramming toward the conceptualization and generation of iPSCs.

Generation of iPSCs set a new landmark in the field of regenerative medicine. The ground-breaking discovery of reprogramming somatic cells to generate iPSCs almost two decades ago has revolutionized stem cell research attracting immense global attention for developing new human disease models, augmenting platform for drug screening, and application of autologous cell-based therapies (Takahashi and Yamanaka, 2006; Yu et al., 2007; Young et al., 2012; Singh et al., 2015; Menon et al., 2016). However, the commonly used conventional reprogramming approaches to generate iPSCs involving retro- and lenti-viral vectors nullify the clinical applicability of these cells. Although these approaches are

robust and efficient, they carry an enormous risk of permanent genetic modifications and tumor formation (Okita et al., 2007; Ben-David and Benvenisty, 2011). In addition, slow reprogramming as well non-stoichiometric expression of reprogramming factors were some of the other major limitations (Takahashi and Yamanaka, 2006; Yu et al., 2007; Takahashi et al., 2007). Overcoming these limitations are crucial as the main purpose of iPSCs are to study drug targets, drug toxicity and for patient-specific regenerative therapies. To evade these safety concerns, tremendous advances have been made in establishing non-integrating viral (adenovirus, adeno-associated viruses and Sendai virus) and non-viral approaches (plasmid transfection, piggybac transposon, mini circle vector, episomal, modified mRNA, microRNAs, recombinant proteins and small molecules) to derive integration-free iPSCs. These methodologies curtail the risk of any genomic alteration and enhance the prospects of these cells from bench-to-bedside.

1.1.2 Different approaches of generating iPSCs

Generation of iPSCs are broadly divided into two approaches: Integrative and Non-integrative approaches. Integrative approaches are those gene delivery approaches which leverages upon the potential of viruses to enter and integrate the gene into the host genome and subsequently multiply, thus imparting a long-term effect. Non-Integrative approaches on the other hand does not involve genomic integration in host genome and thus, circumvents the major limitations imparted by integrative approaches such as genetic modification or transgene reactivation.

1.1.2.1 Integrative approaches

Retroviruses are enveloped RNA viruses of the family *Retroviridae*. They replicate by infecting dividing cells and thus have a very high transduction efficiency and long-term effect in the host system. They are widely used in clinical gene therapy, basic research and also as

gene transfer systems. First approach towards reprogramming using retroviral vector was carried by Yamanaka and colleagues in 2006 for the generation of mouse and human iPSCs from fibroblasts (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). The major disadvantage of this approach is its random integration into the host genome by infecting continuously dividing cells, eventually leading to insertional mutagenesis causing cancer (Shao and Wu, 2010).

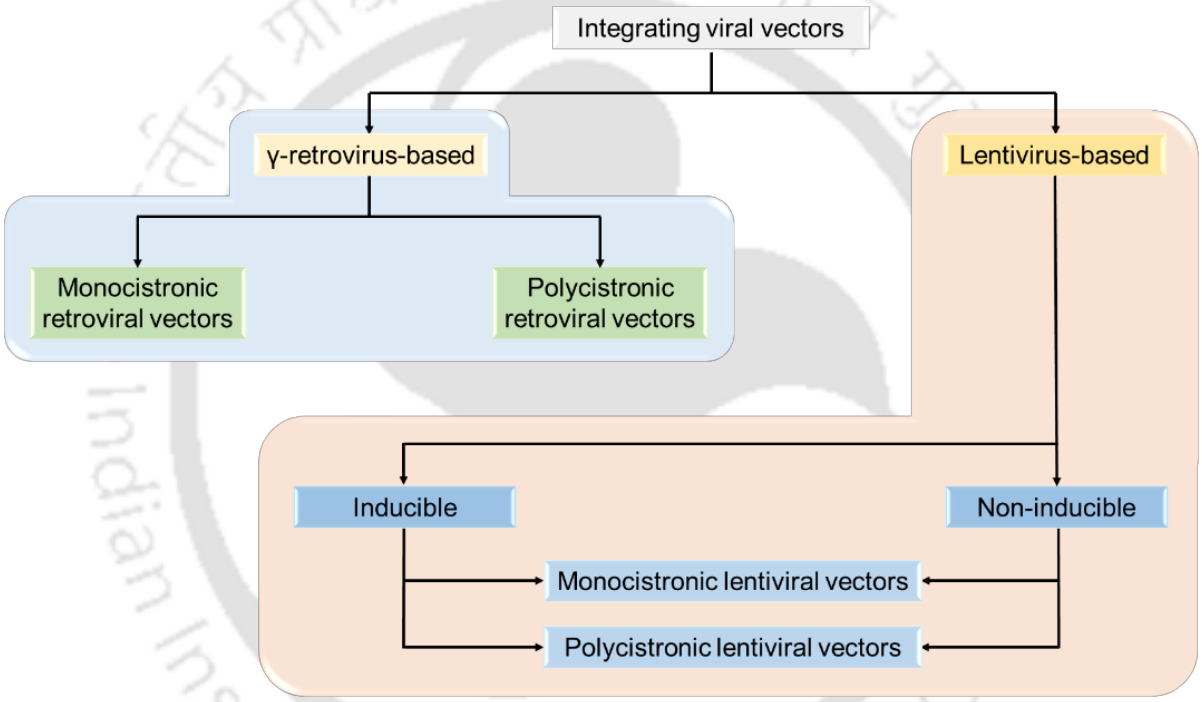


Figure 1.2 Schematic overview of various integrative viral approaches and their sub-divisions.

Lentiviruses are also from the family of *Retroviridae* and is a subclass of retroviruses. They infect both dividing and non-dividing cells, unlike retroviruses. Initial approach for the generation of hiPSCs was carried out by transducing OSNL (Yu et al., 2007). The drawbacks are permanent genomic modifications due to random transgene integration leading to insertional mutagenesis and tumor formation, limiting the clinical applicability of the

generated iPSCs (Okita et al., 2007; Ben-David and Benvenisty, 2011). Also, inefficient silencing and continuous activation of transgenes eventually affect their differentiation potential (Sommer et al. 2009; Kaji et al. 2009; Ramos-Mejia et al. 2012; Kadari et al. 2014). Inefficient gene silencing was tackled by using doxycycline-inducible lentiviruses (Hockemeyer et al., 2008; Maherali et al., 2008). This can also be solved by a transgene excision method where the transgene sequences can be excised out from the iPSCs (Somers et al. 2010; Kadari et al. 2014). However, this excision may leave some residual elements in the genome and result in insertional mutagenesis, and hence makes this technique unsuitable for clinical applications. The overview of the integrative approaches is summarized in Figure 1.2.

1.1.2.2 Non-Integrative approaches

Viral approaches

Non-integrative viral approaches comprise of Sendai viruses, adenoviruses and adeno-associated viruses. To circumvent the serious concerns of genetic mutations and compromised clinical applicability of generated iPSCs, non-integrative safe strategies of reprogramming should be employed to derive integration-free iPSCs. Alternative approaches are explored with minimal or no genetic modifications of cells such as Sendai viral vectors, adenoviral vectors, adeno-associated viral vectors, plasmid transfection, minicircle vectors, episomal vectors, transposon vectors, synthetic messenger RNAs (mRNAs) transfection, microRNAs (miRNAs), small molecules, and recombinant protein transduction (Figure. 1.3). These techniques obviate the chances of insertional mutagenesis and transgene reactivation.

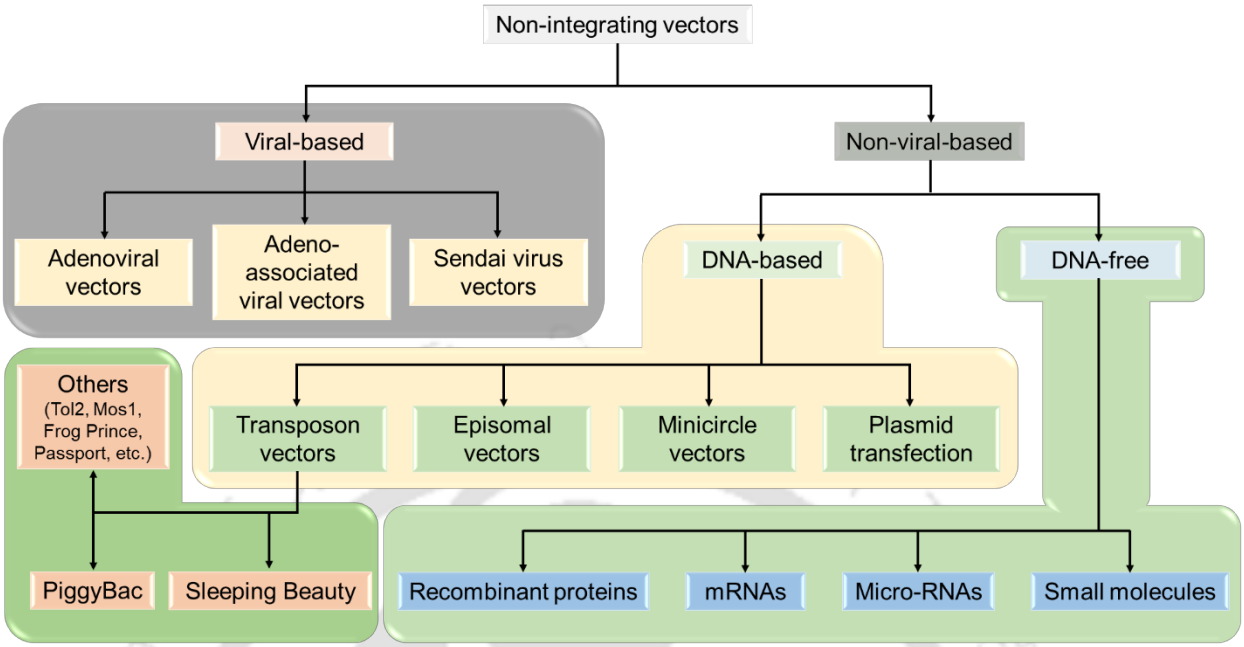


Figure 1.3 Schematic overview of various non-integrative viral and non-viral approaches and their sub-divisions.

Sendai virus vectors

Sendai virus (SeV), a member of the *Paramyxoviridae* family, is successfully used to generate integration-free iPSCs with very high efficiency from a wide variety of cell types. The first study to use SeV vectors showed efficient reprogramming of adult fibroblasts to generate integration-free iPSCs (Fusaki et al., 2009). The second report reprogrammed T cells using a temperature-sensitive mutated SeV vector to derive integration-free iPSCs (Seki et al., 2010). This strategy produced a weaker transgene expression, and this version of the SeV vector could not replicate at typical cell culture conditions. Since then, numerous studies have been reported to generate iPSCs using SeV vectors successfully (Ban et al., 2011; Hamada et al., 2012; Macarthur et al., 2012; Tan et al., 2014, 2018; Jiang et al., 2014; Schlaeger et al., 2014; Trokovic et al., 2014; Fujie., et al. 2014; Kang et al., 2015; Wiley et al., 2016; Haase et al., 2017; Nishimura et al., 2017). The generation of iPSCs in all these studies was accomplished

at around ~25 days with reprogramming efficiencies varying from 0.01 to 4%. The difference in reprogramming efficiencies among these studies could be because of different cell types used for reprogramming, varying stoichiometric expression of reprogramming factors, culture conditions, and/or inclusion of small molecules that promote reprogramming. The SeV vectors have the edge over other viral delivery methods due to the below-listed advantages (Lamb and Parks, 2001; Bitzer et al., 2003; Hosoya et al., 2008; Rao and Malik, 2012; Bayart and Cohen-Haguenauer, 2013; Malik and Rao, 2013; Hu, 2014; Schlaeger et al., 2014; Beers et al., 2015): 1) it is a non-pathogenic virus, 2) it has a single-stranded RNA genome and therefore lacks a DNA phase, which reduces the probability of host genome modifications or gene silencing by epigenetic modifications, 3) it replicates exclusively in the cytoplasm and is incapable of entering the nucleus due to which it does not integrate into the genome of the host cell, 4) its transduction efficiency is very high and has a rapid onset of expression within 24 hours post-transduction, which makes SeV a right gene delivery vehicle, 5) the protein production is very high, which is gradually diluted in approximately ten passages, 6) this vector has a broad tropism as it binds to ubiquitous sialic acid receptors and cellular uptake time is brief, 7) it requires less number of starting cells for infection. Hence, the SeV vectors have been used to derive high-quality iPSCs free of viral contamination; and when compared to other gene delivery strategies, it has the least amount of workload (Schlaeger et al. 2014).

The original SeV vectors have been modified and improved so as to result in better transfection efficiency and improved transgene carrying capacity and notably, these vectors could deliver and express foreign genes in transduced cells and had reduced immunogenicity. Additionally, with the modifications and improved SeV vectors (third generation), the stoichiometry of the reprogramming factors could be modulated precisely to obtain the highest

reprogramming efficiency, and also thus eliminated the problems associated with homologous viral interference (Nishimura et al., 2011; Fujie et al., 2014). Despite these modifications, the biggest concern with utilizing viral delivery systems is the residual presence of viral particles, which might limit its therapeutic use. To overcome this, many strategies have been developed. The first strategy is time-consuming but effective, and it is by passive elimination of the viral RNA genome after few passages (Seki et al., 2010; Ban et al., 2011; Fujie. et al., 2014). The second strategy is to use temperature-sensitive SeV vectors with mutations in the structural genes, which can be eliminated by a simple temperature shift without inducing cytotoxicity (Seki et al., 2010; Ban et al., 2011; Fujie. et al., 2014). An alternative strategy is to use antibodies against Hemagglutinin-Neuraminidase protein, which will allow effective screening of integration-free reprogrammed cells (Fusaki et al., 2009). In addition, small interfering RNA targeting the viral transcription/replication machinery (for, e.g., L gene) can eliminate the viral replicons when desired (Nishimura et al., 2011). A more recent strategy is to utilize the expression of miRNA miR-302 (Nishimura et al., 2017); this miRNA is not expressed in somatic cells but is highly expressed in pluripotent stem cells (Suh et al., 2004; Wilson et al., 2009; Anokye-Danso et al., 2011). These strategies have resulted in the generation of high-quality transgene-free iPSCs that are free of viral particles. In spite of all the developments in the SeV vector, there are still a few concerns that need to be addressed while using this viral delivery method. First of all, being a viral approach and to utilize this viral vector for therapeutic applications, exhaustive screening of iPSC clones is crucial to confirm the absence of any viral remnants and its effect on the genome of iPSCs. Moreover, it is more challenging to work with SeV compared to lentiviruses or γ-retroviruses, as its preparation is highly laborious (Rao and Malik 2012). Additionally, reprogramming using SeV vectors demands

strict biosafety measures, thus increasing the costs of the approach. Furthermore, SeV is fusogenic and immunogenic, but these concerns are ameliorated in the fourth-generation vectors (Yoshizaki et al., 2006; Nishimura et al., 2011). Nevertheless, among all the viral-based reprogramming approaches, the SeV vector is currently the most widespread and versatile for the generation of integration-free iPSCs. This is due to the commercially available SeV-based reprogramming kit, which eliminates the laborious task of virus production, even though it is expensive. Moreover, hiPSCs were generated from a wide range of cell types under xeno-free and/or feeder-free culture conditions using the SeV vector. Notably, this approach results in a low incidence of genomic aberrations and low aneuploidy rates in the reprogrammed cells (Schlaeger et al., 2014; Kang et al., 2015). In conclusion, this reprogramming approach is highly efficient, reliable, has a broad tropism and high adoption rate with the low workload to give rise to genetically stable integration-free iPSCs (Schlaeger et al., 2014).

Adenoviral vectors

Adenoviruses are non-integrating viruses belonging to the family *Adenoviridae*. These viruses are non-enveloped and have an icosahedral nucleocapsid that contains a double-stranded DNA genome (Ginsberg, 2013). They have the property to remain in the epichromosomal form in all cell types, except in egg cells (Shao and Wu 2010). They have a high virus yield and infection efficiency, and an ability to transduce many cell types, including both replicating and non-replicating cells (Carter and Samulski, 2000). These vectors have a high safety profile and account for >400 gene therapy clinical trials (Lee et al., 2017). Using the replication-deficient adenoviral delivery system, transgenes can be delivered and expressed in the host without genomic integration (He et al., 1998; Lee et al., 2017). Replication-incompetent adenoviral

vectors were successfully used to produce mouse iPSCs from tail-tip fibroblasts, hepatocytes and fetal liver cells (Stadtfeld et al., 2008). The expression of reprogramming factors was maintained only for 3–8 days. Therefore, the reprogramming efficiency was extremely low (<0.0001% to 0.001%) as compared to the integrating viral vectors (~0.01 to 0.1%). The low reprogramming efficiency for this method was attributed to low infection efficiency, and to the fact that expression window of reprogramming factors is too narrow (3–8 days) and most cells were unable to retain gene expression for the desired length of time to attain a pluripotent state. iPSCs generated were reported to be free of any viral integration and could contribute to the formation of teratomas and normal post-natal chimeras, but were not able to pass through the germline (Stadtfeld et al., 2008). Concurrently, Yamanaka and colleagues reported the generation of mouse iPSCs by reprogramming hepatocytes using a combinatorial approach of retroviral and adenoviral gene delivery (Okita et al., 2007). iPSCs derived did not exhibit integration of the adenoviral transgene. Surprisingly, the study did not report generation of iPSCs from mouse hepatocytes by introducing the Yamanaka factors in separate adenoviral vectors, probably due to suboptimal individual viral concentrations in each cell. Subsequently, integration-free human iPSCs from embryonic fibroblasts were generated by the introduction of Yamanaka factors with adenoviral vectors (Zhou and Freed, 2009). However, the reprogramming efficiency was very low (0.0002%) even though high viral titers (200–250 pfu/cell) were used. Wu and colleagues linked the four Yamanaka factors to a single reading frame to co-express the genes in the same cells but were unsuccessful to generate iPSCs from MEFs (Shao and Wu 2010). These studies conclude that specific cell types may be responsive to the adenoviral vector infection, whereas other cell types are refractory to reprogramming using the same delivery system. This is mainly due to the varied tropism of these vectors (J

Schneider-Schaulies 2000; Lee et al., 2017) or an immunologic response generated against these vectors by certain cell types (Howarth et al., 2009; Nayerossadat et al., 2012; Lee et al., 2017). Moreover, few concerns need to be addressed before this vector can be routinely used to derive iPSCs. The major drawback limiting the use of adenoviruses in reprogramming is the transient expression due to rapid clearance from dividing cells (Shao and Wu, 2010). The inconsistent and insufficient expression of reprogramming factors in each cell for a prolonged duration due to gene silencing mechanism is another drawback (Shao and Wu, 2010; Lee et al., 2017). To overcome these limitations, use of an inducible or polycistronic adenoviral vector that encodes reprogramming factors under the control of a suitable promoter will improve the induction efficiency (Chen et al., 2015). This is possible due to the large insert capacity (>8 kb) of this vector (Lee et al., 2017). Another major concern is the integrative nature of these vectors. Though integration-free iPSCs have been derived using adenoviral vectors as mentioned earlier (Stadtfeld et al., 2008; Zhou and Freed, 2009), some studies do report rare chances of integration of adenoviral vectors into the mammalian genome (Harui et al., 1999; Zheng et al., 2000; Wang et al., 2005). Moreover, the generation of iPSCs using adenoviruses required high viral titers (Zhou and Freed, 2009), and this can further increase the chance of integration. Additionally, adenoviral vectors upon integration may undergo rearrangements (Harui et al. 1999), making it challenging to detect iPSC clones free of foreign DNA sequences. Therefore, iPSC clones are required to be thoroughly screened using the next generation sequencing technologies to confirm the absence of integration and detect any genomic alterations (Yamanaka, 2009). Also, the toxicity and associated immune responses further restrict their use in iPSC generation (Hartman et al., 2008; Gregory et al., 2011; Lee et al., 2017).

Adeno-associated viral vectors

Adeno-associated viruses (AAV) are non-pathogenic, non-autonomous, single-stranded DNA virus that belongs to the family *Parvoviridae* (Siegl et al., 1985). AAV requires a helper virus for its replication, and in the absence of the helper virus, its genome stays episomal in the host cells. Replication-incompetent AAV vectors derived from AAV are deficient in viral coding sequences and can infect both replicating and non-replicating cells. They are not reported to cause any immune or toxic reactions in the host (Zaiss and Muruve, 2005), and hence make it a promising gene delivery system for clinical applications. To date, >100 gene therapy clinical trials for various diseases with notable successes have been conducted using these vectors (Hirsch et al., 2016). However, the limited packaging capacity (~5 kb) and the occurrence of genomic integration at a very low frequency are the drawbacks of AAV-based gene delivery system (Hirsch et al., 2016; Lee et al., 2017). Fragment AAV vector transduction and split vector approaches (trans-splicing, overlapping and hybrid vectors) were constructed to overcome the packaging limitations (Hirsch et al. 2010, 2016). Generation of mouse iPSCs using recombinant AAV vectors encoding Yamanaka factors has been reported (Weltner et al., 2012). Surprisingly, they could not generate hiPSCs using the same gene delivery vehicle (Weltner et al., 2012), probably due to a high and persistent expression of reprogramming factors that prevents proper stabilization to a primed pluripotent state. Unexpectedly, the study reported frequent stable genomic integration of the transgenes even at a lower multiplicity of infection with a continuous expression of the reprogramming factors in mouse iPSC clones. The reprogramming efficiency was 0.001 to 0.006% when the transgenes were expressed under the regulation of the cytomegalovirus (CMV) promoter and 0.003 to 0.09% when CMV early enhancer/chicken β-actin (CAG) promoter was used. The reason for observed integrations in

iPSC clones even at lower titers may be due to the occurrence of genomic instability during cell reprogramming that results in the acquisition of chromosomal abnormalities such as insertions, deletions and amplifications (Mayshar et al. 2010; Laurent et al. 2011; Gore et al. 2011; Hussein et al. 2011; Martins-Taylor et al. 2011). Various strategies such as reprogramming a somatic cell source of low passage that is devoid of deleterious mutations, inclusion of genomic stability promoting reprogramming factors (e.g. ZSCAN4) in the cocktail, minimizing oxidative stress associated with reprogramming, and small molecules that enhance reprogramming by temporary inhibition of signaling pathways (such as senescence, TGF-β, certain kinases, etc.) have been employed to minimize genetic instability during cell reprogramming (Yoshihara et al., 2016). This will drastically minimize chromosome breakage/deletion during reprogramming and thereby prevent integration of AAV vectors in the target cell genome. Further investigations using modified recombinant AAV constructs such as split AAV vectors (overlapping, trans-splicing, and hybrid trans-splicing) and fragment AAV vectors can be performed as these can overcome the packaging limitations (Hirsch et al. 2010, 2016). Along with modification of AAV vectors, use of small molecules have been reported to enhance their transduction efficiency (Nicolson et al. 2016). Thus, more robust and detailed analysis of the use of AAV vectors in somatic cell reprogramming is required.

Non-viral DNA-based vectors

Plasmid transfection

One of the most elementary methods of exogenous gene expression requires the use of plasmid vectors as they are not vulnerable to exonucleases compared to linear DNA (McLenachan et al., 2007). Yamanaka and colleagues used two separate plasmids: one containing OCT4, SOX2 and KLF4, and the other have c-MYC to derive mouse iPSCs by repeated transfections of the

plasmids over seven days (Okita et al., 2007). Both the plasmid vectors were under the control of a constitutively active CAG promoter. The iPSCs derived were morphologically similar to ESCs and expressed its pluripotency markers at similar levels. Importantly, no ectopic plasmid integration was detected in the code for Thomson reprogramming factors and were adequate to derive iPSCs (Si-Tayeb et al., 2010). This study used plasmid vectors similar to those used to produce lentiviruses (pSin vectors), interestingly, the lack of packaging plasmids from the transfection excluded any possibilities to produce virions. This study generated iPSCs with a reprogramming efficiency of 0.00033% in 4–5 weeks (Si-Tayeb et al. 2010).

Instead of delivering reprogramming factors in separate plasmids into cells, nucleofection of a single polycistronic vector encoding Yamanaka factors under the control of a CAG promoter was reported to derive transgene-free iPSCs from MEF cells (Gonzalez et al. 2009). However, the reprogramming efficiency was very low compared to integrating viral methods. Using site-specific recombination strategy, transgene-free mouse iPSCs were generated from two accessible cell sources (fibroblasts and adipose-derived mesenchymal stem cells) with a reprogramming efficiency of ~0.01% by nucleofection of polycistronic plasmids encoding Yamanaka factors (Karow et al., 2011).

Plasmid transfection is a simple, non-viral approach of episomal nature employed to generate transgene-free iPSCs. Reprogramming factors delivered into cells in separate plasmids are laborious and less efficient as only a few cells receive the complete cocktail of reprogramming factors. The development of polycistronic vectors encoding reprogramming factors in a single vector has made this technique simpler and attractive. The use of picornaviral 2A self-cleaving peptides to link reprogramming factors to allow stoichiometric co-expression of multiple reprogramming factors in a polycistronic construct from the same promoter has

gained acceptance due to its small size and availability of different functional variants (Trichas et al. 2008; Luke et al. 2013). In a cell reprogramming paradigm, a balanced stoichiometric and temporal expression of reprogramming factors is crucial for the induction and maintenance of pluripotency (Sridharan et al., 2009; Papapetrou et al., 2009; Tiemann et al., 2011; Schmitt et al., 2017), and this greatly affects the epigenetic status and the biological properties of the reprogrammed cells (Carey et al., 2009; Tiemann et al., 2011). However, there are various concerns encompassing the use of polycistronic vectors such as: 1) unbalanced expression of each reprogramming factor that could destabilize the stoichiometry and thereby compromise efficient reprogramming (Wen et al., 2016), 2) using cationic lipids as a transfection reagent or by nucleofection, 3) variation in the transfection efficiency from cell type to cell type due to various reasons (Maurisse et al., 2010), 4) reduced transfection efficiency due to its large size; the delivery of large plasmids into mammalian cells is compromised compared to small plasmids (McLenachan et al., 2007; Chabot et al., 2012), 5) due to the transient expression of transgenes, efficient cell reprogramming requires repeated transfections of reprogramming plasmid(s) to generate iPSCs which is very stressful to cells (Maurisse et al. 2010). Multiple transfections are not possible with nucleofection due to their high cytotoxicity and requirement of detachment of adherent cells (Hu, 2014a), 6) the promotion of integration by certain nucleofection reagents by delivering plasmid directly into nuclei (Gonzalez et al. 2009; Montserrat et al. 2011), which eliminates the sole purpose of using plasmid transfection to derive transgene-free iPSCs, 7) another major concern with the use of plasmid transfection is the presence of bacterial backbone sequences that either generates an immune response against unmethylated CpG dinucleotides (Li et al., 1999; Yew et al. 2000, 2002) or result in rapid

silencing of transgenes (Chen et al., 2003). Therefore, minicircle vectors were developed to obviate these limitations and provide longer and stable expression of transgenes.

Minicircle vectors

Minicircle vectors are supercoiled DNA episomal vectors that resemble a standard plasmid but lack both antibiotic resistance gene and origin of replication of the bacterial backbone and contain largely a eukaryotic expression cassette (Darquet et al., 1997). Minicircle vectors are relatively smaller in size, resistant to shear, have superior transfection efficiency and are less prone to transcriptional silencing resulting in prolonged ectopic expression of the transgene compared to conventional plasmids (Darquet et al. 1997; Chen et al. 2003; Catanese et al. 2011; Chabot et al. 2012). In addition, these vectors are non-integrating, easily synthesizable and have precise control over concentration and time of application. They are easy to deliver into cells, even in cells that are refractory to plasmid transfection. However, like conventional plasmids, minicircle vectors cannot self-replicate and therefore the expression time is not as long as for episomal vectors (Jia et al. 2010). Perhaps these vectors get diluted upon cell division and are removed from cells, eventually resulting in the derivation of transgene-free iPSCs. Studies reported successful generation of iPSCs from human adipose stem cells using a polycistronic construct containing green fluorescent protein reporter gene and Thomson reprogramming factors with an efficiency of 0.005% (as compared to the integrating viral vectors (~0.01 to 0.1%)) and reprogramming period of 14–18 days (Jia et al., 2010; Narsinh et al., 2010). The resulting iPSCs were also reported to be transgene-free as confirmed by Southern blot analysis (Jia et al., 2010). However, this method requires multiple transfections that affect cell viability and is marred by low reprogramming efficiencies. To improvise, a codon optimized 4-in-1 minicircle was reported to convert human fibroblasts to iPSCs by a

single transient transfection under feeder-free conditions using a chemically defined media (Diecke et al., 2014). Nevertheless, the reprogramming efficiency was still extremely low (0.005%). The same group later developed a robust transgene expression minicircle vector that encodes for codon optimized Yamanaka factors and short hairpin RNA against the p53 gene to derive mouse and human iPSCs by single transfection, however, this strategy did not markedly increase the reprogramming efficiency (Diecke et al., 2015). iPSCs generated using minicircle vectors must be meticulously screened to confirm non-integration of transgene sequences. Few steps that could promote reprogramming efficiency and allow more cell types to be reprogrammed using minicircle vectors are: 1) use of other efficient delivery techniques such as electropulsation (Chabot et al. 2012), 2) incorporation of a greater number of reprogramming factors and/or microRNAs (Brouwer et al., 2015), 3) inclusion of reprogramming enhancing small molecules (Ma et al., 2017), and 4) reprogramming carried out in hypoxic conditions (Yoshida et al., 2009). Therefore, further refinement is required to make mini-circle technology more appealing and relevant for clinical translation.

Episomal vectors

Episomal vector comprises of two components of Epstein-Barr Virus namely, a sequence encoding a trans-acting factor Epstein-Barr Virus Nuclear Antigen 1 (EBNA1) and a cis-acting viral origin of replication (oriP) element (Yu et al., 2009; Okita et al., 2011, 2013). The EBNA1 encodes a protein that gets expressed from its viral promoter after transduction into somatic cells (van Craenenbroeck et al., 2000). Subsequently, the protein recognizes the oriP and initiates plasmid amplification. The EBNA1 is essential and adequate for stable episomal maintenance and replication of episomal vectors in various established cells (van Craenenbroeck et al., 2000). As a result, episomal plasmids have many advantages compared

to other plasmids: 1) the gene of interest delivered is not subject to regulatory constraints due to non-integration, 2) they have high-level of transgene expression as a result of vector amplification by only single transfection, 3) the maintenance of transgene expression of the replicating episomal plasmids and high protein expression in a short duration, and 4) this approach does not manipulate the genome.

Cell reprogramming carried out using conventional plasmids and minicircle vectors involve multiple transfections due to their inability to replicate in mammalian cells. In contrast, episomal vectors require only a single transfection for long-term stable expression. This vector facilitates easy delivery and replicates autonomously in synchronous with the host genome and remains as an extrachromosomal element without integration in both replicating and non-replicating cells (van Craenenbroeck et al. 2000). By culturing cells in the absence of drug selection, episomes are progressively lost at a rate of ~5% per cell division due to errors in plasmid replication and partition, resulting in the generation of iPSCs free of genomic integration or genetic alterations (Nanbo et al., 2007).

To obviate plasmid dilution due to cell division and requirement of multiple transfections very common to conventional plasmid transfection, human postnatal foreskin fibroblasts were reprogrammed to iPSCs by a single transfection of non-integrating oriP/EBNA1-based episomal vectors (Yu et al. 2009). In addition to the Yamanaka factors, the study used NANOG, LIN28 with Simian Virus 40 Large T antigen (SV40 LT) in three different vector combinations. The proliferative capacity and developmental potential of iPSCs derived were comparable to human ESCs and devoid of the transgene and vector sequences. Nonetheless, the reprogramming efficiency was quite low and required additional chemical compounds to enhance the efficiency (Yu et al. 2009). A major concern with the study was the

use of reprogramming booster SV40 LT in the episomal vector, which is an oncoprotein function primarily by inactivating p53 and retinoblastoma proteins resulting in the formation of iPSCs with higher tumorigenic potential (Ahuja et al., 2005; González et al., 2011). This immensely decreases the safety of iPSCs for their clinical application. Various studies later reported an increase in reprogramming efficiency using this vector system: 1) making modifications in the episomal vector, 2) alterations in the cocktail of reprogramming factors or media conditions, 3) the inclusion of small molecules, or 4) using easily reprogrammable and accessible cell sources for iPSC derivation (Yu et al., 2011; Chen et al., 2011; Slamecka et al., 2016).

In two independent studies, Yamanaka and colleagues used transcription factor L-MYC and p53 suppression along with OCT4, SOX2, KLF4 and LIN28 to boost reprogramming efficiency (Okita et al., 2011, 2013). They replaced c-MYC and NANOG from the cocktail by L-MYC (Okita et al., 2011, 2013), due to its specificity, potency and non-transforming characteristic (Nakagawa et al. 2010). Other studies also reported increase in reprogramming efficiency by modifying the episomal vector containing the five transcription factors, OCT4, SOX2, KLF4, c-MYC and LIN28, together with an additional EBNA1/OriP plasmid for an ephemeral expression of SV40 LT (Hu et al., 2011; Chou et al., 2011). Further, Dowey and co-workers used a single polycistronic episomal vector expressing reprogramming factors as a single entity linked by 2A cleavable peptide sequences to generate human iPSCs (Dowey et al., 2012). However, the reprogramming efficiency using this strategy to derive human iPSCs was low (Dowey et al., 2012), compared to other reported studies (Yu et al., 2009; Chou et al., 2011; Okita et al., 2011, 2013). To improvise on this, Wen et al. reported a ~100-fold enhancement in reprogramming efficiency by fine-tuning the episomal vector and

its respective transcription factor stoichiometry (Wen et al., 2016). Thus, demonstrating the importance of optimal stoichiometry of reprogramming factors for successful and enhanced reprogramming.

Transposon vectors

Transposons are advanced non-viral vectors that can potentially avert the limitations of routinely used integrating viral vectors and naked DNA molecules. They have higher transfection efficiency and low innate immunogenicity than linearized plasmids (Hu, 2014a). They are versatile vehicles for large cargoes (~10 kb) that can stably integrate non-viral constructs into the target cell genome by single transfection and enable robust and persistent expression of desired genes (Tipanee et al., 2017b, a). Stable transposition entails the inclusion of the transposon DNA with the respective transposase gene, mRNA, or protein. These mobile DNA elements are host factor independent and functional in a number of human as well as mouse cell lines (Wu et al. 2006; Kumar et al. 2015). In addition, they are inexpensive, easy to purify and deliver into cells, and allow removal of the transgene cassette without leaving any prominent genetic modifications. Therefore, transposons provide an avenue to be used as a safe, efficient and integration-free approach for derivation of therapeutically safe iPSCs. The most promising transposons currently used are piggyBac (PB) or Sleeping Beauty (SB). Both PB- and SB-based transposons have been employed for stable expression of reprogramming factors and successfully reprogram mouse (Woltjen et al. 2009; Muenthaisong et al. 2012; Grabundzija et al. 2013; Talluri et al. 2014) and human (Woltjen et al., 2009, 2011; Davis et al., 2013) fibroblasts into iPSCs. Other transposons such as Tol2, Mos1, Frog Prince and Passport are also active in mammalian cells (Wu et al., 2006; Kumar et al., 2015), but are still unexplored in the generation of iPSCs.

Typically, these transposon systems consist of a single polycistronic transcript encoding reprogramming factors linked by 2A peptides. Subsequently, this permits post-translational cleavage of the polyprotein into distinct reprogramming proteins to convert somatic cells to iPSCs. The linking of the transgenes with 2A peptides permits stochiometric co-expression of proteins from a single transcript through a ribosomal skipping mechanism, and drastically curtails the number of integration sites in the somatic cells (Hu, 2014a). Importantly, the unique feature of the transposon is to reintroduce the corresponding transposase by transient transfection giving rise to seamless removal of the reprogramming cassette from the generated iPSCs to obtain transgene-free iPSCs (Belay et al., 2012; Kumar et al., 2015). However, it requires multiple rounds of excision during the transposition reaction and reintegration of the transposon into the genome can occur in cells (Wang et al., 2008; Ye et al., 2014). Reintegration can be avoided by utilizing an excision competent/integration defective transposase enzyme (Li et al., 2013). In addition, the transposition reaction is not always accurate. Although 95% of genomic transposon excision events were reported to be precise in mouse ESCs, 5% of the transpositions had genomic alterations (Wang et al. 2008). Frequent transposition events into unknown sites before removal can result into footprint mutations, microdeletions as well as chromosomal rearrangements in the genome of the cells (Geurts et al., 2006; Wang et al., 2008). This results in a laborious screening procedure to identify integration-free iPSCs having an intact genome. As a result, the expression window of transposase should be tightly controlled to achieve traceless excision without inducing any cytotoxicity and genomic alterations (Galla et al. 2011).

PiggyBac (PB) transposon vectors

PB is a highly active transposon vital for efficient gene transfer and stable transgene expression in mammalian cells (Wu et al., 2006; Doherty et al., 2011; Saha et al., 2015; Chen et al., 2015). A combination of non-viral transfection approach comprising of a single polycistronic expression vector encoding Yamanaka factors with a PB transposon system succeeded in reprogramming human embryonic fibroblasts to generate iPSCs (Kaji et al. 2009). However, the reprogramming efficiency using this combinatorial approach was only 0.02–0.05%. This system was then solely employed to successfully derive mouse (Woltjen et al., 2009, 2016; Yusa et al., 2009; Tsukiyama et al., 2011, 2014; Bertin et al., 2015; Behringer et al., 2017) and human (Kaji et al., 2009; Woltjen et al., 2011; Igawa et al., 2014) iPSCs. However, low reprogramming efficiency acts as a persistent problem with the use of PB systems. It has been shown that a single transfection of multiple PB vectors containing early transposon promoter with enhancers of OCT4 and SOX2 and polycistronic doxycycline-inducible factors could generate high-quality iPSCs from mouse somatic cells (Tsukiyama et al., 2011).

Presence of an integrated reprogramming cassette in the derived iPSCs limits the biomedical potential and affects the differentiation potential (Somers et al., 2010; Ramos-Mejia et al., 2012; Kadari et al., 2014). In addition, preserving genomic integrity is vital for therapeutic applications of iPSCs (Simara et al., 2017). Taking this into account, transgene-free iPSCs were derived by transient re-expression of PB transposase, leaving the genome of iPSCs intact without any genomic alteration (Woltjen et al., 2009, 2011; Yusa et al., 2009). Notably, a unique hyperactive PB transposase was engineered to derive mouse transgene-free iPSCs with high efficiency without affecting the genomic integrity of the target cells (Yusa et al. 2011). The hyperactive PB transposase offered 9-fold and 17-fold improvement in

integration and excision, respectively (Yusa et al. 2011), when compared to the mammalian codon optimized mouse PBase (Yusa et al., 2009). This can enhance the removal of the transgene cassette more easily and efficiently. The iPSCs were generated from human dermal fibroblasts using the above system to determine the effect of exogenous reprogramming factors (Igawa et al. 2014). The transgene-free iPSCs were compared with transgene-retaining iPSCs, thereby maintaining the common isogenic genetic background. Further, the study reported that the differentiation potential of transgene-retaining iPSCs into keratinocytes was severely perturbed due to transcriptional reactivation of the transgenes. This suggested that the removal of exogenous genes was critical for proper differentiation into desired derivatives for its clinical application. Recently, a single "All-in-One" PB transposon vector was constructed for drug-inducible transgene regulation incorporating selection markers and different reporter genes to derive and confirm clones or populations of human iPSCs (Kim et al., 2016). Further, combining this system with the potent CRISPR/Cas9 technology, patient-specific iPSCs by gene correction (Xie et al., 2014) and footprint-free iPSCs were generated (Wang et al. 2017). In short, PB performs scarless excision without any footprint to generate genetically unmodified safe and clean iPSCs.

Sleeping Beauty (SB) transposon vectors

An alternative gene delivery system to PB transposon system is SB transposon (Grabundzija et al., 2010). The SB integrates randomly at the genomic level and does not show any propensity towards specific genes and gene-regulatory elements (Mátés et al., 2009). This is due to the absence of SB-like elements within the human genome, curtailing the odds of cross-mobilization of endogenous and exogenously delivered transposon elements and in that way diminishes the chances of insertional oncogenesis (Grabundzija et al., 2010). In addition, the

ITRs have negligible promoter/enhancer activity, and thus they are unable to initiate transcription of genes that flank the integration sites (Ivics and Izsvák 2010). This system is highly efficient even in hard-to-transfect cells.

Exploiting the advantages of this transposon system, mouse fibroblasts were reprogrammed by non-viral transfection of an SB-based transposon containing the polycistronic reprogramming cassette comprising of Yamanaka factors linked with 2A peptides in the presence of a transposase (Muenthaisong et al., 2012). The group utilized a hyperactive transposase SB100X that delivers significantly higher gene transfer efficiencies similar to integrating viral vectors, resulting in efficient integration and expression as reported earlier (Mátés et al., 2009). Notably, the study was able to generate iPSCs from fibroblasts isolated from three different genetic backgrounds, including backgrounds from where the derivation of mouse ESC lines has been challenging (Muenthaisong et al., 2012). Using a similar approach, human iPSCs from fetal fibroblasts with a reprogramming efficiency of ~0.02% were generated (Davis et al., 2013). However, both these studies did not show excision of the integrated reprogramming cassette by *de novo* expression of the transposase or by using recombinase-mediated excision from the target cell genome, to establish derivation of transgene-free iPSCs. To demonstrate this, an elegant study reported the derivation of mouse and human iPSCs using this SB transposon gene delivery system (Grabundzija et al., 2013). The study showed complete removal of the reprogramming cassette by Cre recombinase-mediated elimination in mouse iPSCs. Moreover, the study also reported a targeted integration of a reprogramming cassette into the transposon-tagged insertion locus in mouse iPSCs. This method can permit correction of any defect by a site-specific introduction of a therapeutic gene construct into a safe harbor chromosomal locus in the genome of a patient-derived, autologous

iPSCs. Additionally, the reprogramming efficiency to derive human iPSCs was 0.02% (Grabundzija et al., 2013), similar as reported earlier both using Yamanaka factors and the same SB-transposon system (Davis et al., 2013).

Subsequently, murine fibroblasts isolated from two different genetic backgrounds were reprogrammed to derive iPSCs to compare the reprogramming efficiency of PB- and SB-based transposon systems (Talluri et al., 2014). Both systems were successful in generating iPSCs with comparable efficiencies, 0.027% being the highest. iPSCs were even generated in reprogramming experiments lacking the proto-oncogene c-MYC, but with reduced efficiency. To summarize, both these transposon systems permit the removal of the reprogramming cassette and can be efficiently used for derivation of integration-free iPSCs in a safe, simple and straightforward manner for biomedical applications.

Non-viral DNA free approaches

mRNA transfection

Messenger RNAs (mRNAs) are type of RNA molecules which translates the genetic information to protein from DNA via the processes, namely transcription and translation. Delivery of modified mRNAs encoding reprogramming factors into somatic cells is a straightforward approach to establish genetically stable, integration-free iPSCs (Steichen et al., 2014). The modified mRNAs (also called in vitro transcribed mRNA or synthetic mRNAs) are generated through in vitro transcription (IVT) methods and has become a potent alternative genetic material to study gene expression, and also for therapeutic applications such as protein replacement, vaccine development, and regenerative medicine.

mRNA approach has several advantages over conventional plasmid DNA- and viral-based approaches (Gómez-Aguado et al., 2020; Kowalski et al., 2019; Kwon et al., 2018): 1)

no nuclear machinery required to mediate translation, 2) directly encodes for the protein of interest in the cytoplasm through a single translation step in both dividing and non-dividing cells, 3) rapid expression of the protein of interest, 4) high level of protein expression comparable to viral-based approaches, 5) reasonable control over time and dosage of application, 6) metabolically decaying within a few days, therefore has a transient expression (ideal from a reprogramming perspective), 7) delivered into cells using transfection reagents or nanoparticles obviating the need for electroporation, 8) relatively rapid, simple and easy to synthesize, 9) no requirement of genes encoding for selection (e.g., antibiotic-resistance), 10) no screening procedure required in transfected cells for any remnants due to short half-life, 11) no possible risk of genomic integration and insertional mutagenesis into the host genome, therefore has a better safety profile.

The initial studies used in vitro synthesized non-modified mRNAs to deliver these synthetic mRNAs encoding for reprogramming factors into somatic cells to generate iPSCs (Plews et al. 2010; Yakubov et al. 2010). However, non-modified mRNAs employed triggered an immune response and stimulated cellular defense mechanisms affecting mRNA translation and cell viability (Drews et al., 2012). Interestingly, this was not observed in the case of the delivery of modified mRNAs into somatic cells (Drews et al., 2012). Hence, use of modified mRNAs and inhibition of innate immune response is crucial to facilitate multiple transfections of RNA encoding for reprogramming proteins (Angel and Yanik, 2010; Drews et al., 2012).

Taking these important aspects into consideration, Warren and colleagues modified the mRNAs encoding for reprogramming proteins (Yamanaka factors and Lin28a) to reduce the innate immune response, and improve mRNA stability and translation efficiency (Warren et al., 2010). The authors achieved this by slightly modifying the sequences of the five

reprogramming factors, where uridine was substituted to pseudouridine (psi) and cytidine to 5-methylcytidine (5mC). This modification can promote translation and reduce immunogenicity (Kwon et al., 2018). Various other modifications have been done to enhance the efficiency and reduce the immune response (Warren et al., 2010). Due to these modifications, the authors were able to reprogram human fibroblasts and keratinocytes to iPSCs in two weeks by performing daily transfections of Yamanaka factors (Warren et al., 2010). The reprogramming efficiency achieved was 1.4% when compared to 0.04% for retroviral-based method, and the reprogramming kinetics achieved was 13–15 days compared to 25–29 days for the retroviral-based approach. Further, the authors demonstrated that the inclusion of Lin28a to Yamanaka factors resulted in higher efficiency (i.e., more than 2.5%) at both low (5%) and ambient (20%) oxygen conditions. Using a similar approach (Warren et al., 2010), a subsequent study used six reprogramming factors (Yamanaka factors + LIN28a + NANOG) to derive human iPSCs after 9 days of transfection (Preskey et al., 2016). All these modifications helped in generating iPSCs (Choi et al. 2016; Rohani et al. 2016).

Modified or non-modified in vitro transcribed RNA demands daily transfections that exert stress on the cells during the course of reprogramming. In addition, the use of transfection reagent also induces stress to the cells. Therefore, to significantly reduce the number of transfections to establish pluripotency, Tavernier and co-workers reprogrammed MEFs with only three mRNA transfections (Tavernier et al., 2012). The authors employed cationic lipid mediators to deliver mRNAs encoding Yamanaka factors into MEFs. The emerged cell-clusters were alkaline phosphatase positive and expressed pluripotency markers, NANOG, and SSEA-1.

Many improvements have been carried out by various groups (Figure 1.6) since the first study reporting synthetic mRNA mediated reprogramming of somatic cells (Yakubov et al., 2010). Despite these modifications, there are still a few concerns that need to be addressed (Kwon et al., 2018; Kowalski et al., 2019; Gómez-Aguado et al., 2020). mRNA produced by IVT-method: 1) requires the usage of transfection reagent for delivery, which exerts stress on cells, 2) has limited stability in the cell, therefore requires repeated transfections, 3) is likely to develop an immunogenic response, 4) is expensive to produce, 5) more efficient purification procedures are required to decrease immunogenicity, and 6) may form secondary structure (e.g., in UTR) or can become a target of a miRNA or intracellular defense mechanisms, which can severely affect its translation.

Despite these concerns, mRNA generated via IVT-method is used now by several labs to derive integration-free iPSCs (Mandal and Rossi, 2013; Steichen et al., 2014; Yoshioka and Dowdy, 2017). This approach is highly efficient (Goh et al., 2013; Schlaeger et al., 2014; Warren et al., 2010), and could also derive human iPSCs from somatic cells isolated from healthy subjects and/or patients above 40 years (Goh et al., 2013; Arnold et al., 2012; Heng et al., 2013; Yoshioka and Dowdy, 2017). Notably, the established human iPSC lines derived using this approach showed very low aneuploidy rates, and no significant copy number variations (Schlaeger et al., 2014). One of the other major issues is the requirement of feeder cells and the media comprising of animal-derived components, which brings a risk of transmitting undetected human pathogens. To avert it, numerous labs have established iPSCs efficiently using this approach under xeno-free/feeder-free conditions (Warren et al., 2012; Heng et al., 2013; Warren and Wang, 2013; Goh et al., 2013; Durruthy-Durruthy et al., 2014; Durruthy and Sebastiano, 2015; Kogut et al., 2018), and this can be translated for clinical

translation of iPSCs soon. In conclusion, the mRNA-based reprogramming approach has immense potential to derive genetically stable, foot-print free iPSCs for clinical applications.

MicroRNAs

miRNAs are naturally occurring, non-protein coding, small (around ~22 nucleotides in length) RNAs known to regulate gene expression post-transcriptionally and thereby targeting a wide array of protein-coding mRNAs. They primarily do so either by inhibiting, or degrading them or both. Notably, each miRNA can base-pair and prevent translation of various mRNAs, and on the contrary, numerous different miRNAs can target and collectively regulate a single mRNA target. Therefore, expression of one or more miRNA(s) could drastically alter the gene expression signature and identity of a cell.

Various miRNAs have been reported to act as a barrier to the induction of pluripotency. On the contrary, numerous miRNAs are reported to favor the reprogramming process. In general, the miRNAome of iPSCs is similar to ESCs (Wilson et al., 2009b; Chen et al., 2012), and therefore the miRNAs that are highly expressed in ESCs play a crucial role in the derivation of iPSCs (Hu, 2014). In addition, these miRNAs can enhance the reprogramming efficiency when used along with core reprogramming factors (Leonardo et al., 2012; Lee et al., 2016; Zeng et al., 2018), or act as their substitutes (Pfaff et al., 2011; Judson et al., 2013), or reprogram differentiated cells without incorporation of any exogenous reprogramming factors (Anokye-Danso et al. 2011b; Miyoshi et al. 2011). Importantly, a substantial decrease in reprogramming efficiency was observed upon the suppression of miRNA biogenesis regulators, Dicer, Drosha and Ago2, shedding light on the significance of these regulators in the cellular reprogramming paradigm (Li et al., 2011; Kim et al. 2011; Chen et al., 2012).

Expressing a cluster of miRNAs, two independent labs reported the derivation of mouse and human iPSCs without the need of a single exogenous reprogramming factor (Anokye-Danso et al., 2011a; Miyoshi et al., 2011). Morrisey lab generated iPSCs with higher efficiency and faster kinetics by lentiviral delivery of pluripotent-specific miRNAs, miR-302(b/c/a/d)-miR-367 cluster, in mouse and human fibroblasts (Anokye-Danso et al. 2011a). Notably, numerous other studies had shown that this miRNA cluster or a member of this cluster promoted mesenchymal to epithelial transition and improved efficiency and kinetics when it was used in combination with core reprogramming factors (Card et al., 2008; Judson et al., 2009, 2013; Subramanyam et al., 2011; Liao et al., 2011; Lu et al., 2012; Hu et al., 2013; Zhang and Wu, 2013; Lee et al., 2013; Kogut et al., 2018). Furthermore, this miRNA cluster is a direct target of OCT4 and SOX2 (Card et al., 2008). Importantly, the genetic deletion of this miRNA cluster prevented the derivation of human iPSCs (Zhang et al., 2013), clearly indicating the importance of this cluster for efficient reprogramming.

Reprogramming by a multitude of miRNAs that modulate various biological processes and target lineage-specific or pluripotency-inhibiting genes in somatic cells could contribute to the induction of pluripotency to derive iPSCs. This can be achieved by the delivery of key mature synthetic miRNAs in an integration-free manner using a transfection reagent (Miyoshi et al. 2011). However, the disadvantage of this approach is the stability of the mature synthetic miRNAs and their off-target effects, thereby affecting gene expression in an undesirable manner. Alternatively, miRNA mimics (double-stranded RNA molecules that imitate mature miRNA duplexes) can be employed that imitates endogenous miRNAs to perform gene silencing (Samavarchi-Tehrani et al., 2010; Subramanyam et al., 2011; Judson et al., 2013). In miRNA mimics, chemical modifications (absent in endogenous miRNAs) or nucleotide

alterations in the passenger strands are generally incorporated to increase their stability, and enable guide miRNA loading to RNA-induced silencing complex, and to selectively exclude the passenger strand (Jin et al., 2015; Søkilde et al., 2015). However, miRNA mimics should be used with caution (Jin et al., 2015); therefore, proper controls and validation is required before its use (Søkilde et al. 2015). Using miRNA mimics, the gene silencing of lineage-specific or pluripotency-inhibiting gene(s) in somatic cells can be achieved efficiently to promote reprogramming. Importantly, transfection-based delivery of mature miRNA or miRNA mimics will be transient as the cells will eliminate the delivered miRNAs via the metabolic decay process; therefore, complex elimination procedures (applied in case of integrating viral-based or transposon vectors) can be avoided. However, the disadvantage of transfection-based delivery is that repeated transfections of miRNAs to obtain and maintain the desired gene silencing for a specific duration is essential, which might exert stress to the cells. In addition, the mature miRNA or miRNA mimics delivered could generate an innate immune response. Notably, there is no chance of integration of miRNAs into the genome when delivered using a transfection reagent. Therefore, this approach is ideal to derive integration-free iPSCs due to their small size, non-integrating nature, and easy to synthesize and deliver into cells (Chang and Gregory 2010), leaving no genetic footprint in the iPSCs generated.

Small molecules

Small molecules are chemically synthesized, structurally stable, non-immunogenic, and inexpensive chemical compounds that impart advantages such as easy synthesis and delivery into the cells. In the cell reprogramming paradigm, it imparts complete control over the time and dosage of application to generate *bona fide* iPSCs. The primary function of these chemical compounds lies in modulating the chromatin modifications, targeting signaling pathways,

senescence regulators, metabolism modulators, which are specific to induce pluripotency, thus regulating the cellular functions to favor the cellular reprogramming process. Numerous small molecules have been identified in the context of reprogramming. Each of these can belong to one or more different groups based on their effect in the reprogramming paradigm: 1) small molecules that can enhance reprogramming efficiency and/or kinetics, 2) small molecules that can substitute a reprogramming factor in the reprogramming cocktail, 3) the combination of small molecules that can induce somatic cell reprogramming in the absence of any exogenous reprogramming factors (Feng et al., 2009).

The first study to demonstrate the positive effect of small molecules in a reprogramming paradigm comprised of chromatin modifiers, namely Valproic Acid (VPA) and 5-Azacitidine (5-AZA; a DNA methyltransferase inhibitor (DNMTi)). These chemical compounds enhanced the reprogramming efficiency by 100- and 10-fold, respectively (Huangfu et al., 2008). Furthermore, the DNMTi, 5-AZA, improved the reprogramming efficiency up to four-fold in MEFs by expediting the transition of partially reprogrammed iPSCs to accomplish full pluripotency (Mikkelsen et al., 2008). In addition, other groups have also identified chemical compounds that promote reprogramming efficiency, to some extent via epigenetic alterations. Vitamin C, an antioxidant water-soluble molecule, showed significant enhancement in the reprogramming efficiency and kinetics in both mouse and human somatic cells (Esteban et al., 2010). It worked by removing the reactive oxygen species generated due to reprogramming-induced senescence and thereby alleviating cellular senescence (Esteban et al., 2010). In addition, it also induced DNA demethylation to promote reprogramming (Chung et al., 2010; Gao et al., 2015) and also reported that it induced changes in gene expression and accelerated the conversion of partially reprogrammed colonies to attain

full pluripotency efficiently (Wang et al., 2011). Further, high-throughput screening of small molecule library was performed to identify chemical compounds that can act as an enhancer or a replacer.

Later studies, to eliminate all exogenous reprogramming factors (including OCT4), Deng and colleagues derived iPSCs from MEFs with 0.2% efficiency using a combination of seven chemical compounds (Hou et al., 2013). This was the first study to demonstrate the generation of integration-free iPSCs solely using a cocktail of small molecules obviating the need of reprogramming factors in genetic or non-genetic form. This study also showed that a specific combination of chemical compounds could also substitute the exogenous delivery of master regulator OCT4 (Hou et al., 2013).

Human somatic cells are more complex than the mouse somatic cells, making them complicated to reprogram with only small molecules. To date, no study has reported the generation of human iPSCs exclusively using a cocktail of small molecules. The differences in extracellular signaling pathways (LIF/STAT3 and BMP in case of mouse and FGF in case of human) essential for self-renewal and maintenance of pluripotency of mouse and human pluripotent stem cells could be one reason that the same established protocols cannot be replicated across these species. Hence, the role of a small molecule(s) targeting a specific pathway or responsible for a specific function is to be pursued separately for both mouse and human reprogramming studies.

Although small molecules impart tremendous advantages in reprogramming, the major issues lie due to pleiotropic (off-target) effects, non-specific side effects, and toxicity (Efe and Ding, 2011; Zhang et al., 2012). 5-AZA, an analog of cytosine, undergoes DNA incorporation during the S-phase of the cell-cycle and binds covalently to DNMTs. This cytosine analog is

an effective chemotherapeutic drug, but in renal patients, it can cause mortality (Peterson et al., 1981). This could be due to the introduction of genetic mutations and/or induction of chromosomal instability (Jackson-Grusby et al., 1997; Gaudet et al., 2003), leading to tumor formation and putting patients' lives in danger (Gaudet et al., 2003). BrdU is another synthetic thymidine analog, which incorporates itself into DNA during cell division, and introduces mutations when used at higher concentrations, thereby causing genetic instability (Long et al., 2015). VPA, on the other hand, is a mood-stabilizing and an anti-epileptic drug and has shown organ level side effects (dermatological and abnormalities in the fetus) (Chateauvieux et al., 2010). These side effects are probably due to its ability to downregulate chromatin-associated proteins crucial for the maintenance of proper chromatin structure (Marchion et al., 2005). Importantly, specific small molecules identified to alter the chromatin structure to promote iPSC formation may incorporate undesirable epigenetic modifications causing comprehensive gene dysregulation in the derived iPSCs (Maherali and Hochedlinger, 2008). Therefore, the use of such chemical compounds requires thorough monitoring and evaluation at the genomic level before the therapeutic applications. However, a molecule such as RG108 from the family of nucleoside analog inhibitor does not insert into DNA and hence is less cytotoxic as compared to 5-AZA (Pasha et al. 2011). Moreover, chemical compounds that accelerate cell cycle or those that target a specific signaling pathway holds more importance than epigenetic modifiers in the reprogramming paradigm. Likewise, the identified molecules and established small molecule protocols require elaborate understanding to give rise to *bona fide* iPSCs. Therefore, extensive research is required to identify the ideal combination of chemical compounds, their time of application, and/or their respective dosage to present the full potential of this approach in iPSC generation and patient-specific applications. Taken together, small

molecule-based reprogramming represents a promising and effective approach towards obtaining integration-free, clinical-grade iPSCs for prospective biomedical applications.

Recombinant proteins

Reprogramming factors in the form of pure proteins can be delivered into terminally differentiated cells to generate integration-free iPSCs. These reprogramming proteins can be expressed in microbial or mammalian cells, extracted from them in pure bioactive form, and produced in large amounts for transduction into cells.

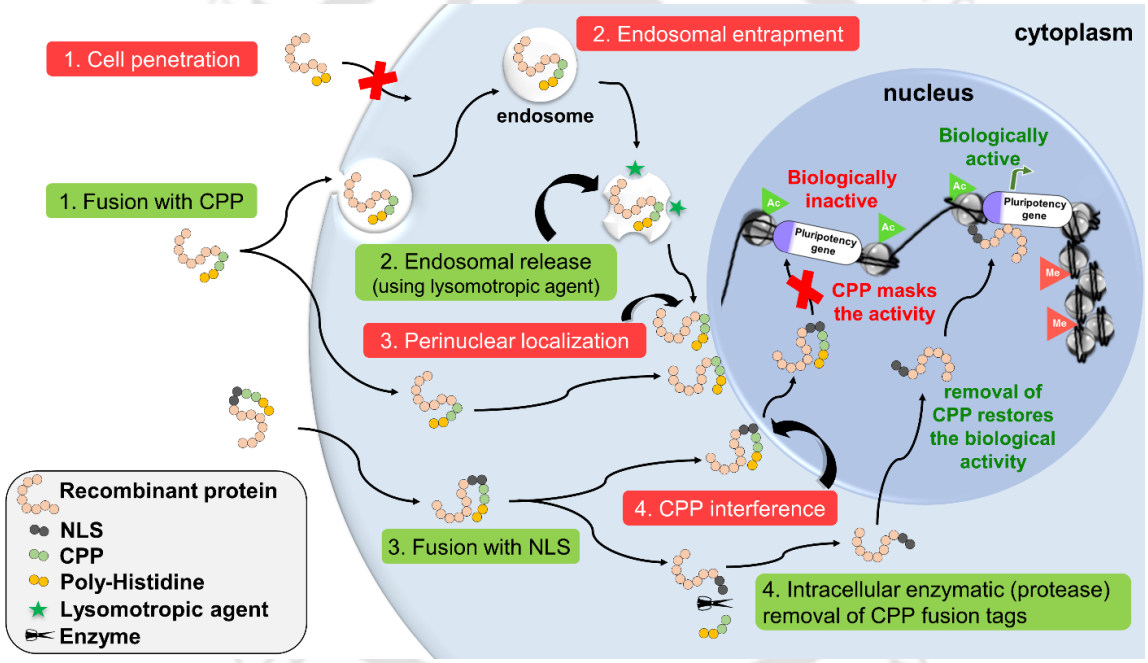


Figure 1.4 A pictorial illustration of bottlenecks associated with recombinant protein transduction (red boxes) and the possible ways to overcome them (green boxes) for the successful generation of clinical-grade iPSCs.

However, the delivery of these pure recombinant proteins into cells are marred by natural barriers of the cell, like cell and nuclear membrane. To overcome these natural barriers, recombinant proteins can be tagged with cell-penetrating peptides (CPPs) and nuclear

localization sequence/signal (NLS; in case of nuclear proteins) to facilitate and enhance sub-cellular and sub-nuclear delivery, respectively (Li et al., 2014; Kaitsuka and Tomizawa, 2015; Seo et al., 2017) (Figure 1.4).

Several groups have generated integration-free iPSCs using the recombinant protein transduction approach. The first study to use recombinant proteins demonstrated the derivation of stable exogene-free iPSCs by successfully reprogramming MEFs (Zhou et al., 2009). To achieve this, the authors used *E. coli* derived Yamanaka factors as pure recombinant proteins conjugated at the C-terminal end with CPP, poly-arginine (11-R), to enhance their entry into MEFs for successful induction of pluripotency. However, even after four rounds of protein transduction, OG2/OCT4-GFP reporter MEF cells did not give rise to any GFP^+ colonies with three or four reprogramming proteins. They obtained stable GFP^+ colonies only when transduction media was supplemented with 1 mM of VPA. Although the study achieved successful protein transduction, the overall reprogramming efficiency and kinetics were abysmal, i.e., 0.006 % after 4-5 weeks. The reason for this inefficiency was because of the endosomal entrapment of the transduced proteins, which resulted in inadequate amount of protein being delivered into the nucleus. Likewise, another group generated integration-free iPSCs by reprogramming human newborn fibroblasts with recombinant proteins (Kim et al., 2009). The authors fused Yamanaka factors with poly-arginine (9-R) CPP to facilitate the intracellular delivery and expressed them in a mammalian expression system, human embryonic kidney 293 (HEK293) cells. Unlike, Zhou et al., they used whole extracts of the cells overexpressing reprogramming factors fused with 9-R to derive iPSCs. However, this study also reported slower kinetics (58 days) and very low efficiency (0.001%) of reprogramming after six rounds of transduction cycles. Concurrently, Cho et al. used ESC

extracts directly to reprogram mouse cardiac fibroblasts to iPSCs with single transduction, resulting in better reprogramming efficiency and kinetics (Cho et al., 2010). Since these extracted proteins lack the cell-penetrating ability, the researchers used streptolysin O-mediated reverse permeabilization to translocate these proteins into the mouse cardiac fibroblasts (Cho et al., 2010). Streptolysin O is a bacterial exotoxin that permeabilizes the cell membrane without affecting the cell viability and facilitates the translocation of biological macromolecules to target cells (Walev et al., 2001). After 4-7 days post-transduction, the authors observed around 5-10 colonies, and the maximum number of colonies were reported only after 20-25 days of induction (Cho et al., 2010). Importantly, they did not observe any pluripotency markers when cardiac fibroblasts were incubated with either cytoplasmic or nuclear extracts, whereas incubation with whole cell extracts of ESCs resulted in the activation of pluripotency genes. This outcome implies that both cytoplasmic and nuclear fractions are essential to derive stable iPSCs from cardiac fibroblasts. Moreover, ESC proteins larger than 100 kDa did not contribute to iPSC generation, whereas proteins larger than 30 kDa did contribute to the induction of pluripotency in adult somatic cells. Interestingly, the inclusion of transcription factor Nanog to the reprogramming cocktail comprising of Yamanaka factors enhanced the reprogramming efficiency significantly in the presence of VPA (Zhang et al. 2012a). In this study, the reprogramming factors were tagged with either poly-arginine (11-R) or Trans-Activator of Transcription (TAT) CPP (a short peptide of HIV-TAT protein) to enhance intracellular delivery. The authors showed efficient intracellular delivery of TAT-fused proteins to the human foreskin fibroblasts (HFFs) compared to the poly-arginine fused proteins. A recent study identified a novel CPP ("nuclear trafficking peptide" (NTP)) that had both intracellular as well as intranuclear translocation activity (Takashina et al., 2018).

Comparatively, this CPP was more efficient than commonly used CPPs (TAT and 11-R) and showed enhanced intracellular as well as intranuclear delivery of artificial transcription factors in MEFs to derive iPSCs (Takashina et al., 2018).

Reprogramming somatic cells using CPP-mediated recombinant protein transduction to generate iPSCs remains highly inefficient in comparison to other non-integrating and the traditional integrating approaches. The reason behind this inefficiency is the reduced nuclear translocation of these transduced proteins. To address this, Nemes and colleagues fused NLS along with CPP to enhance the nuclear delivery of Yamanaka factors (Nemes et al., 2014). Interestingly, they observed that only the outbred MEFs (ICR) were able to generate iPSC colonies, and not the inbred background (C57BL6), signifying that reprogramming depends on the genetic status of the cells to be reprogrammed. Apart from these commonly used CPPs, Khan and coworkers used a non-toxic 1,12-diaminododecane based cationic bolaamphiphile to transduce the reprogramming proteins efficiently into the target cells (Khan et al., 2013). This cationic bolaamphiphile has two hydrophilic groups (amines) at the sides and a hydrophobic core at the center to ensure the stability of carrier-cargo protein complex by hydrogen bonding, electrostatic, and hydrophobic interactions between them (Fuhrhop and Wang, 2004). Out of the four Yamanaka factors, they replaced OCT4 with NR5A2 to facilitate reprogramming with the help of cationic bolaamphiphile, and derived-iPSCs with enhanced efficiency and kinetics in the range equivalent to integrating viral- and non-integrating RNA-based reprogramming, respectively (Khan et al., 2013). These derived cells expressed pluripotency genes and passed all the in vitro and in vivo assays but failed to generate chimeras. Moreover, this study does not explain how the carrier-cargo protein complex dissociates before nuclear localization.

In addition, another fascinating study found a solution to make this approach efficient to derive iPSCs (Lee et al., 2012). Authors demonstrated that inducing the innate immune response, like in case of viral- or modified mRNA-based reprogramming, by stimulating the Toll-like receptor 3 (TLR3) signaling promoted efficient induction of pluripotency. Lack of activation of this crucial signaling in recombinant protein transduced cells might be the reason for low reprogramming efficiency. Therefore, the authors used Polyinosinicpolycytidylic acid (Poly I:C), a TLR3 agonist to stimulate this essential signaling, along with reprogramming factors conjugated with 11-R CPP to reprogram human fibroblasts. As a result, many HDACs (HDAC 1,2,5,7) were downregulated in transduced fibroblasts, thereby leading to efficient generation of iPSCs with enhanced reprogramming kinetics and efficiency. Hence, for efficient protein-based reprogramming, TLR3 stimulation is essential, and this will improve the robustness of this technology.

Although protein-based reprogramming is the safest and a promising alternative (O'Malley et al. 2009; González et al. 2011; Sommer and Mostoslavsky 2013) there are specific challenges associated with this approach. Firstly, cellular reprogramming requires recombinant proteins in large quantities but obtaining them routinely in pure and native (bioactive) form is highly challenging. Secondly, in vitro solubility and stability of reprogramming proteins can be a major obstacle due to variations in physiological conditions. Thirdly, inefficient nuclear localization due to entrapment of the majority of these transduced proteins in endosomes is also a critical issue. To circumvent these major drawbacks, reprogramming factors can be conjugated with 30Kc19 (non-viral and non-toxic CPPs) (Ryu et al., 2016), and cell culture media conditions can be optimized to facilitate solubility and stability (Thier et al. 2010, 2012). Also, pre-treatment or co-induction with lysomotropic

agents (such as chloroquine, DMSO, ammonium chloride, sucrose, or pH-sensitive fusogenic hemagglutinin-2 subunit peptide) promotes the endosomal release of these transduced proteins.

Moreover, conjugating the reprogramming proteins with fusion tags such as CPP and affinity tags (histidine, glutathione-S-transferase, maltose-binding protein, and so forth) might compromise its biological activity. Therefore, intracellular enzymatic removal of these fusion tags from the transduced proteins using a specific protease facilitates the restoration of its function (Konno et al. 2011). Hence, these parameters mentioned above should be considered and pursued vigorously to derive exogene-free protein-based iPSCs. On circumventing these shortcomings, protein-mediated reprogramming can be of immense interest as it offers complete control over dosage and time of application, which could help researchers determine the stage-specific role of pluripotency-inducing genes in the reprogramming paradigm.

1.1.3 Role of reprogramming transcription factors

Transcription factors are those proteins which are associated with the process involving the transcription of DNA to RNA. Transcription factors associated with the process of reprogramming works by altering the epigenetic and transcriptomic landscapes inside the cell, thereby resulting in reprogramming the cell fate to the desired cell type (Iyer and Groves, 2021). Previous studies identified various genes which were responsible for the reconstitution of a particular transcriptional network during the process of reprogramming (Takahashi and Yamanaka, 2016). They identified genes contributing in the maintenance of pluripotency and indefinite proliferation in ESCs and thus, these identified factors were named as ESC associated transcripts. Using these approach few factors such as NANOG, KLF4 and MYC were identified. OCT4 and SOX2 were already well established factors associated with the regulation and expression of pluripotency-associated genes (Takahashi and Yamanaka, 2016).

In the subsequent years using factors OSKM in retroviral form, Yamanaka and group successfully generated mouse and human iPSCs (Takahashi and Yamanaka, 2006; Takahashi et al., 2007) and OSNL in lentiviral form were used to generate human iPSCs by Thomson and group (Yu et al., 2007). Taking into account that OCT4 and SOX2 were the master regulators of pluripotency, it was essential for the scientific community to screen more factors which can either enhance the efficiency of reprogramming or substitute factors like KLF4 and c-MYC due to their oncogenic potential. In the year 2008, Zhao and colleagues reported that UTF1 in combination with OSKM and p53 siRNA can enhance the reprogramming efficiency (Zhao et al., 2008). Additionally, UTF1 could also not only replace c-MYC, but can be used as a potent biomarker for the identification of *bona fide* iPSCs. Further, to maintain the mature iPSCs longer in culture and to replace these factors, another factor GLIS1 was identified by Maekawa and group, where they reported that using GLIS1 in the cocktail could enhance the reprogramming efficiency and generate mature iPSCs (Maekawa et al. 2011). These selected reprogramming factors OCT4, SOX2, UTF1 and GLIS1 is elaborated in the below section.

1.1.3.1 OCT4

OCT4 (also known as POU5F1) is a POU domain containing transcription factor, encoded by the POU5F1 gene, has been reported to be essential for self-renewal and pluripotency. It belongs to the POU domain transcription factor family where members control the target gene by binding to ATGCAAAT octameric sequence of the respective targets (Herr and Cleary, 1995). It has three isoforms namely OCT4A, OCT4B and OCT4B1 (Wang and Dai, 2010). Functionally, OCT4 is very essential for the early development and OCT4-deficient mice die at the time of implantation due to lack of inner cell mass formation (Nichols et al., 1998), thus making OCT4 to be a master regulator of pluripotency. Reducing OCT4 expression by one-

half induces the ESCs to differentiate into trophoblasts and two-fold overexpression leads to ESC differentiation into primitive endoderm and mesoderm (Niwa et al., 2000). Subsequent downregulation of OCT4 upon differentiation in somatic cells has also been reported (Rosner et al., 1990). Upon differentiation the OCT4 locus undergoes a series of repressive epigenetic modifications mediated by Histone methyltransferase G9a and the de novo DNA methyltransferases DNMT3a/b (Feldman et al. 2006). The regulatory elements of large number of OCT4 targets genes contains a composite of OCT4-SOX2 element separated by several nucleotides. OCT4 and SOX2 binds to the element of these genes and synergistically activates the expression of OCT4, SOX2, NANOG, FGF4, UTF1 and ZFP206 which plays a crucial role in maintaining pluripotency (Ambrosetti et al., 1997; Nishimoto et al., 1999; Chew et al., 2005; Rodda et al., 2005; Wang et al., 2007).

In cell reprogramming, OCT4 was identified to be a key member in the transcription factor cocktail to generate iPSCs (Takahashi and Yamanaka 2006; Yu et al. 2007). However, although the exact role of OCT4 in cell reprogramming is not known, it has been reported to be crucial for the de-repression of somatic cell chromatin, induction of the mesenchymal-to-epithelial transition, and establishment of alternative cell states (Radzisheuskaya and Silva, 2014). Furthermore, the dosage of OCT4 is also reported to be critical for efficient reprogramming (Radzisheuskaya and Silva, 2014; Haridhasapavalan et al., 2020). Since the dosage of OCT4 is identified to be crucial in various cells and cellular processes (Niwa et al., 2000; Gidekel et al. 2003; Radzisheuskaya and Silva, 2014; Haridhasapavalan et al. 2020), recombinant OCT4 protein having control over dosage and time of application can serve as an ideal molecular tool to investigate its biological role in these cells and cellular processes.

1.1.3.2 SOX2

SOX2 is a member of the SOXB1 transcription factor family and is the earliest factor of the subfamily to be expressed in mouse (Masui et al., 2007; Sarkar and Hochedlinger, 2013; Zhang and Cui, 2014). It plays a vital role in the embryonic development, where it is expressed in inner cell mass (the source of ESCs), epiblast, germ cells, and the multipotent cells of extraembryonic ectoderm (Avilion et al., 2003). Zygotic deletion of the gene causes early embryonic lethality due to failure in forming pluripotent epiblast cells (Avilion et al., 2003; Sarkar and Hochedlinger, 2013). SOX2 is also expressed in adult stem and progenitor cells (Sarkar and Hochedlinger, 2013), and its major role is observed in the maintenance of neural stem cells and its subsequent differentiation into lineage-specific cell types (Pevny and Nicolis, 2010; Zhang and Cui, 2014). Downregulation of SOX2 showed direct implication on the self-renewal activity in both mouse and human ESCs (Masui et al., 2007; Adachi et al., 2010). Both showed a loss in maintaining pluripotency and subsequently differentiated into trophectodermal lineage. The role of SOX2 was also identified in the generation and maintenance of iPSCs in conjunction with other core reprogramming factors (Takahashi and Yamanaka, 2006; Yu et al., 2007). It played a critical role in regulating the reprogramming network and assisted in the epigenetic reversal of the somatic cells into iPSCs (Zhang and Cui, 2014). Furthermore, the dosage of SOX2 is also reported to be crucial for efficient reprogramming (Haridhasapavalan et al., 2020). Interestingly, the role of SOX2 is also implicated in multiple cancers such as human squamous cell carcinoma, osteosarcoma, glioblastoma, and melanoma (Sarkar and Hochedlinger, 2013). $SOX2^{+}$ cells are marked as a *bona fide* factor in identifying a potential tumor-causing cell and a SOX2-induced drug-resistant cell, where these drug-resistant cells are critical to be identified, especially after cancer therapies, including radiography and chemotherapy (Chuang et al., 2020).

Mechanistically, SOX2 promotes cell proliferation and survival, metastasis through invasion, drug resistance, and cancer stemness, therefore making it a potential anti-cancer target (Wuebben and Rizzino, 2017; Zhang et al., 2020). All these studies implicate the crucial function of SOX2 in various cellular processes and the panoply of diseases. Hence, the generation of recombinant human SOX2 will not only provide an opportunity to understand its stoichiometric and structural relevance with respect to the generation of integration-free human iPSCs and neural stem cells, but also help investigate its function in iPSCs, neural stem cells, and cancer cells.

1.1.3.3 UTF1

Human Undifferentiated embryonic cell transcription factor 1 (UTF1) is a 341 amino acid protein and is expressed in the inner cell mass of the blastocyst (Okuda et al., 1998). The UTF1 gene regulatory elements are comprised of a short TATAAT box lacking promoter sequence and 3' enhancer. The 3' enhancer contains the binding site for both OCT4/SOX2 and presumably NANOG proteins (Nishimoto et al., 1999, 2005). The sequence of UTF1 protein is unique as it shows no homology to any previously reported protein sequences apart from a Myb/SANT domain and the presence of a leucine zipper motif (Fukushima et al., 1998; van den Boom et al., 2007; Kooistra et al., 2009). Furthermore, two highly conserved regions were identified in UTF1 protein, one at the amino-terminal end and the other at the carboxyl-terminal end giving rise to their names, i.e., conserved domain 1 and 2, respectively (Fukushima et al., 1998; van den Boom et al., 2007; Kooistra et al., 2009). The conserved domain 1 is homologous to a Myb/ SANT DNA-binding domain (van den Boom et al., 2007; Kooistra et al., 2009). UTF1 protein is localized to the nucleus (excluded from the nucleoli) and is associated with DNA during all the stages of the cell cycle (van den Boom et al., 2007;

Kooistra et al., 2009; Mouallif et al., 2014). The deletion of conserved domain 1 resulted in mislocalization of the protein to the cytoplasm (van den Boom et al., 2007; Kooistra et al., 2009), indicating that conserved domain 1 is responsible for the nuclear localization of the protein. Moreover, studies showed that both human and mouse UTF1 proteins were tightly bound to the chromatin (van den Boom et al., 2007; Kooistra et al., 2009). The earlier mentioned leucine zipper motif is present in the conserved domain 2 domain and was reported to contribute to the long-term immobilization by tightly binding to the DNA, much like the histones (Okuda et al., 1998; Fukushima et al., 1998; van den Boom et al., 2007; Thummer et al., 2010).

UTF1 is expressed during early embryogenesis. During mouse embryonic development, the expression of UTF1 begins with a rapid onset from lower to higher expression in the early and late blastocyst, respectively (Okuda et al., 1998; Galonska et al., 2014). Genetic deletion of *Utf1* in mice either resulted in a developmental arrest (Bao et al., 2017), or embryonic lethality (Kasowitz et al., 2017), or developmental delay in midgestation embryos and newborn mice and delay in placental growth, possibly due to placental insufficiency (Nishimoto et al., 2013). Notably, the *Utf1* homozygous mutant mice could not survive for more than two days after birth (Nishimoto et al., 2013). On the other hand, the $Utf1^{-/-}$ pups born, irrespective of their genetic backgrounds, were smaller in size than the $Utf1^{+/+}$ and $Utf1^{+/-}$ pups, suggesting that the absence of UTF1 resulted in the developmental delay, which could be explained by placental insufficiency (Nishimoto et al., 2013; Bao et al., 2017; Kasowitz et al., 2017).

The distinguishing functional feature of UTF1 is its activation in the late stage of iPSC induction and is switched off immediately upon differentiation. It has been shown to be

expressed later than NANOG, SSEA1 and OCT4 in the course of iPSC formation and since the expression is found to be low in case of human ESCs, it is thought that UTF1 promoter is a weak promoter (Morshedi et al., 2013). Though UTF1 was listed in the 24 factors cocktail in the pioneering study (Takahashi and Yamanaka, 2006), it did not qualify amongst the final four key reprogramming factor cocktails. However, the importance of UTF1 in cellular reprogramming was soon highlighted when HFFs were reprogrammed with UTF1, small interfering RNAs against p53, and the Yamanaka factors, generating iPSCs with enhanced quality and efficiency (>200-fold higher efficiency) (Zhao et al., 2008). This could be attributed to the improved transition of pre-iPSCs to iPSCs, resulting in enhanced reprogramming efficiency. UTF1 may have aided in establishing a favorable chromatin state or epigenetic profile ideal for switching to pluripotency (Zhao et al., 2008). Additionally, removal of oncogenic c-MYC from this six-factor cocktail enhanced reprogramming efficiency by 100-fold compared to the Yamanaka factor cocktail alone. Also, UTF1 (in the presence of OSK and in the absence of c-MYC and small interfering RNAs against p53) converted pre-iPSCs to iPSCs to give rise to *bona fide* iPSCs (Zhao et al., 2008). Further it was suggested that during reprogramming, UTF1 facilitates the formation of a chromatin structure that is important in pluripotency (Kooistra et al., 2009). UTF1 depleted cells displayed significantly reduced reprogramming efficiency (Schwarz et al., 2018). On the contrary, studies have shown that UTF1 is dispensable in forming mouse iPSCs (De Sousa Lopes et al., 2005; Nishimoto et al., 2013). These studies have reported that UTF1-null mouse fibroblasts could be reprogrammed to iPSCs with typical pluripotency characteristics but with very low reprogramming efficiency. Other studies have reported the importance of UTF1 as a biomarker in iPSC formation. Notably, UTF1 is an early stage indicator for successful reprogramming

(Buganim et al., 2012). Further, its expression is crucial for identifying clones/cells that will attain full pluripotency (Buganim et al., 2012, 2013; Galonska et al., 2014). Furthermore, UTF1-based reporters are highly sensitive in identifying mature pluripotent stem cells and, therefore, can provide a platform to derive and/or maintain homogeneous pluripotent cells with a high degree of pluripotency (Tan et al., 2007; Pfannkuche et al., 2010; Buganim et al., 2012; Morshedi et al., 2013; Galonska et al., 2014; Semmler et al., 2014).

UTF1 expression was reported in a variety of cancer tissues of germ cell and non-germ cell origin. It is highly expressed in embryonic carcinoma cells (van den Boom et al., 2007; Kooistra et al., 2009; Thummer et al., 2012), testicular germ cell neoplasms (Kristensen et al., 2008), and central nervous system germinoma (Pantazis et al., 2014). Elevated expression of UTF1 was also observed in endometrial and prostate cancer (Mouallif et al., 2014), whereas reduced expression was observed in colon, renal, and breast cancer (Mouallif et al., 2014; Xu et al., 2014), when compared to their normal counterparts. Its expression was also found to be overexpressed in endometriosis (Forte et al., 2009), which shares various molecular and cellular resemblances with neoplastic development. This indicated increased chances of possible hormonal regulation of UTF1 expression. The studies reporting its expression in cervical cancer have been conflicting. One study reported that UTF1 is overexpressed in cervical cancer cells, and it is associated with hypermethylation of the UTF1 promoter (Guenin et al., 2012). On the other hand, a different study reported that UTF1 expression is downregulated in cervical cancer cells, and it is primarily due to hypermethylation of the UTF1 promoter (Wu and Zheng, 2013). In brain cancer, UTF1 expression is of potential interest. It could distinguish between grade I–III and grade IV neuroblastoma tumors (Melone et al., 2009) and can serve as an excellent prognostic marker. Thus, the expression of UTF1 in different

cancers has brought into light its function either as an oncogene or as a tumor suppressor (Mouallif et al., 2014), indicating a tissue-specific role of this protein. A simple and convenient recombinant tool with methodological developments can help advance research to understand UTF1 function and regulation in cellular reprogramming towards pluripotency and cancer biology.

1.1.3.4 GLIS1

GLIS1 belongs to the Gli-similar protein family and broadly to the subfamily of Kruppel-like zinc-finger proteins (Kim et al., 2002; Scoville et al., 2017). The GLIS1 gene encodes ~66 kDa protein (620 amino acids) rich in proline residues and has zinc-finger motifs, with the highest sequence homology to different members of Gli and Zic subfamilies of Kruppel-like proteins (Kang et al., 2010; Jetten, 2018). The pivotal role of the nuclear translocation of the GLIS1 protein was due to the contribution of the zinc-finger region in its structure. The ability to bind to the consensus sequence 5′-GACCACCCAC-3′ of the Gli-binding site was also observed for the same (Kim et al., 2002). It has a transactivation domain present at both the N- and C-terminus of the protein, and thus, it acts as a transcription factor that regulates gene expression during specific embryonic developmental stages (Kim et al., 2002). A spatiotemporal expression of GLIS1 during embryonic development contributed to its abundant expression in unfertilized eggs, one- and two-cell embryos, and placenta (Nakashima et al. 2002; Jetten 2018). Also, its abundant expression in the kidney and low expression in the brain, colon, testis, thymus, and adipose tissue was observed in adults (Nakashima et al. 2002; Jetten 2018). The expression profile of GLIS1 protein in ESCs showed a low level of expression, with drastic inhibition in ESC proliferation upon its overexpression (Maekawa et al. 2011).

In reprogramming paradigm, its importance to derive fully reprogrammed iPSCs with high reprogramming efficiency was demonstrated by many studies (Maekawa et al., 2011; Maekawa and Yamanaka, 2011; Yoshioka et al., 2013; Takahashi et al., 2014; Lee et al., 2017b; Wang et al., 2017; Yoshioka and Dowdy, 2017). It was reported that OCT4, SOX2, and KLF4 combined with GLIS1 enhanced the generation of both mouse and human fully reprogrammed iPSCs, along with the successful generation of germline-competent chimeras from mouse iPSCs (Maekawa et al., 2011). In this study, Maekawa et al. demonstrated the generation of iPSCs by substituting c-MYC with GLIS1 from the Yamanaka factor cocktail (Maekawa et al., 2011). The mouse and human iPSCs generated formed *bona fide* ESC-like colonies when compared to the Yamanaka factor cocktail. Specifically, GLIS1 enhanced the generation of fully reprogrammed iPSCs, unlike c-MYC, which promoted the formation of partially reprogrammed clones in a higher proportion. The chimeric mice generated with these mouse iPSCs showed decreased incidence of tumor formation, thus, GLIS1 emerging as a potential promising candidate in the cocktail of reprogramming factors, replacing the oncogene c-MYC (Maekawa et al., 2011). The same combination (OCT4, SOX2, KLF4, and GLIS1) showed enhanced reprogramming efficiency with the non-pathogenic, self-replicating Venezuelan equine encephalitis RNA virus to derive integration-free human iPSCs (Yoshioka et al., 2013). Further, GLIS1, in different reprogramming factor combinations, also efficiently generated iPSCs from mouse and human somatic cells (Lee et al., 2017; Liu et al., 2015; Takahashi et al., 2014; Wang et al., 2017). GLIS1 has multifaceted roles such as the promotion of pro-reprogramming pathways (WNT, NANOG, MYC, LIN28, ESRRB), induction of the expression of the FOXA2 transcription factor, inhibition of epithelial to mesenchymal transition (Maekawa et al., 2011), facilitation of the change in chromatin state, and activation

of a pluripotency-associated gene like SOX2 (Wang et al., 2019). It induces multilevel epigenetic and metabolic remodeling in stem cells, thus, facilitating the induction of pluripotency (Li et al., 2020). All these functions enhanced the reprogramming efficiency and enabled the generation of *bona fide* iPSCs.

Apart from its function in cell reprogramming, embryonic development, and mesodermal cell differentiation during fetal development (Luo et al., 2021; Maekawa et al., 2011; Yasuoka et al., 2020), its role has also been implicated in various ciliated organ diseases of lung, pancreas and kidney, (Yasuoka et al., 2020) breast cancer, and in the late onset of Parkinson's disease (Song et al. 2012). GLIS1 overexpression in breast cancer cells MDA-MB-231 contributed to the increase in the migration and invasion capacities of the cells, possibly through the upregulation of WNT5A (Shimamoto et al., 2020) or by cooperating with CUX1, thus stimulating activity of TCF/β-catenin transcription factor and enhancing cell migration and invasion of breast cancer cells (Vadnais et al., 2014). Hence, the generation of recombinant human GLIS1 will provide an opportunity to understand its stage-specific role in generating *bona fide* iPSCs and also elucidate its function in cancer cells.

1.1.4 Prerequisites of recombinant protein production in a bacterial expression system

Proteins are the workhorses, which regulate the biological functioning of a cell. To understand and prognosticate the putative intracellular mechanism of a biological system, understanding the structure and function of the proteome of a cell, and roles of different therapeutic and/or diagnostic protein targets is of paramount importance (Yadav et al., 2016). Dysfunctional proteins are the reason for multiple pathological conditions such as diabetes, cystic fibrosis, dwarfism, thalassemia, blood clotting disorders, and so forth. Due to the dearth of standardized gene therapy treatments, protein therapy has emerged as a promising alternative. In the

previous decade, advent of iPSCs connected clinical therapies from bench to bedside and opened a new platform for patient-specific therapies. Tapping onto the roles of pluripotency-associated proteins in the generation of iPSCs from terminally differentiated cells, showed another dimension of the use of proteins in reprogramming and eventually differentiation into lineage-specific cells.

Proteins required for this purpose encounters a major limitation that is, the reprogramming proteins are required in bulk for carrying out the reprogramming process. Thus, the advent of recombinant protein technology has facilitated the increase in the generation of proteins in biological systems, which assure the restoration of protein function, cost-effective industrial production and the absence of hazardous contaminants. Recombinant protein production has revolutionized the field of biotechnology as several purified proteins have reached the market in a short span of time. Currently, there are over 400 marketed recombinant products and other 1300 are under development as antibodies, enzymes, hormones, diagnostic kits, and many more (Sanchez-Garcia et al., 2016). Large scale production of recombinant proteins employs a wide range of expression systems such as bacteria (Overton, 2014), algae (Specht et al., 2010), yeast (Mattanovich et al., 2012), mammalian cells (Wurm and Bernard, 1999), and insects (McCarroll and King, 1997) has so far proven to be a complex but an effective process (Chen, 2012). Mammalian expression system is a suitable choice for proteins that require appropriate post-translational modifications for their function. For other proteins, the high production time and costs along with slow growth kinetics and complex nutritional requirements associated with mammalian systems renders them less preferred than bacterial systems (Walsh and Jefferis, 2006). Therefore, for such proteins *Escherichia coli* became an obvious choice for large-scale production due to

their well-established genetic makeup and knowledge of its physiology, reasonable costs and unparalleled growth rate (Huang et al., 2012; Khow and Suntrarachun, 2012; Sahdev et al., 2008).

Heterologous expression of recombinant proteins in bacterial system has several challenges such as codon usage bias, gene toxicity, mRNA instability, poor protein expression, proteolytic cleavage by the host cell, purity, poor solubility and stability (*in vitro*), and protein misfolding, which need to be addressed, thus enhancing its production and purity (Khow and Suntrarachun, 2012). The first and foremost limitation of protein transduction technology is the delivery of the recombinant protein into the cell and nucleus. This can be overcome by tagging cell-penetrating peptides (CPP), nuclear localization signal/sequence at its N or C-terminal of the recombinant reprograming proteins, facilitating its entry into the cell and nucleus (Li et al., 2014; Kaitsuka and Tomizawa, 2015; Seo et al., 2017). Using recombinant protein-mediated cellular reprogramming offers greater control over concentration and time of application, thus, providing ample flexibility to experiment various stoichiometric combinations of reprogramming factors with small molecules using a sliding window approach to unravel their role at different stages during cell reprogramming. Importantly, this technology does not involve any genomic integration and induce an innate immune response (Lee et al., 2012).

Various studies have reported that recombinant protein-based reprogramming results in generation of clinical-grade iPSCs (González et al., 2011; Sommer and Mostoslavsky, 2013).

1.2 Motivation and Scope of the study

iPSC technology in patient-specific therapies and disease modeling has proven to be powerful and instrumental in establishing regenerative therapy in biomedical research. It helped in laying a foundation for unraveling novel pathological mechanisms at cellular level for multiple monogenic diseases and ongoing and emerging clinical trials with iPSC-derived cell products. This technology has sparked immense enthusiasm in the scientific community due to its unprecedented opportunity in drug designing and discovery, personalized organs on chips, gene editing, etc., providing a new dimension in regenerative medicine and cell therapy. The initial approaches used for the generation of iPSCs using integrative methods such as retro- and lenti-viral approaches retrograded their clinical applicability and translation to patient-specific therapies. It is due to insertional mutation, hence, setting a prerequisite to finding approaches that are transient and are integration-free. The current advancements and transition to non-integrative approaches including mRNA, miRNA, Sendai virus, episomal vector, small molecules, and recombinant protein were prioritized (Figure 1.5). Amongst all these approaches, Sendai virus, episomal and mRNA approaches are more pronounced in deriving integration-free iPSCs due to their relative ease of use and high efficiency. However, these approaches despite having comparatively higher efficiency have their drawbacks. For example, in the Sendai virus approach, due to the presence of residual viral components requires greater than 10 passages for their removal, immunogenic response to multiple viral factors, and rare chances of integration. Similarly, episomal plasmids also can cause rare integration and both these approaches require thorough screening for confirming the absence of integration. In the case of mRNA, stability and repeated transfections are the major disadvantages, curbing its utility in iPSC generation. Thus, recombinant

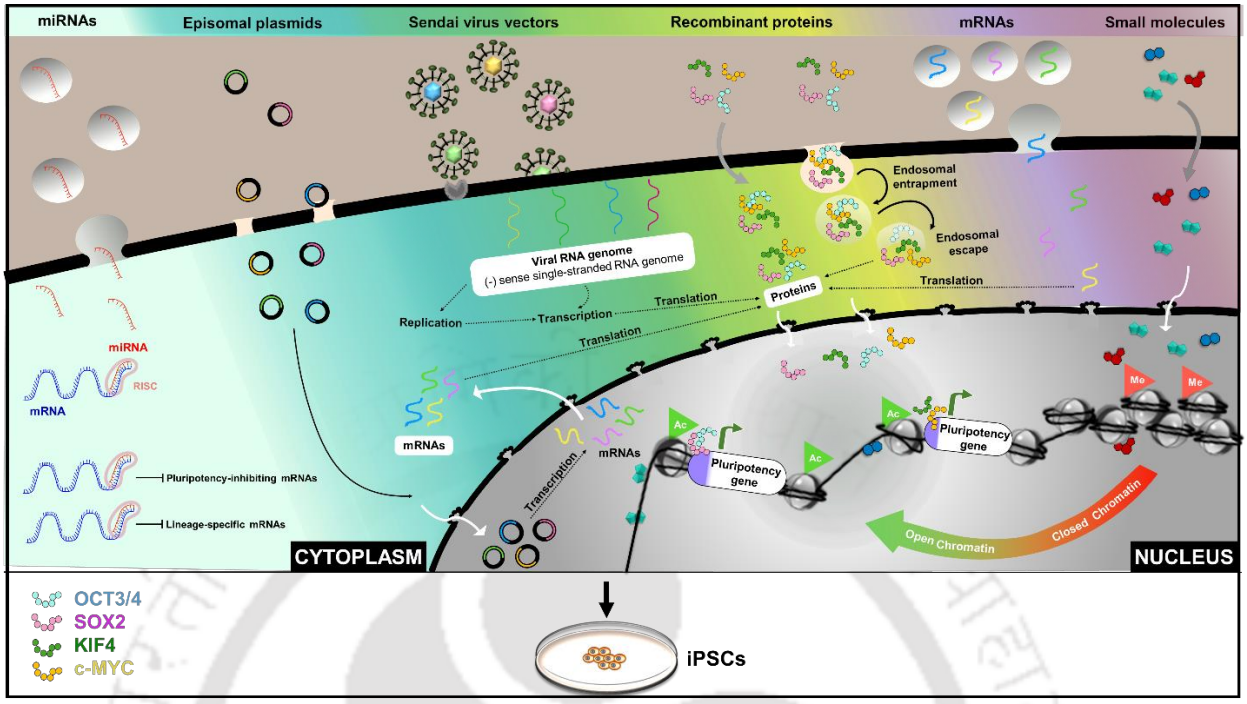


Figure 1.5 A schematic illustration of non-integrative approaches to derive integration-free iPSCs.

proteins herald a way to safest approach toward integration-free reprogramming. However, the crux of this approach is its low efficiency, limiting its applicability in iPSC generation. In prospect of various limitations encompassing recombinant protein production and its potential use in reprogramming, we derived our motivation to tackle the rate-limiting factors such as codon-bias, expression parameters, and establishing successful native purification methods for the selected reprogramming fusion protein (OCT4, SOX2, UTF1, and GLIS1). Literature is filled with details on the importance of these factors in the reprogramming paradigm, as OCT4 and SOX2 are the master regulators, and UTF1 and GLIS1 are the factors helping in substituting oncogenic factors c-MYC and assisting the generation of *bona fide* iPSCs. Establishing a recombinant protein toolbox and providing a clear vision of the critical factors as prerequisites for increasing recombinant reprogramming protein production could provide

a platform to potentially overcome the limitation concerning the efficiency of protein-induced iPSCs in near future. Additionally, individual factors and their multifarious roles in various cellular mechanisms, especially in cancer can also be studied in greater detail, which is unexplored to date. Thus, establishing a recombinant protein toolbox that can substitute the genetic form of reprogramming transcription factors in the generation of integration-free iPSCs along with the use of these proteins to understand their role in multiple cellular aspects will open a plethora of new avenues of research in the near future.

1.3 Objectives

1. Gene cloning and identification of optimal expression parameters to achieve maximal soluble expression of recombinant proteins.

2. Purification, biochemical and biophysical analysis of recombinant proteins.

3. Demonstration of cell penetration, nuclear translocation, and biological activity of recombinant proteins

CHAPTER 2

Gene cloning and identification of optimal expression parameters to achieve maximal soluble expression of recombinant proteins

Brief summary of the chapter

Assessment of codon bias of genes in context to their heterologous expression as well as identification of optimal expression parameters is critical for achieving maximal soluble expression of recombinant proteins. Thus, in this chapter, we demonstrated the construction of recombinant plasmids containing the four selected human reprogramming transcription factors and screened optimal expression parameters for maximal soluble expression. To create the recombinant gene construct, we retrieved their full-length protein-coding sequence from NCBI's RefSeq database and performed codon optimization for the retrieved sequences, followed by validation using online *in silico* tools. Codon optimized fusion tag sequences for cellular and nuclear uptake, and affinity purification, were fused with the gene of interest on either 5' or 3' end of the coding sequence. These fusion gene constructs were synthesized and were cloned into the pET28a(+) expression vector. The integrity of cloning was confirmed using restriction analysis. The recombinant plasmids were next transformed and expressed into the BL21(DE3) strain of *E. coli*. This chapter then deals with the identification of expression parameters such as inducer concentration, optical cell density, induction temperature, and time of expression for maximizing the soluble expression of these recombinant proteins. Further, screening of these parameters emphasized the importance of the position of fusion tags in protein expression along with different expression conditions. Thus, we have selected the C-terminal of all the four factors OCT4-NTH, SOX2-NTH, UTF1-NTH, and GLIS1-NTH for further analysis. In conclusion, this chapter highlights the importance of

identifying the optimal expression conditions and the influence of fusion tags at either terminus on these recombinant protein expressions in terms of quality and quantity.

2.1 Materials and methods

2.1.1 Generation of human recombinant plasmid constructs

The full-length coding sequences of the human reprogramming transcription factors were retrieved from NCBI reference sequence (RefSeq) database (Table 2.1). These sequences were codon optimized for heterologous expression in *E. coli* using ThermoFisher Scientific GeneOptimizer online tool (https://www.thermofisher.com/in/en/home/life-science/cloning/gene-synthesis/geneart-gene-synthesis/geneoptimizer.html). Further, the codon optimized sequences were analyzed and validated using Graphical Codon Usage Analyser 2.0 (http://gcua.schoedl.de) and Genescript Rare Codon Analysis (http://www.genscript.com/tool/rare-codon-analysis) online tools (Figure 2.1). The codon optimized gene coding sequences were fused with the codon optimized tags at either 5'- or 3'-terminal ends of the gene constructs for their expression in bacterial system. The tags used are polyhistidine-tag [8X His; (H)], HIV-Trans-Activator of Transcription [TAT; (T)], and nuclear localization signal/sequence [NLS; (N)], thus resulting in the generation of HTN-*GOI* and *GOI*-NTH. These inserts were synthesized and purchased from GenScript Biotech Corporation, Nanjing, China, and were cloned into the protein expression vector, pET28a(+) (Novagen vector, Merck Millipore), using restriction enzymes *Nco*I (at 5' end) and *Xho*I (3' end) (Fermentas). The resulting plasmid DNA constructs (pET28a(+)-HTN-*GOI* and pET28a(+)-*GOI*-NTH) were primarily analyzed by restriction digestion, followed by DNA sequencing (Eurofins Genomics India Pvt. Ltd., Karnataka, India) for secondary confirmation.

Table 2.1 List of all the genes used in this study and their respective RefSeq accession numbers

Genes	RefSeq accession number	Length of coding sequence (bp)
OCT4	NM_002701.5	1083
SOX2	NM_003106.3	954
UTF1	NM_003577.2	1026
GLIS1	NM_147193.2	1863

2.1.2 Screening for suitable bacterial strain and media for the heterologous expression of recombinant GLIS1 protein

To determine the ideal strain for the expression of human GLIS1 protein, both N- and C-terminal constructs were transformed into *E. coli* BL21(DE3) and Rosetta (DE3) cells. Transformed bacterial cultures with positive clones grown overnight at 37 °C in Luria–Bertani broth (LB) (HiMedia) and Terrific broth (TB) (HiMedia) containing 50 µg/ml of kanamycin (HiMedia) with continuous shaking (Orbital incubator shaker, IKON instruments, India). Further, these cultures were used for inoculating the secondary culture consisting of 2% inoculum in LB and TB (HiMedia) containing 50 µg/ml of kanamycin and 1% glucose (HiMedia) and incubated with continuous shaking at 180 rpm at 37 °C until OD_{600} ~0.5 was reached. Further, the cultures were induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) (HiMedia) at a concentration of 0.5 mM and incubated for 2 hours at 37 °C. Cells were harvested after the incubation period and stored at -80 °C until further use.

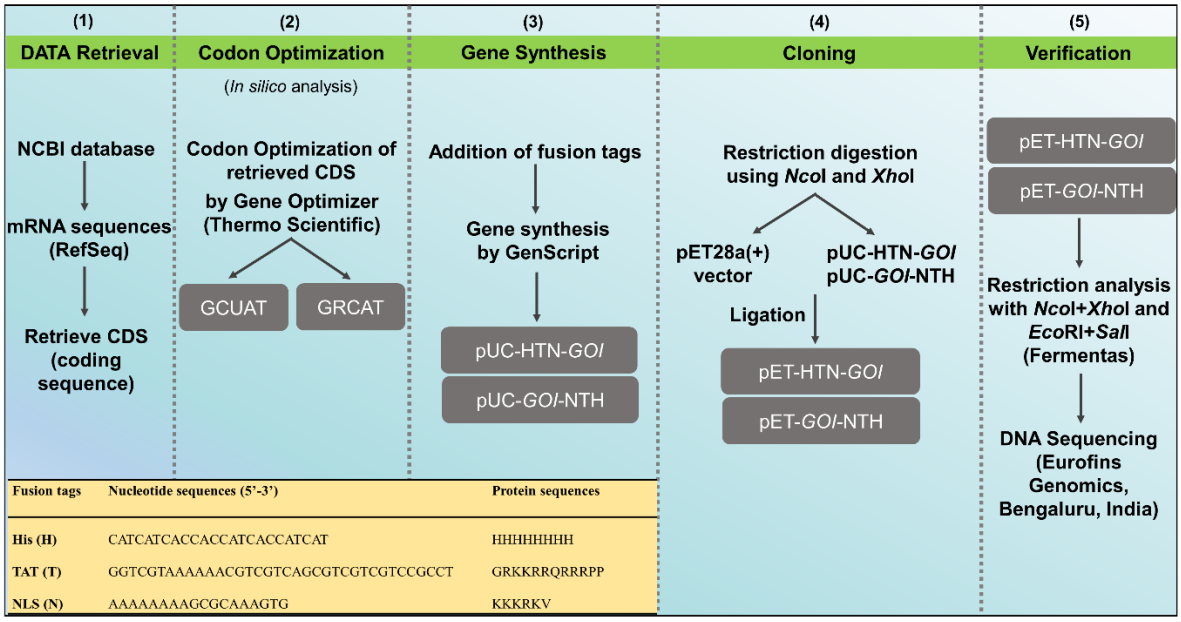


Figure 2.1. Schematic representation of the overall methodology for retrieval of gene sequences and their codon optimization, followed by cloning and sequence confirmation. GOI, gene of interest

2.1.3 Screening for optimal inducer concentrations for the maximum heterologous expression of recombinant fusion proteins in *E. coli*.

To screen and identify the optimal inducer concentrations for maximizing the soluble expression of recombinant fusion proteins (OCT4, SOX2, UTF1 and GLIS1), *E. coli* BL21(DE3) cells were transformed individually with respective pET28a(+) plasmid constructs. Transformed bacterial cultures with positive clones for each of the four constructs were grown overnight at 37 °C in LB containing 50 µg/ml of kanamycin with continuous shaking (Orbital incubator shaker, IKON instruments, India). Further, these cultures were used for inoculating the secondary culture. The secondary culture consisted of 2% inoculum in TB containing 50 µg/ml of kanamycin and 1% glucose and incubated with continuous shaking at

180 rpm at 37 °C until OD_{600} ~ 0.5 was reached. Cultures were then induced with different IPTG concentrations (Table 2.2) and incubated for 2 hours at 37 °C. Cells were harvested after the incubation period and stored at -80 °C until further use.

Table 2.2 Different parameters screening for the identification of optimal expression conditions

Transcription factors	IPTG (mM)	OD_{600} (nm)	Time (hr)	Temperature (°C)
OCT4	0, 0.05, 0.1, 0.25, 0.5	~0.5, 1, 1.5	24, 2	18, 37
SOX2	0, 0.05, 0.1, 0.25, 0.5	~0.5, 1, 1.5	36, 2	18, 37
UTF1	0, 0.05, 0.1, 0.25, 0.5	~0.5, 1, 1.5	24, 2	18, 37
GLIS1	0, 0.05, 0.1, 0.25, 0.5, 2	~0.5, 1, 1.5	32, 16, 8, 4	18, 25, 30, 37

2.1.4 Screening for optimal cell density for the maximum heterologous expression of recombinant fusion proteins in *E. coli.*

To identify the optimal cell density, the transformed clones were induced with respectively screened IPTG concentrations for individual constructs at different cell densities at OD_{600} (Table 2.2) for 2 hours at 37 °C. Cells were harvested after the incubation period and stored at -80 °C until further use.

2.1.5 Screening for optimal induction temperature and time for the maximum heterologous expression of recombinant fusion proteins in *E. coli.*

To identify the optimal induction temperature for maximizing the soluble expression, the transformed clones were induced with optimized IPTG concentrations for individual constructs

at an optimum cell density at OD_{600} for different timepoints at 37 °C and 18 °C for OCT4, SOX2 and UTF1. For GLIS1, different temperatures were screened at different timepoints (Table 2.2). Cells were harvested after the incubation period and stored at -80 °C until further use.

2.1.6 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western blotting

The harvested pellets were dissolved in pre-chilled lysis/resuspension buffer for the respective proteins (Table 2.3) and sonicated on ice using Vibracell™ VCX-130 cell disruptor (Sonics and Materials Inc., Newton, CT, USA). Bradford assay kit (Bio-Rad) was used for measuring the protein concentration with bovine serum albumin as a standard and normalized for soluble expression analysis. Protein samples were mixed with 10% SDS and 4× Laemmli buffer to attain final concentrations of 2% and 1X, respectively. The samples were resolved on 12% (OCT4, SOX2, and UTF1) and 10% (GLIS1) SDS-PAGE gels for visualization. A broad range pre-stained Precision Plus Protein™ Dual Color (Bio-Rad) protein standards were used for molecular size control. Coomassie staining was performed for the resolved protein samples and stained with a staining solution containing 50% (v/v) methanol (Merck Millipore), 10% (v/v) acetic acid (Merck Millipore), and 0.4% (w/v) Coomassie Brilliant Blue G-250 (Merck Millipore) in double distilled deionized water. The image of the stained gel was obtained using a molecular imaging system (ChemiDoc™ XRS+; Bio-Rad, California, USA). The imaging software used for visualizing was Image Lab™ software (Bio-Rad, California, USA). Western blotting was performed for the respective protein samples. Samples were resolved onto 12% (OCT4, SOX2, and UTF1) and 10% (GLIS1) SDS-PAGE gel and transferred onto the

nitrocellulose membrane (Bio-Rad) in Pierce Power Blotter XL System (Thermo Scientific™, Bremen, Germany).

The transferred proteins were confirmed with Ponceau S staining (HiMedia). The stained membrane was washed with 1X Tris-buffer saline (TBS) (HiMedia) and further blocked to avoid non-specific binding using 5% skimmed milk dissolved in 1X TBS with 0.1% Tween-20 (Invitrogen) for 1 hour at room temperature. The blocked nitrocellulose membrane was then washed with 1X TBST and incubated with primary [anti-His (1:5000, BB-AB0010, BioBharati LifeScience Pvt. Ltd., India)] overnight at 4 °C. Further, the membrane was again washed with 1X TBST and incubated with secondary antibody [anti-rabbit IgG-HRP Conjugated (1:5000; BB-SAB01A; BioBharati LifeScience Pvt. Ltd., India)] for 1 hour at room temperature. The immunoblot was then developed post-washing with 1X TBST. The image was recorded using ChemiDoc™ XRS+ (Bio-Rad, California, USA). The imaging software used for visualizing was Image Lab™ software (Bio-Rad, California, USA).

Table 2.3 Composition for Resuspension buffer/Lysis buffer for the respective proteins

Components	Resuspension buffer/Lysis buffer			
	OCT4	**SOX2**	**UTF1**	**GLIS1**
PB (mM)	20	50	20	20
NaCl (mM)	150	150	0	150
Imidazole (mM)	20	20	20	20
Glycerol (%)	20	20	20	0
pH (at RT)	~7.2	~7.8	~7.8	~7.2

PB, sodium phosphate buffer; RT, room temperature

2.1.7 Restriction digestion and agarose gel electrophoresis

The isolated plasmids were quantified using DNA quantification method using Multiskan Go™ (Thermo Scientific). The purity of the plasmid DNA was measured using 260/280 ratio (ideal range 1.8-2.0). The restriction digestion set up was made and loaded onto 0.7% agarose gel prepared in 1 X Tris-EDTA buffer (HiMedia). The gel was visualized using ChemiDoc™ XRS+ (Bio-Rad, California, USA). The imaging software used for visualizing was Image Lab™ software (Bio-Rad, California, USA).

2.2 Results and discussion

2.2.1 Codon optimization of human recombinant transcription factors for maximizing their expression in *E. coli*

The heterologous expression of human genes in *E. coli* presents multiple constraints leading to a compromise in the overall protein expression (Al-Hawash et al., 2017; Burgess-Brown et al., 2008; Maertens et al., 2010). This is mainly contributed by various factors, such as codon bias, generation of destabilizing RNA elements, the formation of RNA secondary structures, presence of cryptic splice-sites, intragenic poly(A)-sites, and so forth (Liu et al., 2018). To determine the importance of codon optimization for heterologous expression of human recombinant transcription factors in *E. coli*, we analyzed the full-length coding sequence of all four factors. We analyzed the sequences retrieved from NCBI RefSeq using two different *in silico* tools Graphical Codon Usage Analyser and GenScript Rare Codon Analysis. The analysis revealed that 7% of the codons for *UTF1*, 9% for *SOX2*, and ≥10% of the codons for *OCT4* and *GLIS1* with codon usage frequency of ≤30 were observed to affect the expression in *E. coli* (Figure 2.2 (*left*, magenta)). Similarly, using the GenScript Rare Codon Analysis

tool, ≥7% of the codons with a relative adaptiveness value (RAV) of ≤30% were identified in all four factors to affect the expression in *E. coli* [Figure. 2.3; (in red; shown by dotted line; Table 2.4]. RAV defines the ratio of the frequency of the used codon to that of the most abundant codon for the same amino acid multiplied by 100 (Fuhrmann et al., 2004).

Thus, removing these rare codons is of great importance for enhancing the overall expression of the proteins in *E. coli*. Upon codon optimization, improvement in the relative adaptiveness index of the sequences was observed, as the rare codons were substituted with codons with frequency ≥30% [Figure 2.2, *right*; Figure 2.3, (blue bar)]. An increase in the codon adaptation index [CAI; CAI is the relative adaptiveness of the codon usage of a protein-coding gene towards the codon use of highly expressed (reference set) genes] was observed (Table 2.4). The CAI value of the codon optimized gene sequences fell in the range of 0.8–1.0, which is considered optimal for gene expression in the desired host. This analysis established that codon optimization is a crucial parameter in heterologous expression of human gene sequences in *E. coli* and also contributes to reducing cell toxicity and increasing overall expression in the *E. coli* system.

2.2.2 Cloning of codon optimized human recombinant transcription factors fused with fusion tags into protein expression vectors

The codon optimized and validated gene sequences were tagged with a set of codon optimized fusion tags either before (to generate HTN-*GOI*; Figure 2.4, *left*) or at the end of the coding sequence (to generate *GOI*-NTH; Figure 2.4, *right*). The reason to have fusion tags at either end is that the placement of these tags can influence expression level, solubility, and stability of human proteins expressed in *E. coli* (Bosnali and Edenhofer, 2008; Braun et al., 2002;

Haridhasapavalan et al., 2020). This strategy enabled the identification and selection of the gene construct with fusion tags, which had no major effect on protein expression and stability. The fusion tags used in the gene constructs were: (i) TAT (T) for intracellular delivery, (ii) NLS (N) for intranuclear delivery, and (iii) poly-His (H) tag to facilitate affinity chromatography mediated purification.

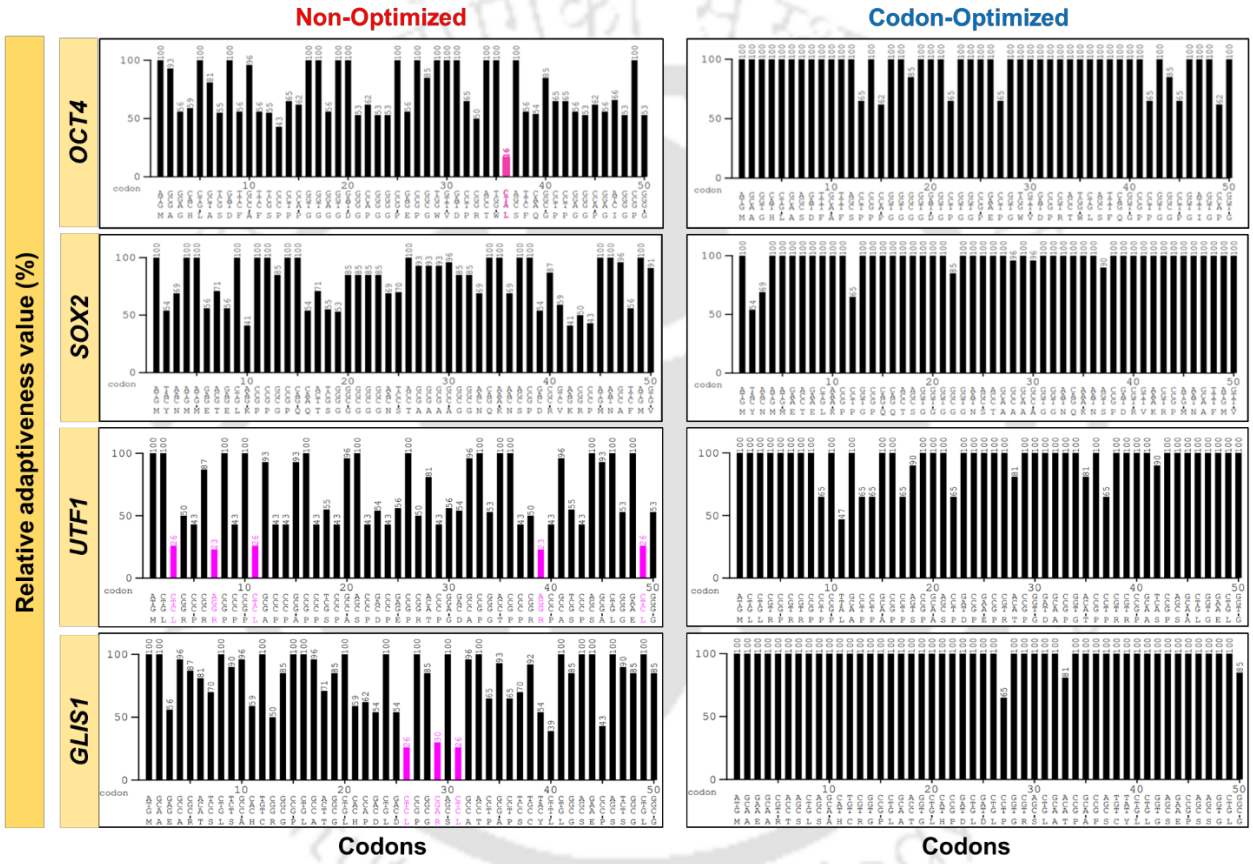


Figure 2.2 Comparative analysis and validation of the non-optimized and codon optimized gene sequences for heterologous expression of recombinant genes in *E. coli* system using Graphical codon usage analyzer tool. The first 50 codons, non-optimized (*left*) and codon optimized (*right*), represented all four genes. Below the threshold of ≤30 % relative adaptiveness index value (*magenta*) represents the codons unsuitable for protein expression.

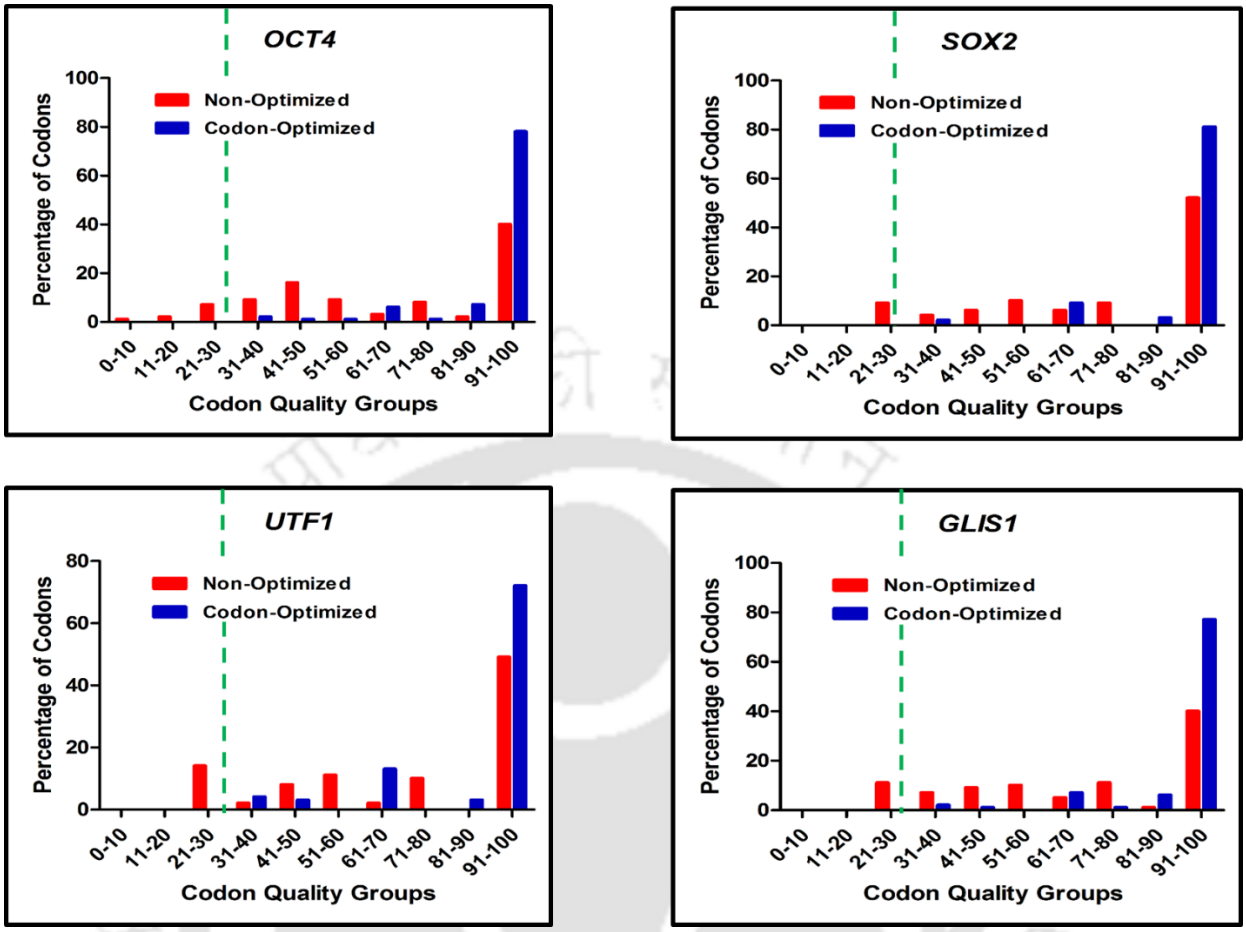


Figure 2.3 Comparative analysis and validation of the non-optimized and codon optimized gene sequences for heterologous expression of recombinant genes in *E. coli* system using GenScript rare codon usage analyzer tool. The bar graph shows the percent codon distribution for the codon quality group for non-optimized (*red*) and codon optimized (*blue*). Codons having highest frequency of usage for a given amino acid for the required expression host are assigned a value of 100. Below the threshold of ≤30 are likely to compromise the expression of recombinant protein.

Table 2.4 Summary of the parameters of GenScript rare codon analysis for non-optimized and codon optimized sequences of the genes

Genes	Parameters	Non-optimized	Codon optimized	Ideal Range
OCT4	Codon Adaptation Index (CAI)	0.61	0.90	0.8-1.0
	GC Content	60.02%	51.04%	30%-70%
	Codon Frequency Distribution (CFD)	10%	0%	<30%
SOX2	Codon Adaptation Index (CAI)	0.70	0.91	0.8-1.0
	GC Content	66.42%	51.01%	30%-70%
	Codon Frequency Distribution (CFD)	9%	0%	<30%
UTF1	Codon Adaptation Index (CAI)	0.66	0.86	0.8-1.0
	GC Content	78.22%	60.33%	30%-70%
	Codon Frequency Distribution (CFD)	14%	0%	<30%
GLIS1	Codon Adaptation Index (CAI)	0.64	0.90	0.8-1.0
	GC Content	66.14%	53.57%	30%-70%
	Codon Frequency Distribution (CFD)	11%	0%	<30%

Earlier studies have also reported a similar strategy to enable subcellular and subnuclear delivery of various transcription factors in mammalian cells (Bosnali and Edenhofer, 2008; Haridhasapavalan et al., 2021; Münst et al., 2016; Narayan et al., 2021b; Peitz et al., 2014; Thier et al., 2010). To achieve the heterologous expression of these four recombinant transcription factors, the gene inserts were cloned under the control of a strong inducible T7 promoter of the pET28a(+) vector. pET28a(+) is a standard expression

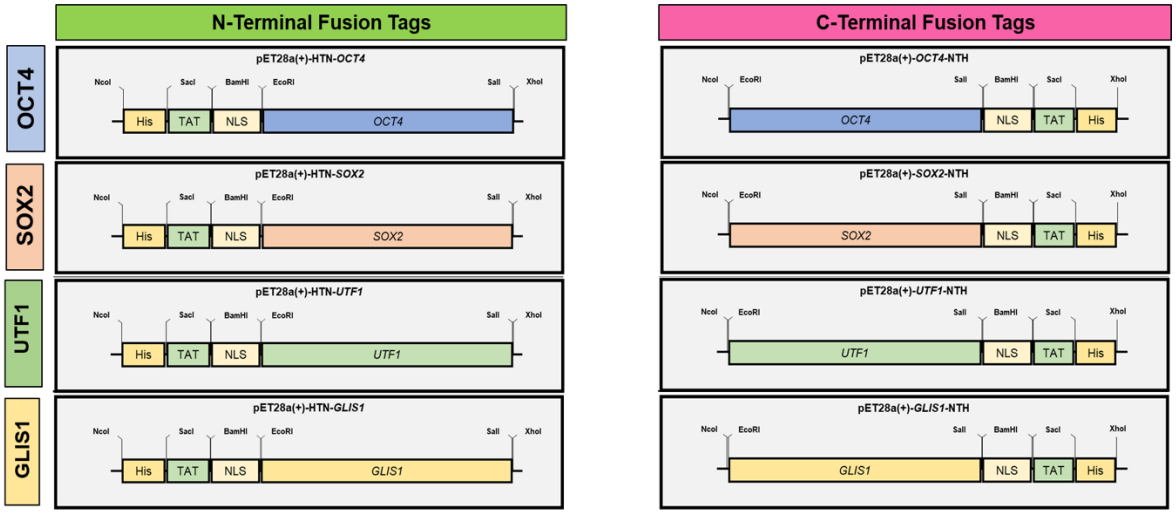


Figure 2.4 The schematic of the gene constructs (HTN-*GOI* and GOI-NTH). Genes cloned in pET28a(+) vector, with fusion tags cloned at both 5' and 3' ends. GOI, gene of interest vector with various features to facilitate heterologous protein expression in bacteria. These gene inserts were cloned between *Nco*I and *Xho*I restriction sites of this vector (Figure 2.4). The empty vector, pET28a(+) only, was also taken as a control to confirm the absence of the gene of interest (data not shown). The resulting plasmids pET28a(+)-HTN-*GOI* and pET28a(+)-*GOI*-NTH were verified using restriction digestion analysis (Figure 2.5). Further, the fidelity of the sequence was confirmed using DNA sequencing.

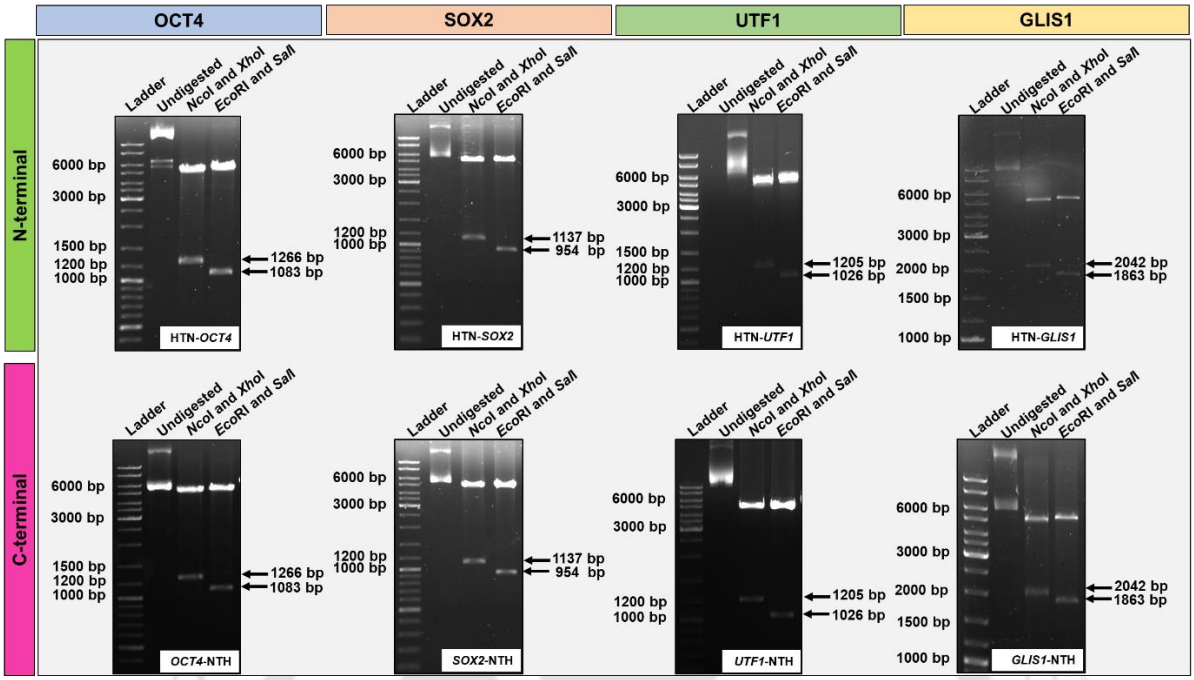


Figure 2.5 Confirmation of cloning of the gene of interests in expression vector. Restriction digestion analysis of the 5'- and 3'-terminal tagged genes in pET28a(+) expression vector. The genes were cloned using *Nco*I, and *Xho*I restriction endonucleases. Fusion tags: N; NLS- Nuclear localization signal, T; TAT- Transactivation of transcription, H; His- Histidine.

2.2.3 Selection of strain and media conditions played a crucial role in heterologous expression of GLIS1 fusion protein

GLIS1 is a ~73 kDa protein and failed to show substantial expression in TB media unlike other proteins (OCT4, SOX2 and UTF1). Hence, to maximize the overall expression we screened for optimal strain and media to express the protein. Two *E. coli* strains, BL21(DE3) and Rosetta (BL21 derivatives designed to improve the expression of eukaryotic proteins that contain codons rarely used in *E. coli*) were used for this study. These two strains were selected because BL21(DE3) has a strong T7 promoter system, lacks *lon* and *Omp*T proteases, and is compatible with the pET expression vector (Rosano and Ceccarelli, 2014). On the other hand, Rosetta is a

highly engineered strain and a derivative of BL21(DE3) containing plasmid pRARE, used to express genes containing rare codons (Berrow et al., 2006; Kaur et al., 2018). Selecting these two strains would give us an idea of the presence of any rare codon in the sequences, compromising the expression of the gene of interest. GLIS1 has been previously reported that expressing higher molecular weight protein is a tedious process and leads to lower protein expression (Francis and Page, 2010). Thus, Rosetta was chosen for the initial expression analysis along with BL21(DE3) for expressing HTN-GLIS1 and GLIS1-NTH fusion protein. Rosetta strain is reported to increase the expression and purification of human recombinant proteins containing rare codons (Tegel et al., 2010). Although the gene sequences were codon optimized, to rule out the possibility of the presence of any rare codons, Rosetta was chosen for comparison. In addition, two culture media, LB and TB (more nutritionally rich), were also used in the first screening process. However, post-induction, an intense band at ~73 kDa GLIS1-NTH fusion protein, was observed in BL21(DE3) transformed clones compared to Rosetta in both LB and TB media (Figure 2.6 A, *top* and *bottom*; 2.6 B *top* and *bottom*). However, HTN-GLIS1 showed no expression in any of the conditions (Figure 2.6 A, *top* and *bottom*; 2.6 B, *top* and *bottom*).

Interestingly, degradation was observed for HTN-GLIS1 in BL21(DE3) grown in TB media (Figure. 2.6 B, *bottom*). No significant difference in the overall cell biomass was observed in BL21(DE3) transformed with GLIS1-NTH in both LB and TB at 37 °C (data not shown). This implied that codon optimization was efficiently performed, and the low expression in TB media compared to LB media was due to reasons other than the presence of rare codons. In general, a similar observation of high expression in BL21(DE3) strain and low expression in Rosetta strain was also made by a study purifying Cas9 protein (Pannunzio

Carmignotto and Rodrigues Azzoni, 2019) and few other proteins from a set of 68 human proteins (Tegel et al., 2010). We speculate that the extra metabolic burden due to an additional plasmid pRARE and the presence of chloramphenicol might have resulted in the low GLIS1 gene expression in Rosetta, similar to what was reported earlier with other recombinant proteins (Pannunzio Carmignotto and Rodrigues Azzoni, 2019; Søgaard and Nørholm, 2016; Tegel et al., 2010).

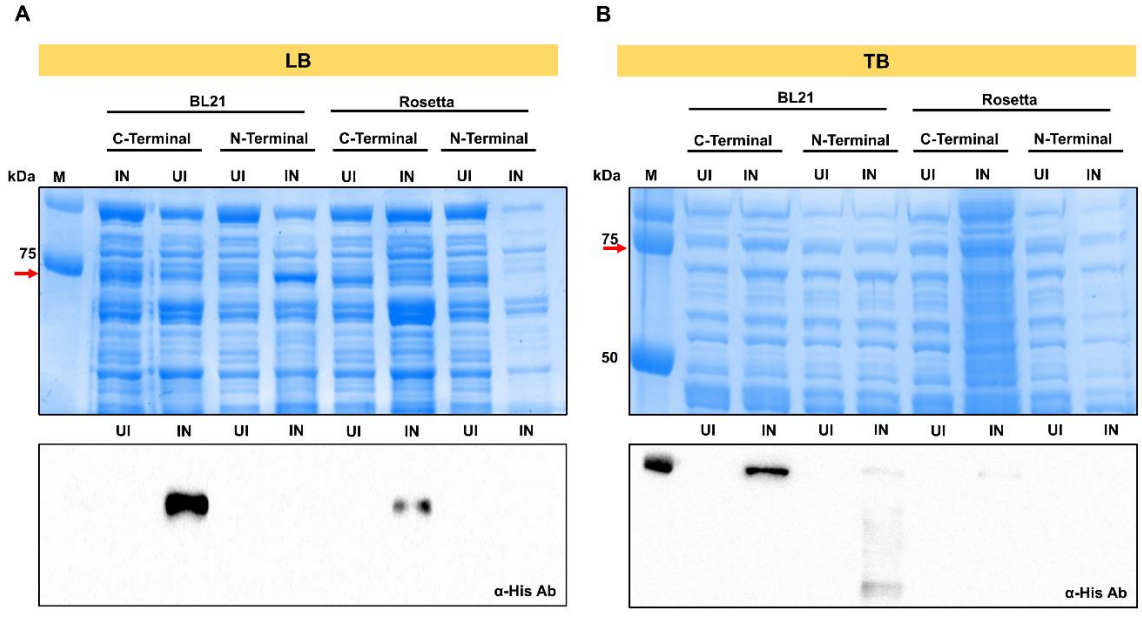


Figure 2.6 Screening for suitable expression host strain, media and appropriate gene construct for maximum expression of recombinant GLIS1 protein. (**A and B**) The HTN-GLIS1 and GLIS1-NTH were expressed in both BL21(DE3) and Rosetta in LB and TB media at 37 °C with 0.5 mM IPTG for 2 hours. The harvested cells were lysed and total lysates were run on SDS-PAGE. Resolved gels were stained with Coomassie Brilliant Blue (*top*) and Western blotting was performed for the same samples with anti-His antibody (*bottom*). All the experiments were performed n=2 times. M, protein marker; UI, Uninduced (total cell lysate);

IN, Induced. Red arrows indicate the molecular weight at which the protein is expressed (~73 kDa).

In agreement with earlier studies, our study showed that the inclusion of plasmids containing extra copies of low abundance of tRNA genes and the presence of chloramphenicol may cause negative metabolic side effects decreasing the expression levels of recombinant proteins, but this may not be the case for other recombinant proteins and detailed screening will be required. Also, the higher expression of GLIS1-NTH in the case of LB compared to TB was observed. We speculate that TB being a more nutritionally rich medium (composed of increased concentrations of peptone, yeast extract, and glycerol as a carbon source) might have contributed to the accumulation of acetate in the culture, presumably affecting the overall expression of GLIS1-NTH fusion protein compared to LB media. Earlier studies have reported that acetate acts as an inhibitor in biomass production and, thus, reduces recombinant protein production (De Mey et al., 2007; Kleman and Strohl, 1994). Although we did not see any difference in the overall cell biomass between LB and TB at 37 °C, the expression of GLIS1-NTH was affected in the case of TB compared to LB. This could be due to overflow metabolism (the Crabtree effect) producing acetate because of the presence of excess carbon sources in the TB medium (Åkesson et al., 2001). Other studies have also reported that *E. coli*, in the presence of excess carbon sources like glucose and glycerol, produces acetate as a by-product due to acidic fermentation (Eiteman and Altman, 2006; Martínez-Gómez et al., 2012). Therefore, we confirmed that GLIS1-NTH showed maximum expression in *E. coli* strain BL21(DE3). GLIS1-NTH also shows comparatively higher expression when expressed in LB media. This study confirmed the role of screening strains and media conditions for maximizing

protein expression to be one of the critical factors in heterologous recombinant protein expression, especially for higher molecular weight proteins.

2.2.4 Identification of optimal induction parameters for obtaining maximum heterologous expression in *E. coli*

E. coli is an ideal expression host due to following characteristics: well-known genetics, high transformation efficiency, rapid growth, and economical protein production (Burgess-Brown et al., 2008). Numerous studies have reported the identification of optimal expression parameters to achieve maximal expression of biologically active recombinant proteins from *E. coli* in a soluble form (Azaman et al., 2010; Galloway et al., 2003; Rabhi-Essafi et al., 2007; Ryan and Henehan, 2013; San-Miguel et al., 2013; Sørensen and Mortensen, 2005; Vasina and Baneyx, 1997). *E. coli* strain BL21(DE3) was chosen due to various reasons: (i) commonly used for recombinant protein production, (ii) compatible with pET28a(+) vector, (iii) engineered to produce T7 RNA polymerase, (iv) deficient in *lon* and *OmpT* proteases, and (v) allows high-level of stable expression of the protein of interest. Therefore, to achieve high expression of fusion proteins (HTN-*GOI* and *GOI*-NTH) in *E. coli*, the gene constructs were transformed into *E. coli* strain BL21(DE3) and screened for inducer concentration (IPTG), optical density (OD), and induction time.

Several studies have reported that reducing the inducer concentration, time, and temperature helps maximize the solubility of the protein, reduces the metabolic burden, and facilitates protein folding (Baneyx and Mujacic, 2004; Kaur et al., 2018; Sørensen and Mortensen, 2005). This step is crucial to avoid protein purification from inclusion bodies that contain either misfolded or partially folded proteins. Hence, it demands solubilization with strong detergents and refolding to the native state (Baneyx and Mujacic, 2004).

Based on these studies, we first screened the minimum IPTG concentration for all the four transcription factors and investigated the impact of different IPTG concentrations on their overall expression in *E. coli*. The results showed the difference in expression based on varying IPTG concentrations for each transcription factor (Table 2.5). Recombinant OCT4 and UTF1 fusion protein showed maximum expression at 0.25 mM IPTG concentration, whereas recombinant SOX2 showed maximum expression at 0.1 mM and GLIS1 at 0.05 mM IPTG concentration. Thus, this study implies that maximum expression is achieved when the gene is induced with minimum optimal inducer concentration (Table 2.5). The bacterial growth phase also plays a critical role in the soluble expression of recombinant proteins. Several studies have reported that the maximum expression of the protein was achieved at the early to mid-log phase (OD_{600} 0.1–0.5) (Jadeja et al., 2016; San-Miguel et al., 2013). Higher cell densities lead to exhaustion of nutrients leading to nutrient deprivation, production of acetate, reduced dissolved oxygen, and increased carbon dioxide post-induction.

These factors contribute to the decrease in the expression of recombinant proteins (Hu et al., 2015; Kaur et al., 2018; Olaofe et al., 2010; Paulina Balbás, 2001). Upon screening we found that the maximum protein expression for all the four recombinant transcription factors were observed at OD_{600} ~0.5 when induced with optimized minimum inducer concentration (Table 2.5).

In addition, studies have demonstrated that the reduction in temperature will enhance solubility and prevent aggregation of the protein of interest (Ryan and Henehan, 2013; San-Miguel et al., 2013; Sørensen and Mortensen, 2005; Vasina and Baneyx, 1997). Therefore, different temperatures (37 °C and 18 °C) for OCT4, SOX2 and UTF1 fusion proteins (37, 30,

25, and 18 °C) were screened for GLIS1-NTH fusion proteins to obtain the maximal amount of soluble protein to allow us to purify under native conditions.

Table 2.5 Optimized induction parameters for maximum expression of recombinant fusion proteins in *E. coli*.

Transcription factors	IPTG (mM)	OD_{600} (Abs)	Time (hour)	Temperature (°C)
OCT4	0.25	~0.5-0.6	2	37
SOX2	0.25	~0.5-0.6	2	37
UTF1	0.25	~0.5-0.6	2	37
GLIS1	0.05	~0.5-0.6	4	37

Abs, Absorbance

This step is critical to circumvent purification from inclusion bodies containing partially folded or misfolded proteins. Therefore, it necessitates solubilization with a strong detergent and refolding to the native state (Baneyx and Mujacic, 2004). Refold proteins to their native state is a cumbersome process, and hence, we screened for the optimal temperature for maximum soluble expression of our recombinant proteins. We also investigated the influence of the position of tags on the solubility of the proteins and hence, analyzed both HTN-GOI and GOI-HTN tagged protein expression. Recombinant OCT4 fusion proteins (HTN-OCT4 and OCT4-NTH) showed expression at both temperatures. However, OCT4-NTH showed increased expression compared to HTN-OCT4 at 37 °C, whereas HTN-OCT4 showed soluble expression at 18 °C than OCT4-NTH. Despite lowering the temperature, no protein in the soluble fraction was observed for OCT4-NTH fusion protein (Figure 2.7 A). In the case of SOX2, HTN-SOX2 fusion protein showed increased expression compared to SOX2-NTH at

37 °C. Similar to 37 °C, at a lower temperature (18 °C), soluble expression of HTN-SOX2 was more compared to its counterpart SOX2-NTH (Figure 2.7 B). In addition to the full-length intact protein, faint truncated protein fragments were also observed for both OCT4-NTH and SOX2-NTH fusion proteins, as observed in Western blotting (Figure 2.7 A and B). The inclusion of protease inhibitors in the solutions did not decrease proteolysis (data not shown). The truncated fragments may be due to proteolysis at specific sites in some protein molecules during expression and purification (Ryan and Henehan, 2013). It provides evidence that truncations were due to the presence of internal sites prone to cleavage during expression (Ryan and Henehan, 2013) or due to the presence of intragenic sequences mimicking *E. coli* ribosomal entry sites (Maertens et al., 2010). A similar observation of faint truncated protein fragments was also reported with mouse OCT4 protein expressed in *E. coli* in an earlier study (Bosnali and Edenhofer, 2008). Also, it was observed that both the fusion constructs showed variable results in the amount of soluble protein at their respective temperatures (Figure 2.7 A and B). These results established that the placement of fusion tags at different ends can affect the expression and solubility of the protein of interest. Moreover, it is also in agreement with an earlier study that reported similar observations with mouse OCT4 and SOX2 protein (Bosnali and Edenhofer, 2008). Thus, for further purification, both the N- and C-terminal tagged OCT4 and SOX2 constructs were used. UTF1-NTH showed expression of the protein in the soluble fraction at both temperatures (Figure 2.7 C). HTN-UTF1 failed to express at any given condition and hence was not used for further experiments. Interestingly, GLIS1-NTH upon optimization of induction temperature, expressed only at 37 °C. Thus, we next sought to understand whether GLIS1-NTH expresses protein in soluble fraction at 37 °C.

Upon analysis, soluble expression was observed at 37 °C (Figure 2.7 D). Both UTF1-NTH and GLIS1-NTH did not show any truncation unlike OCT4 and SOX2 proteins.

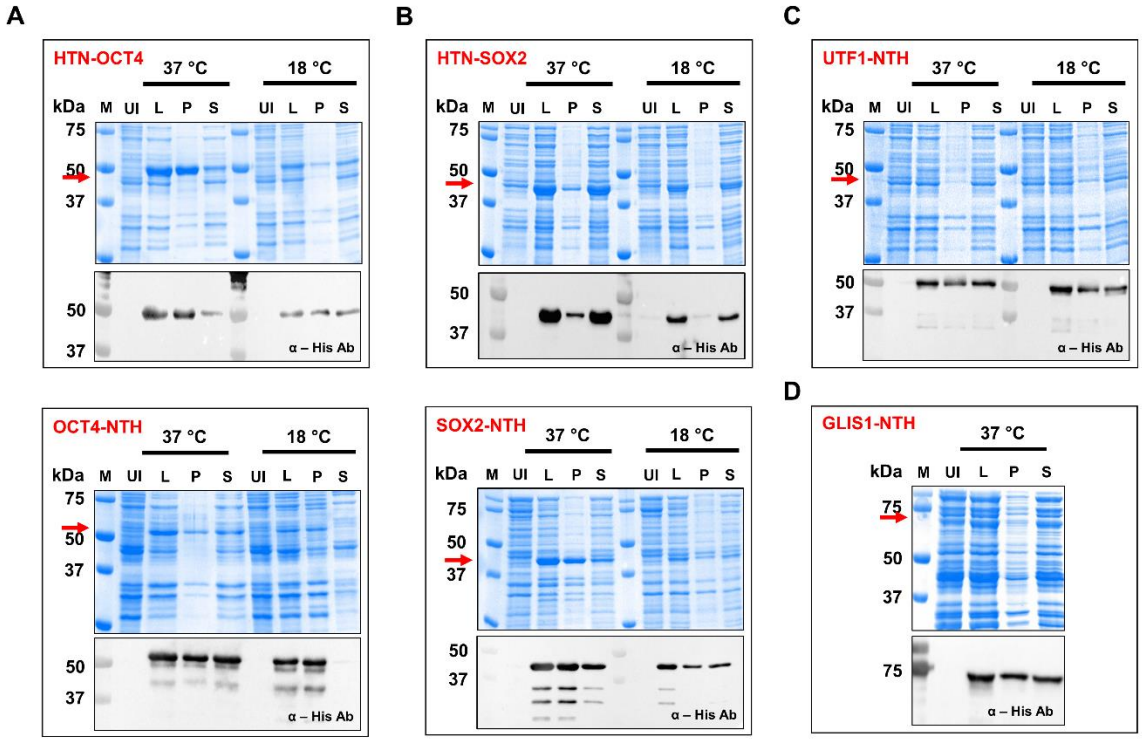


Figure 2.7 Screening and identification of optimal induction parameters for obtaining maximal soluble expression for recombinant human fusion proteins in *E. coli*. The proteins were induced at optimal induction parameters and the harvested cells were then lysed to obtain total cell lysate (L) and further centrifuged to obtain pellet (P) and supernatant (S). The L, P, and S fractions were then normalized based on the respective L fraction and analyzed using SDS-PAGE gel. An equal volume of P and S fractions were loaded corresponding to their L fraction to analyze the distribution of protein in L, P, and S. The gels were either stained with Coomassie Brilliant Blue (*top*) or Western blotting was performed with anti-His antibody (*bottom*). (**A**) Recombinant OCT4 fusion proteins tagged at both N-terminal and C-terminal (**B**) Recombinant SOX2 fusion proteins tagged at both N-terminal and C-terminal (**C**)

Recombinant UTF1-NTH fusion protein (**D**) Recombinant GLIS1-NTH fusion protein. All the experiments were performed n=3 times. M, protein marker; UI, Uninduced (total cell lysate); L, Lysate; P, Pellet; S, Supernatant (soluble) fraction. Red arrows indicate the molecular weight and position at which the protein is expressed.

In addition, induction time also influences the recombinant protein expression thus, we screened and identified the optimal induction time for the maximal soluble expression of these recombinant fusion proteins. The results showed that OCT4, SOX2, and UTF1 show maximum expression when induced for 2 hours at 37 °C, and GLIS1 shows maximum expression when induced for 4 hours at 37 °C (Table 2.5). These results confirm the influence of the position of fusion tags playing a significant role in the soluble expression of the protein. The time and temperature also contribute in maximizing their soluble expression. Our previously published papers (ETS2, GATA4, MESP1, HAND2, TBX5, PDX1 and NGN3) have also confirmed the critical role of optimal induction parameters in the heterologous expression of recombinant proteins in *E. coli* (Haridhasapavalan et al., 2020, 2021a, 2022b, 2022a; Kumar Haridhasapavalan et al., 2021; Narayan et al., 2021b, 2021a). From these results, we confer that HTN-OCT4 and OCT4-NTH expressed at 37 °C; HTN-SOX2 and SOX2-NTH expressed at both 37 °C and 18 °C; UTF1-NTH at 37 °C and 18 °C GLIS1-NTH at 37 °C are the most suitable genetic constructs. These constructs were further used for the soluble expression and purification of these fusion proteins under the optimized parameters.

2.3 Conclusion

This chapter highlights the significant change in the RAV and CAI value of the genes after codon optimization and identification of optimal induction parameters for maximal soluble expression of four heterologously expressed recombinant fusion proteins in *E. coli*. We

demonstrate the influence of bacterial strain and media conditions on the expression of a large protein such as GLIS1. In addition to the induction parameters (IPTG, OD_{600}, temperature, and time), it also provides an insight into how the position of tags influences the total protein expression and their solubility. The optimization of maximal soluble expression showed that all the recombinant proteins (OCT4, SOX2, UTF1 and GLIS1) show maximum expression at 37 °C compared to 18 °C. We observed that GLIS1-NTH is expressed in both BL21(DE3) and Rosetta, with less expression in the Rosetta strain grown in TB media compared to LB media. Whereas, HTN-GLIS1 failed to express at any given strain and condition. Interestingly, HTN-OCT4, OCT4-NTH; HTN-SOX2, SOX2-NTH and UTF1-NTH were expressed in both the LB and TB media. Similar to HTN-GLIS1, HTN-UTF1 also failed to express at any given optimized induction conditions. With respect to the position of tags and their influence on the expression and solubility of the protein, the study revealed that indeed changing the position changes the expression pattern. As a proof of concept, both OCT4 and SOX2 were tagged at both the terminals. Although both proteins showed soluble expression, the difference in overall expression was evident based on the position of tags. However, in the case of UTF1-NTH, both 37 °C and 18 °C showed soluble expression, and no expression was observed in the case of HTN-UTF1. Similarly, GLIS1-NTH only showed soluble expression at 37 °C. Hence, this study not only established a protocol to successfully express these human recombinant fusion proteins in *E. coli* but, also provides us with an understanding of the position of tags in quality, quantity, and solubility of the protein.

CHAPTER 3

Native purification, biochemical and biophysical analysis of recombinant proteins

Brief summary of the chapter

Protein purification under native conditions provides a better chance of retaining protein functionality by maintaining the native-like structure of the purified proteins. In this chapter, we demonstrate and establish a protocol for homogeneous purification of these reprogramming recombinant proteins from *E. coli*. To achieve this, we have performed affinity purification using immobilized metal ion affinity chromatography (IMAC) from the supernatant fraction of the total cell lysate. The purification conditions differed for each protein based on their overall expression in the soluble fraction, isoelectric point, and interaction with buffer composition such as salt and glycerol. Based on these critical parameters we established that OCT4, SOX2 and GLIS1 require ionic interaction for successful purification, however, UTF1 requires no salt condition for purification. We confirmed the protein identity using Western blotting analysis with both anti-His and respective protein-specific antibodies. The secondary structure analysis of the purified proteins was determined using far ultraviolet circular dichroism spectroscopy. Upon analysis, the spectra obtained confirmed that the proteins retained their secondary structure conformation. In conclusion, this chapter provides an insight into the importance of different parameters in protein purification and provides a methodology to successfully purify recombinant proteins.

3.1 Materials and methods

3.1.1 Protein purification of recombinant fusion proteins using immobilized metal ion chromatography (IMAC) method

Native purification was carried out using immobilized metal ion affinity chromatography (IMAC) method for the recombinant fusion proteins. Large volume of cultures (1.2 L) were induced using the optimized induction parameters (elaborately explained in Table 2.2 of chapter 2). Prior to resuspension, the wet cell pellet weight was recorded for determining the resin volume to be used for purification. The induced culture pellets were resuspended in 20 ml for UTF1 and 40 ml for OCT4, SOX2 and GLIS1 with pre-chilled lysis buffers as per Table 2.3. Resuspended pellets were then sonicated using ultrasonication at an amplitude of 33% (Pulse ON: 5 secs and OFF: 25 secs) using Vibracell^{TM} VCX-130 cell disruptor Sonics and Materials Inc., Newtown, CT, The USA on ice for 15-50 minutes depending on the protein of interest. The cell suspension was clarified by centrifugation at 8500-9000 rpm, 4 °C for 30 minutes to separate lysate and supernatant fraction for purification.

Column preparation: Prior to binding of the supernatant fraction with the nickel-nitrilotriacetic acid (Ni-NTA) resin, the column was washed with double distilled deionized water followed by equilibration using lysis/equilibration buffer (Table 3.1).

Purification: The supernatant fraction was incubated with Ni-NTA resin beads for 2 hours in shaking conditions at 4 °C. The resin and supernatant were then loaded onto the 20 ml purification column (Bio-Rad) and the sample for flow-through was collected. The column was subsequently washed thrice with gradually increasing concentrations of wash buffers (50/100/150/200 mM imidazole washes) (Table 3.1). The bound proteins were eluted using

respective elution buffers (Table 3.1). The eluted proteins were stored at 4 °C and checked immediately onto SDS-PAGE and identity was confirmed using western blotting.

Table 3.1 Native purification buffers used for the respective human recombinant fusion proteins.

Proteins	Buffer components	Resuspension buffer	Washes				Elution buffer
			1	2	3	4	
OCT4	PB (mM)	20	20	20	20	20	20
	NaCl (mM)	150	150	150	150	150	150
	Imidazole (mM)	20	50	100	150	200	500
	Glycerol (%)	20	20	20	20	20	20
	pH (at RT)	~7.2	~7.2	~7.2	~7.2	~7.2	~7.2
SOX2	PB (mM)	20	20	20	20	-	20
	NaCl (mM)	150	150	150	150	-	150
	Imidazole (mM)	20	50	100	150	-	500
	Glycerol (%)	20	20	20	20	-	20
	pH (at RT)	~7.8	~7.8	~7.8	~7.8	-	~7.8
UTF1	PB (mM)	20	20	20	20	-	20
	NaCl (mM)	0	0	0	0	-	150
	Imidazole (mM)	20	50	100	150	-	500
	Glycerol (%)	20	20	20	20	-	20
	pH (at RT)	~7.8	~7.8	~7.8	~7.8	-	~7.8
GLIS1	PB (mM)	20	20	20	20	-	20
	NaCl (mM)	150	150	150	150	-	150
	Imidazole (mM)	20	50	100	150	-	500
	Glycerol (%)	-	-	-	-	-	-
	pH (at RT)	~7.2	~7.2	~7.2	~7.2	-	~7.2

PB, sodium phosphate buffer; RT, room temperature

The purified proteins were further desalted using PD10 column (GE healthcare) against DMEM cell culture media as per the manufacturer's protocol. The desalted proteins were then flash frozen and stored in -80 °C until further use.

3.1.2 Mass spectrometry

3.1.2.1 In gel digestion

The purified recombinant fusion protein was run on SDS-PAGE gel and stained with staining solution containing 50% (v/v) methanol, 10% (v/v) acetic acid, 0.4% (w/v) Coomassie Brilliant Blue G-250 in deionized water. The desired band was excised and destained with 40 mM ammonium bicarbonate (HiMedia) in a 40% (v/v) acetonitrile solution (Merck Millipore). The destained gel was then treated with reduction solution (5 mM dithiothreitol (Sigma-Aldrich) in 40 mM ammonium bicarbonate) for 30 minutes at 60 ºC, followed by alkylation solution (20 mM iodoacetamide (Sigma-Aldrich) in 40 mM ammonium bicarbonate) for 10 minutes at room temperature (dark). The excised gel slice was dehydrated by adding 100% (v/v) acetonitrile and then digested with trypsin (Promega). After overnight digestion, the peptides were extracted using extraction buffer (5% formic acid (Merck Millipore) and 40% acetonitrile (Merck Millipore)) and were vacuum-dried using SpeedVac and stored at -80 °C until further use.

3.1.2.2 LC-MS/MS analysis

For liquid chromatography-tandem mass spectrometry (LC–MS/MS), the dried peptides of the OCT4-NTH fusion protein were reconstituted in 0.1% formic acid (Merck Millipore) and analyzed using Q Exactive™ (Thermo Scientific™) mass spectrometer coupled with the Proxeon Easy nLC system (Thermo Scientifc™). For peptide enrichment, protein fragments

were passed onto an Acclaim™ PepMap™ trap column (Michrom Biosciences Inc.). Peptides were separated on an analytical column employing a linear gradient of 7–30% acetonitrile for 80 minutes. MS and MS/MS scan acquisitions were executed in the quadrupole Orbitrap mass analyzer. For data analysis, the MS data were searched against NCBI HsRefSeq81 (human protein database; version 81) and analyzed using the SequestHT program and Proteome Discoverer software (Thermo Fisher Scientific; version 1.4.0.288)

3.1.3 Far ultraviolet circular dichroism spectroscopy

The secondary structure of the purified recombinant fusion proteins (HTN-GOI and GOI-NTH) were determined using far ultraviolet (UV) circular dichroism (CD) spectroscopy (J-815/J-1500 spectropolarimeter (Jasco, Japan)), equipped with a thermoelectric cooling-based temperature control unit. The spectral analysis was recorded as an average of 10 accumulations from wavelength 260-190 nm in quartz cuvette with pathlength 0.1 cm at a scan rate of 100 nm/min with a data integration time of 1 sec. The background noise due to protein elution media (DMEM) was subtracted from the final spectrum. The spectra of the respective proteins were analyzed using in silico Beta Structure Selection (BeStSel) online tool (http://bestsel.elte.hu/index.php) to estimate the secondary structure of the purified protein (Micsonai et al., 2018), an online tool commonly used to estimate the secondary structure content of purified proteins.

3.1.4 SDS-PAGE and Western blotting analysis

SDS-PAGE and Western blotting analysis were performed for the purified recombinant fusion proteins as per the described methodology in Chapter 2.

3.2 Results and discussion

3.2.1 Purification of human recombinant fusion proteins

Native purification using IMAC to purify recombinant proteins has been proven to be one of the most widely used, cheapest and facile approach. It relies on the interaction between the transition metal ion Ni^{2+} immobilized onto resin beads. The imidazole group of histidine forms coordinate bonds with the immobilized transition metal (Bornhorst and Falke, 2000). We expressed the proteins in their soluble forms and then purified human recombinant fusion proteins (HTN-GOI and GOI-NTH) using one-step affinity purification method. Several studies have reported that purification of recombinant proteins under native purification method using IMAC retains their native-like secondary structure conformation, and are potentially biologically active in nature (Bosnali and Edenhofer, 2008; Haridhasapavalan et al., 2022b, 2022a; Münst et al., 2016; Narayan et al., 2021b; Peitz et al., 2014; Stock et al., 2010; Thier et al., 2012). We have used this approach for purifying these proteins tagged with poly-histidine tags.

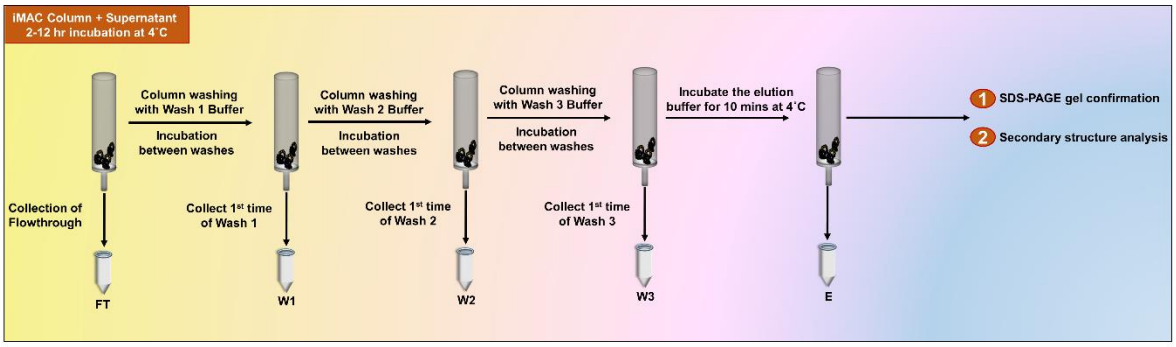


Figure 3.1 Schematic representation of native purification workflow. FT, flow-through; W1/2/3, wash buffers 1/2/3; E, elution.

The proteins were expressed as per the optimized induction parameters (Chapter 2, Table 2.5) in large volumes and the supernatant fraction after further sonication and centrifugation were used for native purification. The schematic explaining the workflow of native purification is shown in Figure 3.1.

3.2.1.1 Purification of human recombinant OCT4 fusion proteins

Based on the previous studies (Bosnali and Edenhofer, 2008; Echigoya et al., 2020; Lee et al., 2012; Pan et al., 2010; Zhou et al., 2009), it was crucial to develop a method to purify recombinant human OCT4 fusion protein in its native-like state, as earlier methods included purification from inclusion bodies. Previous methods are extremely cumbersome, time-consuming and expensive; hence, we purified the N- and C-terminal-tagged human OCT4 (HTN-OCT4 and OCT4-NTH) recombinant proteins using IMAC under native conditions in a facile manner. The soluble protein fractions of both HTN-OCT4 and OCT4-NTH were purified, and the purity was confirmed using SDS-PAGE analysis (Figure 3.2 A (*right*); Figure 3.2 B (*left*)). Although faint truncated protein fragments of OCT4 were observed at around ~43 kDa, most of it was retained as a full-length protein of ~45 kDa (Figure 3.2 A, B). The identity of the OCT4 fusion proteins was confirmed with Western blotting using an anti-His antibody (OCT4-NTH; Figure 3.2 A (*middle*)), and an anti-OCT4 antibody (HTN-OCT4, Figure 3.2 A (*bottom*); OCT4-NTH, Figure 3.2 B (*bottom*)). Based on the overall expression of the recombinant OCT4 fusion proteins, the total yield of HTN-OCT4 and OCT4-NTH was 3.4 mg/L and 6 mg/L (summarized in Table 3.2). Despite purifying full-length OCT4 fusion proteins, the overall yield was low and significant amount of protein loss was observed in the flow-through and wash 1 (Figure 3.2 A and B). Considerable amount of protein was also detected in the resin fraction which could not be eluted, thus contributing to the decrease in

total yield of the proteins. To overcome the loss of protein in flow-through/unbound fraction, wash 1 and to elute tightly bound proteins from the resin, we used different salt concentrations in lysis buffers and higher imidazole concentration in the elution buffer (Haridhasapavalan et al., 2022b, 2022a).

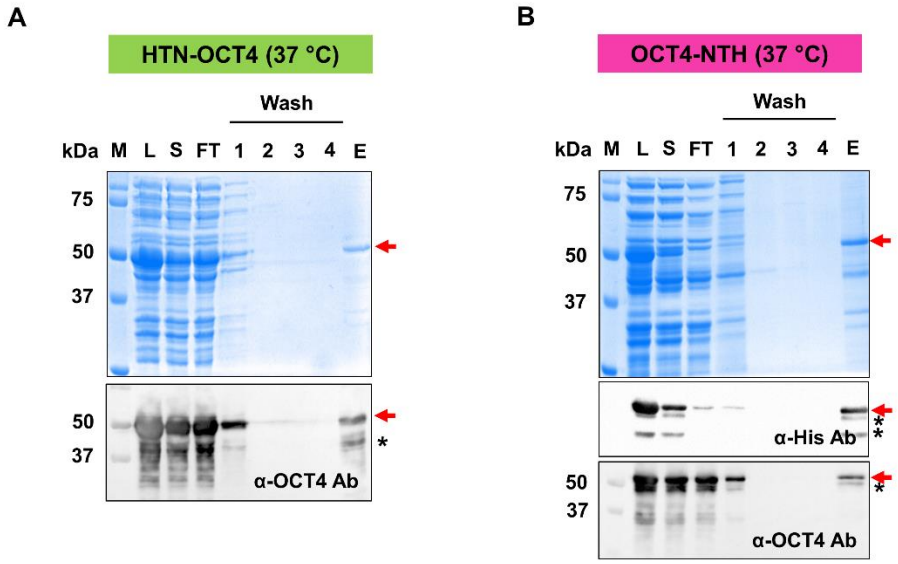


Figure 3.2 Purification of human OCT4 fusion proteins using native affinity purification. **(A)** Recombinant human HTN-OCT4 fusion protein purification. **(B)** Recombinant human OCT4-NTH fusion protein purification. M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; FT, flow-through fraction; W, wash fraction; E, eluted fraction; *truncated fragments of recombinant OCT4 fusion proteins. Red arrows indicate the molecular weight at which the protein is expressed (~45 kDa).

However, no change was observed in the yield as the protein loss was still observed in the unbound fraction and tightly bound proteins were observed in resin. Although yield is less, the SDS-PAGE and Western blotting confirmed the successful purification of OCT4 fusion proteins. In this study, we are the first to establish a method for purifying full-length human recombinant

Table 3.2 Purification summary of recombinant fusion proteins.

Protein	Steps	Total protein (mg)[b]	Target protein (mg)[c]	Yield (%)	Purity (%)[d]
HTN-OCT4	Crude lysate [a]	110.35	16.81	100	15.23
	Cleared lysate (soluble)	92.66	13.61	80.99	14.69
	IMAC (pooled)	0.56	0.53	3.19	95.65
OCT4-NTH	Crude lysate [a]	93.51	12.07	100	12.91
	Cleared lysate (soluble)	91.67	10.82	89.67	11.8
	IMAC (pooled)	0.97	0.89	7.42	92.32
HTN-SOX2	Crude lysate [a]	77.2	60.37	100	78.2
	Cleared lysate (soluble)	49.5	37.14	61.52	75.03
	IMAC (pooled)	0.11	0.10	0.17	96.05
SOX2-NTH	Crude lysate [a]	84.65	76.61	100	90.53
	Cleared lysate (soluble)	68.7	58.92	76.88	85.76
	IMAC (pooled)	0.29	0.28	0.37	97.1
UTF1-NTH (induced at 37 °C)	Crude lysate [a]	93.28	7.37	100	7.9
	Cleared lysate (soluble)	81.28	5.31	72.02	6.53
	IMAC (pooled)	0.15	0.14	1.83	90.07
UTF1-NTH (induced at 18 °C)	Crude lysate [a]	95.36	7.88	100	8.26
	Cleared lysate (soluble)	89.06	6.25	79.37	7.02
	IMAC (pooled)	0.09	0.08	1.06	93.1
GLIS1-NTH	Crude lysate [a]	39.6	2.61	100	6.58
	Cleared lysate (soluble)	30	1.59	61.25	5.32
	IMAC (pooled)	0.11	0.10	3.86	94.92

[a] From 1 g wet weight of induced *E. coli* BL21(DE3) cell pellet (from 0.15-0.35 L of culture).

[b] Protein concentration determined by Bradford assay using BSA as a standard protein.

[c] Determined from total protein concentration and purity.

[d] Purity determined from SDS-PAGE analysis using ImageJ software.

OCT4 fusion proteins under native conditions and further these purified proteins were used for further analysis.

3.2.1.2 Purification of human recombinant SOX2 fusion proteins

Previous studies have reported the purification of SOX2 protein from *E. coli* (Chen et al., 2016; Hu et al., 2012; Mirakhori et al., 2015; Pan et al., 2010; Pouya et al., 2020). However, all these studies have reported purification from inclusion bodies, therefore making the solubilization of the purified protein dependent on the use of harsh detergents, making refolding tedious and time-consuming, and commonly ended with a low yield of biologically active protein in its native conformation (Tsumoto et al., 2003)**.** Therefore, it was important to establish a method to purify human recombinant SOX2 fusion protein under native conditions. Hence, to purify HTN-SOX2 and SOX2-NTH, the cultures were induced at 37 °C as per the optimized expression parameters (Chapter 2; Table 2.5). The soluble fraction was then subjected to IMAC purification, which is a versatile technique employed to purify polyhistidine tagged proteins and also help accomplish a high yield of proteins with nearly 95% purity (Bornhorst and Falke, 2000). Purified full-length SOX2 protein band was observed at ~41 kDa along with lower truncated bands. Based on the position of tags, SOX2-NTH yielded more protein after purification compared to HTN-SOX2. The total yield of HTN-SOX2 was 0.7 mg/L and SOX2-NTH was 2.5 mg/L, summarized in Table 3.2. The number of truncations observed in SOX2-NTH was more than HTN-SOX2, and thus we speculate that the position of the tags contributed to the difference in total yield as well as the protein folding. The identity of the SOX2 fusion proteins was confirmed with Western blotting using an anti-His antibody and an anti-SOX2 antibody (HTN-SOX2; Figure 3.3 A (*left*); SOX2-NTH, Figure 3.3 B (*right*)). Similar to OCT4, SOX2 protein loss in flow-through/unbound fraction as well

in wash 1 was observed. Loss of protein in these fractions led to the reduction in the overall protein yield in both the tagged constructs. Although yield is less, but the SDS-PAGE and Western blotting confirmed the successful purification of SOX2 fusion proteins. In this study, we are the first to establish a method for purifying the full-length human recombinant SOX2 fusion proteins under native conditions and these purified fusion proteins were used for analysis.

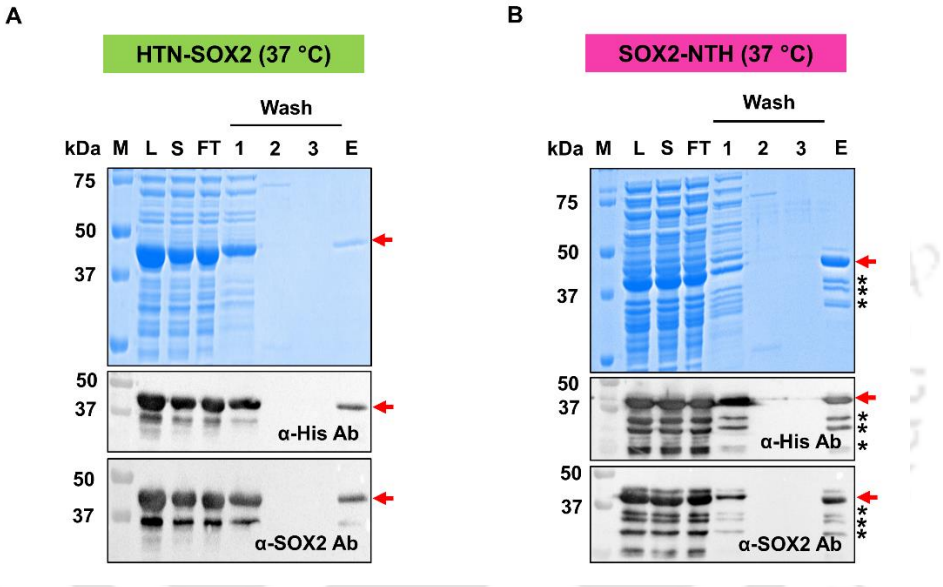


Figure 3.3 Purification of human SOX2 fusion proteins using native affinity purification. **(A)** Recombinant human HTN-SOX2 fusion protein purification. (B) Recombinant human SOX2-NTH fusion protein purification. M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; FT, flow-through fraction; W, wash fraction; E, eluted fraction; *truncated fragments of recombinant SOX2 fusion proteins. Red arrows indicate the molecular weight at which the protein is expressed (~43 kDa).

3.2.1.3 Purification of human recombinant UTF1 fusion protein

UTF1 protein is one of the four factors of our recombinant protein toolbox. It has many applications in generating *bona fide* iPSCs, cancer cells, embryonic development and spermatogenesis. However, to date no report has been published establishing a method to successfully purify UTF1 protein. In this study we have purified the recombinant UTF1-NTH protein from the soluble fraction of the total lysate using Ni^{2+}-NTA affinity purification method. The expression of protein in the soluble fraction was screened based on different salt concentrations.

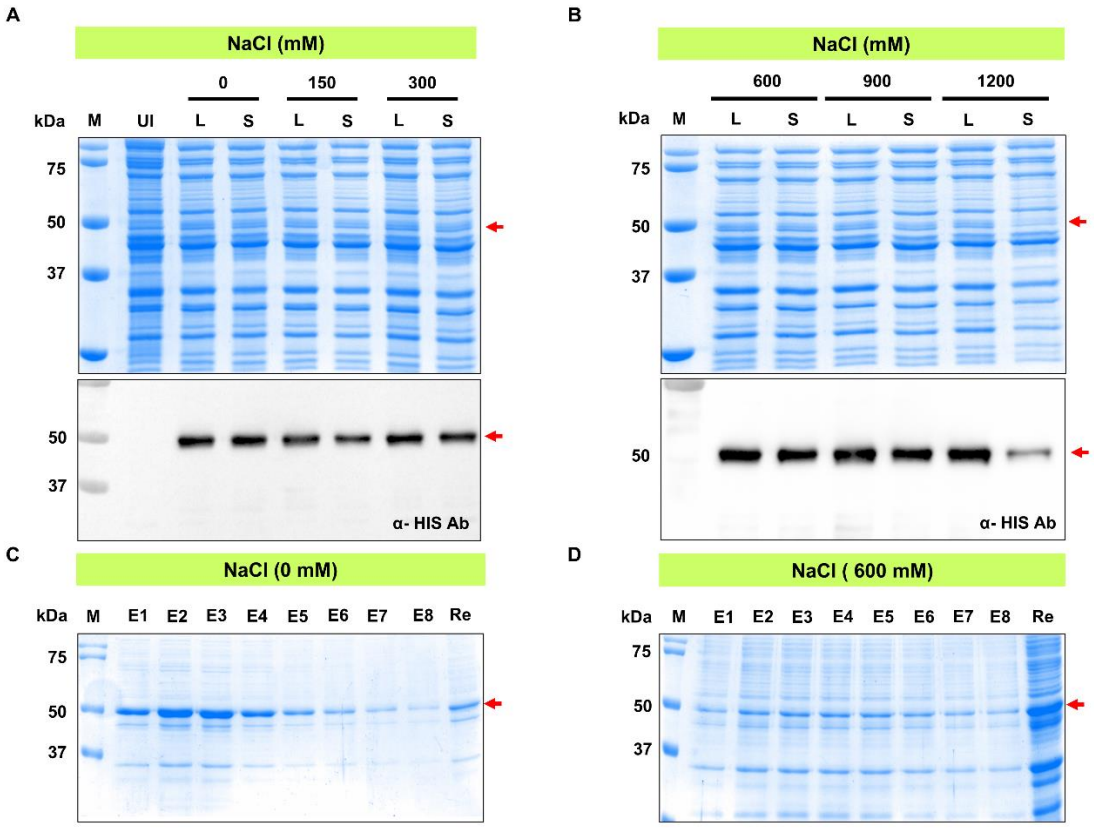


Figure 3.4 Effect of different salt concentrations on the solubility and purification of UTF1-NTH fusion protein. (**A** and **B**) SDS-PAGE and Western blotting analysis of

recombinant UTF1-NTH under gradually increasing salt concentrations in their lysis buffer for understanding the effect of different salt concentration on protein solubility. **(C)** SDS-PAGE analysis of the purification of human UTF1-NTH fusion protein under 0 mM salt concentration. **(D)** SDS-PAGE analysis of the purification of human UTF1-NTH fusion protein under 600 mM salt concentration. M, protein marker; UI, uninduced; L, cell lysate; S, supernatant (soluble) fraction; E, elution fractions; Re, resin. Red arrows indicate the molecular weight at which the protein is expressed (~43 kDa).

This revealed that salt concentration did not play any role in increasing the solubility of the protein, however it had an influence in purification of UTF1-NTH fusion proteins (Figure 3.4 A and B). Upon comparison with 0 mM and 600 mM salt concentration throughout the purification buffers (Figure 3.4 C and D), the results showed full-length protein purification of human UTF1-NTH protein in 0 mM salt condition. Truncations were observed when purified with 600 mM salt buffer. Thus, purification was performed in the absence of salt. Further, the soluble fractions of UTF1-NTH protein expressed at both 37 °C and 18 °C were successfully purified (Figure 3.5 A and B; *top, middle and bottom*). The total protein yield for 37 °C was 0.81 mg/L and 18 °C was 0.54 mg/L (Table 3.2). The total protein yield of the protein was less because of the overall low expression of the protein as well as loss in flow-through and wash 1. Although yield is less, but the SDS-PAGE and Western blotting confirmed the successful purification of UTF1-NTH fusion proteins. In this study, we are the first to establish a method for purifying full-length human recombinant UTF1 fusion protein under native conditions and this purified fusion protein was used for further analysis.

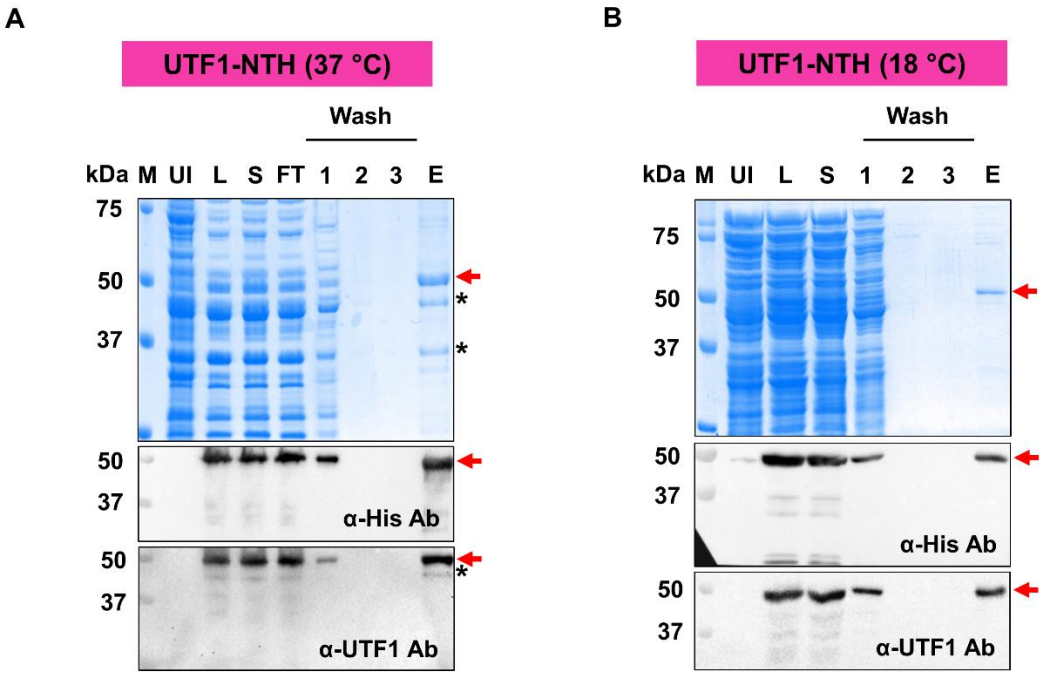


Figure 3.5 Purification of human UTF1 fusion proteins using native affinity purification. **(A)** Recombinant human UTF1-NTH fusion protein purification at 37 °C. (B) Recombinant human UTF1-NTH fusion protein purification at 18 °C. M, protein marker; UI, uninduced; L, cell lysate; S, supernatant (soluble) fraction; FT, flow-through fraction; W, wash fraction; E, eluted fraction; *truncated fragments of recombinant UTF1 fusion proteins. Red arrows indicate the molecular weight at which the protein is expressed (~43 kDa).

3.2.1.4 Purification of human recombinant GLIS1 fusion protein

To date, no study has reported a method to purify GLIS1 protein in its native-like form, thus, we established a method to purify GLIS1 protein under native conditions.

GLIS1-NTH was expressed using identified optimal expression conditions, and the supernatant fraction was used for purification to purify under native conditions. 2.4 L culture expressing recombinant GLIS1-NTH was expressed according to the optimized culture

conditions (Table 2.5). The supernatant was loaded onto the pre-equilibrated Ni^{2+}-NTA column. The purified GLIS1-NTH was observed corresponding to the expected molecular weight of ~73 kDa (calculated molecular weight: 73.017 kDa) in Coomassie-stained SDS-PAGE gel (eluted fraction) and immunoblot (Figure 3.6, *top* and *bottom*). However, the loss of protein was observed in the flow-through fraction in both Coomassie and Western blotting (Figure 3.6, *top* and *bottom*). A very faint-truncated GLIS1 fusion protein at around 45 kDa was also observed in Coomassie-stained SDS-PAGE gel (eluted fraction) and Western blotting using anti-His antibody (Figure 3.6, *top* and *bottom*). GLIS1-NTH being a higher molecular weight protein had overall less expression, and thus less total protein yield. The yield of the protein was 0.52 mg/L as summarized in Table 3.2. The SDS-PAGE and Western blotting confirmed the successful purification of GLIS1-NTH fusion protein. In this study, we are the first to establish a method for purifying full-length human recombinant GLIS1 fusion protein under native conditions and this purified protein was used for further analysis.

3.2.2 Determination of secondary structure of these purified human recombinant fusion proteins

To determine whether these purified fusion proteins retained their secondary structure conformation, we performed far UV CD spectroscopic analysis. This technique is used for secondary structure analysis of proteins purified from cells/tissues (Greenfield, 2006), particularly for proteins whose secondary structure content is unknown, such as OCT4, SOX2,

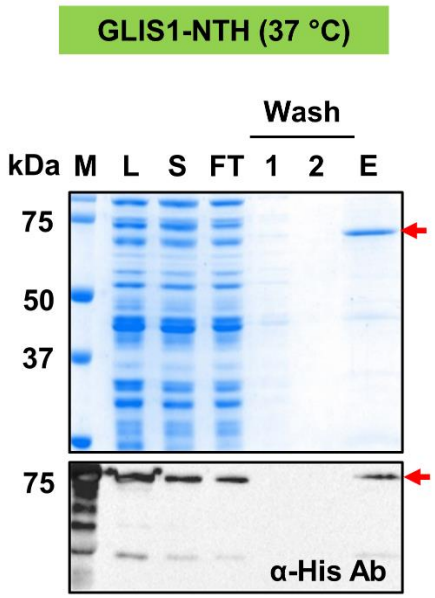


Figure 3.6 Purification of human GLIS1 fusion protein using native affinity purification. Recombinant human GLIS1-NTH fusion protein purification at 37 °C. M, protein marker; L, cell lysate; S, supernatant (soluble) fraction; FT, flow-through fraction; W, wash fraction; E, eluted fraction. Red arrows indicate the molecular weight at which the protein is expressed (~73 kDa).

UTF1 and GLIS1. The characteristic shape and magnitude of the far UV CD spectrum indicate different secondary structures, such as α-helix, β-sheet, turn and random coil (Greenfield, 2006; Kelly et al., 2005). α-helix has a positive peak at 193 nm and negative peaks at 222 nm and 208 nm, while β-sheet has a positive peak at 195 nm and a negative peak at 218 nm. Additionally, the random coil has a positive peak at 210 nm and a negative peak of about 195 nm (Greenfield, 2006). These purified proteins were subjected to de-salting procedure for the removal of salt and imidazole using size exclusion chromatography with PD10 column. The CD spectra analysis of OCT4-NTH showed a positive peak at ~192 nm and negative peaks at ~209 and ~220 (Figure 3.7 A). However, HTN-OCT4 showed a sharp negative peak at ~200

nm (Figure 3.7 A). Based on the literature, only crystal structure of DNA binding domain of mouse OCT4 protein is available, stating the DNA binding domain of OCT4 comprised of majorly α-helix (Esch et al., 2013). The observed spectral pattern for our purified OCT4-NTH fusion protein shows similar spectral pattern corresponding to the α-helix spectra. Minute differences in the spectra were observed due to the presence of fusion tags, influencing the protein conformation as previously reported (Haridhasapavalan et al., 2020, 2021a, 2022b, 2022a; Haridhasapavalan et al., 2021; Narayan et al., 2021b, 2021a). On the other hand HTN-OCT4 CD spectra showed a complete disordered secondary structure conformation (Zhang et al., 2016). Thus, confirming that indeed fusion tags has great influence on the protein conformation. Although, these spectral patterns confirm that the purified proteins maintained their secondary structure but, the major limitation was that these *in silico* tools provide presumptive analysis based of the spectral patterns obtained, protein concentration, molecular weight of the protein and path length. Therefore, a confirmative protein structure for these recombinant protein can be obtained by using X-ray crystallographic analysis in near future.

Secondary structure analysis of human OCT4 fusion proteins

The far UV CD spectra (plotted using BeStSel result) of recombinant HTN-OCT4 protein revealed secondary structure content containing mostly the random coils (~53%) (Figure 3.8 A (*right*)). However, OCT4-NTH confirmed the retention of secondary structure, predominantly comprising of α-helices (~ 41%) and random coils (~28%), and a significant contribution of β-sheets (~17%) and turns (~14%) (Figure 3.8 A (*left*)).

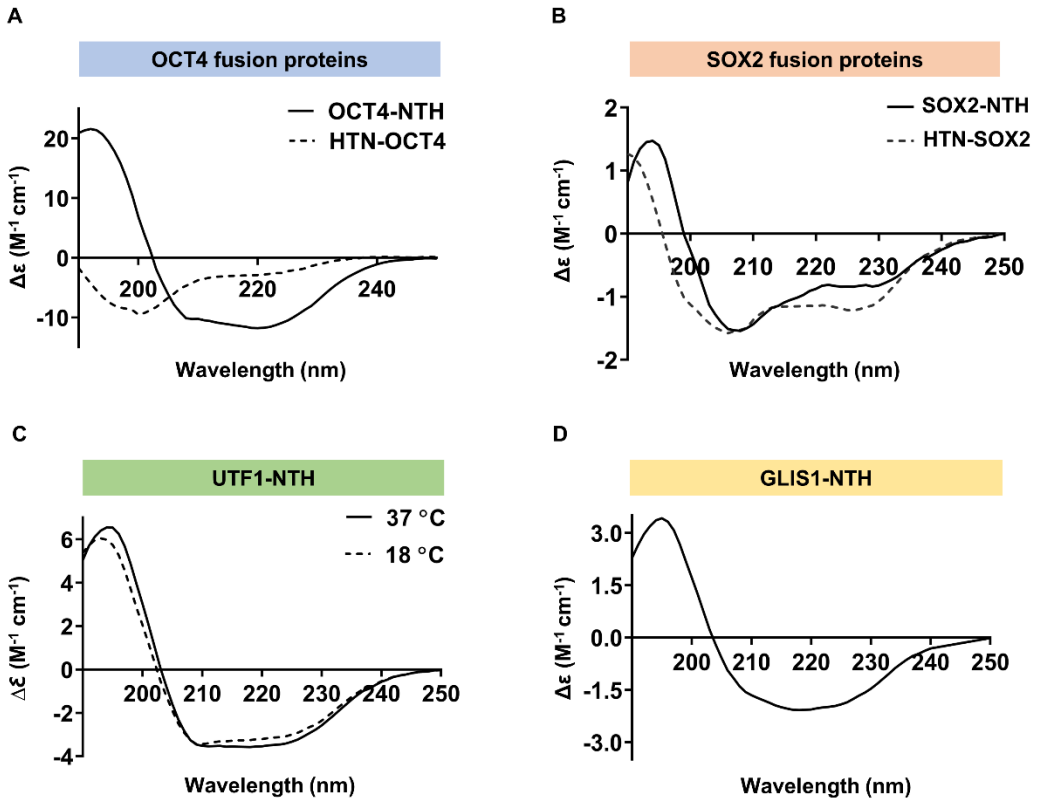


Figure 3.7 Determination of secondary structure of the recombinant fusion proteins using far UV CD spectroscopy. The purified human recombinant fusion proteins were desalted using size-exclusion chromatography for the analysis using CD spectroscopy. Spectral measurements were analyzed using BeStSel online server. The spectra were plotted wavelength (nm) on X-axis and delta epsilon ($M^{-1}cm^{-1}$) on Y-axis. (**A**) Far UV CD spectral analysis of HTN-OCT4 and OCT4-NTH fusion proteins purified under native conditions at 37 °C. (**B**) Far UV CD spectral analysis of HTN-SOX2 and SOX2-NTH fusion proteins purified under native conditions at 37 °C. (**C**) Far UV CD spectral analysis of UTF1-NTH fusion proteins purified under native conditions at 37 °C and 18 °C. (**D**) Far UV CD spectral analysis of GLIS1-NTH fusion proteins purified under native conditions at 37 °C.

Mouse OCT4 protein comprises a 75 amino acids POU-specific domain with four α-helices and a 60 amino acids POU-homeodomain with three α-helices (Esch et al., 2013). These domains are linked by linker region, also having α-helical conformation (Esch et al., 2013). These domains are highly conserved in mouse and human OCT4 proteins and constitute half of the total length of the protein (Radzisheuskaya and Silva, 2014). Our far UV CD data of OCT4-NTH are in agreement with the mouse OCT4 crystallographic structure, indicating that this protein primarily has α-helical conformation. Thus, purified OCT4-NTH maintained its structure and possess better chance of bioactivity.

Secondary structure analysis of human SOX2 fusion proteins

The crystal structure of full-length human SOX2 protein or its secondary structure content has not been reported to date. The CD spectral analysis of SOX2-NTH showed a positive peak at ~194 nm and two negative peaks at ~208 and ~226 nm, similarly HTN-SOX2 showed a positive peak at ~190 nm and two negative peaks at ~206 and ~226 nm (Figure 3.7 B). These spectral patterns are suggestive of a mixed secondary structure conformation. Upon confirmation using BeStSel online tool, the CD spectra and secondary structure content for SOX2-NTH were α-helices (~11%), β-sheets (~28%), turns (~15%), and random coils (~46%) (Figure 3.8 B (*right*)). Similar to SOX2-NTH, HTN-SOX2 fusion proteins showed secondary structure with α-helices (~10%), β-sheets (~28%), turns (~16%), and random coils (~46%) (Figure 3.8 B (*left*)). Notably, these results established that the purified fusion proteins majorly comprised of random coils and β-sheets and a good proportion of turns and α-helices. This data established that the recombinant SOX2 fusion proteins had maintained their secondary structure.

Secondary structure analysis of human UTF1 fusion proteins

To date, no reports are there on the crystal structure of full-length human UTF1 protein or its secondary structure. The CD spectral analysis of UTF1-NTH purified from the soluble fraction induced at 37 °C showed a positive peak at ~194 nm and two negative peaks at ~211 and ~218 nm, similarly UTF1-NTH purified from the soluble fraction induced at 18 °C showed positive peak at ~193 nm and two negative peaks at ~208 and ~214 nm (Figure 3.7 C). These spectral patterns are suggestive of a mixed secondary structure conformation. Upon confirmation using BeStSel online tool, the CD spectrum and secondary structure content for 37 °C UTF1-NTH were α-helices (~29%), β-sheets (~12%), turns (~13%), and random coils (~46%) (Figure 3.8 C). Notably, these results established that the purified fusion proteins majorly comprised of random coils and mixture of α-helices, β-sheets and turns. This data established that the recombinant UTF1 fusion proteins had maintained the secondary structure. However, as the yield of UTF1-NTH expressed at 37 °C is higher compared to UTF1-NTH expressed at 18 °C, thus we proceeded with UTF1-NTH expressed at 37 °C for further functional assays.

Secondary structure analysis of human GLIS1 fusion protein

To determine the secondary structure of purified recombinant GLIS1-NTH fusion protein, far UV CD spectroscopy was performed similar to other purified proteins OCT4, SOX2 and UTF1. The spectral results showed a positive peak at ~195 nm and a negative peak at ~218 nm (Figure 3.7 D), which are corresponding to the peaks for β-sheets (Greenfield, 2006; Kelly et al., 2005). The GLIS1-NTH structure was observed to be majorly composed of random coils (~ 46%) and β-sheets (~27%) and a substantial contribution of α-helix (~14%)

and turns (~13%) determined using the BestSel method (Figure 3.8 D). These results confirm that recombinant GLIS1-NTH fusion protein has maintained its secondary structure.

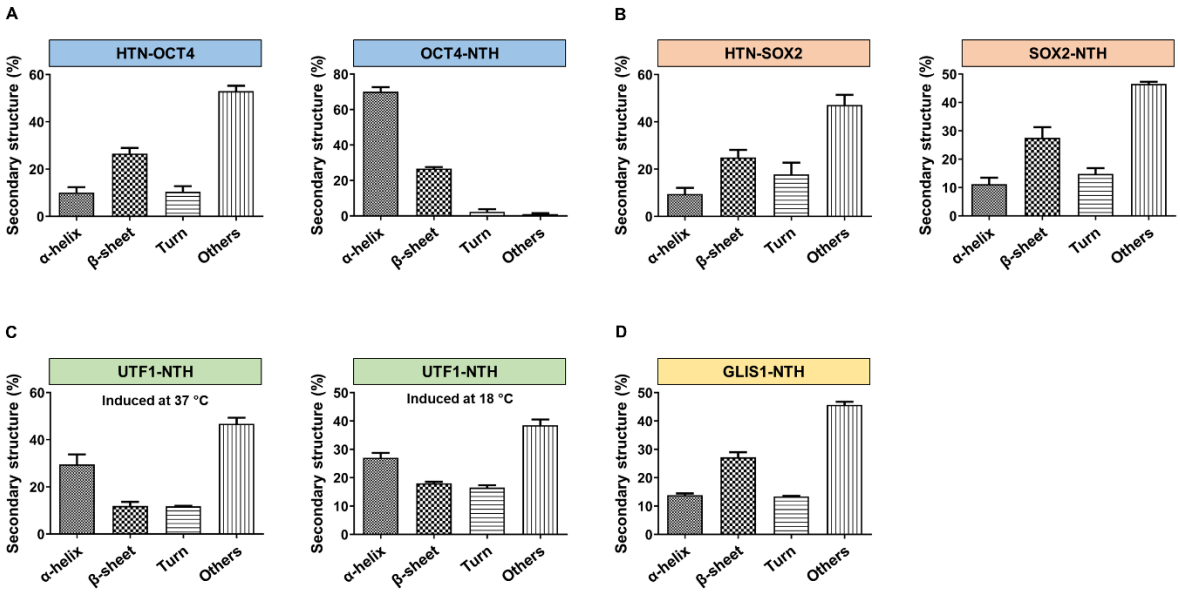


Figure 3.8 Quantification of the far UV CD spectra of purified human recombinant fusion proteins. The secondary structure quantification of the purified proteins was analyzed using BeStSel online server. The secondary structures (α-helices, β-sheets, turns and others) were represented using the bar graphs. The quantitative data shown are mean ± SEM (n≥3).

3.2.3 Mass spectrometric analysis of recombinant human OCT4-NTH fusion protein

The recombinant human OCT4-NTH fusion protein was purified and characterized using liquid chromatography-tandem mass spectrometry (LC-MS/MS). LC-MS/MS is a technique used for the identification and quantification of the peptides in a given sample or the complete proteome of organism (Karpievitch et al., 2010).

In our study, the Orbitrap LC-MS/MS technique was used to confirm the identity of our recombinant OCT4-NTH fusion protein. The purified protein was digested into small

manageable peptide fragments. These fragments were then separated, ionized and fed into mass spectrometer (Karpievitch et al., 2010). The LC–MS/MS study of digested peptides generated a match with human OCT4 protein, and from which, five unique peptide sequences (including two pairs of overlapping sequences) were identified (Figure 3.9 A) to be a part of human OCT4 (Accession Number: NP_002692.2) using SequestHT algorithm and Proteome Discoverer software.

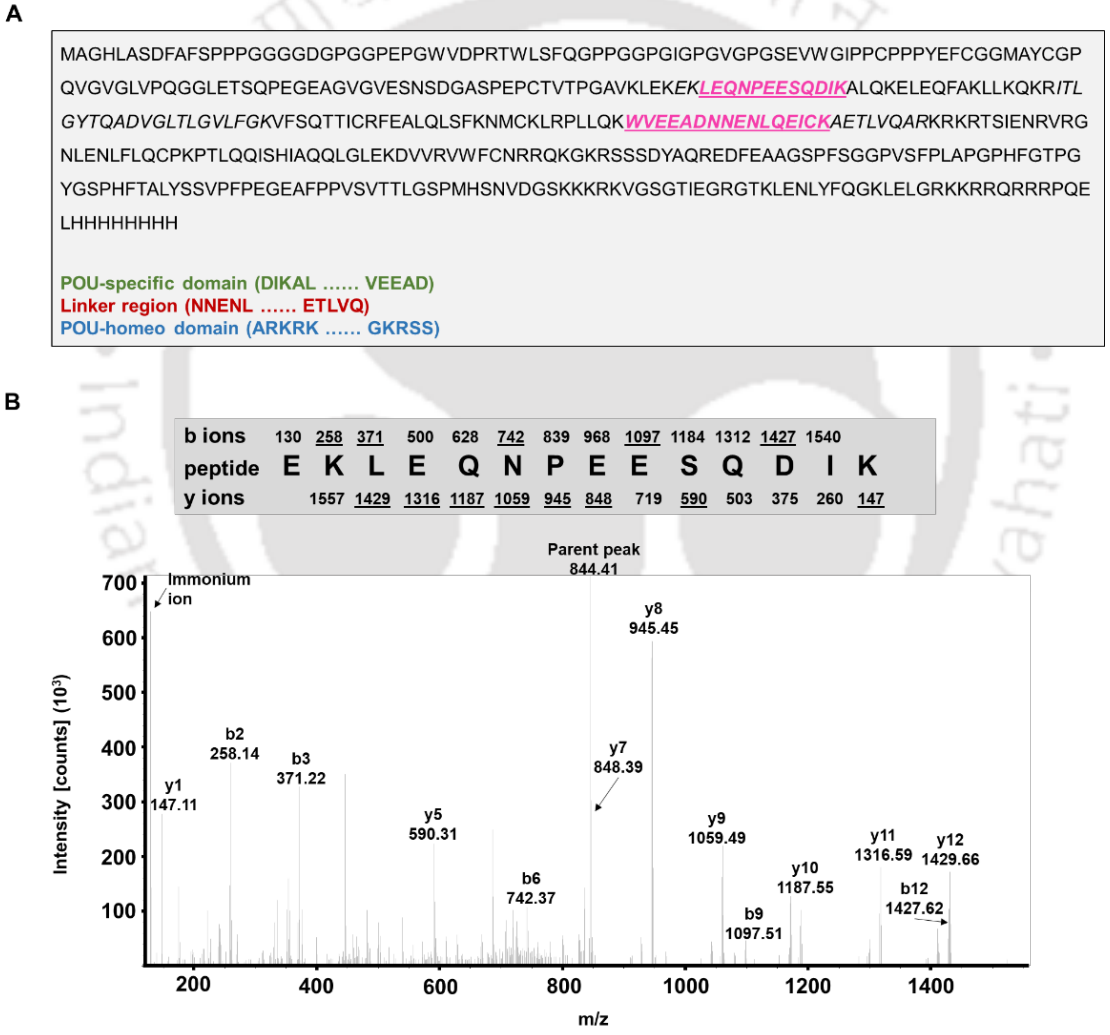


Figure 3.9 Confirmation of the identity of recombinant human OCT4-NTH protein using mass spectrometry. **(A)** The desired human OCT4 band was excised from the Coomassie-

stained polyacrylamide gel and then destained, processed, followed by trypsin digestion. The resulting peptide fragments were then analyzed using LC–MS/MS, and the identified five unique peptide sequences were highlighted in the full-length amino acid sequence of the OCT4-NTH protein. **(B)** From the identified peptide sequences, annotated spectra have been shown for one of these five unique peptides, EKLEQNPEESQDIK, belonging to the POU-specific domain of the human OCT4 protein. Italics case letters are peptide sequences detected by LC-MS/MS and the underlined letters are overlapping peptide sequences detected by LC-MS/MS.

The identified peptides either belong to the POU-specific domain of human OCT4 or adjacent to this critical domain or the linker region. The annotated spectrum of one of these five unique peptides (EKLEQNPEESQDIK) is shown in Figure 3.9 B. Thus, Western blotting and mass spectrometry confirmed the identity of recombinant OCT4-NTH protein.

3.3 Conclusion

Native purification of heterologously expressed recombinant proteins depends on essential prerequisites such as soluble expression of the protein, buffer conditions (salt and imidazole concentration, pH, percentage of glycerol) and so forth (Remans et al., 2022; Wingfield, 2015). These are the critical steps for determining the successful purification of the recombinant proteins.

In this chapter, we demonstrate the optimization of these key parameters encompassing the purification of these four recombinant human proteins OCT4, SOX2, UTF1 and GLIS1. We developed a cost-effective, simple, facile and reproducible methodology to successfully purify these proteins from the soluble fraction using Ni^{2+} affinity purification, hence,

circumventing the limitations imparted by purification from inclusion bodies. The major crux of using inclusion bodies is the requirement for refolding of the denatured protein, which is time-consuming, cumbersome and potentially compromises the biological activity. Furthermore, these optimized purification protocols have certain roadblocks concerning protein loss in flow-through and resin, which compromises the overall yield, hence, requires further improvement. In this chapter we have also shown the importance of fusion tags and their influence on purification, protein stability, yield and structural integrity. Upon purification of OCT4, it was observed that although OCT4-NTH and HTN-OCT4 both were purified successfully using affinity purification, HTN-OCT4 upon secondary structure determination using far UV CD spectroscopy showed a completely disordered structure. On the other hand, HTN-SOX2 showed significantly lesser yield compared to its counterpart SOX2-NTH fusion protein. This confirmed the influence of the position of fusion tags in purification, structure and function of the protein. This chapter also discusses the role of salt and glycerol and their impact on purification. We observed that the presence of salt in case of OCT4, SOX2 and GLIS1 did not hamper the purification, however, in case of UTF1, the protein elution was compromised. This result is suggestive of UTF1 being an ion-sensitive protein, thus the presence of salt can promote unfavorable interactions between salt and protein, resulting in salt-dependent aggregation during the purification procedure (Tsumoto et al., 2007), thereby affecting the protein yield. The secondary structure analysis of these proteins revealed that recombinant human OCT4 is a predominantly α-helical structure, whereas recombinant SOX2, UTF1 and GLIS1 are predominantly random coils and β-sheets. These purified proteins have great prospect in the field of iPSCs as these can be used for the generation of integration-free human iPSCs, overcoming the major limitation of random

genomic integration associated with integrative approaches. These proteins also potentiate new possibilities for exploring the individual roles of the proteins in cancer paradigm.

CHAPTER 4

Demonstration of cell penetration, nuclear translocation, and functional activity of purified recombinant proteins

Brief summary of the chapter

In the previous chapter, we have demonstrated the purification of the recombinant fusion proteins OCT4, SOX2, UTF1 and GLIS1 from soluble fraction using Ni-NTA affinity chromatography and have shown that the purified recombinant fusion proteins retained their secondary structure, however, further detailed examination is required to attest to their biological activity. In that aspect, protein stability is one of the first important criteria for validating their function at cell culture conditions. To further validate the ability of the protein of interest to transduce and translocate into human cells and nucleus, we performed immunocytochemistry. We showed that these purified recombinant fusion proteins successfully penetrated the cell membrane of both human normal (fibroblasts) as well as cancer (HeLa) cells using the immunocytochemistry technique. However, the nuclear translocation of these proteins varied depending on different cell lines. We further examined the role of these proteins; SOX2, UTF1 and GLIS1 in various cancers and also assessed the role of OCT4 overexpression in normal cells. To further confirm the transcriptional activity of this purified OCT4 fusion proteins, a reporter system (OCT4-GFP-2A-Puro) was electroporated into HeLa cells. Transduction of GLIS1-NTH fusion protein on breast cancer cells (MDA-MB-231) and SOX2-NTH and UTF1-NTH fusion proteins on cervical cancer cells (HeLa) showed their functional role in cell proliferation and cell migration. Post-treatment of HeLa cells with UTF1-NTH fusion protein showed changes in cell's clonogenic potential. Further, the bioactivity of the respective recombinant fusion proteins was confirmed through real-time-quantitative PCR for their respective target genes. In conclusion, this chapter demonstrated the

stability, transduction potential, and biological activity of these purified recombinant fusion proteins.

4.1 Materials and methods

4.1.1 Stability analysis of purified recombinant fusion proteins at cell culture conditions

Flash frozen purified recombinant fusion proteins were diluted in cell culture media consisting of high-glucose Dulbecco's modified eagle medium (DMEM; Invitrogen) along with 5% fetal bovine serum (FBS; Invitrogen) and 1% penicillin-streptomycin (P/S; Invitrogen) to a final concentration of 200 nM for OCT4-NTH and SOX2-NTH and 400 nM for UTF1-NTH and GLIS1-NTH. The protein containing transduction medium for the respective proteins were filtered using 0.22 µm (HiMedia) membrane filter and incubated at standard cell culture conditions (37 °C in a humidified chamber containing 5% CO_2) for 0, 24 and 48 hours. Samples were collected after every time point and stored at -20 °C until further analysis. The samples were centrifuged at 10,000-14,000 rpm at 4 °C for 10 minutes. The supernatant of the centrifuged samples was taken and analyzed by Western blotting analysis using protein-specific primary antibodies [anti-OCT4 (1:1000, sc-5279); anti-SOX2 (1:2500, sc-365823)] and anti-UTF1 antibody (1:500, R&D systems-AF3958) and secondary antibodies [anti-rabbit IgG antibody (1:5000, Invitrogen-31460), anti-mouse IgG-HRP Conjugated (1:5000, Invitrogen-31430), anti-Goat IgG-HRP (1:5000, Invitrogen)].

4.1.2 Mammalian cell culture

HFF (BJ) (ATCC® CRL2522™) were obtained from the American Type Culture Collection, USA and were cultured in DMEM supplemented with 10% FBS, 1% P/S, and 1X non-essential amino acids (NEAA) (Invitrogen). Human dermal fibroblasts (HDF) (HiMedia) and cancer cell lines: HeLa and MDA-MB-231 (National Centre for Cell Science, Pune, India) cells were

cultured in DMEM with 10% FBS, plus 1% P/S. Cells were cultured at standard cell culture conditions. (mentioned in previous section). Cells were passaged at a confluency of 70–80% with 0.25% trypsin–EDTA (Invitrogen) for further culture conditions.

4.1.3 Protein transduction, immunocytochemistry and microscopy

Protein Transduction: Cells were plated onto 6 well and 24 well plates at a cell density of 0.2×10^6 HFF/ 6 well, 0.07×10^6 HFF/ 24 well and 0.1×10^6 HeLa/ 12 well for protein transduction studies. The cells were cultured until 50–60% confluency at 37 °C with 5% CO_2 under humidified conditions. Upon reaching the required confluency, the medium was replaced with sterile protein transduction medium [DMEM, 5% FBS, 1% P/S, and 200 nM (OCT4-NTH), 200 nM (SOX2-NTH), 400 nM (UTF1-NTH and GLIS1-NTH) of purified recombinant protein or transduction media without the protein with equal volume of DMEM or glycerol buffer (vehicle control)]. The cells were incubated with the protein for 12 hours and later washed with Phosphate-buffered saline (PBS) for further analysis.

Immunocytochemistry and microscopy: Cells were first fixed with 2% paraformaldehyde (Merck Millipore) for 10 minutes and then permeabilized with 0.1% Triton™ X-100 (Sigma-Aldrich) in PBS for 10 minutes. Further, the cells were blocked with PBS+ (0.5% bovine serum albumin and 0.15% glycine in PBS) for 1 hour at room temperature. After blocking, the cells were washed twice with PBS and incubated with primary antibody [anti-OCT4 (1:100, sc-5279, Santa Cruz Biotechnology Inc., USA); anti-SOX2 (1:200, sc-365823); anti-UTF1 (1:200, AF-3958 (R&D systems)] for 2 hours. Next, the cells were washed thrice with PBS and incubated with anti-mouse IgG Alexa Fluor 488 secondary antibody (1: 2000, A11029, Invitrogen, USA) for 1 hour at room temperature in a moist chamber. Further, the cells were washed thrice with PBS, and the nucleus was counterstained with 4′,6-diamidino-2-

phenylindole (DAPI) dye at a dilution of (1: 15,000; Sigma-Aldrich) for 10 minutes or Hoechst 33342 (1: 10,000; Invitrogen) for 5 minutes at room temperature. Cells were washed thrice with PBS to remove excess DAPI staining. The nuclear stain DAPI was excited at 345 nm, and Alexa 488 was excited at 499 nm, and further visualized using an inverted fluorescent microscope (IX83 Olympus, Japan), equipped with a DP80 CCD camera. Samples were illuminated using a pE-300 white CoolLED light source. Image stacks were acquired using 20x/0.45NA objective at 1.57 µm intervals. Image analysis was performed by CellSens dimension (Olympus) and further processed using Image J software.

4.1.4 Cell proliferation and cell migration assays

Cell proliferation (MTT assay): HFF (BJ) cells were seeded at a density of 2000 cells/well and HeLa was seeded at a cell density of 4000 cells/well in a 96 well plate in triplicates every day, consecutively for 6 days (Figure 4.1). After overnight incubation, the medium was removed next day and changed to sterile protein medium. The cells were incubated in the respective medium for 2 hours (OCT4-NTH) and then changed to the fibroblast growth medium (10% FBS, 1% P/S, and 1% NEAA in DMEM). For UTF1-NTH, the cells were incubated with the protein transduction media and replenished every 24 hours. The treatment was carried for each well for 6 consecutive days. Cell proliferation was measured using MTT (3- [4, 5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide; thiazolyl blue, Sigma-Aldrich, USA) at a working concentration of 0.25 mg/ml. The cells were incubated for 2 hours in MTT solution, and later the formazan crystals formed were dissolved with 200 µl dimethyl sulfoxide (DMSO) (HiMedia), and absorbance was measured at 540/575 nm using Multiskan GO (Thermo Scientific) with correction at 695 nm. Tests were performed in quadruplicate. The absorbance obtained is directly proportional to the number of live cells in each well.

Day	-1	0	1	2	3	4	5	6
6	S	T	T	T	T	T	T	T
5	-	S	T	T	T	T	T	T
4	-	-	S	T	T	T	T	T
3	-	-	-	S	T	T	T	T
2	-	-	-	-	S	T	T	T
1	-	-	-	-	-	S	T	T
0	-	-	-	-	-	-	S	T

Figure 4.1 Schematic for cell seeding for cell proliferation assay using MTT

Cell proliferation (cell counting method): HeLa cells were seeded at a cell density of 0.2×10^6 cells/well of a 6 well plate to achieve a cell density of 30-40%. Cells were treated with SOX2-NTH and its respective vehicle control media every 24 hours. Upon attaining 80-90% confluency, wells were washed thrice with PBS and trypsinized using 0.25% Trypsin-EDTA solution (incubated at 37 °C for 3-4 minutes). The cells were diluted using trypan blue solution (1:1) and counted using Hemocytometer through visualization under inverted brightfield microscope (ZOE Fluorescent Cell Imager, Bio-Rad, California, USA) at 20X magnification. The cells were re-seeded again at the same cell-density in a 6 well plate and the treatment cycle was repeated for 10 days. The cumulative cell count was then analyzed using Graphpad Prism software.

Cell migration (in vitro scratch assay): HFFs (BJ) were adjusted to 0.1×10^6 cells/well of 12-well and 0.05×10^6 cells/well of 24 well plate, HeLa and HDF cells were plated at a cell density of 0.08×10^6 cells of 24 well plate and MDA-MB-231, seeded at a density of 0.085×10^6 cells/well of a 24 well plate. The cells were incubated at standard cell culture conditions. Upon achieving 80-90% confluency, clean scratches were drawn using 10 µL sterile pipette tip. The

media was aspirated out, followed by washes with PBS and replaced with 200 nM (OCT4-NTH, SOX2-NTH and GLIS1-NTH) and 400 nM (UTF1-NTH) fusion proteins along with their vehicle control media. Protein media was replaced and replenished every 24 hours and for BJ cells, the protein media was changed to fibroblast growth media (DMEM supplemented with 10% FBS and 1% P/S) after 4 hours. Imaging was performed at 0, 24 and 48 h time intervals using an inverted brightfield microscope (ZOE Fluorescent Cell Imager, Bio-Rad, California, USA) at 20X magnification. The migration rate at 48 h was analyzed and calculated using ImageJ software. The rate of migration is defined as rate at which the cells migrates to fill the scratched area in a stipulated period of time. The area method of assessing the migration percentage was calculated using the formula.

$$Migration\ (\%) = \frac{(initial\ area - final\ area)}{initial\ area} \times 100$$

4.1.5 Colony assay formation

HeLa cells were counted and diluted in complete growth media, 400 cells/well of a 6 well plate was plated and incubated at standard cell culture conditions. After 24 hours, the cell culture media was replaced with 400 nM UTF1-NTH protein transduction media and its respective vehicle control. The spent protein media was replaced with fresh media after every 48 hours. The treatment cycle was followed for 10-12 days until distinct colonies were visible. The obtained colonies were visualized and counted using ImageJ software upon fixing with methanol/acetic acid solution (3:1) for 20 minutes and stained with 5% crystal-violet solution.

4.1.6 Relative gene expression using quantitative RT-PCR (RT-qPCR)

HeLa cells: 0.4×10^6 cells/well were plated for the treatment of 24 hours and 0.1×10^6 cells/well for 72 hours treatment of a 6-well plate. MDA-MB-231 cells were plated at a cell density of 0.2×10^6 cells/well for 72 hours treatment of a 6 well plate. After 24 hours the culture media was removed and substituted with protein transduction media and their respective vehicle controls (media without the protein). Protein transduction media was changed every 24 hours until the stipulated treatment time duration.

RNA isolation: The cells were washed thrice with ice cold PBS. RNA isolation was carried out from the vehicle control and the protein treated wells using TRIzol reagent (Invitrogen) as per the manufacturer's protocol.

cDNA synthesis: Approximately 1 µg of total RNA obtained was reverse transcribed using iScript cDNA synthesis kit (Bio-Rad) as per the manufacturer's instructions.

RT-qPCR reaction: The RT-qPCR reaction mixture containing iTaq Universal SYBR Green Supermix (2X concentration; Thermo Scientific) was prepared as per the lab optimized protocol. Template cDNA was amplified with gene-specific primers (Table 4.1) with an AriaMx Real-Time PCR System (Agilent). Each sample was assessed in triplicates with GAPDH taken as an internal control. The relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method, with normalization to GAPDH as a reference gene.

Table 4.1 Primers used for RT-qPCR

Genes	Primer Sequences (5' – 3')	AT (°C)	PS (bp)
WNT5A	F – AGACGTTTCGGCTACAGACC R - CCCAGTTCATTCACACCACA	53.9	96
$p27^{Kip1}$	F - CTGCCCTCCCCAGTCTCTCT R - CAAGCACCTCGGATTTT	52.6	101
GAPDH	F - GTCTCCTCTGACTTCAACAGCG R - ACCACCCTGTTGCTGTAGCCAA	57	131

AT, annealing temperature; PS, product size; F, forward; R, reverse

4.1.7 Reporter assay

HeLa cells at a cell density of 0.5×10^6 cells were loaded onto a 100 µL Neon tip and were electroporated with 4 µg of OCT4-GFP-2A-PURO (Addgene ID: 52379) (Zimmerlin et al., 2016) and 1 µg of pCXLE-eGFP (Addgene ID: 27082) plasmid using Neon Transfection system (Thermo Scientific) using listed parameters in Table 4.2. Post electroporation (Figure 4.2), the cells were seeded on a 24 well plate with culture media containing (DMEM supplemented with 20% FBS and 1% P/S). Fresh media was supplemented every 24 hours. Upon reaching a confluency of 40-50%, the selection media was changed with 400 nM OCT4-NTH protein transduction media and vehicle control media, respectively. Protein media was replenished after every 24 hours and imaging was performed once GFP^+ cells were observed using an inverted brightfield microscope (ZOE Fluorescent Cell Imager, Bio-Rad, California, USA) at 20X magnification.

Table 4.2 Parameters for Neon transfection in HeLa cells

Cell line	Pulse (Volts)	Pulse (ms)	Pulse No.	Cell count	Tip
HeLa	1005	35	2	0.5×10^6	100µL

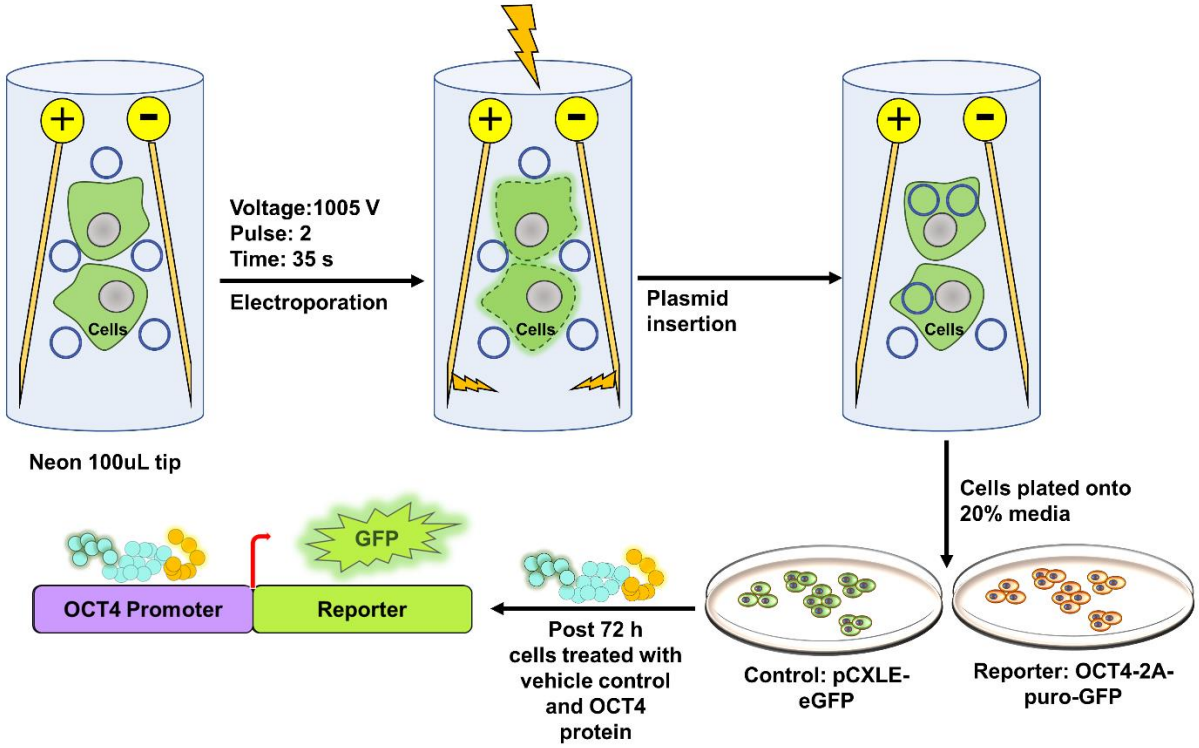


Figure 4.2 Workflow of reporter plasmid transfection in HeLa cells

4.1.8 Western blotting

HeLa cells were plated at a cell density of 0.1×10^6 cells/well of a 6-well plate. The seeded cells were treated with 200 nM of SOX2-NTH protein transduction media and its respective vehicle control for 72 hours. Post treatment, the cells were washed thrice with ice cold PBS and subjected to 500 µL of RIPA lysis buffer in ice. The scraped cells were collected and stored at -80 °C until further analysis.

For sample preparation, the samples were thawed on ice and homogenized (Vortex: 10 secs). The samples were then centrifuged and the supernatant was loaded on SDS-PAGE gel. The samples were assessed for the expression of phospho-GSK-3β and total GSK-3β expression along with β-actin. β-actin levels were used as a loading control. The protein-specific antibodies [phospho-GSK-3β (1:5000; Cell Signaling, D85E12), Total GSK-3β (1:5000; Cell Signaling, D5C5Z), anti-β-actin (1:5000; BioBharati, BB-AB0024)]. Western blotting was performed according to the detailed protocol already discussed in Chapter 2.

4.1.9 Statistical analysis

The experimental data obtained in this section were analyzed using student unpaired *t*-test or multiple *t*-test using GraphPad Prism 8 software. Values are expressed as mean ± SE. $P < 0.05$ was considered significant.

4.2 Results and discussion

4.2.1 Stability and translocation ability of purified recombinant fusion proteins

The solubility and stability of the protein in transduction media under cell culture conditions play a pivotal role in successful protein transduction (Bosnali and Edenhofer, 2008; Thier et al., 2010). Hence, we have analyzed the *in vitro* solubility and stability of the purified recombinant fusion proteins in a transduction medium under standard cell culture conditions. Western blotting with protein-specific antibodies showed clearly that the OCT4-NTH and SOX2-NTH proteins were stable for up to 24 hours (Figure 4.3). OCT4 protein in transduction media showed a 50% reduction in stability at 48 hours, whereas SOX2 did not show reduction until 48 hours. Similarly, Western blotting with anti-UTF1 antibody showed UTF1-NTH was

stable till 48 hours in cell culture media (Figure 4.3). In the case of GLIS1-NTH, we were not able to analyze the stability due to the unavailability of commercial GLIS1 antibody.

The recombinant protein-based reprogramming approach requires multiple rounds of transductions, with one round for 4–24 hours per day (Bosnali and Edenhofer, 2008). Thus, our purified OCT4, SOX2 and UTF1 fusion proteins fulfill the first and foremost criterion of protein stability in transduction media for its further use for reprogramming applications in the near future.

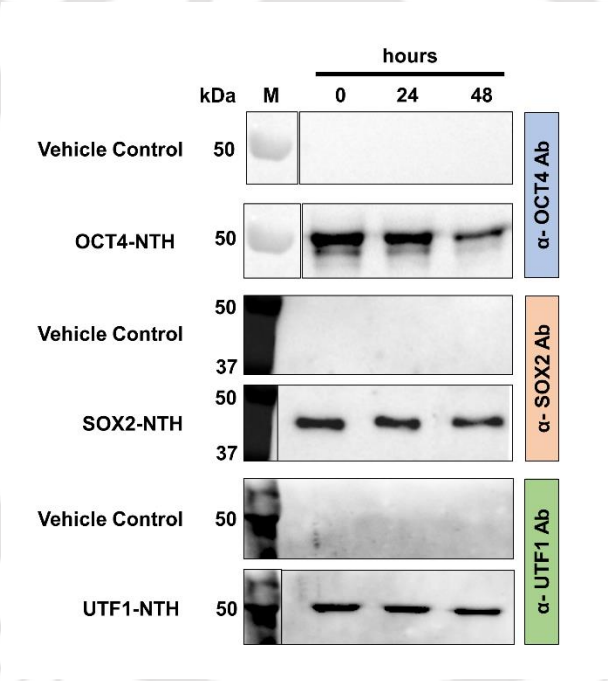


Figure 4.3 Assessment of protein stability under cell culture conditions. Purified recombinant fusion proteins were diluted in low serum containing transduction media with final concentration of 200-400 nM for 48 hours. The Western blotting analysis showed OCT4-NTH to be stable for 24 hours and SOX2-NTH and UTF1-NTH to be at least stable for 48 hours. Experiments were performed n=3.

We next sought to validate our fusion strategy for the cell penetration and nuclear translocation ability of these purified fusion proteins in the target mammalian cells. HFFs and HeLa cell lines were used for this study because of their differing protein transduction efficiency (Weill et al., 2008) and also due to absence of endogenous expression of these proteins of interest (Bosnali and Edenhofer, 2008; Cantz et al., 2008). Moreover, HFF (BJ) cells are one of the most widely used cell lines for cellular reprogramming. Hence, for this study, we have treated HFFs and HeLa cells with OCT4, SOX2 fusion proteins and HeLa cells with UTF1 fusion protein. Protein transduction media without the protein (Vehicle control) was taken as a negative control for the experiment. The fluorescence microscopy results showed no signal in the vehicle control, thus confirming the absence of endogenous OCT4, SOX2 and UTF1 expression in both the cell lines as well as negating any false-positive signals during imaging. An efficient cytoplasmic translocation was observed for OCT4-NTH and SOX2-NTH fusion protein in both HFFs and HeLa cells (Figure 4.4 and Figure 4.5) and UTF1-NTH in HeLa cells (Figure 4.6). The transduced cells showed both intense and faint localization into the nucleus (Figure 4.4, 4.5 and 4.6: inset image). Immunocytochemistry analysis for the translocation of GLIS1-NTH protein was not carried out due to non-availability of protein-specific antibody.

It is a fact that nuclear translocation of transcription factors holds a key role in their functionality for various molecular processes. Thus, these purified fusion proteins being transcription factors (Mehravar et al., 2021), their nuclear entry is one of the most critical aspects for their functionality. Our strategy of tagging these fusion proteins with TAT and NLS facilitated the delivery of the protein in fibroblasts and HeLa cells, both in the cytosol and nucleus. Tagging transcription factors with TAT and NLS have previously been reported to

have facilitated cell penetration and nuclear entry for other stem cell as well as β-cell and cardiac-specific factors (Bosnali and Edenhofer, 2008; Haridhasapavalan et al., 2022a, 2022b, 2021a, 2021b, 2020; Kumar Haridhasapavalan et al., 2021; Münst et al., 2016; Narayan et al., 2021b, 2021a; Peitz et al., 2014; Thier et al., 2010).

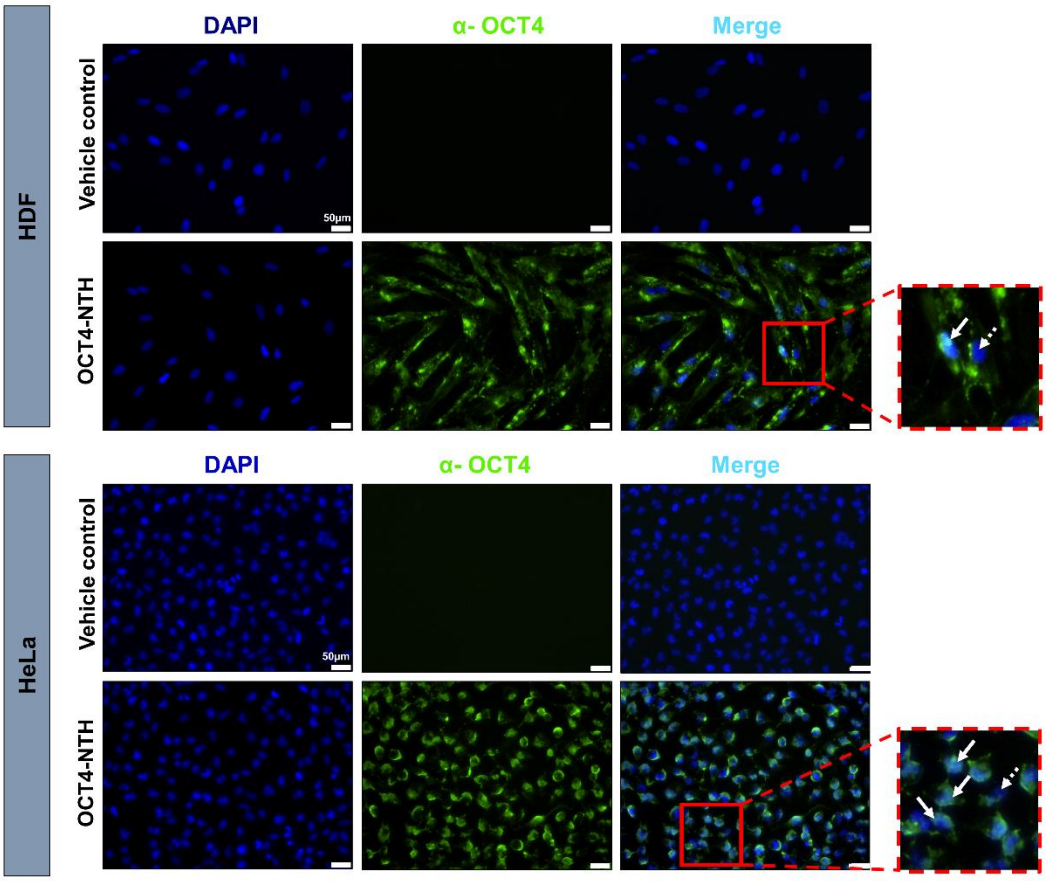


Figure 4.4 Cell transduction ability of purified OCT4-NTH fusion protein in HFFs and HeLa cells. Both the cell lines were treated with OCT4-NTH protein or vehicle control for 12 hours and incubated at 37 °C in 5% CO_2 under humidified conditions. Cellular and nuclear translocation was analyzed using an inverted fluorescent microscope. Z-sections were analyzed for colocalization of OCT4 staining and DAPI signal. The inset image shows the nuclear

uptake, depicted with solid white arrows and non-transduced cells depicted with white dotted arrows.

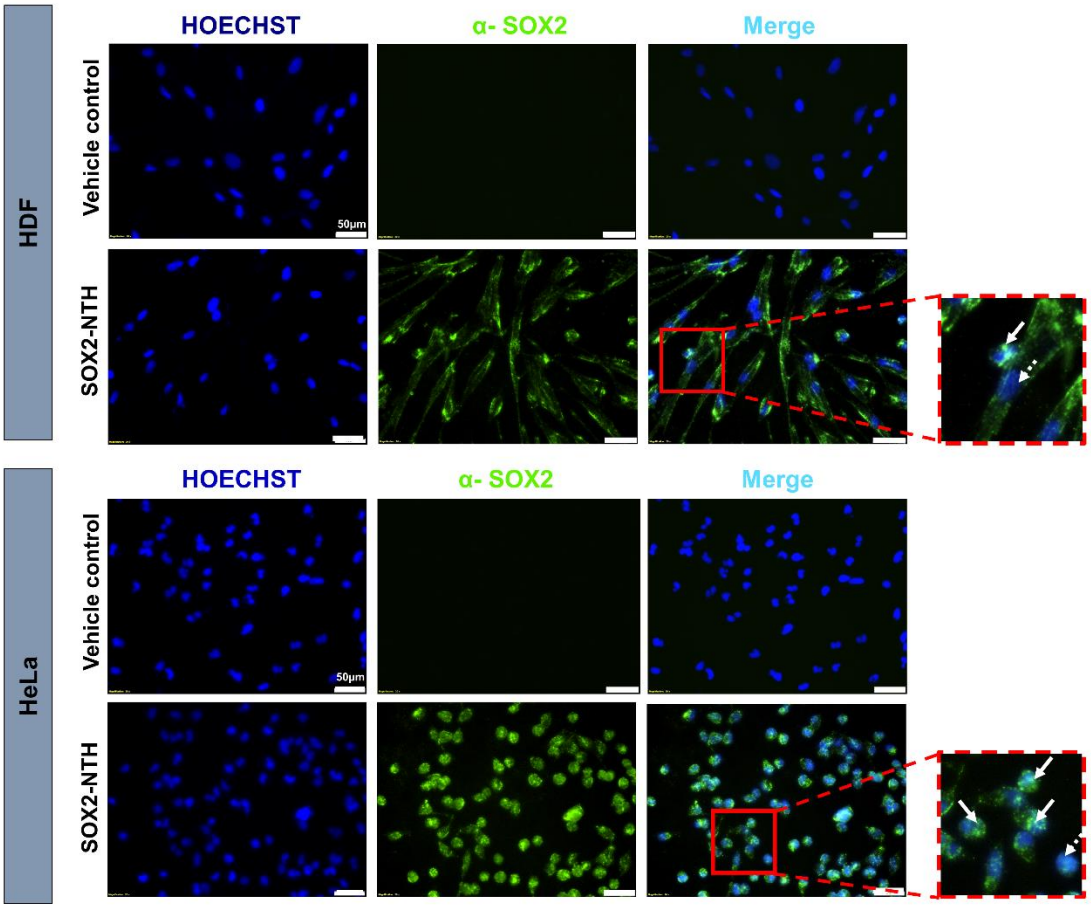


Figure 4.5 Cell transduction ability of purified SOX2-NTH fusion protein in HFFs and HeLa cells. Both the cell lines were treated with SOX2-NTH protein or vehicle control for 12 hours and incubated at 37 °C in 5% CO_2 under humidified conditions. Cellular and nuclear translocation was analyzed using an inverted fluorescent microscope. Z-sections were analyzed for colocalization of SOX2 staining and Hoechst signal. The inset image shows the nuclear uptake, depicted with solid white arrows and non-transduced cells depicted with white dotted arrows.

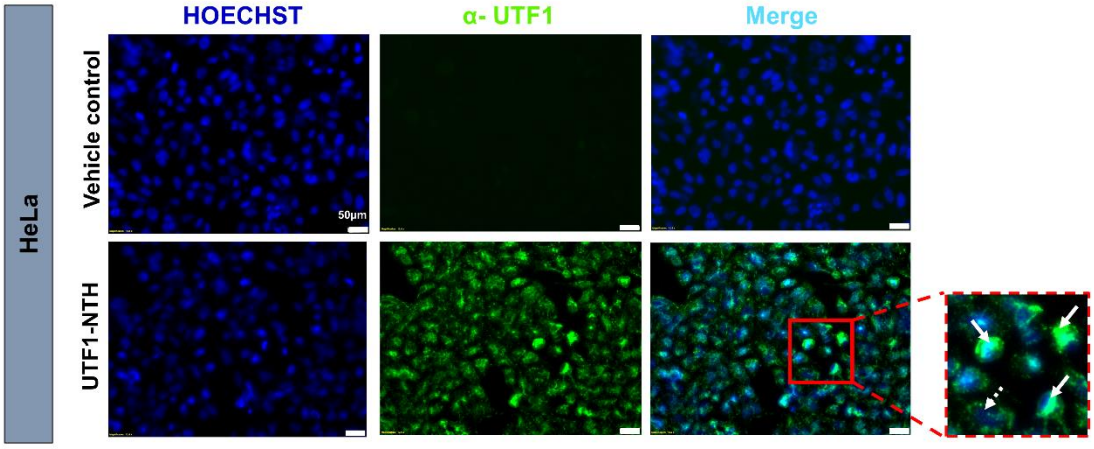


Figure 4.6 Cell transduction ability of purified UTF1-NTH fusion protein in HeLa cells. HeLa cells were treated with UTF1-NTH protein or vehicle control for 12 hours and incubated at 37 °C in 5% CO_2 under humidified conditions. Cellular and nuclear translocation was analyzed using an inverted fluorescent microscope. Z-sections were analyzed for colocalization of UTF1 staining and Hoechst signal. The inset image shows the nuclear uptake, depicted with solid white arrows and non-transduced cells depicted with white dotted arrows.

These studies also reported that the presence of these tags does not compromise the biological activity of the protein of interest and is a promising alternative to various integrating as well as non-integrating approaches. Notably, in this study, no protein transduction reagent was used for additionally facilitating the translocation of OCT4, SOX2 and UTF1 fusion proteins into the cytosol and nucleus of both the cell lines. Thus, our strategy to tag all these proteins with TAT and NLS has been successful in delivering the protein into the cells and is most likely biologically active.

4.2.2 Assessment of functional potential of the purified recombinant fusion proteins

The purified recombinant fusion proteins retained their secondary structure and immunocytochemistry analysis demonstrated the ability of the purified fusion proteins to successfully penetrate cell membrane and translocate into the nucleus. This potentiates the proteins to be functionally active. Thus, we sought to assess their bioactivity potential in various cellular functions of mammalian cells.

4.2.2.1 Biological activity of purified recombinant OCT4-NTH protein

Effect of the purified recombinant OCT4-NTH fusion protein on cell proliferation and migration rate of HFFs

To determine the biological activity and the role of OCT4 expression in fibroblast cells, 2000 cells/well were seeded in a 96 well plate every day, as shown in the schematic diagram (Figure. 4.1). The cells were seeded in triplicates for both the vehicle control and the protein for six days, and the analysis was done on the last day using MTT assay (Santos et al., 2011). The absorbance of the formazan product was measured at 540 nm, and the results observed clearly showed a gradual increase in the vehicle control-treated cells as compared to the cells treated with OCT4-NTH (Figure. 4.7). Pertaining to the fact that the absorbance is directly proportional to the number of live cells (Janani et al., 2018), it is clear that cell proliferation was slowed down in fibroblasts upon protein treatment for 6 days, whereas vehicle control showed no reduction in cell proliferation.

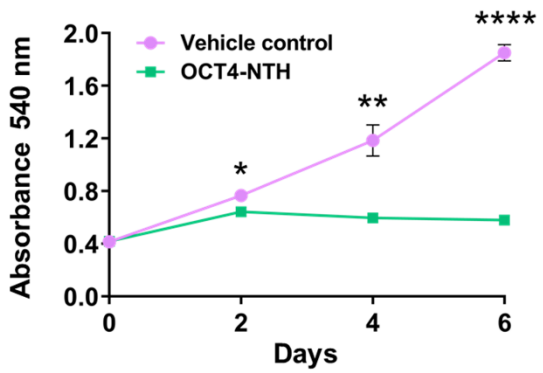


Figure 4.7 Effect of purified OCT4-NTH protein on cell proliferation of HFFs. The absorbance was measured at 540 nm. Y-axis signifies absorbance at 540 nm, and X-axis signifies days of treatment. The data shows significant reduction in cell proliferation with OCT4-NTH protein treated wells compared to vehicle control from Day 2 (**P*=0.0163), Day 4 (***P*=0.0081), Day 6 (*****P* < 0.0001). The quantitative data shown are mean ± SEM (n = 3). **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

Similar results were observed when mouse embryonic fibroblast cells were overexpressed with OCT4, leading to up two-fold reduction in cell proliferation (Hochedlinger et al., 2005). In congruence with the previous report, we have also observed the reduction in cell proliferation in fibroblasts upon treatment with purified recombinant OCT4-NTH fusion protein. In addition, an earlier study has also reported the over-expression of human OCT4 and its effect on cell migration, and has consequently shown that OCT4 overexpressing human dermal fibroblasts showed a slower migration rate as compared to GFP overexpressing human dermal fibroblasts (Peskova et al., 2019). Based on this study, we have further analyzed the effect of migration rate in fibroblasts upon transduction of purified OCT4-NTH fusion protein using scratch (migration) assay. The results showed that the migration rate of HFFs was significantly lower in the presence of OCT4-NTH compared to the vehicle control (*P*<0.05)

(Figure. 4.8 A and 4.8 B). The analysis and calculation for the migration rate were done following the same formula as stated in an earlier study (Grada et al., 2017). This result correlated with the fact that cell proliferation is directly related to the migration rate of cells; thus, slower cell proliferation is directly proportional to the slower migration rate of OCT4-NTH protein-treated HFFs.

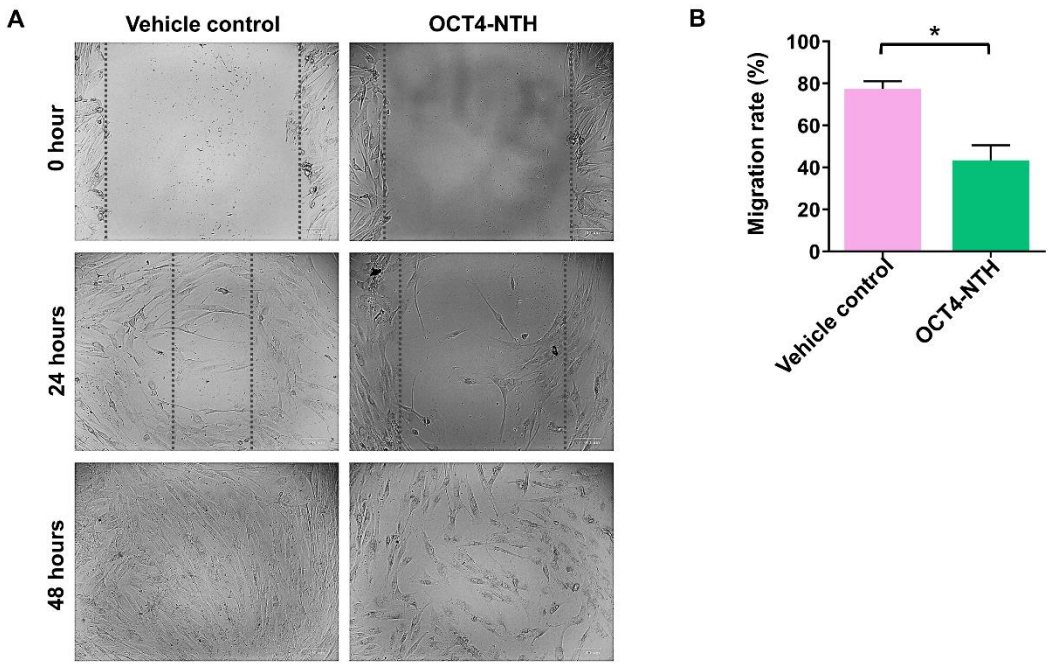


Figure 4.8 Effect of purified OCT4-NTH protein on migration rate of HFFs. (A) Scratch assay performed on HFF (BJ) cells for 48 hours. **(B)** Graphical representation of the changes in the migration rate upon treatment with OCT4-NTH and vehicle control. The data shows significant reduction in migration rate with OCT4-NTH protein treated wells compared to vehicle control (**P*=0.0136). The quantitative data shown are mean ± SEM (n = 3). **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

Activation of OCT4-GFP-2A-PURO reporter system upon recombinant OCT4-NTH fusion protein treatment on HeLa cells

Further investigation was carried to substantiate our findings that the purified recombinant OCT4-NTH fusion proteins is biologically active. It is reported that OCT4 binds to its own promoter and activates itself and its target genes. Exploiting this ability of OCT4, we confirmed its transcriptional activity by stably transfecting OCT4-GFP-2A-PURO plasmid in HeLa cells using Neon transfection system.

Previously we have assessed the translocation efficiency on both HeLa and HFF (BJ) cells and the results showed successful nuclear translocation of the protein. However, for further confirming the transcriptional activity of OCT4, HeLa cells were chosen as they are reported to be robust, highly dividing and has highest transfection efficiency (Hornstein et al., 2016). This study also highlights the correlation between plasmid vector size and transfection efficiency. The maximal plasmid transfection efficiency was observed with plasmid of 727 bp and up to ~3913 bp. Additionally, nuclear entry of vector is highest in dividing cells compared to non-dividing cells (Hornstein et al., 2016). HFF (BJ) cells however has low transfection efficiency compared to HeLa cells and OCT4-GFP-2A-PURO plasmid is ~21 kbp size hence we selected HeLa cells for nucleofection study.

Previously, ESCs were transfected with OCT4-GFP-2A-PURO plasmid vector to identify naïve and primed hiPSCs expressing endogenous OCT4 (Gafni et al., 2013). Subsequently, another group stably nucleofected naïve LIF-3i/MEF C2 fibro-hiPSCs using OCT4-GFP-2A-PURO for identifying endogenous OCT4 expression (Zimmerlin et al., 2016). Based on these literatures, HeLa cells were stably transfected with the full length OCT4-GFP-2A-PURO plasmid construct. Upon electroporation, the cells were treated with purified 400

nM OCT4-NTH fusion protein and vehicle control, respectively. After 48 hours, protein treated wells expressed GFP^+ cells, whereas no GFP^+ cells were observed in vehicle control treated cells (Figure 4.9). These results showed that purified OCT4-NTH fusion protein has successfully bound to the OCT4 promoter sequence, thus expressing the downstream GFP, giving rise to GFP^+ cells. This confirms that we have successfully purified full-length OCT4 fusion protein in native form and the protein is biologically active.

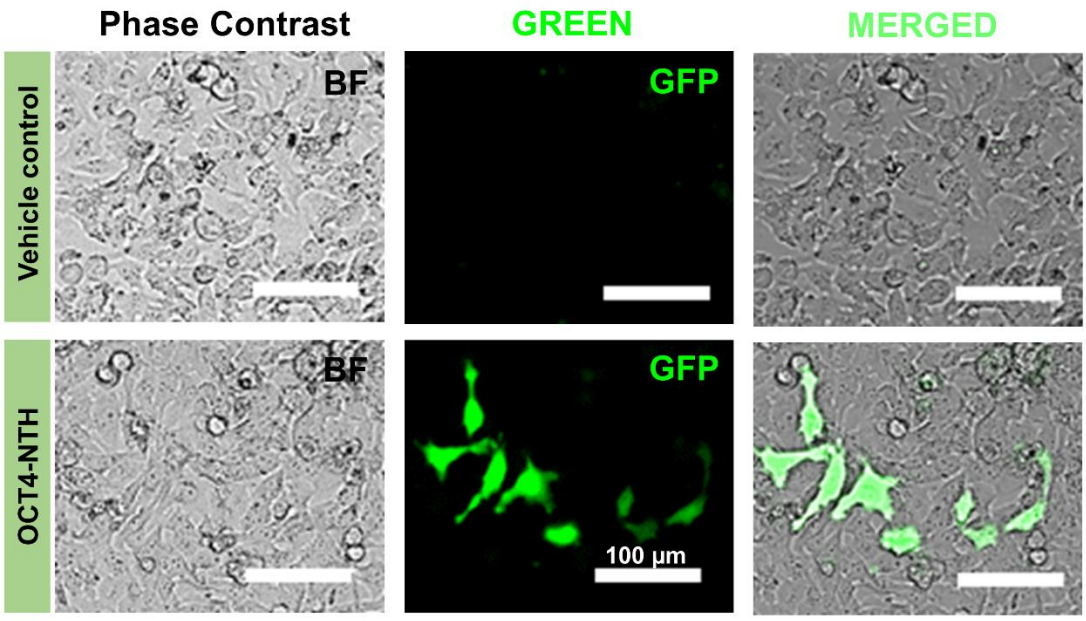


Figure 4.9 OCT4-NTH induced activation of OCT4-GFP-24-PURO reporter in HeLa cells. HeLa cells were transduced with OCT4-GFP-24-PURO plasmid vector. Subsequently, they were treated with purified OCT4-NTH protein (induced either at 37 °C) or vehicle control. After 72 hours, images were captured and analyzed. All images were taken with identical camera settings. Scale bar: 100 µm.

4.2.2.2 Biological activity of purified recombinant SOX2-NTH protein

Effect of the purified recombinant SOX2-NTH fusion protein on cell proliferation and migration rate of HeLa cells

Previous reports have shown that overexpression of SOX2 by transfecting pIRSE2-SOX2 vector in the cervical cancer cells (SiHa and HeLa) increased cell proliferation, clonogenicity and tumorigenicity compared to control pIRSE2-eGFP vector (Ji and Zheng, 2010). Thus, we sought to understand the impact of transducing purified SOX2-NTH fusion protein in HeLa cells and its effect on the cell proliferation. SOX2-NTH proteins were transduced onto HeLa cells and the results observed showed a significant increase in the cell proliferation from day 6 of protein-treated cells as compared to the cells treated with vehicle control ($P<0.05$) (Figure 4.10 A). Similarly, overexpression of SOX2 on breast cancer cell MDA-MB-231 was reported to significantly increase migration rate compared to the control (Li et al., 2013). Based on this study we have further analyzed the effect of ectopically expressing SOX2-NTH on HeLa cells to assess its effect on its migration rate. Upon treatment it was observed that the rate of migration significantly increased for SOX2-NTH treated wells compared to the vehicle control (Figure 4.10 B and 4.10 C). As shown in the figure the scratch was almost filled within 24 hours for SOX2 treated cells, however, the migration was significantly retarded in control cells, thus indicating that ectopic expression of SOX2-NTH protein promoted migration potential of HeLa cells, similar to MDA-MB-231 cells. These results correlated with the fact that cell proliferation is directly related to the migration rate of a cell and indicated that SOX2-NTH protein also promoted tumorigenic potential of cervical cancer cells and attested that the protein is biologically active.

Effect of the purified recombinant SOX2-NTH fusion protein on the regulation of phospho-GSK-3β expression in HeLa cells

SOX2 has been reported to have diverse role in cancer paradigm, especially in maintenance of cancer stem cells (CSCs), thus contributing to cancer recurrence, cell survival and tumorigenicity (Zhang et al., 2020). Essentially, overexpression of SOX2 is associated with the CSC characteristics of colorectal cancer (Lundberg et al., 2016) and ectopic expression of the same transcription factor was responsible for the increase in CSCs in lungs and ovarian cancer (Bareiss et al., 2013; Nakatsugawa et al., 2011). In context of cell proliferation, SOX2 was observed to enhance the proliferation rate in esophageal squamous cell carcinoma (ESCC) in a cell cycle independent manner by activating AKT/mTORC1 pathway (Gen et al., 2013). Its expression was also deemed important in maintaining and enhancing the proliferation rate in plethora of different cancers, namely laryngeal and ESCCs (Gen et al., 2010), osteosarcoma (Maurizi et al., 2018), breast cancer (Liu et al., 2017) and ovarian cancer (Wang et al., 2017). Recently, it has been reported that overexpression of SOX2 in ESCCs under serum starvation increases the expression level of phospho-AKT and its downstream phospho-GSK-3β leading to the promotion of its cell survival, corroborating the previous study on activation of AKT/mTORC2 pathway, corresponding to increase in cell proliferation (Terasaki et al., 2021).

To substantiate our findings and to understand whether ectopically expressing SOX2-NTH protein upregulated phospho-GSK-3β gene expression in HeLa cells similar to ESCCs, we transduced the purified protein and vehicle control in HeLa cells and the western blotting analysis revealed that there is an increase in the expression of phospho-GSK-3β compared to the control, whereas no difference in total GSK-3β was observed upon treatment (Figure 4.10

D). These results corroborated with the previously published report (Terasaki et al., 2021). Thus, we speculate that

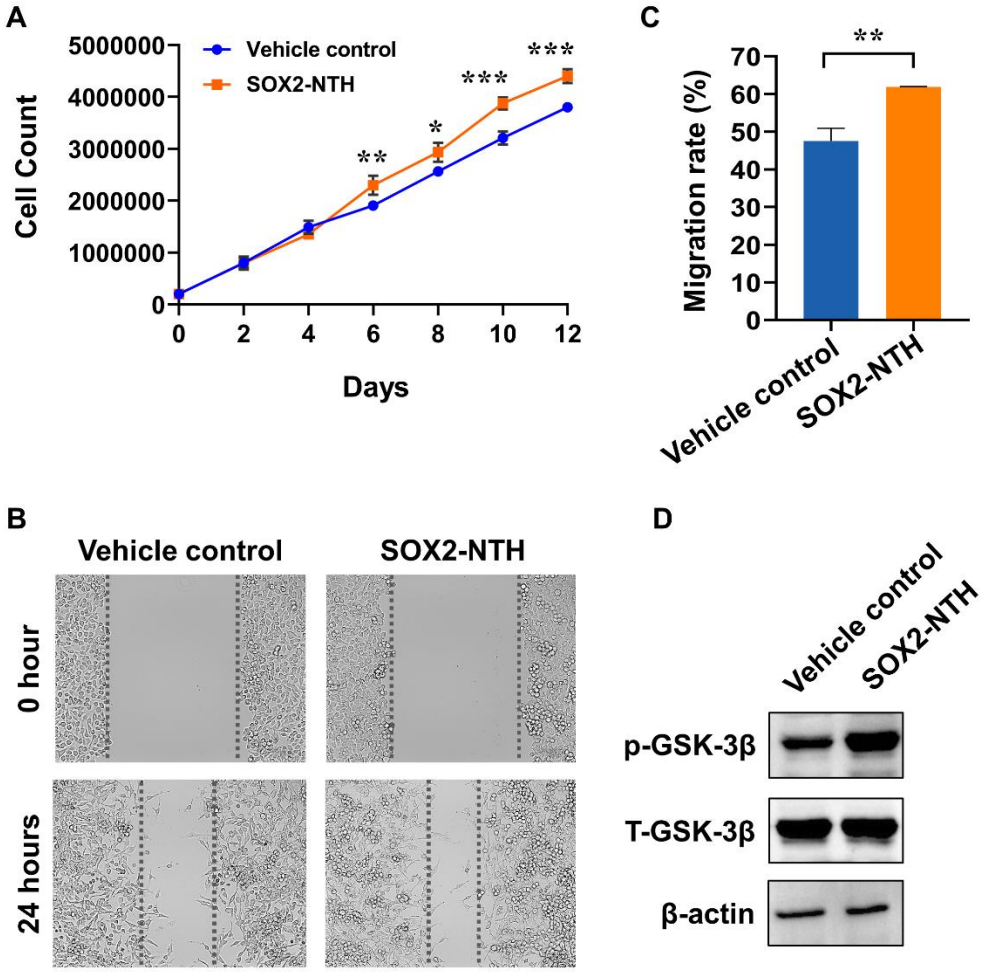


Figure 4.10 Effect of purified SOX2-NTH fusion protein on HeLa cells. (A) The cell proliferation was performed by counting cells for time duration of 12 days. Y-axis signifies cell count, and X-axis signifies days of treatment. The data shows significant reduction in cell proliferation with OCT4-NTH protein treated wells compared to vehicle control from Day 6 (**P=0.0065), Day 8 (*P=0.0110), Day 10 (***P = 0.0002), Day 12 (***P = 0.0002) (n = 4). **(B)** Scratch assay performed on HeLa cells for 24 hours. **(C)** Graphical representation of the changes in the migration rate upon treatment with SOX2-NTH and vehicle control. The data

shows significant reduction in migration rate with SOX2-NTH protein treated wells compared to vehicle control (**P=0.0022). **(D)** The western blotting analysis on the expression of phospho-GSK-3β (P-GSK-3β), total-GSK-3β (T-GSK-3β) and reference gene β-actin for the SOX2-HTN and vehicle control treated HeLa for 72 hours. The quantitative data shown are mean ± SEM (n = 3). *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

similar to ESCCs, expression of SOX2 on HeLa cells impacts the cells survival by enhancing cell proliferation through the activation of AKT/GSK-3β pathway. It also suggested that PTEN/PI3K/PDK1 and mTORC2 pathways might also be involved in the activation of SOX2 in HeLa cells similar to ESCCs. However, detailed analysis is required for unraveling the cell signaling pathways concerning expression of SOX2 and cell survival, cell migration and cell proliferation in HeLa cells.

4.2.2.3 Biological activity of purified recombinant UTF1-NTH protein

Effect of the exogenously delivered recombinant UTF1 fusion protein on cell proliferation and clonogenic potential and the rate of migration in HeLa cells

Previous studies have deemed UTF1 as a tumor suppressor. Upon expressing UTF1 in HeLa and overexpressing UTF1 in SiHa cells (Cervical cancer cell lines), they demonstrated a significant reduction in cell proliferation and MTT assays compared to their control treated cells. Also, the clonogenic potential of UTF1 expressing cells were observed to have dramatically reduced upon treatment. Thus, the previous report demonstrated that ectopic expression of UTF1 suppresses cancer cell growth *in vitro* (Wu and Zheng, 2013). The question was whether purified UTF1-NTH protein when transduced in HeLa cells will also lead to the reduction in cell proliferation and clonogenic potential of the cells. The cells were

seeded in triplicates for both the vehicle control and the protein for six days, and the analysis was done on the last day using MTT assay (Santos et al., 2011). The absorbance of the formazan product was measured at 575 nm, and the results observed clearly showed a significant decrease in the UTF1-NTH treated cells as compared to the vehicle control treated cells (Figure 4.11 A). Pertaining to the fact that the absorbance is directly proportional to the number of live cells (Janani et al., 2018), it is clear that cell proliferation was slowed down in HeLa upon protein treatment for 6 days, whereas vehicle control showed no abrogation/reduction in cell proliferation.

We sought to understand whether UTF1-NTH protein abrogates the clonogenic potential of HeLa cells. Upon treatment of cells for 10 days, the results showed a drastic decline in the number of foci formed in UTF1-NTH treated cell compared to the vehicle control (Figure 4.11 B). Both these observations corroborated the previous, results demonstrating ectopically expressing UTF1 retarded cell proliferation and inhibited colony forming potential in HeLa cells (Wu and Zheng, 2013). Although, earlier study shows G1/S transition arrest in UTF1 overexpressed HeLa cells (Wu and Zheng, 2013), however, we did not observe any significant difference in the G1/S phase in UTF1-NTH treated cells after 72 hours compared to vehicle control (data not shown). The reason might be due to the duration of treatment, which might not sufficient to significantly affect the cell cycle in HeLa cells.

We were curious to assess the migration potential of HeLa cells upon UTF1-NTH treatment for 48 hours and understand whether it also reduces the migration potential of the cells. Earlier reports have demonstrated that although cell proliferation and cell migration follow two different cell signaling pathways, however, cell proliferation and cell migration are interconnected process where they work in sequence in the process of wound healing (Donatis

et al., 2010). Interestingly, UTF1-NTH treated cells showed increase in the rate of migration compared to vehicle control treated cells (Figure 4.11 C and 4.11 D). This result does not correlate with the decrease in cell proliferation upon UTF1-NTH treatment, however, we speculated that UTF1 increased the rate of migration in the first 48 hours but longer exposure of the protein on HeLa cells will eventually attenuate the migratory potential similar to cell proliferation. Further in detail understanding of the migratory genes and the impact of short and long-term expression of UTF1-NTH on the invasiveness of HeLa can be assessed in the near future.

Effect of ectopic expression of purified UTF1-NTH protein on expression of $p27^{Kip1}$ on HeLa cells

Wu and colleagues screened the potential genes responsible in the arresting cell cycle at G1/S transition phase, consequently leading to decrease in cell proliferation and clonogenic potential of HeLa cells. They identified that amongst many genes, mRNA levels of $p27^{Kip1}$ was elevated in UTF1 overexpressing HeLa cells. They proposed that UTF1 acts as a tumor suppressor on cervical cancer cells through the activation of $p27^{Kip1}$ (Wu and Zheng, 2013). Thus, to further investigate the potential role of UTF1-NTH protein in regulating the expression of $p27^{Kip1}$ gene, we transduced the cells for 24 hours and 72 hours with purified UTF1-NTH protein, the relative gene expression analysis showed significant increase in the mRNA levels with increase in time of treatment in comparison to vehicle control treated cells (Figure 4.11 E) similar to the previously published report where ~1.5 fold increase in the relative expression of the gene in UTF1 treated cells were observed. However, further understanding of the detailed mechanism involved in tumor suppressor activity of UTF1 in cancer paradigm, especially in cervical cancer cells is important.

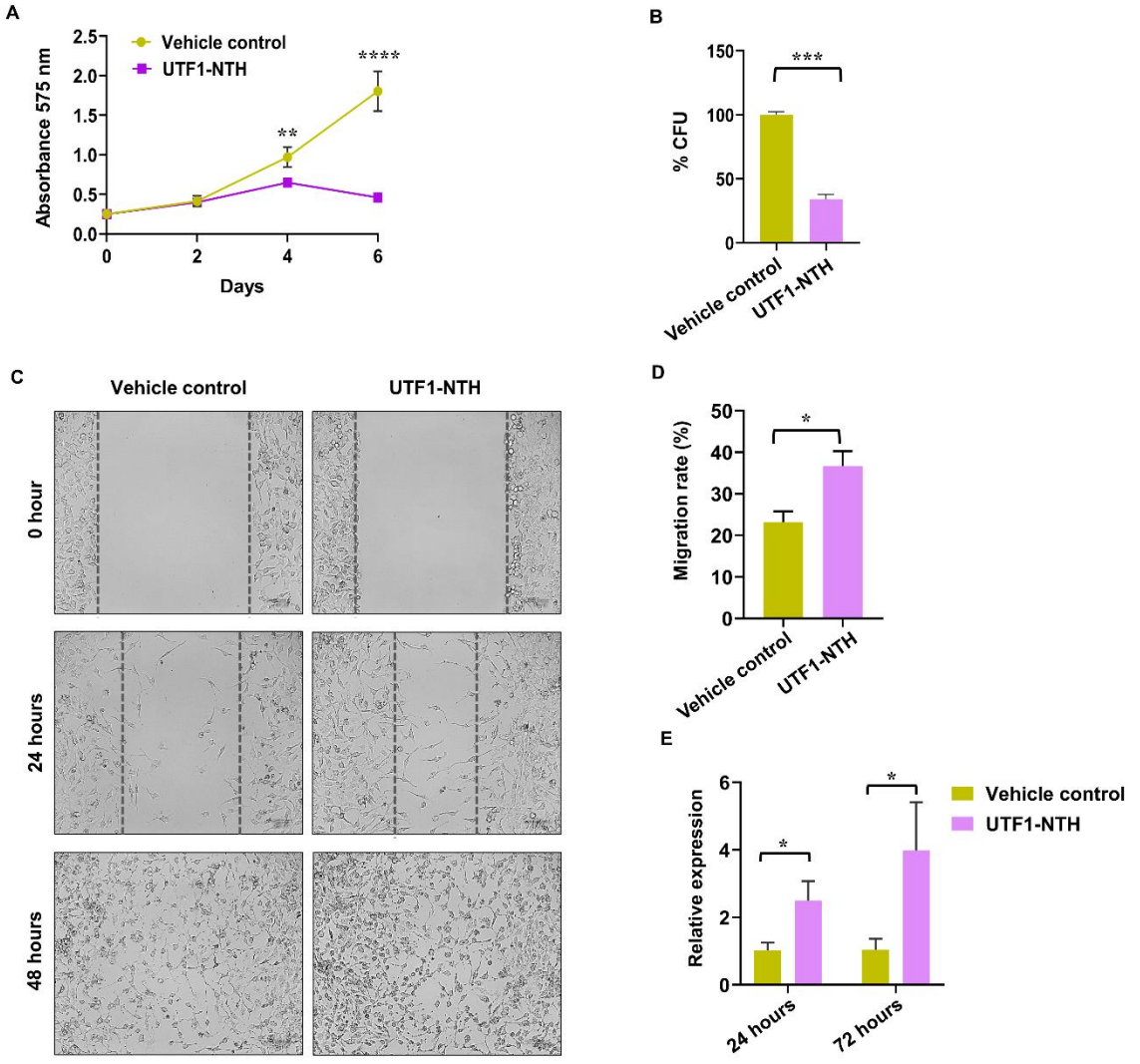


Figure 4.11 Effect of purified UTF1-NTH fusion protein on HeLa cells. (A) The absorbance was measured at 575 nm. Y-axis signifies absorbance at 575 nm, and X-axis signifies days of treatment. The data shows significant reduction in cell proliferation with UTF1-NTH protein treated wells compared to vehicle control from Day 4 (**P=0.0083), Day 6 (****P=0.0001) (n = 3). **(B)** Quantitative analysis of the clonogenic potential of UTF1-NTH and Vehicle control treated HeLa cells (***P=0.0002) **(C)** Scratch assay performed on HeLa cells for 48 hours. **(D)** Graphical representation of the changes in the migration rate upon treatment with UTF1-NTH and vehicle control. The data shows significant reduction in

migration rate with UTF1-NTH protein treated wells compared to vehicle control (*P=0.0380). **(E)** The relative gene expression analysis of $p27^{Kip1}$ on control and UTF1-NTH transduced HeLa cells, analyzed using RT-qPCR method [24 hours (*P=0.0144); 72 hours (*P=0.0257). The quantitative data shown are mean ± SEM (n = 3). *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

4.2.2.4 Biological activity of purified recombinant GLIS1-NTH protein

Effect of the exogenously delivered recombinant GLIS1 fusion protein on the rate of migration of MDA-MB-231 cells

Overexpression of GLIS1 in breast cancer MDA-MB-231 cells contributed to the increase in the migration rate of these cells (Shimamoto et al., 2020). Therefore, to determine the functionality of the purified GLIS1-NTH fusion protein, the rate of migration of the breast cancer cells MDA-MB-231 was determined using migration assay. Upon protein transduction in MDA-MB-231 cells at every 24 hours, the results showed that the migration rate of cells treated with purified GLIS1-NTH protein migrated faster compared to the vehicle control (Figure 4.12 A and 4.12B). This increase in the rate of migration was in congruence with the recent study (Shimamoto et al., 2020). The calculation for the migration rate was performed as reported earlier (Grada et al., 2017). The migration assay was also performed for HFF (BJ) cells, and the results showed no significant difference in the migration rate (Figure 4.11 C), indicating that the observation was specific to breast cancer cells. GLIS1 overexpression in breast cancer cells MDA-MB-231 contributed to the increase in the migration and invasion capacities of the cells, possibly through the upregulation of WNT5A (Shimamoto et al., 2020) or by cooperating with CUX1, thus stimulating activity of TCF/β-catenin transcription factor

and enhancing cell migration and invasion of breast cancer cells (Vadnais et al., 2014). In our study, we also observed an increase in migration rate when the same MDA-MB-231 cells were treated with the GLIS1 fusion protein. Interestingly, no significant difference in migration rate was observed when BJ cells were treated with the GLIS1 fusion protein. This might be due to the fact that GLIS1 alone has no prominent effect on somatic cells, but in combination with other transcription factors like OCT4, SOX2, and KLF4, it pushes the cell fate towards the generation of iPSCs.

Effect of the exogenously delivered recombinant GLIS1 fusion protein on expression of WNT5A in MDA-MB-231 cells

GLIS1 overexpression in breast cancer MDA-MB-231 cells contributed to the increase in the migration and invasion capacities of the cells, possibly through the upregulation of WNT5A. This played an advanced role in estrogen-independent breast cancer progression leading to poorer prognosis. It was reported that GLIS1 have myriads of downstream targets, however, WNT5A and SNAI2 were upregulated in GLIS1 expressing cells. Additionally, inhibition of WNT5A downregulated the expression of GLIS1, suggesting a regulatory feedback loop (Shimamoto et al., 2020).

In our study, we also observed an increase in migration rate when the same MDA-MB-231 cells were treated with the GLIS1 fusion protein and to further understand whether exogenously expressing GLIS1-NTH protein on MDA-MB-231 cells will affect the gene regulation of WNT5A, we transduced the cells with the protein for 72 hours and the relative expression analysis showed 2-fold increase in the WNT5A expression compared to the control (Figure 4.12 D), similar to the previous report stating ~2.5-fold increase in the expression of WNT5A upon transfection of FLAG-

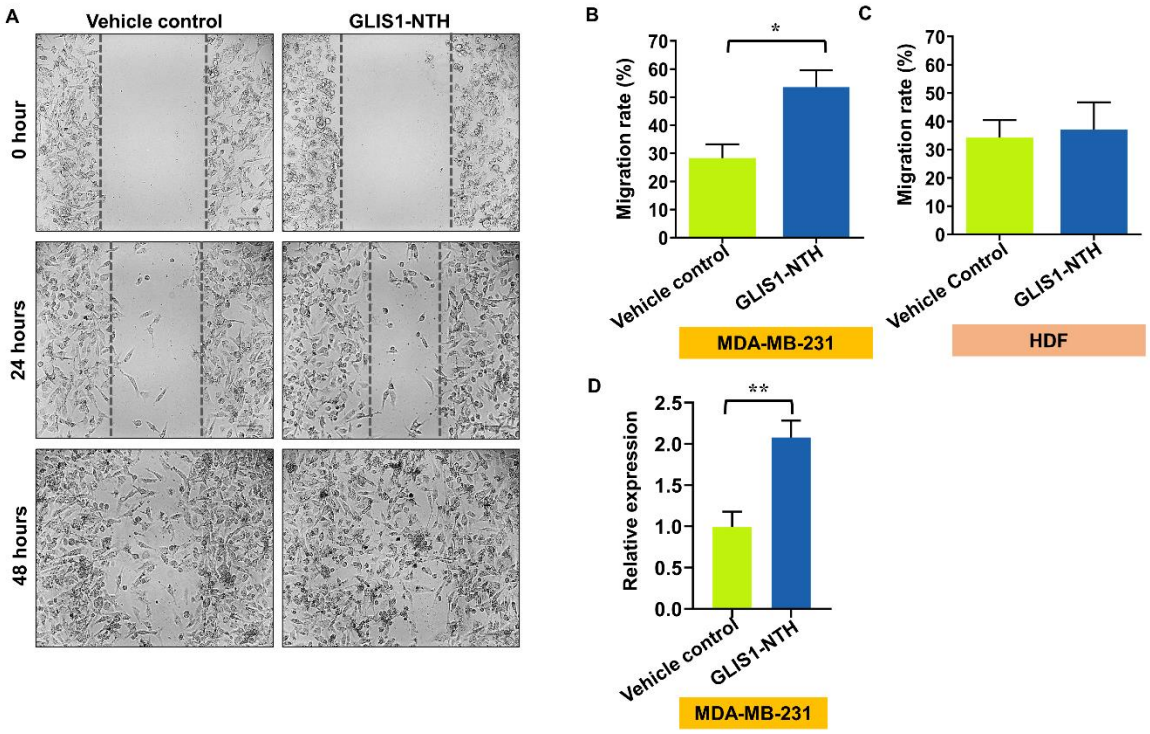


Figure 4.12 Effect of purified GLIS1-NTH fusion protein on MDA-MB-231 cells. (A) Scratch assay performed on MDA-MB-231 cells for 48 hours. **(B and C)** Quantitative analysis of the changes in the migration rate upon treatment with GLIS1-NTH and vehicle control. The data shows significant reduction in migration rate with GLIS1-NTH protein treated wells compared to vehicle control (**P*=0.0170) on MDA-MB-231 cells, whereas HDF showed no significant change in the rate of migration rate. **(D)** The relative gene expression analysis of WNT5A on control and UTF1-NTH transduced MDA-MB-231 cells, analyzed using RT-qPCR method [72 hours (***P*=0.0079). The quantitative data shown are mean ± SEM (n = 3). **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

GLIS1 compared to its control. The difference in fold change between GLIS1-NTH transduced cells and FLAG-GLIS1 transfected cells might be due to the limitation imparted by delivery of purified protein into mammalian cells such as endosomal entrapment, protein stability and

the amount of protein translocating into the nucleus. However, our results indicate that the generated protein is biologically active and to has great potential in providing a controlled time and dose dependent platform for deciphering the multifarious roles it possesses in cancer paradigm as well as in somatic cells.

4.3 Conclusion

In this chapter, we have demonstrated that the four purified recombinant fusion proteins OCT4, SOX2, UTF1 and GLIS1 are stable at cell culture conditions. OCT4 showed stability for 24 hours, whereas SOX2 and UTF1 were stable for 48 hours. GLIS1 being the largest molecular weight protein was partially stable for 24 hours. These proteins (OCT4 and SOX2) with the fusion tags showed successful cytoplasmic and nuclear translocation in both HFF (BJ) and HeLa cells and (UTF1) in HeLa cells after protein transduction. Interestingly, translocation studies also revealed the difference in the protein uptake and translocation between HeLa and HFF (BJ) cells. Further we demonstrated the functional role of these purified fusion proteins in mammalian cells. OCT4 showed significant reduction in cell proliferation and cell migration rate in HFF (BJ) cells and the reporter (OCT4-GFP-2A-PURO) confirmed that the protein is transcriptionally active, thus giving rise to GFP^{+} cells when transduced into HeLa cells containing the reporter. SOX2 fusion protein upon transduction showed enhanced cell proliferation and cell migration in HeLa cells whereas UTF1 fusion protein showed reduction in cell proliferation and clonogenic potential in HeLa cells. The relative gene expression analysis also revealed that UTF1 enhanced the mRNA levels of $p27^{Kip1}$, consequently causing reduced cell proliferation, thus acting as a tumor suppressor protein. Moreover, SOX2 enhanced the expression of phospho-GSK-3β protein in HeLa cells, thus potentially assisting its cell survival. Hence, these results proved that SOX2 also acted as a tumor progressor protein

in HeLa cells. GLIS1, on the other hand, showed functional role in breast cancer cells (MDA-MB-231). Expressing GLIS1 fusion protein led to increased cell migration rate and enhanced the mRNA gene expression of WNT5A, thus, showing tumorigenic potential.

These purified fusion proteins have tremendous potential to substitute for its viral and genetic counterparts in the cellular reprogramming process and can aid in the successful generation of integration-free cells and potentially can be used to derive various cell lineages in a dose- and time-dependent manner (Mitchell et al., 2014). It will allow the scientific community to understand its detailed function in different stage-specific developmental processes, its function, and downstream implications in various cancers. Additionally, it will also help in understanding other disease models, molecular interactions with reprogramming factors and signaling pathways, nuances of the mechanism during cell-fate transitions, and substituting its viral/ genetic counterpart in the generation of integration-free iPSCs. Importantly, this tool can be used to gauge the specific spatiotemporal expression of these reprogramming factors in different combinations during the generation of integration-free iPSCs and understand its expression and function in different diseases in the near future.

CHAPTER 5

Conclusions and future directions

The detailed work carried out in this thesis work established that the strategy we formulated with the codon optimized gene constructs with fusion tags, led to the successful generation of biologically active reprogramming transcription factors. We have established a facile, efficient, and cost-effective approach to achieve the expression and purification of these recombinant fusion proteins under native conditions. We have extensively screened the optimal expression parameters for each of the recombinant transcription factors. Amongst N- and C-terminal tagged OCT4 and SOX2 fusion proteins, HTN-OCT4 showed disordered secondary structure upon native purification, whereas HTN-SOX2 showed comparatively less protein yield than its counterpart. Hence, C-terminal tagged OCT4 and SOX2 were further used for functional assays. Further, protein expression was observed only in C-terminal tagged UTF1 and GLIS1 fusion proteins compared to their counterparts, which confirmed that the position of fusion tags influences the overall expression and folding conformation of the protein. The purified full-length recombinant reprogramming fusion proteins have retained their secondary structure and have shown successful transduction ability into both cell and nucleus of the cells, attributed by the addition of fusion tags CPP and NLS with the protein. OCT4-NTH protein showed a reduction in migration rate and cell proliferation upon transduction in human fibroblasts. The reporter assay performed in HeLa cells also demonstrated that the purified OCT4-NTH fusion protein can bind to its own promoter placed upstream of GFP, thus giving rise to GFP^+ cells. This signified that OCT4 is transcriptionally active, and can be used

as a substitute for its genetic counterpart in cell reprogramming experiments. Similarly, SOX2-NTH protein promoted cell migration and proliferation in HeLa cells. It also showed increased expression of phospho-GSK-3β. AKT/ GSK-3β pathway activation and subsequent cell survival confirms the role of SOX2 in HeLa cells as a tumor progressor, however further detailed molecular analysis on the role of SOX2 in cancer paradigm is required in the near future. UTF1-NTH on the other hand when transduced into HeLa cells reduced clonogenic potential and cell proliferation. It also showed enhanced relative gene expression of $p27^{Kip1}$, thus potentially acting as a tumor suppressor protein. These results show immensely diverse roles of both UTF1 and SOX2 in cervical cancer cells. It also paves a platform for plethora of opportunities to elucidate the role of these two proteins singularly or in combination in cervical cancer cells. GLIS1-NTH protein when transduced on human fibroblast cells showed no effect. However, an enhanced migration rate in MDA-MB-231 cells upon protein transduction was observed along with an enhanced WNT5A mRNA levels, suggesting its potential contribution to enhancing cancer invasion. Remarkably, all these proteins have diverse effects in normal as well as the cancerous cellular environment, which requires detailed research and understanding in the future. These proteins, however, have the potential to replace their genetic counterparts for understanding their role in reprogramming. The advantages imparted by the use of recombinant proteins toolbox are control over dosage and time, ease of generation both in quantity and quality and provides immense scope to unravel the importance of these transcription factors in the generation of integration-free iPSCs.

Highlights

1. Identified suitable genetic construct and optimal expression conditions for each transcription factor.

2. Established a one-step homogenous purification protocol for the recombinant fusion proteins under native conditions.

3. Confirmed that the purified recombinant fusion proteins are stable under cell culture conditions and have cell penetration and nuclear translocation ability.

4. Validated the functionality of the purified recombinant fusion proteins and confirmed the fusion strategy did not affect their bioactivity.

The established recombinant protein toolbox holds immense future potential in both the reprogramming and cancer paradigm. These factors individually or in combination provides an opportunity to understand their prospective roles in the generation of induced pluripotent stem cells in different phases of reprogramming by replacing their genetic counterparts from the reprogramming cocktail using a sliding window approach. It will also provide a way to understand the detailed importance of stoichiometry of these factors in reprogramming (Figure 5.1).

To further elaborate on the follow-up work involving sliding-window approach, the next step includes the generation of reprogramming transcription factor cocktail in retroviral form. Using retroviruses for reprogramming will provide the positive control for the sliding window approach. The viral forms of the factors will be substituted with our generated proteins (for e.g. OCT4-NTH + pMX-SOX2 + pMX-KLF4 + pMX-cMYC).

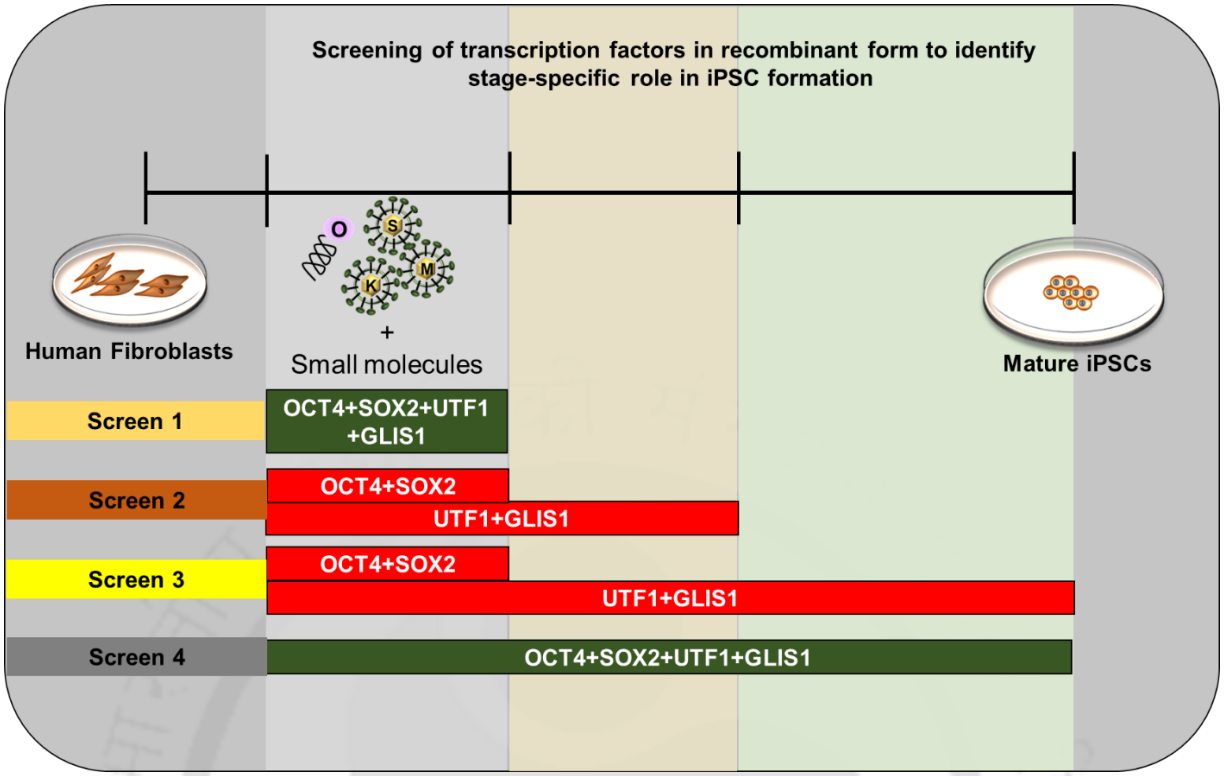


Figure 5.1 Schematic representation of the prospective future applications of recombinant protein toolbox

Generated iPSCs, both from retroviral approach as well as sliding window approach will be subjected to karyotyping analysis, expression of surface and nuclear markers in iPSCs as well as source cells (Fibroblasts and/or PBMCs) using immunocytochemistry, to provide confirmation on the conversion from one cell type towards the pluripotent stage. mRNA and protein level expressions for pluripotency and somatic cell associated genes will be carried out via Western blotting and RT-qPCR method. Using all four factors in recombinant protein form for the generation of human iPSCs, our ultimate aim is to differentiate these integration-free human iPSCs into desired differentiated derivatives and use CRISPR-Cas9 technology to generate various disease models.

These factors have also shown significant role in cancer and normal cellular metabolic functions and thus far, the role of UTF1 and GLIS1 proteins are essentially under explored. They open up avenues to decipher various role of these proteins in different cancer environment, stem cells and in normal cells. Following up to this idea of delineating the roles of UTF1 and GLIS1 proteins in various cellular functions, we aim to generate knockout cell lines using CRISPR-Cas9 technology and evaluate the effect of absence, presence and overexpression of these proteins in the cells. Thus, emphasizing on understanding the synergistic effects of these proteins in similar or different cancer environments which is of paramount importance and a great prospect for finding cancer targets leading to future remedies.

References

Adachi, K., Suemori, H., Yasuda, S., Nakatsuji, N., Kawase, E., 2010. Role of SOX2 in maintaining pluripotency of human embryonic stem cells. Genes to Cells. 15, 455–470.

Ahuja, D., Sáenz-Robles, M.T., Pipas, J.M., 2005. SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. Oncogene. 24(52), 7729–7745.

Åkesson, M., Hagander, P., Axelsson, J.P., Bioeng, B., 2001. Avoiding Acetate Accumulation in *Escherichia coli* Cultures Using Feedback Control of Glucose Feeding. Biotechnol. Bioeng. 73(3), 223-230.

Al-Hawash, A.B., Zhang, X., Ma, F., 2017. Strategies of codon optimization for high-level heterologous protein expression in microbial expression systems. Gene Rep. 9, 46–53.

Ambrosetti, D.-C., Basilico, C., Dailey, L., 1997. Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol Cell Biol. 17, 6321–6329.

Angel, M., Yanik, M.F., 2010. Innate Immune Suppression Enables Frequent Transfection with RNA Encoding Reprogramming Proteins. PLoS One. 5, e11756.

Anokye-Danso, F., Trivedi, C.M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P.J., Epstein, J.A., Morrisey, E.E., 2011a. Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency. Cell Stem Cell. 8, 376–388.

Arnold, A., Naaldijk, Y.M., Fabian, C., Wirth, H., Binder, H., Nikkhah, G., Armstrong, L., Stolzing, A., 2012. Reprogramming of human huntington fibroblasts using mRNA. Int. sch. res. Notices.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., Lovell-Badge, R., 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 17, 126–140.

Azaman, S.N.A., Ramanan, R.N., Joo Shun, T., Rahim, R., Abdullah, M., Ariff, A., 2010. Screening for the optimal induction parameters for periplasmic producing interferon-α2b in *Escherichia coli*. African J. Biotechnol. 9, 6345–6354.

Ban, H., Nishishita, N., Fusaki, N., Tabata, T., Saeki, K., Shikamura, M., Takada, N., Inoue, M., Hasegawa, M., Kawamata, S., Nishikawa, S.I., 2011. Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. Proc. Natl. Acad. Sci. U S A. 108, 14234–14239.

Baneyx, F., Mujacic, M., 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nat. Biotechnol. 22, 1399–1408.

Bao, Q., Morshedi, A., Wang, F., Bhargy, S., Pervushin, K., Yu, W.P., Dröge, P., 2017. Utf1 contributes to intergenerational epigenetic inheritance of pluripotency. Sci Rep. 7(1), 1–11.

Bareiss, P.M., Paczulla, A., Wang, H., Schairer, R., Wiehr, S., Kohlhofer, U., Rothfuss, O.C., Fischer, A., Perner, S., Staebler, A., Wallwiener, D., Fend, F., Fehm, T., Pichler, B., Kanz, L., Quintanilla-Martinez, L., Schulze-Osthoff, K., Essmann, F., Lengerke, C., 2013. SOX2 expression associates with stem cell state in human ovarian carcinoma. Cancer Res. 73, 5544–5555.

Bayart, E., Cohen-Haguenauer, O., 2013. Technological Overview of iPS Induction from Human Adult Somatic Cells. Curr Gene Ther. 13(2), 73-92.

Beers, J., Linask, K.L., Chen, J.A., Siniscalchi, L.I., Lin, Y., Zheng, W., Rao, M., Chen, G., 2015. A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. Sci Rep. 5(1), 1–9.

Behringer, R., Gertsenstein, M., Vintersten Nagy, K., Nagy, A., 2017. Integrating piggyBac Transposon Transgenes into Mouse Fibroblasts by Electroporation. Cold Spring Harb Protoc. 2017(10).

Belay, E., Dastidar, S., VandenDriessche, T., K.L. Chuah, M., 2012. Transposon-Mediated Gene Transfer into Adult and Induced Pluripotent Stem Cells. Curr Gene Ther. 11, 406–413.

Ben-David, U., Benvenisty, N., 2011. The tumorigenicity of human embryonic and induced pluripotent stem cells. Nat Rev Cancer. 11, 268–277.

Berrow, N.S., Büssow, K., Coutard, B., Diprose, J., Ekberg, M., Folkers, G.E., Levy, N., Lieu, V., Owens, R.J., Peleg, Y., Pinaglia, C., Quevillon-Cheruel, S., Salim, L., Scheich, C., Vincentelli, R., Busso, D., 2006. Recombinant protein expression and solubility screening in *Escherichia coli*: A comparative study. Acta Crystallogr. Sect. D Biol. Crystallogr. 62, 1218–1226.

Bertin, E., Piccoli, M., Franzin, C., Nagy, A., Mileikovsky, M., de Coppi, P., Pozzobon, M., 2015. Reprogramming of mouse amniotic fluid cells using a PiggyBac transposon system. Stem Cell Res. 15, 510–513.

Bitzer, M., Armeanu, S., Lauer, U.M., Neubert, W.J., 2003. Sendai virus vectors as an emerging negative-strand RNA viral vector system. J Gene Med. 5, 543–553.

Borgohain, M.P., Narayan, G., Kumar, H.K., Dey, C., Thummer, R.P., 2018. Maximizing Expression and Yield of Human Recombinant Proteins from Bacterial Cell Factories for Biomedical Applications, Advances in Microbial Biotechnology. Apple Academic Press, pp. 447–486.

Bornhorst, J.A., Falke, J.J., 2000. Purification of proteins using polyhistidine affinity tags, in: Methods in Enzymology. Elsevier, 245–254.

Bosnali, M., Edenhofer, F., 2008. Generation of transducible versions of transcription factors Oct4 and Sox2. Biol. Chem. 389, 851–861.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., LaBaer, J., 2002. Proteome-scale purification of human proteins from bacteria. Proc. Natl. Acad. Sci. U. S.

A. 99, 2654–2659.

Briggs, R., King, T.J., 1951. Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. Proc. Natl. Acad. Sci. U S A 167, 232.

Brouwer, M., Zhou, H., Nadif Kasri, N., 2015. Choices for Induction of Pluripotency: Recent Developments in Human Induced Pluripotent Stem Cell Reprogramming Strategies. Stem Cell Reviews and Reports 12(1), 54–72.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., Jaenisch, R., 2012. Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase. Cell 150, 1209–1222.

Buganim, Y., Faddah, D.A., Jaenisch, R., 2013. Mechanisms and models of somatic cell reprogramming. Nat Rev Genet. 14(6), 427–439.

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.

Cantz, T., Key, G., Bleidiβel, M., Gentile, L., Han, D.W., Brenne, A., Schöler, H.R., 2008. Absence of OCT4 Expression in Somatic Tumor Cell Lines. Stem Cells. 26, 692–697.

Card G., D.A., Hebbar, P.B., Li, L., Trotter, K.W., Komatsu, Y., Mishina, Y., Archer, T.K., 2008. Oct4/Sox2-Regulated miR-302 Targets Cyclin D1 in Human Embryonic Stem Cells. Mol Cell Biol. 28, 6426–6438.

Carey, B.W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M., Jaenisch, R., 2009. Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc Natl Acad Sci U S A. 106, 157–162.

Carter, P.J., Samulski, R.J., 2000. Adeno-associated viral vectors as gene delivery vehicles. Int J Mol Med. 6, 17–27.

Catanese, D.J., Fogg, J.M., Schrock, D.E., Gilbert, B.E., Zechiedrich, L., 2011. Supercoiled Minivector DNA resists shear forces associated with gene therapy delivery. Gene Ther. 19(1), 94–100.

Chabot, S., Orio, J., Schmeer, M., Schleef, M., Golzio, M., Teissié, J., 2012. Minicircle DNA electrotransfer for efficient tissue-targeted gene delivery. Gene Ther. 20(1), 62–68.

Chateauvieux, S., Morceau, F., Dicato, M., Diederich, M., 2010. Molecular and Therapeutic Potential and Toxicity of Valproic Acid. J Biomed Biotechnol.

Chen, F., Zhang, G., Yu, L., Feng, Y., Li, X., Zhang, Z., Wang, Y., Sun, D., Pradhan, S., 2016. High-efficiency generation of induced pluripotent mesenchymal stem cells from human dermal fibroblasts using recombinant proteins. Stem Cell Res. Ther. 7(1), 1-14.

Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden, S.E., Diol, N.R., Propson, N.E., Wagner, R., Lee, G.O., Antosiewicz-Bourget, J., Teng, J.M.C., Thomson, J.A., 2011. Chemically defined conditions for human

iPSC derivation and culture. Nat Methods. 8, 424–429.

Chen, H., Wang, D., Xia, R., Mao, Q., Xia, H., 2015. A novel adenoviral vector carrying an all-in-one Tet-On system with an autoregulatory loop for tight, inducible transgene expression. BMC Biotechnol. 15, 1–8

Chen, J., Wang, G., Lu, C., Guo, X., Hong, W., Kang, J., Wang, J., 2012. Synergetic Cooperation of microRNAs with Transcription Factors in iPS Cell Generation. PLoS One. 7, e40849.

Chen, R., 2012. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnol.Adv. 30, 1102–1107.

Chen, X., Cui, J., Yan, Z., Zhang, H., Chen, Xian, Wang, N., Shah, P., Deng, F., Zhao, C., Geng, N., Li, M., Denduluri, S.K., Haydon, R.C., Luu, H.H., Reid, R.R., He, T.C., 2015. Sustained high level transgene expression in mammalian cells mediated by the optimized piggyBac transposon system. Genes Dis. 2, 96–105.

Chen, Z.Y., He, C.Y., Ehrhardt, A., Kay, M.A., 2003. Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo. Mol. Ther. 8, 495–500.

Chew, J.-L., Loh, Y.-H., Zhang, W., Chen, X., Tam, W.-L., Yeap, L.-S., Li, P., Ang, Y.-S., Lim, B., Robson, P., Ng, H.-H., 2005. Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells . Mol Cell Biol. 25, 6031–6046.

Cho, H.-J., Lee, C.-S., Kwon, Y.-W., Paek, J.S., Lee, S.-H., Hur, J., Lee, E.J., Roh, T.-Y., Chu, I.-S., Leem, S.-H., 2010. Induction of pluripotent stem cells from adult somatic cells by protein-based reprogramming without genetic manipulation. Blood. 116, 386–395.

Choi, H., Lee, T., Yang, G., Oh, J., Won, J., … J.H.-J. of C., 2016, undefined, n.d. Efficient mRNA delivery with graphene oxide-polyethylenimine for generation of footprint-free human induced pluripotent stem cells. J Control Release. 222-235.

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

Chuang, H.-M., Huang, M.-H., Chen, Y.-S., Harn, H.-J., 2020. SOX2 for Stem Cell Therapy and Medical Use: Pros or Cons? Cell Transplant. 29, 0963689720907565.

Chung, T.L., Brena, R.M., Kolle, G., Grimmond, S.M., Berman, B.P., Laird, P.W., Pera, M.F., Wolvetang, E.J., 2010. Vitamin C Promotes Widespread Yet Specific DNA Demethylation of the Epigenome in Human Embryonic Stem Cells. Stem Cells. 28, 1848–1855.

Chuva De Sousa Lopes, S.M., van den Driesche, S., Carvalho, R.L.C., Larsson, J., Eggen, B., Surani, M.A., Mummery, C.L., 2005. Altered primordial germ cell migration in the absence of transforming growth factor β signaling via ALK5. Dev Biol. 284, 194–203.

Darquet, A.M., Cameron, B., Wils, P., Scherman, D., Crouzet, J., 1997. A new DNA vehicle for nonviral gene delivery: supercoiled minicircle. Gene Ther. 4(12), 1341–1349.

Davis, R.P., Nemes, C., Varga, E., Freund, C., Kosmidis, G., Gkatzis, K., de Jong, D., Szuhai, K., Dinnyés, A., Mummery, C.L., 2013. Generation of induced pluripotent stem cells from human foetal fibroblasts using the Sleeping Beauty transposon gene delivery system. Differentiation. 86, 30–37.

De Donatis, A., Ranaldi, F., Cirri, P., 2010. Reciprocal control of cell proliferation and migration. Cell Commun. Signal. 8, 1–4.

De Mey, M., De Maeseneire, S., Soetaert, W., Vandamme, E., 2007. Minimizing acetate formation in *E. coli* fermentations. J. Ind. Microbiol. Biotechnol. 34(11), 689-700

Diecke, S., Lisowski, L., Kooreman, N.G., Wu, J.C., 2014. Second generation codon optimized minicircle (CoMiC) for nonviral reprogramming of human adult fibroblasts. Methods Mol Biol. 1181, 1–13.

Diecke, S., Lu, J., Lee, J., Termglinchan, V., Kooreman, N.G., Burridge, P.W., Ebert, A.D., Churko, J.M., Sharma, A., Kay, M.A., Wu, J.C., 2015. Novel codon-optimized mini-intronic plasmid for efficient, inexpensive and xeno-free induction of pluripotency. Sci Rep. 5(1), 1–9.

Doherty, J.E., Huye, L.E., Yusa, K., Zhou, L., Craig, N.L., Wilson, M.H., 2011. Hyperactive piggyBac Gene Transfer in Human Cells and In Vivo. Hum.Gene Ther, 23(3), 311-320.

Dowey, S.N., Huang, X., Chou, B.K., Ye, Z., Cheng, L., 2012. Generation of integration-free human induced pluripotent stem cells from postnatal blood mononuclear cells by plasmid vector expression. Nat. Protoc. 7(11), 2013–2021.

Drews, K., Tavernier, G., Demeester, J., Lehrach, H., de Smedt, S.C., Rejman, J., Adjaye, J., 2012. The cytotoxic and immunogenic hurdles associated with non-viral mRNA-mediated reprogramming of human fibroblasts. Biomaterials. 33, 4059–4068.

Durruthy-Durruthy, J., Briggs, S.F., Awe, J., Ramathal, C.Y., Karumbayaram, S., Lee, P.C., Heidmann, J.D., Clark, A., Karakikes, I., Loh, K.M., Wu, J.C., Hoffman, A.R., Byrne, J., Reijo Pera, R.A., Sebastiano, V., 2014. Rapid and Efficient Conversion of Integration-Free Human Induced Pluripotent Stem Cells to GMP-Grade Culture Conditions. PLoS One. 9, e94231.

Durruthy, J.D., Sebastiano, V., 2015. Derivation of GMP-compliant integration-free hiPSCs using modified mRNAs. Methods Mol Biol. 1283, 31–42.

Echigoya, K., Koyama, M., Negishi, L., Takizawa, Y., Mizukami, Y., Shimabayashi, H., Kuroda, A., Kurumizaka, H., 2020. Nucleosome binding by the pioneer transcription factor OCT4. Sci. Rep. 10(1), 1-11.

Efe, J.A., Ding, S., 2011. The evolving biology of small molecules: controlling cell fate and identity. Philosophical Transactions of the Royal Society B: Biol Sci. 366, 2208–2221.

Eiteman, M.A., Altman, E., 2006. Overcoming acetate in *Escherichia coli* recombinant protein fermentations. Trends Biotechnol. 24(11), 530-536.

Esch, D., Vahokoski, J., Groves, M.R., Pogenberg, V., Cojocaru, V., Vom Bruch, H., Han, D., Drexler, H.C.A., Araúzo-Bravo, M.J., Ng, C.K.L., Jauch, R., Wilmanns, M., Schöler, H.R., 2013. A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat. Cell Biol. 15, 295–301.

Esteban, M.A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S., Chen, K., Li, Y., Liu, X., Xu, J., Zhang, S., Li, F., He, W., Labuda, K., Song, Y., Peterbauer, A., Wolbank, S., Redl, H., Zhong, M., Cai, D., Zeng, L., Pei, D., 2010. Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. Cell Stem Cell. 6, 71–79.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., Bergman, Y., 2006. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol. 8(2) , 188–194.

Feng, B., Ng, J.H., Heng, J.C.D., Ng, H.H., 2009. Molecules that Promote or Enhance Reprogramming of Somatic Cells to Induced Pluripotent Stem Cells. Cell Stem Cell. 4, 301–312.

Forte, A., Schettino, M.T., Finicelli, M., Cipollaro, M., Colacurci, N., Cobellis, L., Galderisi, U., 2009. Expression pattern of stemness-related genes in human endometrial and endometriotic tissues. Mol Med. 15, 392–401.

Francis, D.M., Page, R., 2010. Strategies to Optimize Protein Expression in *E. coli*. Curr. Protoc. Protein Sci., 61(1), 5-24.

Fuhrhop, J.-H., Wang, T., 2004. Bolaamphiphiles, Chem. Rev. 104(6), 2901-2938.

Fuhrmann, M., Hausherr, A., Ferbitz, L., Scho¨dlscho¨dl, T., Heitzer, M., Hegemann, P., 2004. Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. Plant Mol. Biol. 55(6), 869-881.

Fujie., Y., Fusaki, N., Katayama, T., Hamasaki, M., Soejima, Y., Soga, M., Ban, H., Hasegawa, M., Yamashita, S., Kimura, S., Suzuki, S., Matsuzawa, T., Akari, H., Era, T., 2014. New Type of Sendai Virus Vector Provides Transgene-Free iPS Cells Derived from Chimpanzee Blood. PLoS One. 9, e113052.

Fukushima, A., Okuda, A., Nishimoto, M., Seki, N., Hori, T.A., Muramatsu, M., 1998. Characterization of functional domains of an embryonic stem cell coactivator UTF1 which are conserved and essential for potentiation of ATF-2 activity. JBC. 273, 25840–25849.

Fusaki, N., Nishiyama, A., Saeki Ã1, K., Hasegawa, M., 2009. Effcient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc. Jpn. Acad., Ser. B. 85(8), 348-362.

Gafni, O., Weinberger, L., Alfatah Mansour, A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., Rais, Y., Shipony, Z., Mukamel, Z., Krupalnik, V., Zerbib, M., Geula, S., Caspi, I., Schneir, D., Shwartz, T., Gilad, S., Amann-Zalcenstein, D., Benjamin, S., Amit, I., Tanay, A., Massarwa, R., Novershtern, N., Hanna,

J.H., 2013. Derivation of novel human ground state naive pluripotent stem cells. Nature. 504(7479).

Galla, M., Schambach, A., Falk, C.S., Maetzig, T., Kuehle, J., Lange, K., Zychlinski, D., Heinz, N., Brugman, M.H., Göhring, G., Izsvák, Z., Ivics, Z., Baum, C., 2011. Avoiding cytotoxicity of transposases by dose-controlled mRNA delivery. Nucleic Acids Res. 39, 7147–7160.

Galloway, C., Sowden, M., Smith, H., 2003. Increasing the Yield of Soluble Recombinant Protein Expressed in *E. coli* by Induction during Late Log Phase. Biotechniques. 34, 524–526.

Galonska, C., Smith, Z.D., Meissner, A., 2014. In Vivo and In Vitro Dynamics of Undifferentiated Embryonic Cell Transcription Factor 1. Stem Cell Rep. 2, 245–252.

Gao, Y., Han, Z., Li, Q., Wu, Y., Shi, X., Ai, Z., Du, J., Li, W., Guo, Z., Zhang, Y., 2015. Vitamin C induces a pluripotent state in mouse embryonic stem cells by modulating microRNA expression. FEBS J. 282, 685–699.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., Jaenisch, R., 2003. Induction of tumors in mice by genomic hypomethylation. Science. 300, 489–492.

Gen, Y., Yasui, K., Nishikawa, T., Yoshikawa, T., 2013. SOX2 promotes tumor growth of esophageal squamous cell carcinoma through the AKT/mammalian target of rapamycin complex 1 signaling pathway. Cancer Sci. 104, 810–816.

Gen, Y., Yasui, K., Zen, Y., Zen, K., Dohi, O., Endo, M., Tsuji, K., Wakabayashi, N., Itoh, Y., Naito, Y., Taniwaki, M., Nakanuma, Y., Okanoue, T., Yoshikawa, T., 2010. SOX2 identified as a target gene for the amplification at 3q26 that is frequently detected in esophageal squamous cell carcinoma. Cancer Genet. Cytogenet. 202, 82–93.

Geurts, A.M., Collier, L.S., Geurts, J.L., Oseth, L.L., Bell, M.L., Mu, D., Lucito, R., Godbout, S.A., Green, L.E., Lowe, S.W., Hirsch, B.A., Leinwand, L.A., Largaespada, D.A., 2006. Gene Mutations and Genomic Rearrangements in the Mouse as a Result of Transposon Mobilization from Chromosomal Concatemers. PLoS Genet. 2(9), e156.

Gidekel, S., Pizov, G., Bergman, Y., Pikarsky, E., 2003. Oct-3/4 is a dose-dependent oncogenic fate determinant. Cancer Cell. 4, 361–370.

Ginsberg, H., 2013. The Adenoviruses. Springer science and business media.

Goh, P.A., Caxaria, S., Casper, C., Rosales, C., Warner, T.T., Coffey, P.J., Nathwani, A.C., 2013. A Systematic Evaluation of Integration Free Reprogramming Methods for Deriving Clinically Relevant Patient Specific Induced Pluripotent Stem (iPS) Cells. PLoS One. 8, e81622.

Gómez-Aguado, I., Rodríguez-Castejón, J., Vicente-Pascual, M., Rodríguez-Gascón, A., Solinís, M.Á., Del Pozo-Rodríguez, A., 2020. Nanomedicines to Deliver mRNA: State of the Art and Future Perspectives. Nanomater. 10, 364.

González, F., Boué, S., Belmonte, J.C.I., 2011. Methods for making induced pluripotent stem

cells: reprogramming à la carte. Nat. Rev. Genet. 12(4), 231–242.

Gonzalez, F., Monasterio, M.B., Tiscornia, G., Pulido, N.M., Vassena, R., Morera, L.B., Piza, I.R., Belmonte, J.C.I., 2009. Generation of mouse-induced pluripotent stem cells by transient expression of a single nonviral polycistronic vector. Proc Natl Acad Sci U S A. 106, 8918–8922.

Gore, A., Li, Z., Fung, H.L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E., Lee, J.H., Loh, Y.H., Manos, P.D., Montserrat, N., Panopoulos, A.D., Ruiz, S., Wilbert, M.L., Yu, J., Kirkness, E.F., Belmonte, J.C.I., Rossi, D.J., Thomson, J.A., Eggan, K., Daley, G.Q., Goldstein, L.S.B., Zhang, K., 2011. Somatic coding mutations in human induced pluripotent stem cells. Nature. 471(7336), 63–67.

Grabundzija, I., Irgang, M., Mátés, L., Belay, E., Matrai, J., Gogol-Döring, A., Kawakami, K., Chen, W., Ruiz, P., Chuah, M.K.L., Vandendriessche, T., Izsvák, Z., Ivics, Z., 2010. Comparative Analysis of Transposable Element Vector Systems in Human Cells. Mol Ther. 18, 1200–1209.

Grabundzija, I., Wang, J., Sebe, A., Erdei, Z., Kajdi, R., Devaraj, A., Steinemann, D., Szuhai, K., Stein, U., Cantz, T., Schambach, A., Baum, C., Izsvák, Z., Sarkadi, B., Ivics, Z., 2013. Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. Nucleic Acids Res. 41, 1829–1847.

Grada, A., Otero-Vinas, M., Prieto-Castrillo, F., Obagi, Z., Falanga, V., 2017. Research Techniques Made Simple: Analysis of Collective Cell Migration Using the Wound Healing Assay. J. Invest. Dermatol. 137(2), e11-e16.

Greenfield, N.J., 2006. Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. 1, 2876–2890.

Gregory, S.M., Nazir, S.A., Metcalf, J.P., 2011. Implications of the innate immune response to adenovirus and adenoviral vectors.Future Virol. 6(3), 357-374.

Guenin, S., Mouallif, M., Deplus, R., Lampe, X., Krusy, N., Calonne, E., Delbecque, K., Kridelka, F., Fuks, F., Ennaji, M.M., Delvenne, P., 2012. Aberrant Promoter Methylation and Expression of UTF1 during Cervical Carcinogenesis. PLoS One. 7, e42704.

Gurdon, J.B., 1962. The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. J Embryol Exp Morphol. 622-640

Haase, A., Göhring, G., Martin, U., 2017. Generation of non-transgenic iPS cells from human cord blood CD34+ cells under animal component-free conditions. Stem Cell Res. 21, 71–73.

Hamada, M., Horiuchi, Y., Matsuo-Takasaki, Y., Imoto, M., 2012. Generation of Induced Pluripotent Stem Cells from Human Nasal Epithelial Cells Using a Sendai Virus Vector. PLoS One. 7, 42855.

Haridhasapavalan, K.K., Das, N.J., Thummer, R.P., 2022a. Generation of a transducible version of a bioactive recombinant human TBX5 transcription factor from *E. Coli*. Curr.

Res. Biotechnol. 4, 66–77.

Haridhasapavalan, K.K., Raina, K., Dey, C., Adhikari, P., Thummer, R.P., 2020. An Insight into Reprogramming Barriers to iPSC Generation. Stem Cells Rev Rep. 16, 56–81.

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(6), 2363–2376.

Haridhasapavalan, K.K., Sundaravadivelu, P.K., Joshi, N., Das, N.J., Mohapatra, A., Voorkara, U., Kaveeshwar, V., Thummer, R.P., 2022b. Generation of a recombinant version of a biologically active cell-permeant human HAND2 transcription factor from *E. coli*. Sci. Rep. 12(1), 1–16.

Haridhasapavalan, K.K., Sundaravadivelu, P.K., Thummer, R.P., 2020. Codon Optimization, Cloning, Expression, Purification, and Secondary Structure Determination of Human ETS2 Transcription Factor. Mol. Biotechnol. 62, 485–494.

Haridhasapavalan, K.K., Sundaravadivelu, P.K., Bhattacharyya, S., Ranjan, S.H., Raina, K., Thummer, R.P., 2021. Generation of cell-permeant recombinant human transcription factor GATA4 from *E. coli*. Bioprocess Biosyst. Eng. 44(6), 1131-1146.

Hartman, Z.C., Appledorn, D.M., Amalfitano, A., 2008. Adenovirus vector induced innate immune responses: Impact upon efficacy and toxicity in gene therapy and vaccine applications. Virus Res. 132, 1–14.

Harui, A., Suzuki, S., Kochanek, S., Mitani, K., 1999. Frequency and Stability of Chromosomal Integration of Adenovirus Vectors. J Virol. 73, 6141–6146.

Heng, B.C., Heinimann, K., Miny, P., Iezzi, G., Glatz, K., Scherberich, A., Zulewski, H., Fussenegger, M., 2013a. mRNA transfection-based, feeder-free, induced pluripotent stem cells derived from adipose tissue of a 50-year-old patient. Metab Eng. 18, 9–24.

Herr, W., Cleary, M.A., 1995. The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. Gen. Dev. 9914, 1679-1693.

He, T.C., Zhou, S., da Costa, L.T., Yu, J., Kinzler, K.W., Vogelstein, B., 1998. A simplified system for generating recombinant adenoviruses. Proc Natl Acad Sci U S A. 95, 2509–2514.

Hirsch, M.L., Agbandje-Mckenna, M., Samulski, R.J., 2010. Little vector, big gene transduction: Fragmented genome reassembly of adeno-associated virus. Molecular Therapy 18, 6–8.

Hirsch, M.L., Wolf, S.J., Samulski, R.J., 2016. Delivering transgenic DNA exceeding the carrying capacity of AAV vectors. Methods Mol Biol. 1382, 21–39.

Hochedlinger, K., Yamada, Y., Beard, C., Jaenisch, R., 2005. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. Cell. 121(3), 465-477.

Hockemeyer, D., Soldner, F., Cook, E.G., Gao, Q., Mitalipova, M., Jaenisch, R., 2008. A Drug-

Inducible System for Direct Reprogramming of Human Somatic Cells to Pluripotency. Cell Stem Cell. 3, 346–353.

Hornstein, B.D., Roman, D., Aré valo-Soliz, L.M., Engevik, M.A., Zechiedrich, L., 2016. Effects of Circular DNA Length on Transfection Efficiency by Electroporation into HeLa Cells. PloS one. 11(12), e0167537.

Hosoya, N., Miura, T., Kawana-Tachikawa, A., Koibuchi, T., Shioda, T., Odawara, T., Nakamura, T., Kitamura, Y., Kano, M., Kato, A., Hasegawa, M., Nagai, Y., Iwamoto, A., 2008. Comparison between Sendai virus and adenovirus vectors to transduce HIV-1 genes into human dendritic cells. J Med Virol. 80, 373–382.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K., Ge, J., Xu, J., Zhang, Q., Zhao, Y., Deng, H., 2013. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science. 341, 651–654.

Howarth, J.L., Lee, Y.B., Uney, J.B., 2009. Using viral vectors as gene transfer tools (Cell Biology and Toxicology Special Issue: ETCS-UK 1 day meeting on genetic manipulation of cells). Cell Biol. Toxicol. 26(1), 1–20.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., Melton, D.A., 2008. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nature Biotechnology 2008 26(7), 795–797.

Hu, J.H., Wang, F., Liu, C.Z., 2015. Development of an efficient process intensification strategy for enhancing Pfu DNA polymerase production in recombinant Escherichia coli. Bioprocess Biosyst. Eng. 38, 651–659.

Hu, K., 2014a. All roads lead to induced pluripotent stem cells: The technologies of iPSC generation. Stem Cells Dev. 23(12), 1285-1300.

Hu, K., Yu, J., Suknuntha, K., Tian, S., Montgomery, K., Choi, K.-D., Stewart, R., Thomson, J.A., Slukvin, I.I., 2011. Efficient generation of transgene-free induced pluripotent stem cells from normal and neoplastic bone marrow and cord blood mononuclear cells.Blood, Am. J. Hematol. 117(140), 109-119.

Hu, P.F., Guan, W.J., Li, X.C., Ma, Y.H., 2012. Construction of recombinant proteins for reprogramming of endangered Luxi cattle fibroblast cells. Mol. Biol. Rep. 39, 7175–7182.

Hu, S., Wilson, K.D., Ghosh, Z., Han, L., Wang, Y., Lan, F., Ransohoff, K.J., Burridge, P., Wu, J.C., 2013. MicroRNA-302 Increases Reprogramming Efficiency via Repression of NR2F2. Stem Cells. 31, 259–268

Huang, C.-J., Lin, H., Yang, X., 2012. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. J. Ind. Microbiol. Biotechnol. 39, 383–399.

Hussein, S.M., Batada, N.N., Vuoristo, S., Ching, R.W., Autio, R., Narvä, E., Ng, S., Sourour, M., Hämälä, R., Olsson, C., Lundin, K., Mikkola, M., Trokovic, R., Peitz, M., Brüstle, O., Bazett-Jones, D.P., Alitalo, K., Lahesmaa, R., Nagy, A., Otonkoski, T., 2011. Copy number variation and selection during reprogramming to pluripotency. Nature. 471, 58–62.

Igawa, K., Kokubu, C., Yusa, K., Horie, K., Yoshimura, Y., Yamauchi, K., Suemori, H., Yokozeki, H., Toyoda, M., Kiyokawa, N., Okita, H., Miyagawa, Y., Akutsu, H., Umezawa, A., Katayama, I., Takeda, J., 2014. Removal of Reprogramming Transgenes Improves the Tissue Reconstitution Potential of Keratinocytes Generated From Human Induced Pluripotent Stem Cells. Stem Cells Transl Med. 3, 992–1001.

Ivics, Z., Izsvák, Z., 2010. The expanding universe of transposon technologies for gene and cell engineering. Mobile DNA. 1(1), 1–15.

Iyer AA., Groves AK., 2021. Transcription factor reprogramming in the inner ear: turning on cell fate switches to regenerate sensory hair cells. Front Cell Neurosci. 29(15), 660748

Jackson-Grusby, L., Laird, P.W., Magge, S.N., Moeller, B.J., Jaenisch, R., 1997. Mutagenicity of 5-aza-2′-deoxycytidine is mediated by the mammalian DNA methyltransferase. Proc Natl Acad Sci U S A. 94, 4681–4685.

Jadeja, D., Dogra, N., Arya, S., Singh, Gurpreet, Singh, Gurdyal, Kaur, J., 2016. Characterization of LipN (Rv2970c) of *Mycobacterium Tuberculosis* H37Rv and its Probable Role in Xenobiotic Degradation. J. Cell. Biochem. 117, 390–401.

Janani, M.K., Jaichandran, V., Madhavan, H.N.R., Vijaya, L., George, R.J., Ambastha, P.S., Manu, A., Malathi, J., 2018. Effect of Lignocaine Concentration on Human Fibroblasts Growth in Eyes Undergoing Trabeculectomy: An *in vitro* Stu Biomed Hub. 3(3), 1-10.

Jetten, A.M., 2018. GLIS1–3 transcription factors: critical roles in the regulation of multiple physiological processes and diseases. CMLS. 75(19), 3473-3494.

Ji, J., Zheng, P.S., 2010. Expression of Sox2 in human cervical carcinogenesis. Hum. Pathol. 41(10), 1438–1447.

Jia, F., Wilson, K.D., Sun, N., Gupta, D.M., Huang, M., Li, Z., Panetta, N.J., Chen, Z.Y., Robbins, R.C., Kay, M.A., Longaker, M.T., Wu, J.C., 2010. A nonviral minicircle vector for deriving human iPS cells. Nat Methods. 7, 197–199.

Jiang, G., di Bernardo, J., Maiden, M.M., Villa-Diaz, L.G., Mabrouk, O.S., Krebsbach, P.H., O'shea, K.S., Kunisaki, S.M., 2014. Human Transgene-Free Amniotic-Fluid-Derived Induced Pluripotent Stem Cells for Autologous Cell Therapy. Stem Cells Dev. 23, 2613–2625.

Jin, H.Y., Gonzalez-Martin, A., Miletic, A. v., Lai, M., Knight, S., Sabouri-Ghomi, M., Head, S.R., Macauley, M.S., Rickert, R.C., Xiao, C., 2015. Transfection of microRNA mimics should be used with caution. Front Genet. 6, 340.

Judson, R.L., Babiarz, J.E., Venere, M., Blelloch, R., 2009. Embryonic stem cell–specific microRNAs promote induced pluripotency. Nat Biotechnol. 27(5), 459–461.

Judson, R.L., Greve, T.S., Parchem, R.J., Blelloch, R., 2013. MicroRNA-based discovery of barriers to dedifferentiation of fibroblasts to pluripotent stem cells. Nat Struct Mol Biol. 20(10), 1227–1235.

Kadari, A., Lu, M., Li, M., Sekaran, T., Thummer, R.P., Guyette, N., Chu, V., Edenhofer, F., 2014a. Excision of viral reprogramming cassettes by Cre protein transduction enables

rapid, robust and efficient derivation of transgene-free human induced pluripotent stem cells, Stem Cell Res. Ther. 5(2), 1-10.

Kaitsuka, T., Tomizawa, K., 2015. Cell-Penetrating Peptide as a Means of Directing the Differentiation of Induced-Pluripotent Stem Cells. Int. J. Mol. Sci. 16(11), 26667-26676.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., Woltjen, K., 2009. Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature. 458, 771–775.

Kang, H.S., Zeruth, G., Lichti-Kaiser, K., Vasanth, S., Yin, Z., Kim, Y.-S., Jetten, A.M., 2010. Gli-similar (Glis) Krüppel-like zinc finger proteins: insights into their physiological functions and critical roles in neonatal diabetes and cystic renal disease.Histol. Histopathol. 25(110), 1481.

Kang, X., Yu, Q., Huang, Y., Song, B., Chen, Y., Gao, X., He, W., Sun, X., Fan, Y., 2015. Effects of integrating and non-integrating reprogramming methods on copy number variation and genomic stability of human induced pluripotent stem cells. PLoS One. 10(7), e0131128.

Karow, M., Chavez, C.L., Farruggio, A.P., Geisinger, J.M., Keravala, A., Jung, W.E., Lan, F., Wu, J.C., Yanru, C.T., Calos, M.P., 2011. Site-Specific Recombinase Strategy to Create Induced Pluripotent Stem Cells Efficiently with Plasmid DNA. Stem Cells. 29, 1696–1704.

Karpievitch, Y. V, Polpitiya, A.D., Anderson, G.A., Smith, R.D., Dabney, A.R., 2010. Liquid chromatography mass spectrometry-based proteomics: biological and technological aspects. Ann Appl Stat. 4, 1797–1823.

Kasowitz, S.D., Luo, M., Ma, J., Leu, N.A., Wang, P.J., 2017. Embryonic lethality and defective male germ cell development in mice lacking UTF1. Sci Rep. 7(1), 1–10.

Kaur, Jashandeep, Kumar, A., Kaur, Jagdeep, 2018. Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. Internaional J. Biol. Macromol. 106, 803–822.

Kelly, S.M., Jess, T.J., Price, N.C., 2005. How to study proteins by circular dichroism. Biochim. Biophys. Acta - Proteins Proteom. 1751, 119–139.

Khan, M., Narayanan, K., Lu, H., Choo, Y., Du, C., Wiradharma, N., Yang, Y.-Y., Wan, A.C.A., 2013. Delivery of reprogramming factors into fibroblasts for generation of non-genetic induced pluripotent stem cells using a cationic bolaamphiphile as a non-viral vector. Biomaterials. 34, 5336–5343.

Khow, O., Suntrarachun, S., 2012. Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pac. J. Trop. Biomed. 2, 159–162.

Kim, D., Kim, C.-H., Moon, J.-I., Chung, Y.-G., Chang, M.-Y., Han, B.-S., Ko, S., Yang, E., Cha, K.Y., Lanza, R., 2009. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell Stem Cell. 4, 472.

Kim, S. il, Oceguera-Yanez, F., Sakurai, C., Nakagawa, M., Yamanaka, S., Woltjen, K., 2016. Inducible transgene expression in human iPS cells using versatile all-in-one piggybac transposons. Methods mol. biol. 1357, 111–131.

Kim, S.Y., Kim, M.J., Jung, H., Kim, W.K., Kwon, S.O., Son, M.J., Jang, I.S., Choi, J.S., Park, S.G., Park, B.C., Han, Y.M., Lee, S.C., Cho, Y.S., Bae, K.H., 2011. Comparative Proteomic Analysis of Human Somatic Cells, Induced Pluripotent Stem Cells, and Embryonic Stem Cells., Stem Cells Dev. 21(8), 1271-1286.

Kim, Y.S., Lewandoski, M., Perantoni, A.O., Kurebayashi, S., Nakanishi, G., Jetten, A.M., 2002. Identification of Glis1, a novel Gli-related, Krüppel-like zinc finger protein containing transactivation and repressor functions. JBC. 277, 30901–30913.

Kleman, G.L., Strohl, W.R., 1994. Acetate Metabolism by *Escherichia coli* in High-Cell-Density Fermentation, Appl. Environ. Microbiol. 60(11), 3952-2958.

Kogut, I., McCarthy, S.M., Pavlova, M., Astling, D.P., Chen, X., Jakimenko, A., Jones, K.L., Getahun, A., Cambier, J.C., Pasmooij, A.M.G., Jonkman, M.F., Roop, D.R., Bilousova, G., 2018. High-efficiency RNA-based reprogramming of human primary fibroblasts. Nat Commun. 9(1), 1–15.

Konno, M., Masui, S., Hamazaki, T.S., Okochi, H., 2011. Intracellular reactivation of transcription factors fused with protein transduction domain. J Biotechnol. 154, 298–303.

Kooistra, S.M., Thummer, R.P., Eggen, B.J.L., 2009. Characterization of human UTF1, a chromatin-associated protein with repressor activity expressed in pluripotent cells. Stem Cell Res. 2, 211–218.

Kowalski, P.S., Rudra, A., Miao, L., Anderson, D.G., 2019. Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery. Mol. Ther. 27, 710–728.

Kristensen, D.M., Nielsen, J.E., Skakkebaek, N.E., Graem, N., Jacobsen, G.K., Rajpert-De Meyts, E., Leffers, H., 2008. Presumed pluripotency markers UTF-1 and REX-1 are expressed in human adult testes and germ cell neoplasms.Hum. Reprod. 23(4), 775-782.

Kumar, D., Talluri, T.R., Anand, T., Kues, W.A., 2015. Transposon-based reprogramming to induced pluripotency. Histol. histopathol. cell. mol. biol. 30(12), 1397-1409.

Kwon, H., Kim, M., Seo, Y., Moon, Y.S., Lee, H.J., Lee, K., Lee, H., 2018. Emergence of synthetic mRNA: *In vitro* synthesis of mRNA and its applications in regenerative medicine. Biomaterials. 156, 172–193.

Lamb, R., 2001. Paramyxoviridae : the viruses and their replication. Fields Virology.

Laurent, L.C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., Lynch, C., Harness, J. v., Lee, S., Barrero, M.J., Ku, S., Martynova, M., Semechkin, R., Galat, V., Gottesfeld, J., Belmonte, J.C.I., Murry, C., Keirstead, H.S., Park, H.S., Schmidt, U., Laslett, A.L., Muller, F.J., Nievergelt, C.M., Shamir, R., Loring, J.F., 2011. Dynamic Changes in the Copy Number of Pluripotency and Cell Proliferation Genes in Human ESCs and iPSCs during Reprogramming and Time in Culture. Cell Stem Cell. 8, 106–118.

Lee, C.S., Bishop, E.S., Zhang, R., Yu, X., Farina, E.M., Yan, S., Zhao, C., Zeng, Z., Shu, Y.,

Wu, X., Lei, J., Li, Y., Zhang, W., Yang, C., Wu, K., Wu, Y., Ho, S., Athiviraham, A., Lee, M.J., Wolf, J.M., Reid, R.R., He, T.C., 2017. Adenovirus-mediated gene delivery: Potential applications for gene and cell-based therapies in the new era of personalized medicine. Genes Dis. 4, 43–63.

Lee, J., Sayed, N., Hunter, A., Au, K.F., Wong, W.H., Mocarski, E.S., Pera, R.R., Yakubov, E., Cooke, J.P., 2012. Activation of innate immunity is required for efficient nuclear reprogramming. Cell. 151, 547–558.

Lee, M.R., Prasain, N., Chae, H.D., Kim, Y.J., Mantel, C., Yoder, M.C., Broxmeyer, H.E., 2013. Epigenetic Regulation of Nanog by MiR-302 Cluster-MBD2 Completes Induced Pluripotent Stem Cell Reprogramming. Stem Cells. 31, 666–681.

Lee, S.-Y., Noh, H. Bin, Kim, H.-T., Lee, K.-I., Hwang, D.-Y., 2017. Glis family proteins are differentially implicated in the cellular reprogramming of human somatic cells. Oncotarget. 8(44), 77041.

Lee, Y.J., Ramakrishna, S., Chauhan, H., Park, W.S., Hong, S.H., Kim, K.S., 2016. Dissecting microRNA-mediated regulation of stemness, reprogramming, and pluripotency. Cell Regen. 5(1), 1–10.

Leonardo, T.R., Schultheisz, H.L., Loring, J.F., Laurent, L.C., 2012. The functions of microRNAs in pluripotency and reprogramming. Nat Cell Biol. 14(11), 1114–1121.

Liao, B., Bao, X., Liu, L., Feng, S., Zovoilis, A., Liu, W., Xue, Y., Cai, J., Guo, X., Qin, B., Zhang, R., Wu, J., Lai, L., Teng, M., Niu, L., Zhang, B., Esteban, M.A., Pei, D., 2011. MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. JBC. 286, 17359–17364.

Li, L., Chen, K., Wang, T., Wu, Y., Xing, G., Chen, M., Hao, Z., Zhang, C., Zhang, J., Ma, B., Liu, Zihuang, Yuan, H., Liu, Zijian, Long, Q., Zhou, Y., Qi, J., Zhao, D., Gao, M., Pei, D., Nie, J., Ye, D., Pan, G., Liu, X., 2020. Glis1 facilitates induction of pluripotency via an epigenome–metabolome–epigenome signalling cascade. Nat. Metab. 2, 882–892.

Li, S., Wu, S.P., Whitmore, M., Loeffert, E.J., Wang, L., Watkins, S.C., Pitt, B.R., Huang, L., 1999. Effect of immune response on gene transfer to the lung via systemic administration of cationic lipidic vectors. Am J Physiol Lung Cell Mol Physiol. 276.

Li, X., Burnight, E.R., Cooney, A.L., Malani, N., Brady, T., Sander, J.D., Staber, J., Wheelan, S.J., Joung, J.K., McCray, P.B., Bushman, F.D., Sinn, P.L., Craig, N.L., 2013. PiggyBac transposase tools for genome engineering. Proc Natl Acad Sci U S A. 110, 2279–2287.

Li, X., Zhang, P., Wei, C., Zhang, Y., 2014. Generation of pluripotent stem cells via protein transduction. Int. J. Dev. Biol. 58(1), 21-27

Li, Xuefei, Xu, Y., Chen, Y., Chen, S., Jia, X., Sun, T., Liu, Y., Li, Xiru, Xiang, R., Li, N., 2013. SOX2 promotes tumor metastasis by stimulating epithelial-to-mesenchymal transition via regulation of WNT/β-catenin signal network. Cancer Lett. 336, 379–389.

Li, Z., Yang, C.S., Nakashima, K., Rana, T.M., 2011. Small RNA-mediated regulation of iPS cell generation. EMBO J. 30, 823–834.

Liu, B., Kong, Q., Zhang, D., Yan, L., 2018. Codon optimization significantly enhanced the expression of human 37-kDa iLRP in *Escherichia coli*. 3 Biotech. 8(4), 1-7.

Liu, J., Han, Q., Peng, T., Peng, M., Wei, B., Li, D., Wang, X., Yu, S., Yang, J., Cao, S., Huang, K., Hutchins, A.P., Liu, H., Kuang, J., Zhou, Z., Chen, Jing, Wu, Haoyu, Guo, L., Chen, Yongqiang, Chen, You, Li, X., Wu, Hongling, Liao, B., He, W., Song, H., Yao, H., Pan, G., Chen, Jiekai, Pei, D., 2015. The oncogene c-Jun impedes somatic cell reprogramming. Nat. Cell Biol. 17, 856–867.

Liu, K., Xie, F., Gao, A., Zhang, R., Zhang, L., Xiao, Z., Hu, Q., Huang, W., Huang, Q., Lin, B., Zhu, J., Wang, H., Que, J., Lan, X., 2017. SOX2 regulates multiple malignant processes of breast cancer development through the SOX2/miR-181a-5p, miR-30e-5p/TUSC3 axis. Mol. Cancer. 16, 1–14.

Long, Y., Wang, M., Gu, H., Xie, X., 2015. Bromodeoxyuridine promotes full-chemical induction of mouse pluripotent stem cells. Cell Res. 25(10), 1171–1174.

Lu, D., Davis, M.P.A., Abreu-Goodger, C., Wang, W., Campos, L.S., Siede, J., Vigorito, E., Skarnes, W.C., Dunham, I., Enright, A.J., Liu, P., 2012. MiR-25 Regulates Wwp2 and Fbxw7 and Promotes Reprogramming of Mouse Fibroblast Cells to iPSCs. PLoS One. 7, e40938.

Luke, G.A., Escuin, H., Felipe, P. de, Ryan, M.D., 2013. 2A to the Fore – Research, Technology and Applications.Biotechnol. Genet. Eng. Rev. 26(1), 223-260.

Lundberg, I. V., Edin, S., Eklöf, V., Öberg, Å., Palmqvist, R., Wikberg, M.L., 2016. SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer. BMC Cancer. 16, 1–11.

Luo, R., Zhang, X., Wang, L., Zhang, L., Li, G., Zheng, Z., 2021. GLIS1, a potential candidate gene affect fat deposition in sheep tail. Mol. Biol. Rep. 48, 4925–4931.

Ma, X., Kong, L., Zhu, S., 2017. Reprogramming cell fates by small molecules. Protein Cell 8, 328–348.

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.

Maekawa, M., Yamanaka, S., 2011. Glis1, a unique pro-reprogramming factor, may facilitate clinical applications of iPSC technology. Cell Cycle. 10(21), 3613-3614.

Maertens, B., Spriestersbach, A., von Groll, U., Roth, U., Kubicek, J., Gerrits, M., Graf, M., Liss, M., Daubert, D., Wagner, R., 2010. Gene optimization mechanisms: a multi-gene study reveals a high success rate of full-length human proteins expressed in *Escherichia coli*. Protein Sci. 19, 1312–1326.

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., Hochedlinger, K., 2008. A High-Efficiency System for the Generation and Study of Human Induced Pluripotent Stem Cells. Cell Stem Cell. 3, 340–345.

Maherali, N., Hochedlinger, K., 2008. Guidelines and techniques for the generation of induced

pluripotent stem cells. Cell Stem Cell. 3, 595–605.

Malik, N., Rao, M.S., 2013. A review of the methods for human iPSC derivation. Methods mol biol. 997, 23–33.

Mandal, P., protocols, D.R.-N., 2013, undefined, n.d. Reprogramming human fibroblasts to pluripotency using modified mRNA. Nat Protoc. 8(3), 568-582.

Marchion, D.C., Bicaku, E., Daud, A.I., Sullivan, D.M., Munster, P.N., 2005. Valproic Acid Alters Chromatin Structure by Regulation of Chromatin Modulation Proteins. Cancer Res. 65, 3815–3822.

Martinez, N.J., Gregory, R.I., 2010. MicroRNA-induced pluripotent stem cells. Cell Stem Cell. 7, 31.

Martínez-Gómez, K., Flores, N., Castañeda, H.M., Martínez-Batallar, G., Hernández-Chávez, G., Ramírez, O.T., Gosset, G., Encarnación, S., Bolivar, F., 2012. New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. Microb.Cell Factories. 11(1), 1-21

Martin, G.R., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A. 78, 7634–7638.

Martins-Taylor, K., Nisler, B.S., Taapken, S.M., Compton, T., Crandall, L., Montgomery, K.D., Lalande, M., Xu, R.H., 2011. Recurrent copy number variations in human induced pluripotent stem cells. Nat Biotechnol. 29:6 29, 488–491

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol. 9, 625–635.

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., Porro, D., 2012. Recombinant protein production in yeasts. Methods Mol. Biol. 824, 329–358.

Mátés, L., Chuah, M.K.L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., Vandendriessche, T., Ivics, Z., Izsvák, Z., 2009. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nature Genet. 41:6 41, 753–761.

Maurisse, R., de Semir, D., Emamekhoo, H., Bedayat, B., Abdolmohammadi, A., Parsi, H., Gruenert, D.C., 2010. Comparative transfection of DNA into primary and transformed mammalian cells from different lineages. BMC Biotechnol. 10, 1–9.

Maurizi, G., Verma, N., Gadi, A., Mansukhani, A., Basilico, C., 2018. Sox2 is required for tumor development and cancer cells proliferation in osteosarcoma. Oncogene. 37, 4626.

Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J.C., Yakir, B., Clark, A.T., Plath, K., Lowry, W.E., Benvenisty, N., 2010. Identification and Classification of Chromosomal

Aberrations in Human Induced Pluripotent Stem Cells. Cell Stem Cell. 7, 521–531.

McCarroll, L., King, L.A., 1997. Stable insect cell cultures for recombinant protein production. Curr. Opin. Biotechnol. 8, 590–594.

McLenachan, S., Sarsero, J.P., Ioannou, P.A., 2007. Flow-cytometric analysis of mouse embryonic stem cell lipofection using small and large DNA constructs. Genomics. 89, 708–720.

Mehravar, M., Ghaemimanesh, F., Poursani, E.M., 2021. An Overview on the Complexity of OCT4: at the Level of DNA, RNA and Protein. Stem Cell Rev Rep. 17(4), 1121-1136.

Melone, M.A.B., Giulano, M., Squillaro, T., Alessio, N., Casale, F., Mattioli, E., Cipollaro, M., Giordano, A., Galderisi, U., 2009. Genes involved in regulation of stem cell properties: studies on their expression in a small cohort of neuroblastoma patients.Cancer Biol. Ther. 8(13), 1300-1306.

Menon, S., Shailendra, S., Renda, A., Longaker, M., Quarto, N., 2016. An overview of direct somatic reprogramming: the ins and outs of iPSCs. Int J Mol Sci. 17, 141.

Micsonai, A., Wien, F., Bulyáki, É., Kun, J., Moussong, É., Lee, Y.-H., Goto, Y., Réfrégiers, M., Kardos, J., 2018. BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. Nucleic Acids Res. 46, W315–W322.

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

Mirakhori, F., Zeynali, B., Rassouli, H., Salekdeh, G.H., Baharvand, H., 2015. Direct conversion of human fibroblasts into dopaminergic neural progenitor-like cells using TAT-mediated protein transduction of recombinant factors. Biochem. Biophys. Res. Commun. 459, 655–661.

Mitchell, R., Szabo, E., Shapovalova, Z., Aslostovar, L., Makondo, K., Bhatia, M., 2014. Molecular evidence for OCT4-induced plasticity in adult human fibroblasts required for direct cell fate conversion to lineage specific progenitors. Stem Cells. 32(8), 2178-2187.

Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D.L., Kano, Y., Nishikawa, S., Tanemura, M., Mimori, K., Tanaka, F., Saito, T., Nishimura, J., Takemasa, I., Mizushima, T., Ikeda, M., Yamamoto, H., Sekimoto, M., Doki, Y., Mori, M., 2011. Reprogramming of Mouse and Human Cells to Pluripotency Using Mature MicroRNAs. Cell Stem Cell. 8, 633–638.

Montserrat, N., Garreta, E., González, F., Gutiérrez, J., Eguizábal, C., Ramos, V., Borrós, S., Belmonte, J.C.I., 2011. Simple generation of human induced pluripotent stem cells using poly-β-amino esters as the non-viral gene delivery system. JBC. 286, 12417–12428.

Morshedi, A., Soroush Noghabi, M., Dröge, P., 2013. Use of UTF1 Genetic Control Elements as iPSC Reporter. Stem Cell Rev Rep. 9, 523–530.

Mouallif, M., Albert, A., Zeddou, M., Ennaji, M.M., Delvenne, P., Guenin, S., 2014.

Expression profile of undifferentiated cell transcription factor 1 in normal and cancerous human epithelia. Int J Exp Pathol. 95, 251–259.

Muenthaisong, S., Ujhelly, O., Polgar, Z., Varga, E., Ivics, Z., Pirity, M.K., Dinnyes, A., 2012. Generation of mouse induced pluripotent stem cells from different genetic backgrounds using Sleeping beauty transposon mediated gene transfer. Exp Cell Res. 318, 2482–2489.

Münst, B., Thier, M.C., Winnemöller, D., Helfen, M., Thummer, R.P., Edenhofer, F., 2016. Nanog induces suppression of senescence through downregulation of p27KIP1 expression. J. Cell Sci. 129, 912–920.

Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T., Yamanaka, S., 2010. Promotion of direct reprogramming by transformation-deficient Myc. Proc Natl Acad Sci U S A. 107, 14152–14157.

Nakashima, M., Tanese, N., Ito, M., Auerbach, W., Bai, C., Furukawa, T., Toyono, T., Akamine, A., Joyner, A.L., 2002. A novel gene, GliH1, with homology to the Gli zinc finger domain not required for mouse development.Mech. Dev. 119(1), 21-34.

Nakatsugawa, M., Takahashi, A., Hirohashi, Y., Torigoe, T., Inoda, S., Murase, M., Asanuma, H., Tamura, Y., Morita, R., Michifuri, Y., Kondo, T., Hasegawa, T., Takahashi, H., Sato, N., 2011. SOX2 is overexpressed in stem-like cells of human lung adenocarcinoma and augments the tumorigenicity. Lab. Investig. 91(12), 1796–1804.

Nanbo, A., Sugden, A., Sugden, B., 2007. The coupling of synthesis and partitioning of EBV's plasmid replicon is revealed in live cells. EMBO J. 26, 4252–4262.

Narayan, G., Agrawal, A., Joshi, N., Gogoi, R., Nagotu, S., Thummer, R.P., 2021a. Protein Production and Purification of a Codon-Optimized Human NGN3 Transcription Factor from E. coli. Protein J. 40, 891–906.

Narayan, G., Sundaravadivelu, P.K., Agrawal, A., Gogoi, R., Nagotu, S., Thummer, R.P., 2021b. Soluble expression, purification, and secondary structure determination of human PDX1 transcription factor. Protein Expr. Purif. 180.

Narsinh, K.H., Jia, F., Robbins, R.C., Kay, M.A., Longaker, M.T., Wu, J.C., 2010. Generation of adult human induced pluripotent stem cells using nonviral minicircle DNA vectors. Nat Protoc. 6:1 6, 78–88.

Nayerossadat, N., Maedeh, T., Ali, P.A., 2012. Viral and nonviral delivery systems for gene delivery. Adv Biomed Res. 1, 27.

Nemes, C., Varga, E., Polgar, Z., Klincumhom, N., Pirity, M.K., Dinnyes, A., 2014. Generation of mouse induced pluripotent stem cells by protein transduction. Tissue Eng Part C Methods. 20, 383–392.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., Smith, A., 1998. Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. Cell. 95, 379–391.

Nicolson, S.C., Li, C., Hirsch, M.L., Setola, V., Samulski, R.J., 2016. Identification and

Validation of Small Molecules That Enhance Recombinant Adeno-associated Virus Transduction following High-Throughput Screens. J Virol. 90, 7019–7031.

Nishimoto, M., Fukushima, A., Okuda, A., Muramatsu, M., 1999. The Gene for the Embryonic Stem Cell Coactivator UTF1 Carries a Regulatory Element Which Selectively Interacts with a Complex Composed of Oct-3/4 and Sox-2. Mol Cell Biol. 19, 5453–5465.

Nishimoto, M., Katano, M., Yamagishi, T., Hishida, T., Kamon, M., Suzuki, A., Hirasaki, M., Nabeshima, Yoko, Nabeshima, Yo ichi, Katsura, Y., Satta, Y., Deakin, J.E., Graves, J.A.M., Kuroki, Y., Ono, R., Ishino, F., Ema, M., Takahashi, S., Kato, H., Okuda, A., 2013. In Vivo Function and Evolution of the Eutherian-Specific Pluripotency Marker UTF1. PLoS One. 8, e68119.

Nishimoto, M., Miyagi, S., Yamagishi, T., Sakaguchi, T., Niwa, H., Muramatsu, M., Okuda, A., 2005. Oct-3/4 Maintains the Proliferative Embryonic Stem Cell State via Specific Binding to a Variant Octamer Sequence in the Regulatory Region of the UTF1 Locus . Mol Cell Biol. 25, 5084–5094.

Nishimura, K., Ohtaka, M., Takada, H., Kurisaki, A., Tran, N.V.K., Tran, Y.T.H., Hisatake, K., Sano, M., Nakanishi, M., 2017. Simple and effective generation of transgene-free induced pluripotent stem cells using an auto-erasable Sendai virus vector responding to microRNA-302. Stem Cell Res. 23, 13–19.

Nishimura, K., Sano, M., Ohtaka, M., Furuta, B., Umemura, Y., Nakajima, Y., Ikehara, Y., Kobayashi, T., Segawa, H., Takayasu, S., Sato, H., Motomura, K., Uchida, E., Kanayasu-Toyoda, T., Asashima, M., Nakauchi, H., Yamaguchi, T., Nakanishia, M., 2011. Development of defective and persistent Sendai virus vector: A unique gene delivery/expression system ideal for cell reprogramming. J. Biol. Chem. 286, 4760–4771.

Niwa, H., Miyazaki, J.I., Smith, A.G., 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet. 24, 372–376.

Okita, K., Ichisaka, T., Yamanaka, S., 2007. Generation of germline-competent induced pluripotent stem cells. Nature. 448, 313–317.

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 (763), 409–412.

Okita, K., Yamakawa, T., Matsumura, Y., Sato, Y., Amano, N., Watanabe, A., Goshima, N., Yamanaka, S., 2013. An Efficient Nonviral Method to Generate Integration-Free Human-Induced Pluripotent Stem Cells from Cord Blood and Peripheral Blood Cells. Stem Cells. 31, 458–466.

Okuda, A., Fukushima, A., Nishimoto, M., Orimo, A., Yamagishi, T., Nabeshima, Y., Kuro-o, M., Nabeshima, Y.I., Boon, K., Keaveney, M., Stunnenberg, H.G., Muramatsu, M., 1998. UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. EMBO J. 17, 2019–2032.

Olaofe, O.A., Burton, S.G., Cowan, D.A., Harrison, S.T.L., 2010. Improving the production

of a thermostable amidase through optimising IPTG induction in a highly dense culture of recombinant *Escherichia coli*. Biochem. Eng. J. 52, 19–24.

O'Malley, J., Woltjen, K., Kaji, K., 2009. New strategies to generate induced pluripotent stem cells. Curr. Opin. Biotechnol. 20, 516–521.

Overton, T.W., 2014. Recombinant protein production in bacterial hosts. Drug Discov. Today 19(5), 590-601

Pan, C., Lu, B., Chen, H., Bishop, C.E., 2010. Reprogramming human fibroblasts using HIV-1 TAT recombinant proteins OCT4, SOX2, KLF4 and c-MYC. Mol. Biol. Rep. 37, 2117–2124.

Pannunzio Carmignotto, G., Rodrigues Azzoni, A., 2019. On the expression of recombinant Cas9 protein in *E. coli* BL21(DE3) and BL21(DE3) Rosetta strains. J.Biotechnol. 306, 62-70.

Pantazis, G., Harter, P.N., Capper, D., Kohlhof, P., Mittelbronn, M., Schittenhelm, J., 2014. The embryonic stem cell factor UTF1 serves as a reliable diagnostic marker for germinomas. Pathol. 46, 225–229.

Papapetrou, E.P., Tomishima, M.J., Chambers, S.M., Mica, Y., Reed, E., Menon, J., Tabar, V., Mo, Q., Studer, L., Sadelain, M., 2009. Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. Proc. Natl. Acad. Sci. U S A. 106, 12759–12764.

Pasha, Z., Haider, H.K., Ashraf, M., 2011. Efficient Non-Viral Reprogramming of Myoblasts to Stemness with a Single Small Molecule to Generate Cardiac Progenitor Cells. PLoS One. 6, e23667.

Paulina Balbás, 2001. Understanding the Art of Producing Protein and Nonprotein Molecules in *Escherichia coli*. Mol. Biotechnol. 19(3), 251-267.

Peitz, M., Münst, B., Thummer, R.P., Helfen, M., Edenhofer, F., 2014. Cell-permeant recombinant Nanog protein promotes pluripotency by inhibiting endodermal specification. Stem Cell Res. 12, 680–689.

Peskova, L., Cerna, K., Oppelt, J., Mraz, M., Barta, T., 2019. Oct4-mediated reprogramming induces embryonic-like microRNA expression signatures in human fibroblasts. Sci. Rep. 9(1), 1-13.

Peterson, B.A., Collins, A.J., Vogelzang, N.J., Bloomfield, C.D., 1981. 5-Azacytidine and renal tubular dysfunction. Blood. 57, 182–185.

Pevny, L.H., Nicolis, S.K., 2010. Sox2 roles in neural stem cells. Int. J. Biochem. Cell Biol. 42, 421–424.

Pfaff, N., Fiedler, J., Holzmann, A., Schambach, A., Moritz, T., Cantz, T., Thum, T., 2011. miRNA screening reveals a new miRNA family stimulating iPS cell generation via regulation of Meox2. EMBO Rep. 12, 1153–1159.

Pfannkuche, K., Fatima, A., Gupta, M.K., Dieterich, R., Hescheler, J., 2010. Initial Colony

Morphology-Based Selection for iPS Cells Derived from Adult Fibroblasts Is Substantially Improved by Temporary UTF1-Based Selection. PLoS One. 5, e9580.

Plews, J.R., Li, J.L., Jones, M., Moore, H.D., Mason, C., Andrews, P.W., Na, J., 2010. Activation of Pluripotency Genes in Human Fibroblast Cells by a Novel mRNA Based Approach. PLoS One. 5, e14397.

Pouya, A., Rassouli, H., Rezaei-Larijani, M., Salekdeh, G.H., Baharvand, H., 2020. SOX2 protein transduction directly converts human fibroblasts into oligodendrocyte-like cells. Biochem. Biophys. Res. Commun. 525, 1–7.

Preskey, D., Allison, T.F., Jones, M., Mamchaoui, K., Unger, C., 2016. Synthetically modified mRNA for efficient and fast human iPS cell generation and direct transdifferentiation to myoblasts. Biochem. Biophys. Res. Commun. 473, 743–751.

Rabhi-Essafi, I., Sadok, A., Khalaf, N., Fathallah, D.M., 2007. A strategy for high-level expression of soluble and functional human interferon a as a GST-fusion protein in *E.coli*. Protien eng. des. sel. 20(5), 201-209.

Radzisheuskaya, A., Silva, J.C.R., 2014. Do all roads lead to Oct4? The emerging concepts of induced pluripotency. Trends Cell Biol. 24(5), 275-284.

Ramos-Mejia, V., Fraga, M.F., Menendez, P., 2012. IPSCs from cancer cells: Challenges and opportunities. Trends Mol. Med. 18 (5), 245-247.

Rao, M.S., Malik, N., 2012. Assessing iPSC reprogramming methods for their suitability in translational medicine. J. Cell Biochem. 113, 3061–3068.

Remans, K., Lebendiker, M., Abreu, C., Maffei, M., Sellathurai, S., May, M.M., Vaněk, O., de Marco, A., 2022. Protein purification strategies must consider downstream applications and individual biological characteristics. Microb. Cell Fact. 21, 1–16.

Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., Robson, P., 2005. Transcriptional regulation of Nanog by OCT4 and SOX2. J. Biol. Chem. 280, 24731–24737.

Rohani, L., Fabian, C., Holland, H., Naaldijk, Y., Dressel, R., Löffler-Wirth, H., Binder, H., Arnold, A., Stolzing, A., 2016. Generation of human induced pluripotent stem cells using non-synthetic mRNA. Stem Cell Res. 16, 662–672.

Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front. Microbiol. 5, 172.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirie, F., Rigby, P.W.J., Staudt, L.M., 1990. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature 345, 686–692.

Ryan, B.J., Henehan, G.T., 2013. Overview of approaches to preventing and avoiding proteolysis during expression and purification of proteins. Curr. Protoc. Protein Sci. 71, 5–25.

Ryu, J., Park, H.H., Park, J.H., Lee, H.J., Rhee, W.J., Park, T.H., 2016. Soluble expression and

stability enhancement of transcription factors using 30Kc19 cell-penetrating protein. Appl. Microbiol. Biotechnol. 100, 3523–3532.

Saha, S., Woodard, L.E., Charron, E.M., Welch, R.C., Rooney, C.M., Wilson, M.H., 2015. Evaluating the potential for undesired genomic effects of the piggyBac transposon system in human cells. Nucleic Acids Res. 43, 1770–1782.

Sahdev, S., Khattar, S.K., Saini, K.S., 2008. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. Mol. Cell. Biochem. 307, 249–264.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., Wrana, J.L., 2010. Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming. Cell Stem Cell. 7, 64–77.

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. 2(1), 1-4.

Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., Villaverde, A., 2016. Recombinant pharmaceuticals from microbial cells a 2015 update, Microb. Cell Fact. 15(1), 1-7.

Santos, L., Ferraz, M.P., Shirosaki, Y., Lopes, M.A., Fernandes, M.H., Osaka, A., Santos, J.D., 2011. Degradation studies and biological behavior on an artificial cornea material. Investig. Ophthalmol. Vis. Sci. 52.

Sarkar, A., Hochedlinger, K., 2013. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell. 12, 15–30.

Schlaeger, T.M., Daheron, L., Brickler, T.R., Entwisle, S., Chan, K., Cianci, A., DeVine, A., Ettenger, A., Fitzgerald, K., Godfrey, M., Gupta, D., McPherson, J., Malwadkar, P., Gupta, M., Bell, B., Doi, A., Jung, N., Li, X., Lynes, M.S., Brookes, E., Cherry, A.B.C., Demirbas, D., Tsankov, A.M., Zon, L.I., Rubin, L.L., Feinberg, A.P., Meissner, A., Cowan, C.A., Daley, G.Q., 2014. A comparison of non-integrating reprogramming methods. Nat. Biotechnol. 33(1), 58–63.

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, 1–9.

Schneider-Schaulies., 2000. Cellular receptors for viruses: links to tropism and pathogenesis. J.Gen.Virol, 81(6), 1413-1429.

Schwarz, B.A., Cetinbas, M., Clement, K., Walsh, R.M., Cheloufi, S., Gu, H., Langkabel, J., Kamiya, A., Schorle, H., Meissner, A., Sadreyev, R.I., Hochedlinger, K., 2018. Prospective Isolation of Poised iPSC Intermediates Reveals Principles of Cellular Reprogramming. Cell Stem Cell 23, 289-305.e5.

Scoville, D.W., Kang, H.S., Jetten, A.M., 2017. GLIS1-3: Emerging roles in reprogramming,

stem and progenitor cell differentiation and maintenance. Stem Cell Investig. 2017 (4).

Seki, T., Yuasa, S., Oda, M., Egashira, T., Yae, K., Kusumoto, D., Nakata, H., Tohyama, S., Hashimoto, H., Kodaira, M., Okada, Y., Seimiya, H., Fusaki, N., Hasegawa, M., Fukuda, K., 2010. Cell Stem Cell Brief Report Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. Stem Cell. 7, 11–14.

Semmler, J., Lehmann, M., Pfannkuche, K., Reppel, M., Hescheler, J., Nguemo, F., 2014. Functional Expression and regulation of hyperpolarization-activated cyclic nucleotide-gated channels (HCN) in mouse iPS cell-derived cardiomyocytes after UTF1-Neo selection. Cell. Physiol. Biochem. 34, 1199–1215.

Seo, B.J., Hong, Y.J., Do, J.T., 2017. Cellular reprogramming using protein and cell-penetrating peptides. Int. J. Mol. Med. 18, 552.

Shao, L., Wu, W.S., 2010. Gene-delivery systems for iPS cell generation. Expert Opin. Biol. Ther. 10, 231–242.

Shimamoto, K., Tanimoto, K., Fukazawa, T., Nakamura, H., Kanai, A., Bono, H., Ono, H., Eguchi, H., Hirohashi, N., 2020. GLIS1, a novel hypoxia-inducible transcription factor, promotes breast cancer cell motility via activation of WNT5A. Carcinogenesis. 41, 1184–1194.

Siegl, G., Bates, R.C., Berns, K.I., Carter, B.J., Kelly, D.C., Kurstak, E., Tattersall, P., 1985. Characteristics and Taxonomy of *Parvoviridae*. Intervirology. 23, 61–73.

Simara, P., Tesarova, L., Rehakova, D., Matula, P., Stejskal, S., Hampl, A., Koutna, I., 2017. DNA double-strand breaks in human induced pluripotent stem cell reprogramming and long-term in vitro culturing. Stem Cell Res. Ther. 8, 1–13.

Singh, V.K., Kalsan, M., Kumar, N., Saini, A., Chandra, R., 2015. Induced pluripotent stem cells: applications in regenerative medicine, disease modeling, and drug discovery. Front. Cell Dev. Biol. 3, 2.

Si-Tayeb, K., Noto, F.K., Sepac, A., Sedlic, F., Bosnjak, Z.J., Lough, J.W., Duncan, S.A., 2010. Generation of human induced pluripotent stem cells by simple transient transfection of plasmid DNA encoding reprogramming factors. BMC Dev. Biol. 10, 1–10.

Slamecka, J., Salimova, L., McClellan, S., van Kelle, M., Kehl, D., Laurini, J., Cinelli, P., Owen, L., Hoerstrup, S.P., Weber, B., 2016. Non-integrating episomal plasmid-based reprogramming of human amniotic fluid stem cells into induced pluripotent stem cells in chemically defined conditions.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., Rogers, D., 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature. 336(6200), 688–690.

Søgaard, K.M., Nørholm, M.H.H., 2016. Side effects of extra tRNA supplied in a typical bacterial protein production scenario. Protein Sci. 25, 2102–2108.

Søkilde, R., Newie, I., Persson, H., Borg, Å., Rovira, C., 2015. Passenger strand loading in overexpression experiments using microRNA mimics. RNA Biol. 12, 787–791.

Somers, A., Jean, J.-C., Sommer, C.A., Omari, A., Ford, C.C., Mills, J.A., Ying, L., Gianotti Sommer, A., Jean, J.M., Smith, B.W., Lafyatis, R., Demierre, M.-F., Weiss, D.J., French, D.L., Gadue, P., Murphy, G.J., Mostoslavsky, G., Kotton, D.N., 2010. Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. Stem Cells. 1728–1740.

Sommer, C.A., Gianotti Sommer, A., Longmire, T.A., Christodoulou, C., Thomas, D.D., Gostissa, M., Alt, F.W., Murphy, G.J., Kotton, D.N., Mostoslavsky, G., 2009. Excision of Reprogramming Transgenes Improves the Differentiation Potential of iPS Cells Generated with a Single Excisable Vector. Stem Cells. 28 (1), 64-74.

Sommer, C.A., Mostoslavsky, G., 2013. The evolving field of induced pluripotency: recent progress and future challenges. J. Cell. Physiol. 228, 267–275.

Song W., Cheng, Y.P., Huang, R., Chen, K., Pan, P.L., Li, J., Yang, Y., Shang, H.F., 2012. GLIS1 : an increased risk factor for late-onset Parkinson's disease in the Han Chinese population. Eur. Neurol. 68, 89–92.

Sørensen, H.P., Mortensen, K.K., 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microb. Cell Fact. 4(1), 1-8.

Specht, E., Miyake-Stoner, S., Mayfield, S., 2010. Micro-algae come of age as a platform for recombinant protein production. Biotechnol. Lett. 32, 1373–1383.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., Plath, K., 2009. Role of the Murine Reprogramming Factors in the Induction of Pluripotency. Cell. 136, 364–377.

Stadtfeld, M., Hochedlinger, K., 2010. Induced pluripotency: history, mechanisms, and applications. Genes Dev. 24, 2239–2263.

Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G., Hochedlinger, K., 2008. Induced pluripotent stem cells generated without viral integration. Science. 322, 945–949.

Steichen, C., Luce, E., Maluenda, J., Tosca, L., Moreno-Gimeno, I., Desterke, C., Dianat, N., Goulinet-Mainot, S., Awan-Toor, S., Burks, D., Marie, Joëlle, Weber, A., Tachdjian, G., Melki, J., Dubart-Kupperschmitt, A., El′, E., Erˆomeerˆ, J.′, Maluenda, E., Jo¨, J., Marie, Jo¨elle, Erard Tachdjian, G.′, 2014. Messenger RNA- Versus Retrovirus-Based Induced Pluripotent Stem Cell Reprogramming Strategies: Analysis of Genomic Integrity. Stem Cells Transl. Med. 3, 686–691.

Stock, K., Nolden, L., Edenhofer, F., Quandel, T., Brüstle, O., 2010. Transcription factor-based modulation of neural stem cell differentiation using direct protein transduction. Cell. Mol. Life Sci. 67, 2439–2449.

Subramanyam, D., Lamouille, S., Judson, R.L., Liu, J.Y., Bucay, N., Derynck, R., Blelloch, R., 2011. Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. Nat. Biotechnol. 29(5), 443–448.

Suh, M.R., Lee, Y., Kim, J.Y., Kim, S.K., Moon, S.H., Lee, J.Y., Cha, K.Y., Chung, H.M., Yoon, H.S., Moon, S.Y., Kim, V.N., Kim, K.S., 2004. Human embryonic stem cells

express a unique set of microRNAs. Dev. Biol. 270, 488–498.

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.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K., Yamanaka, S., 2014. Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nat. Commun. 5(1), 1-9.

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

Takahashi K, Yamanaka S., 2016. A decade of transcription factor-mediated reprogramming to pluripotency. Nat. Rev. Mol. Cell Biol. 17(3), 183-193

Takashina, T., Koyama, T., Nohara, S., Hasegawa, M., Ishiguro, A., Iijima, K., Lu, J., Shimura, M., Okamura, T., Sakuma, T., Yamamoto, T., Ishizaka, Y., 2018. Identification of a cell-penetrating peptide applicable to a protein-based transcription activator-like effector expression system for cell engineering. Biomaterials. 173, 11–21.

Talluri, T.R., Kumar, D., Glage, S., Garrels, W., Ivics, Z., Debowski, K., Behr, R., Kues, W.A., 2014. Non-viral reprogramming of fibroblasts into induced pluripotent stem cells by Sleeping Beauty and piggyBac transposons. Biochem. Biophys. Res. Commun. 450, 581–587.

Tan, H.-K., Toh, C.-X.D., Ma, D., Yang, B., Liu, T.M., Lu, J., Wong, C.-W., Tan, T.-K., Li, H., Syn, C., Tan, E.-L., Lim, B., Lim, Y.-P., Cook, S.A., Loh, Y.-H., 2014. Human finger-prick induced pluripotent stem cells facilitate the development of stem cell banking. Stem Cells Transl. Med. 3, 586–598.

Tan, S.M., Wang, S.T., Hentze, H., Dröge, P., 2007. A UTF1 -based selection system for stable homogeneously pluripotent human embryonic stem cell cultures. Nucleic Acids Res. 35, e118–e118.

Tan, X., Dai, Qingli, Guo, T., Xu, J., Dai, Qingyuan, 2018. Efficient generation of transgene- and feeder-free induced pluripotent stem cells from human dental mesenchymal stem cells and their chemically defined differentiation into cardiomyocytes. Biochem. Biophys. Res. Commun. 495, 2490–2497.

Tavernier, G., Wolfrum, K., Demeester, J., De Smedt, S.C., Adjaye, J. and Rejman, J., 2012. Activation of pluripotency-associated genes in mouse embryonic fibroblasts by non-viral transfection with in vitro-derived mRNAs encoding Oct4, Sox2, Klf4 and cMyc. Biomaterials, 33(2), 412-417.

Tegel, H., Tourle, S., Ottosson, J., Persson, A., 2010. Increased levels of recombinant human proteins with the *Escherichia coli* strain Rosetta(DE3). Protein Expr. Purif. 69(2),159-167.

Terasaki, K., Gen, Y., Iwai, N., Soda, T., Kitaichi, T., Dohi, O., Taketani, H., Seko, Y., Umemura, A., Nishikawa, T., Yamaguchi, K., Moriguchi, M., Konishi, H., Naito, Y., Itoh,

Y., Yasui, K., 2021. SOX2 enhances cell survival and induces resistance to apoptosis under serum starvation conditions through the AKT/GSK-3β signaling pathway in esophageal squamous cell carcinoma. Oncol. Lett. 21(4),1-1.

Thier, M., Münst, B., Edenhofer, F., 2010. Exploring refined conditions for reprogramming cells by recombinant Oct4 protein. Int. J. Dev. Biol, 54 (11-12), 1713-1721.

Thier, M., Münst, B., Mielke, S., Edenhofer, F., 2012. Cellular reprogramming employing recombinant sox2 protein. Stem Cells Int. 2012.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., Jones, J.M., 1998. Embryonic stem cell lines derived from human blastocysts. Science. 282, 1145–1147.

Thummer, R.P., Drenth-Diephuis, L.J., Carney, K.E., Eggen, B.J.L., 2010. Functional Characterization of Single-Nucleotide Polymorphisms in the Human Undifferentiated Embryonic-Cell Transcription Factor 1 Gene. DNA Cell Biol. 29 (5), 241-248.

Thummer, R.P., Drenth-Diephuis, L.J., Eggen, B.J., 2012. Constitutive GFP-UTF1 Expression Interferes with ES and EC Cell Differentiation Stem Cell. J. Stem Cell Res. Ther. 2 (127), 1-7.

Tiemann, U., Sgodda, M., Warlich, E., Ballmaier, M., Schöler, H.R., Schambach, A., Cantz, T., 2011. Optimal reprogramming factor stoichiometry increases colony numbers and affects molecular characteristics of murine induced pluripotent stem cells. Cytometry Part A 79A, 426–435.

Tipanee, J., Chai, Y.C., Driessche, T. vanden, Chuah, M.K., 2017a. Preclinical and clinical advances in transposon-based gene therapy. Biosci. Rep. 37 (6).

Tipanee, J., VandenDriessche, T., Chuah, M.K., 2017b. Transposons: Moving forward from preclinical studies to clinical trials. Hum. Gene Ther. 28, 1087–1104.

Trichas, G., Begbie, J., Srinivas, S., 2008. Use of the viral 2A peptide for bicistronic expression in transgenic mice. BMC Biol. 6, 1–13.

Trokovic, R., Weltner, J., Nishimura, K., Ohtaka, M., Nakanishi, M., Salomaa, V., Jalanko, A., Otonkoski, T., Kyttälä, A., 2014. Advanced feeder-free generation of induced pluripotent stem cells directly from blood cells. Stem Cells Transl. Med. 3, 1402–1409.

Tsukiyama, T., Asano, R., Kawaguchi, T., Kim, N., Yamada, M., Minami, N., Ohinata, Y., Imai, H., 2011. Simple and efficient method for generation of induced pluripotent stem cells using piggyBac transposition of doxycycline-inducible factors and an EOS reporter system. Genes to Cells 16, 815–825.

Tsukiyama, T., Kato-Itoh, M., Nakauchi, H., Ohinata, Y., 2014. A Comprehensive System for Generation and Evaluation of Induced Pluripotent Stem Cells Using piggyBac Transposition. PLoS One 9, e92973.

Tsumoto, K., Ejima, D., Kumagai, I., Arakawa, T., 2003. Practical considerations in refolding proteins from inclusion bodies. Protein Expr. Purif. 28, 1–8.

Tsumoto, K., Ejima, D., Senczuk, A.M., Kita, Y., Arakawa, T., 2007. Effects of salts on protein–surface interactions: applications for column chromatography. J. Pharm. Sci. 96, 1677–1690.

Vadnais, C., Shooshtarizadeh, P., Rajadurai, C. V., Lesurf, R., Hulea, L., Davoudi, S., Cadieux, C., Hallett, M., Park, M., Nepveu, A., 2014. Autocrine Activation of the Wnt/β-Catenin Pathway by CUX1 and GLIS1 in Breast Cancers. Biol. Open 3, 937–946.

Van Craenenbroeck, K., Vanhoenacker, P., Haegeman, G., 2000. Episomal vectors for gene expression in mammalian cells. Eur. J. Biochem. 267, 5665–5678.

van den Boom, V., Kooistra, S.M., Boesjes, M., Geverts, B., Houtsmuller, A.B., Monzen, K., Komuro, I., Essers, J., Drenth-Diephuis, L.J., Eggen, B.J.L., 2007. UTF1 is a chromatin-associated protein involved in ES cell differentiation. J. Cell Biol. 178, 913–924.

Vasina, J.A., Baneyx, F., 1997. Expression of Aggregation-Prone Recombinant Proteins at Low Temperatures: A Comparative Study of the *Escherichia coli* cspA and tac Promoter Systems. Protein Expr. Purif. 9, 211–218.

Walev, I., Bhakdi, S.C., Hofmann, F., Djonder, N., Valeva, A., Aktories, K., Bhakdi, S., 2001. Delivery of proteins into living cells by reversible membrane permeabilization with streptolysin-O. Proc. Natl. Acad. Sci. U S A. 98, 3185–3190.

Walsh, G., Jefferis, R., 2006. Post-translational modifications in the context of therapeutic proteins. Nat. Biotechnol. 24(10), 1241–1252.

Wang, B., Wu, L., Li, D., Liu, Y., Guo, J., Li, C., Yao, Y., Wang, Y., Zhao, G., Wang, X., Fu, M., Liu, H., Cao, S., Wu, C., Yu, S., Zhou, C., Qin, Y., Kuang, J., Ming, J., Chu, S., Yang, X., Zhu, P., Pan, G., Chen, J., Liu, J., Pei, D., 2019. Induction of Pluripotent Stem Cells from Mouse Embryonic Fibroblasts by Jdp2-Jhdm1b-Mkk6-Glis1-Nanog-Essrb-Sall4. Cell Rep. 27, 3473-3485.

Wang, H., Shayakhmetov, D.M., Leege, T., Harkey, M., Li, Q., Papayannopoulou, T., Stamatoyannopolous, G., Lieber, A., 2005. A capsid-modified helper-dependent adenovirus vector containing the β-globin locus control region displays a nonrandom integration pattern and allows stable, erythroid-specific gene expression. J. Virol. 79, 10999–11013.

Wang, L., Chen, Y., Guan, C., Zhao, Z., Li, Q., Yang, J., Mo, J., Wang, B., Wu, W., Yang, X., Song, L., Li, J., 2017. Using low-risk factors to generate non-integrated human induced pluripotent stem cells from urine-derived cells. Stem Cell Res. Ther, 8(1), 1-13.

Wang, T., Chen, K., Zeng, X., Yang, J., Wu, Y., Shi, X., Qin, B., Zeng, L., Esteban, M.A., Pan, G., Pei, D., 2011. The Histone Demethylases Jhdm1a/1b Enhance Somatic Cell Reprogramming in a Vitamin-C-Dependent Manner. Cell Stem Cell 9, 575–587.

Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., Liu, P., 2008. Chromosomal transposition of PiggyBac in mouse embryonic stem cells. Proc. Natl. Acad. Sci. U S A. 105, 9290–9295.

Wang, X., Dai, J., 2010. Concise Review: Isoforms of OCT4 Contribute to the Confusing Diversity in Stem Cell Biology. Stem Cells 28, 885–893.

Wang, Z.-X., Hui-Leng Teh, C., L Kueh, J.L., Lufkin, T., Robson, P., Stanton, L.W., 2007. Oct4 and Sox2 directly regulate expression of another pluripotency transcription factor, Zfp206, in embryonic stem cells. J. Biol. Chem. 282, 12822–12830.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., Daley, G.Q., Brack, A.S., Collins, J.J., Cowan, C., Schlaeger, T.M., Rossi, D.J., 2010. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell. 7, 618–630.

Warren, L., Ni, Y., Wang, J., Guo, X., 2012. Feeder-free derivation of human induced pluripotent stem cells with messenger RNA. Sci. Rep. 2 (1), 1–7.

Warren, L., Wang, J., 2013. Feeder-free reprogramming of human fibroblasts with messenger RNA. Curr. Protoc. Stem Cell Biol. 27(1), 4A-6.

Weill, C.O., Biri, S., Adib, A., Erbacher, P., 2008. A practical approach for intracellular protein delivery. Cytotechnology. 56, 41–48.

Weltner, J., Anisimov, A., Alitalo, K., Otonkoski, T., Trokovic, R., 2012. Induced Pluripotent Stem Cell Clones Reprogrammed via Recombinant Adeno-Associated Virus-Mediated Transduction Contain Integrated Vector Sequences. J Virol. 86, 4463–4467.

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

Wiley, L.A., Burnight, E.R., Deluca, A.P., Anfinson, K.R., Cranston, C.M., Kaalberg, E.E., Penticoff, J.A., Affatigato, L.M., Mullins, R.F., Stone, E.M., Tucker, B.A., 2016. cGMP production of patient-specific iPSCs and photoreceptor precursor cells to treat retinal degenerative blindness. Sci Rep. 6(1), 1–16.

Wilmut I, A. E. Schnleke*, J. McWhlr, A. J. Kind*, K. H. S. Campbell, 1997. 1997_Viable offspring derived from fetal and adult mammalian cells. Nature. 385, 6619.

Wilson, K.D., Venkatasubrahmanyam, S., Jia, F., Sun, N., Butte, A.J., Wu, J.C., 2009b. MicroRNA Profiling of Human-Induced Pluripotent Stem Cells.Stem Cells Dev. 18(5), 749-757.

Wingfield, P.T., 2015. Overview of the Purification of Recombinant Proteins. Curr. Protoc. Protein Sci. 80(1), 6-1.

Woltjen, K., Hämäläinen, R., Kibschull, M., Mileikovsky, M., Nagy, A., 2011. Transgene-free production of pluripotent stem cells using piggyBac transposons. Methods Mol. Biol. 767, 87–103.

Woltjen, K., Kim, S. il, Nagy, A., 2016. The piggyBac transposon as a platform technology for somatic cell reprogramming studies in mouse. Methods Mol Biol. 1357, 1–22.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling,

R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H.K., Nagy, A., 2009. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature. 458(7239), 766–770.

Wu, S.C.Y., Meir, Y.J.J., Coates, C.J., Handler, A.M., Pelczar, P., Moisyadi, S., Kaminski, J.M., 2006. piggyBac is a flexible and highly active transposon as compared to Sleeping Beauty, Tol2, and Mos1 in mammalian cells. Proc Natl Acad Sci U S A. 103, 15008–15013.

Wu, X.L., Zheng, P.S., 2013. Undifferentiated embryonic cell transcription factor-1 (UTF1) inhibits the growth of cervical cancer cells by transactivating p27kip1. Carcinogenesis. 34, 1660–1668.

Wuebben, E.L., Rizzino, A., 2017. The dark side of SOX2: cancer-a comprehensive overview. Oncotarget. 8, 44917–44943.

Wurm, F., Bernard, A., 1999. Large-scale transient expression in mammalian cells for recombinant protein production. Curr. Opin. Biotechnol. 10, 156–159.

Xie, F., Ye, L., Chang, J.C., Beyer, A.I., Wang, J., Muench, M.O., Kan, Y.W., n.d. Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac.Genome Res. 24(9), 1526-1533.

Xu, C., Zhou, Y., Chen, W., 2014. Expression of undifferentiated embryonic cell transcription factor-1 (UTF1) in breast cancers and their matched normal tissues. Cancer Cell Int. 14, 1–4.

Yadav, D.K., Yadav, N., Yadav, S., Haque, S., Tuteja, N., 2016. An insight into fusion technology aiding efficient recombinant protein production for functional proteomics. Arch. Biochem. Biophys. 612, 57–77.

Yakubov, E., Rechavi, G., Rozenblatt, S., Givol, D., 2010. Reprogramming of human fibroblasts to pluripotent stem cells using mRNA of four transcription factors. Biochem Biophys Res Commun. 394, 189–193.

Yamanaka, S., 2009. A Fresh Look at iPS Cells. Cell. 137, 13–17.

Yasuoka, Y., Matsumoto, M., Yagi, K., Okazaki, Y., 2020. Evolutionary History of GLIS Genes Illuminates Their Roles in Cell Reprograming and Ciliogenesis. Mol Biol Evol. 37, 100–109.

Ye, L., Wang, J., Beyer, A.I., Teque, F., Cradick, T.J., Qi, Z., Chang, J.C., Bao, G., Muench, M.O., Yu, J., Levy, J.A., Kan, Y.W., 2014. Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Δ32 mutation confers resistance to HIV infection. Proc Natl Acad Sci U S A. 111, 9591–9596.

Yew, N.S., Zhao, H., Przybylska, M., Wu, I.H., Tousignant, J.D., Scheule, R.K., Cheng, S.H., 2002. CpG-Depleted Plasmid DNA Vectors with Enhanced Safety and Long-Term Gene Expression in Vivo. Molecular Therapy. 5, 731–738.

Yew, N.S., Zhao, H., Wu, I.H., Song, A., Tousignant, J.D., Przybylska, M., Cheng, S.H., 2000. Reduced Inflammatory Response to Plasmid DNA Vectors by Elimination and Inhibition

of Immunostimulatory CpG Motifs. Mol Ther. 1, 255–262.

Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T., Yamanaka, S., 2009. Hypoxia Enhances the Generation of Induced Pluripotent Stem Cells. Cell Stem Cell. 5(3), 237-241.

Yoshihara, M., Hayashizaki, Y., Murakawa, Y., 2016. Genomic Instability of iPSCs: Challenges Towards Their Clinical Applications. Stem Cell Rev Rep. 13(1), 7–16.

Yoshioka, N., Dowdy, S.F., 2017. Enhanced generation of iPSCs from older adult human cells by a synthetic five-factor self-replicative RNA. PLoS One. 12(7) e0182018.

Yoshioka, N., Gros, E., Li, H.R., Kumar, S., Deacon, D.C., Maron, C., Muotri, A.R., Chi, N.C., Fu, X.D., Yu, B.D., Dowdy, S.F., 2013. Efficient generation of human iPSCs by a synthetic self-replicative RNA. Cell Stem Cell. 13, 246–254.

Yoshizaki, M., Hironaka, T., Iwasaki, H., Ban, H., Tokusumi, Y., Iida, A., Nagai, Y., Hasegawa, M., Inoue, M., 2006. Naked Sendai virus vector lacking all of the envelope-related genes: reduced cytopathogenicity and immunogenicity. J Gene Med. 8, 1151–1159.

Young, W., D'Souza, S.L., Lemischka, I.R., Schaniel, C., 2012. Patient-specific Induced Pluripotent Stem Cells as a Platform for Disease Modeling. Drug Discovery and Precision Personalized Medicine. J Stem Cell Res Ther S. 10(2)

Yu, J., Chau, K.F., Vodyanik, M.A., Jiang, J., Jiang, Y., 2011. Efficient Feeder-Free Episomal Reprogramming with Small Molecules. PLoS One. 6, e17557.

Yu, J., Kejin, H., Kim, S.O., Shulan, T., Stewart, R., Slukvin, I.I., Thomson, J.A., 2009. Human induced pluripotent stem cells free of vector and transgene sequences. Science. 324, 797–801.

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., 2007. Induced pluripotent stem cell lines derived from human somatic cells. Science. 318, 1917–1920.

Yusa, K., Rad, R., Takeda, J., Bradley, A., 2009. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. Nature Methods, 6(5), 363–369.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., Craig, N.L., 2011. A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A. 108, 1531–1536.

Zaiss, A.K., Muruve, D.A., 2005. Immune Responses to Adeno-Associated Virus Vectors. Curr Gene Ther. 5, 323–331.

Zeng, Z.L., Lin, X. long, Tan, L.L., Liu, Y.M., Qu, K., Wang, Z., 2018. MicroRNAs: Important Regulators of Induced Pluripotent Stem Cell Generation and Differentiation. Stem Cell Rev Rep. 14, 71–81.

Zhang, H., Ma, Y., Gu, J., Liao, B., Li, J., Wong, J., Jin, Y., 2012. Reprogramming of somatic cells via TAT-mediated protein transduction of recombinant factors. Biomaterials. 33, 5047–5055.

Zhang, L., Kondo, H., Kamikubo, H., Kataoka, M., Sakamoto, W., 2016. VIPP1 Has a Disordered C-Terminal Tail Necessary for Protecting Photosynthetic Membranes against Stress. Plant Physiol. 171, 1983–1995.

Zhang, S., Cui, W., 2014. Sox2, a key factor in the regulation of pluripotency and neural differentiation. World J Stem Cells. 6, 305–311.

Zhang, S., Xiong, X., Sun, Y., 2020. Functional characterization of SOX2 as an anticancer target. Signal Transduct. Target. Ther. 5, 1–17.

Zhang, Y., Li, W., Laurent, T., Ding, S., 2012. Small molecules, big roles – the chemical manipulation of stem cell fate and somatic cell reprogramming. J Cell Sci. 125, 5609–5620.

Zhang, Z., Wu, W.S., 2013. Sodium Butyrate Promotes Generation of Human Induced Pluripotent Stem Cells Through Induction of the miR302/367 Cluster. Stem Cells Dev. 22(16), 2268-2277.

Zhang, Z., Xiang, D., Heriyanto, F., Gao, Y., Qian, Z., Wu, W.S., 2013. Dissecting the Roles of miR-302/367 Cluster in Cellular Reprogramming Using TALE-based Repressor and TALEN. Stem Cell Rep. 1, 218–225.

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.

Zheng, C., Baum, B.J., Iadarola, M.J., O'Connell, B.C., 2000. Genomic integration and gene expression by a modified adenoviral vector. Nat Biotechnol. 18(2), 176–180.

Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., 2009. Generation of induced pluripotent stem cells using recombinant proteins. Cell Stem Cell. 4(5), 381–384.

Zhou, W., Freed, C.R., 2009. Adenoviral Gene Delivery Can Reprogram Human Fibroblasts to Induced Pluripotent Stem Cells. Stem Cells. 27, 2667–2674.

Zimmerlin, L., Park, T.S., Huo, J.S., Verma, K., Pather, S.R., Talbot, C.C., Agarwal, J., Steppan, D., Zhang, Y.W., Considine, M., Guo, H., Zhong, X., Gutierrez, C., Cope, L., Canto-Soler, M.V., Friedman, A.D., Baylin, S.B., Zambidis, E.T., 2016. Tankyrase inhibition promotes a stable human naïve pluripotent state with improved functionality. Dev. 143(23), 4368–4380.

Publications from Ph.D. Thesis

Research articles

1 **Dey C**, Raina K, Thummer RP (**2023**) Production of a bioactive recombinant human UTF1 protein from *E. coli*. Healthcare Research and Related Technologies - Proceedings of NERC 2022 (Springer Nature).

2 **Dey C**, Venkatesan V, Thummer RP (**2022**) Identification of optimal expression parameters and purification of a codon-optimized human GLIS1 transcription factor from *Escherichia coli*. Molecular Biotechnology, 64(1):42-56 (Springer Nature).

3 **Dey C***, Thool M*, Bhattacharyya S, Sudhagar S, Thummer RP (**2021**) Generation of biologically active recombinant human OCT4 protein from *E. coli*. 3 Biotech, 11(5), 207 (Springer Nature). *Equal contribution.

4 Thool M*, **Dey C***, Sudhagar S, Thummer RP (**2021**) Generation of a recombinant stem cell-specific human SOX2 protein from *Escherichia coli* under native conditions. Molecular Biotechnology, 63(4), 327-338 (Springer Nature). *Equal contribution.

Review articles / Book chapters

1 **Dey C***, Raina K*, Thool M*, Adhikari P, Haridhasapavalan KK, Sundaravadivelu PK, Venkatesan V, Gogoi R, Sudhagar S, Thummer RP (**2021**) Auxiliary pluripotency-associated genes and their contributions in the generation of induced pluripotent stem cells. Molecular Players in iPSC Technology, Volume 12. Elsevier: Academic Press. ISBN: 9780323900591 *Equal contribution, Invited Book Chapter.

2 Raina K, **Dey C**, Thool M, Sudhagar S, Thummer RP (**2021**) An insight into the role of UTF1 in development, stem cells, and cancer. Stem Cell Reviews and Reports. (Springer Nature).

3 **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. *Equal contribution, Invited Book Chapter

4 Haridhasapavalan KK*, Raina K*, **Dey C***, Adhikari P, Thummer RP (**2020**) An insight into reprogramming barriers to iPSC generation. Stem Cell Reviews and Reports. 16(1): 56-81. (Springer Nature). *Equal contribution.

5 Borgohain MP*, Haridhasapavalan KK*, **Dey C**, Adhikari P, Thummer RP (**2019**) An insight into DNA-free reprogramming approaches to generate integration-free induced pluripotent stem cells for prospective biomedical applications. Stem Cell Reviews and Reports. 15(2): 286-313. (Springer Nature). *Equal contribution.

6 Haridhasapavalan KK*, Borgohain MP*, **Dey C**, Saha B, Narayan G, Kumar S, Thummer RP (**2019**) An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells. Gene. 686:146-159. (Elsevier). *Equal contribution.

7 Saha B, Borgohain MP, **Dey C**, Thummer RP (**2018**) iPS cell generation: Current and future challenges. Annals of Stem Cell Research and Therapy. 1(2): 1007.

8 Borgohain MP*, Narayan G*, Krishna Kumar H, **Dey C**, Thummer RP (**2018**) Maximizing expression and yield of human recombinant proteins from bacterial cell factories for biomedical applications. Advances in Microbial Biotechnology: Current Trends and Future Prospects. Apple Academic Press. ISBN: 9781771886673. *Equal contribution.

9 **Dey C***, Narayan G*, Krishna Kumar H*, Borgohain MP*, Lenka N, Thummer RP (**2017**) Cell-penetrating peptides as a tool to deliver biologically active recombinant proteins to generate transgene-free induced pluripotent stem cells. Studies on Stem Cells Research and Therapy 3(1): 006-015. *Equal contribution.

10 Ronima KR*, **Dey C***, Thummer RP. An insight into the role of GLIS1 in development, stem cells, and cancer. Manuscript to be submitted.

Publications from other collaborative research

Research Article

1 Chetia M, Sarkar T, **Dey C**, Thummer RP, Chatterjee S. Salt-tolerant and protease resistant lipopeptide AMPs and antimicrobial hydrogels. Manuscript submitted.

Conferences / workshops attended

1 Poster presentation in North-East Research Conclave 2022 on “Production of bioactive recombinant human UTF1 transcription factor from *E. coli*” organized by IIT Guwahati.

2 Poster presentation in Research Conclave 2018 on “Generation of transgene-free human induced pluripotent stem cells using non-genetic approaches for biomedical applications” organized by IIT Guwahati.

3 Poster presentation in Research Conclave 2017 on “Review of generation of transgene-free human induced pluripotent stem cells using non-genetic approaches for biomedical applications” organized by IIT Guwahati.

Awards

1 **Travel grant**: Awarded a travel grant for "**Training program in Generation and Maintenance of Human iPS cells**" organized by "The Accelerating the application of Stem cell technology in Human Diseases (ASHD)" program supported by Department of Biotechnology (DBT), Government of India and The Centre for iPS Cell Research and Application (CiRA), Kyoto University, Japan from 7th to 13th November, 2018.