INDUCED PLURIPOTENT STEM CELLS: FACTORS THAT DETERMINE PLURIPOTENCY

SUMITHA. P. B

Ph.D. THESIS

2016

SREE CHITRA TIRUNAL INSTITUTE

FOR

MEDICAL SCIENCES AND TECHNOLOGY, TRIVANDRUM

Thiruvananthapuram

INDUCED PLURIPOTENT STEM CELLS: FACTORS THAT DETERMINE PLURIPOTENCY

A THESIS PRESENTED BY

SUMITHA. P. B

TO

THE SREE CHITRA TIRUNAL INSTITUTE FOR
MEDICAL SCIENCES AND TECHNOLOGY, TRIVANDRUM
Thiruvananthapuram

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF

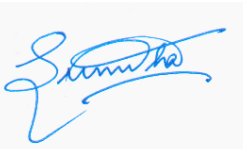
DOCTOR OF PHILOSOPHY

2016

CERTIFICATE

I, **Sumitha. P. B**., hereby certify that I had personally carried out the work depicted in the thesis entitled, "**INDUCED PLURIPOTENT STEM CELLS: FACTORS THAT DETERMINE PLURIPOTENCY**".

No part of the thesis has been submitted for the award of any other degree or diploma prior to this date.



Sumitha. P. B

17.05.2016

Institute for Stem Cell Biology and Regenerative Medicine
An autonomous institute of Department of Biotechnology, Government of India.
inStem

Centre for Stem Cell Research

A unit of inStem, Bengaluru, in collaboration with DBT and CMC, Vellore

17.05.2016

This is to certify that **Sumitha. P. B**. in the **Department of Haematology** and **Centre for Stem Cell Research, Christian Medical College** has fulfilled the requirements prescribed for the Ph.D degree of the Sree Chitra Tirunal Institute for Medical Sciences and Technology, Thiruvananthapuram.

The thesis entitled, "**INDUCED PLURIPOTENT STEM CELLS: FACTORS THAT DETERMINE PLURIPOTENCY**" was carried out under my direct supervision. No part of the thesis was submitted for the award of any degree or diploma prior this date.

*Clearance was obtained from the Institutional Ethics Committee / Institutional Animal Ethics for carrying out the study.

Dr. Shaji. R. V
Professor, Department of Haematology,
Adjunct Scientist, Centre for Stem Cell Research,
Christian Medical College, Bagayam,
Vellore, India - 632002.

Christian Medical College Campus,
Bagayam, Vellore - 632002. INDIA.

Phone: +91 - 416 - 307 5101 / 5107
Fax: +91 - 416 - 307 5103.
E-mail: cscr@cmcvellore.ac.in

The thesis entitled

INDUCED PLURIPOTENT STEM CELLS: FACTORS THAT DETERMINE PLURIPOTENCY

Submitted by

SUMITHA. P. B

for the degree of

Doctor of Philosophy

of

SREE CHITRA TIRUNAL INSTITUTE

FOR

MEDICAL SCIENCES AND TECHNOLOGY, TRIVANDRUM

Thiruvananthapuram

is evaluated and approved by

Dr. Shaji R. V.

Professor, Dept. of Haematology
Adjunct Scientist
Centre for Stem Cell Research
Christian Medical College
Vellore, India
(Guide)

Dr. Jagan Pongubala

Professor, Head
Dept. of Animal Sciences
School of Life Sciences
University of Hyderabad
Hyderabad, India
(Thesis examiner)

ACKNOWLEDGEMENTS

I am profoundly indebted to a large number of people who have provided me with invaluable support and assistance, both personal and professional, throughout my PhD research.

I take it as a privilege to express my deep sense of gratitude to my mentor and guide, Dr. Shaji. R. V. for expressing confidence in me and helping me step in to the field of hiPSC technology. He trained me in every aspect of research work, monitored the progress of my work and was available at anytime when I needed his help. I am immensely grateful to him for his efforts to drive my research work forward.

With respect and gratitude, I acknowledge Dr. Alok Srivastava, Director of CSCR and Dr. Vikram Mathew, HOD of Haematology Department, CMC for all the facilities and help extended to me to carry out my research work at CSCR and CMC.

I am grateful to my labmates for their valuable inputs to my research work. I take this opportunity to thank Janaki for always being there to rely on. I thank Dhavapriya for extending helping hands whenever I was in need of assistance at the bench. I thank Kannan, Musheer and Thiyagaraj for the scientific discussions we had and their valuable suggestions at work. I acknowledge Nancy for her assistance in FA project. My special thanks go to Saravanan R, Aneesha, Abhirup and Kasthuri for their excellent support during the final phase of my PhD work.

I would like to thank the entire team of CSCR for their constant help and support throughout my PhD work. I am grateful to the technical help rendered by Ms. Saranya, Ms. Samrajyan and Mr. Vaidyanathan at FACS facility, Mr. Satheesh and Ms. Esther at Histopathology Lab, and Dr. Sumathi, Dr. Pratheesh, Dr. Haemantha and Ms. Pavithra at Animal facility. I express my gratitude to the office staffs, Mr. Muthukrishnan, Mr. Tamil, Ms. Geetha and Ms. Selvi for always willing to help.

I express my sincere thanks to Dr. Molly Jacob and Dr. Sanjay Kumar, the members of my Doctoral Advisory Committee for their suggestions during my research tenure. I thank Dr. Rekha Samuel for her words of appreciations and encouragements. I take this opportunity to thank Dr. Poonkuzhali and Dr. Eunice of Haematology department for their scientific and technical assistance.

I sincerely acknowledge CSIR for my JRF/SRF research fellowship.

I am fortunate to have great friends like Joz, Edna, Balu and Anish, who have always stood by my side as a pillar of support and strength. I fall in short of words to thank Naveen for his constant support and motivation. I thank my friends next door, Franklin, Abisha and Delvin for their affection and encouragement. I fondly relish the moments spent with my dear friends, Sreeja and Rekha, during my

stay in Vellore. This work would have been impossible without the motivation and encouragement of all these people and I duly express my sincere gratitude for all.

At this juncture, I fondly remember my dear teachers – Ms. Shobha (Eden School, Quilandy), Ms. Indira (English Tutor, Kasaragod), Ms. Akkamma (TIHSS, Kasaragod), Dr. Velayudhan Pillai (Mentor, Talent Search Programme), Ms. Tanuja (K V No 1, Kasaragod), Dr. Prajwal Lobo (St Aloysius, Mangalore), Dr. Fathima Benazir (GRD, Coimbatore), Dr. Hashim (CSIR coaching class) - for their constant inspiration and support during my academic life. They laid the basis for my academic and professional achievements.

I am forever indebted to my grand parents, parents and sisters for their love, patience and support. I am grateful to my aunt, Beena and my cousin Sonu for standing by my side through thick and thin. Last, but not the least, I am grateful to God, for making me what I am and giving me all that I deserve.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SYNOPSIS

Background of the study

Reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) involves epigenetic resetting of somatic cells to a state similar to that of embryonic stem cells (ESCs). During the process, somatic cells transverse through multiple stages that include changes in the cell cycle profile, shut down of the somatic cell specific program, mesenchymal to epithelial transition, the activation of the endogenous pluripotency network and exogenous reprogramming factor independence. The progression through these stages is hindered by multiple molecular barriers, which result in lower efficiency and slower pace of the process. Overexpression (OE) and knockdown (KD) studies have identified several factors capable of influencing reprogramming process and the subsequent functional validation studies have revealed their specific roles at different stages of reprograming. These factors include genes involved in epigenetic regulation, mesenchymal to epithelial transition (MET), epithelial to mesenchymal transition (EMT), signaling pathways, cell cycle kinetics and DNA repair pathways. Even though regulation of the expression of these identified factors improved pluripotency induction, majority of the cells still remain refractory to reprogramming, indicating the presence of several unidentified key regulators of pluripotency and reprogramming.

To identify new factors playing important roles in pluripotency and reprogramming, candidate genes associated with specific molecular events or pathways that were less explored were selected and their role in reprogramming and pluripotency were assessed by regulating their expression in reprogramming cells by adopting different approaches. Mesenchymal to epithelial transition is one of the important initial events in reprogramming and inhibition of MET- and overexpression of EMT- associated genes and pathways were found to adversely affect the reprogramming process. Even though Grainy Head Like 2 (GRHL2) is a major transcription factor involved in regulation of epithelial genes and a partner in OCT4 interactome, their role in hiPSC reprogramming has not been investigated so far. Despite epigenetic changes being the most prominent event through out the reprogramming process, the involvement of many

epigenetic factors in reprogramming is still unknown. Signaling pathways critical for maintaining pluripotency and self renewal properties of embryonic stem cells are well studied. But, the signaling pathways involved in pluripotency induction and reprogramming process have not been explored. It has been reported that somatic cells defective of Fanconi Anemia pathway is refractory to reprogramming unless the defective gene is corrected. The significance of this DNA damage repair pathway in pluripotency induction and reprogramming has not been investigated.

Objectives of the study

1. *Establishment of reprogramming strategies for the generation of induced pluripotent stem cells (iPSCs) from human somatic cells.*
2. *Understanding the expression kinetics of molecular markers during the reprogramming process and assessment of their ability to identify reprogramming intermediates and hiPSCs.*
3. *Derivation of Fanconi anaemia (FA) patient-specific hiPSCs for elucidating the role of Fanconi anemia pathway in reprogramming and for disease modeling.*
4. *Identification of promoters that show robust activity in different stages of reprogramming.*
5. *Identification of novel factors involved in somatic cell reprogramming by gene RNA interference (RNAi).*

Experimental methods used in the study

1. Generation of hiPSCs from human dermal fibroblasts by delivering OSKM using retroviral, lentiviral, sendai viral and episomal vectors.
2. Characterization of established hiPSC lines to confirm their pluripotency by using the following methods:
   a) Immunofluorescence and real time PCRs for analyzing the expression of pluripotency markers

b) Bisulfite sequencing for analyzing methylation status of pluripotency gene promoters

c) Evaluation of *in-vitro* and *in-vivo* differentiation of hiPSCs by embryoid body (EB) mediated differentiation and teratoma formation.

3. Identification of promoters of the lentiviral vectors suitable for reprogramming studies by transducing fibroblasts with lentiviral vectors with CBA, UbC, SFFV, mCMV, hCMV, mELF1α and hELF1α promoters and evaluating transgene silencing during reprogramming.
4. Generation of lentiviral overexpression and knockdown vectors for regulation of target gene expression during reprogramming.
5. Estimation of reprogramming kinetics and efficiency by FACS and immunofluorescence analysis of pluripotency markers.
6. Isolation of cells at different stage of reprogramming by FACS based on coexpression pattern of fibroblast and pluripotency markers.
7. RNAi screen of signaling pathway genes using a shRNA library and identification of enriched and depleted shRNAs in partially and fully reprogrammed states by next generation sequencing.
8. Functional evaluation of Fanconi Anemia pathway in reprogramming and patient specific iPSC lines.

Major findings

1. The first goal of the project was to standardize the efficient methods for generation of hiPSCs. OCT4, SOX2, KLF4 AND c-MYC mediated reprogramming of fibroblasts by different gene delivery methods using retroviruses, lentiviruses, episomal plasmids and sendai viruses could be achieved. All the methods for the generation, identification, isolation and characterization of hiPSCs could be successfully established in the laboratory. HiPSC lines were fully characterized for their expression of pluripotency markers, acquisition of hypomethylated *OCT4* and *NANOG* promoters, *in vitro* differentiation potential by EB

formation and *in vivo* differentiation potential by teratoma formation. The difference in the reprogramming efficiencies achieved with different methods were estimated to assess their suitability in the subsequent experiments. This was the first study for establishing successful reprogramming methods in India.

2. During reprogramming of fibroblasts, it was observed that the initial culture conditions and the culture media of the somatic cells influence the reprogramming efficiency. Fibroblasts maintained in the medium that favoured higher proliferation rate, infection efficiency and transgene expression levels exhibited faster kinetics and higher efficiency of reprogramming. Based on these observations, it was confirmed that preconditioning of somatic cells is an important factor for that determine efficient reprogramming.
3. Even though hiPSC identification based on morphology, pluripotency marker expression and retroviral transgene silencing are widely employed; a comparison on the suitability of these three markers in identifying true hiPSCs has not been well established. By systematic monitoring of morphology, pluripotency marker expression and retroviral RFP silencing in the colonies generated during reprogramming, significant correlation was found between the hESC-like morphology of the hiPSC clones and retroviral transgene silencing which was measured by a retrovirally expressed fluorescence protein. These clones expressed all the pluripotency markers. A few colonies that lacked hESC-like morphology showed RFP silencing and/or pluripotency marker expression. Through the routine observation of reprogramming cells, it was found that emerging hiPSC colonies can be easily identified based on their morphological changes during the course of reprogramming. They emerge dislodging feeder cells around them radially forming flat colonies with defined boundaries and symmetric shape. It was observed that inclusion of retrovirally expressed RFP as a marker could greatly aid in the identification of such hiPSC colonies as they appeared as patches of RFP^- colonies surrounded by RFP^+ non-iPSCs. Thus by describing the kinetics of acquisition of hESC like morphology, pluripotency marker expression and retroviral RFP silencing by the

reprogramming cells, it was shown that morphology of hiPSCs serves as a more reliable marker in the identification of hiPSC clones. Through this study it was shown that retroviral transgene silencing is a highly reliable marker for isolation of pluripotent cells on reprogramming dish although morphology is the best criterion for the isolation of pluripotent clones.

4. Sorting of reprogramming cells based on a fibroblast marker, CD13, and pluripotency markers, SSEA-4 and TRA-1-60, could facilitate isolation of the cells at different stages of reprogramming. Gene expression analysis and shRNA library screening assays performed on these sorted cells confirmed the suitability of these markers in defining the stages of reprogramming.
5. In search for promoters of viral vectors that do not get silenced during the entire course of reprogramming, lentiviral vectors with different promoters that drive the expression of GFP were evaluated. It was found that mouse EF1α and human EF1α promoters did not exhibit transgene silencing and they are the most suitable promoters for consistent knockdown and overexpression experiments in reprogramming.
6. The role of GRHL1, GRHL2 and GRHL3 was evaluated based on the effect of their knockdown on reprogramming efficiency and kinetics. Analysis of the cells sorted from different stages of reprogramming revealed that GRHL2 show a constant increase in expression as the cells underwent reprogramming, and this expression pattern was similar to that of pluripotency markers. The reprogramming efficiency was found to be reduced on GRHL2 knockdown.
7. The role of the epigenetic factors was evaluated based on the effect of their knockdown on reprogramming efficiency and kinetics. The epigenetic factors which showed variations in their expression levels during reprogramming were selected after bioinformatics analysis of the gene expression data available from previous publications. The expression levels of the selected epigenetic factors were validated in the flow sorted cells from different stages of

reprogramming. By knocking down the genes individually with shRNAs, their involvement in reprogramming could be assessed.

8. For elucidating the role of various signaling pathways in the reprogramming process, human dermal fibroblasts were infected with lentiviral vectors to express shRNAs against 5000 human signaling pathway genes and then they were reprogrammed with OSKM. The partially (SSEA-4^{+}TRA-1-60^{-}) and fully (SSEA-4^{+}TRA-1-60^{+}) reprogrammed cells were sorted from the reprogramming cells and were analyzed for incorporation of shRNAs by massive parallel sequencing. Based on the specific shRNAs enriched or depleted in the sorted fractions of cells, specific signaling pathways involved in reprogramming were identified.
9. The role of Fanconi Anemia pathway genes in pluripotency and reprogramming was investigated using an inducible gene expression system for complementing defective *FANCA* gene in patient derived cells. HiPSC lines were derived from FANCA deficient fibroblast complemented with functional *FANCA* gene under an inducible promoter and the lines were characterized for their pluripotency. The formation of teratoma by the hiPSC line, without FANCA induction during its *in vivo* growth in SCID mice, confirmed their pluripotency. By conditional complementation of FANCA in these hiPSC lines, the functional significance of Fanconi Anemia pathway could be assed in these cells. In the complete absence of *FANCA* gene and functional fanconi anemia pathway, hiPSCs failed to survive, there by mimicking the disease phenotype. These hiPSC lines can serve as an excellent tool for elucidating the role of FA pathway in pluripotency and disease pathogenesis.

Significance and implications of the findings

Reprogramming of somatic cells to hiPSCs is a dynamic process with a low efficiency and slow kinetics of pluripotency induction wherein cells transverse through different stages by overcoming multiple molecular barriers. Even though several factors that regulate reprogramming and play significant roles in the process have been identified, majority of the

cells fail to achieve pluripotent state indicating the presence of additional regulators of pluripotency induction. In this study, the potential role of factors in regulating molecular events associated with reprogramming was assessed. The findings from this study could improve our understanding about the molecular mechanism of reprogramming and the knowledge gained will be helpful in developing reprogramming strategies to achieve higher efficiency. In addition, Fanconi Anemia patient specific hiPSC line derived in the lab showed disease phenotypes and hence can serve as an excellent tool to study the role of FA pathway in pluripotency and disease pathogenesis.

Chapter 1

INTRODUCTION

1.1. Background of the study

Induced pluripotent stem cells (iPSCs) are generated by inducing pluripotency in somatic cells by manipulating the epigenetic state of the cells through overexpression and knockdown of specific factors (Takahashi *et al.,* 2007; Bayart and Cohen-Haguenauer, 2013). iPSCs can serve as an unlimited source of pluripotent cells, as well as specialized cells that are derived from them, for basic science research, clinical translational research and regenerative medicine (Robinton and Daley, 2012). They are potential replacements for embryonic stem cells with least ethical issues associated with their generation. iPSCs have been derived from multiple cell types and organisms with the aid of different factor combinations that are delivered employing either integrative or non-integrative factor delivery modes (González *et al.,* 2011).

One of the major limitations in the reprogramming of somatic cells to iPSCs is the slow pace and low efficiency of the process, with an average success rate of >1% and period of 3-4 weeks for generation of iPSCs (Bayart and Cohen-Haguenauer, 2013). Also, the reprogramming cells and colonies exhibit huge heterogeneity at the molecular level and, therefore, identification of true iPSC clones that have achieved all features of pluripotent cells is technically challenging. The currently used human induced pluripotent stem cell (hiPSC) identification methods based on pluripotency marker expression and morphology have met with varying degree of success (Chan *et al.,* 2009). This makes it inevitable to subject isolated iPSC colonies to extensive molecular and functional characterisation for deriving iPSC lines suitable for downstream applications (Stadtfeld *et al.,* 2010). Thus, due to the low efficiency and high cellular heterogeneity associated with the reprogramming process, generation of individual specific iPSC lines from somatic cells is tedious, expensive and time-consuming. Refining the current reprogramming strategies may help in improving the efficiency of pluripotency induction with respect to the number of colonies formed as well as their quality. To achieve this, it is critical to understand the molecular mechanism of pluripotency induction during reprogramming.

Under the influence of reprogramming factors and their downstream targets, the

somatic cells subjected to reprogramming undergo continuous transitions in their molecular states (González and Huangfu, 2015). Based on the analysis on reprogramming intermediates, it was reported that the cells are driven through multiple routes that culminate in an aborted, a partially or a bona fide reprogrammed state (Buganim *et al.,* 2013; David and Polo 2014; González and Huangfu, 2015). The major events identified in successful reprogramming process include inactivation of somatic cell specific gene network, mesenchymal to epithelial transition (MET), activation of endogenous pluripotency network and silencing of exogenous factors (Ho *et al.,* 2011; Liang and Zhang, 2012; David and Polo 2014; González and Huangfu, 2015). To understand the molecular mechanism of successful reprogramming, the pathways that favour and inhibit reprogramming need to be identified and their complex interplay need to be dissected. The role of several novel factors in reprogramming could be identified by the molecular interventions of the pathways by overexpression, knockdown or chemical inhibition of specific genes (Mancarci *et al.,* 2012; Zhang *et al.,* 2012). By manipulating the expression of these genes, their role in pluripotency induction could be identified. The stage-specific role of such factors was identified by subsequent validation studies, and reprogramming efficiency could be improved by appropriate gene regulation. The factors that could favour reprogramming included the genes that are part of the pathways maintaining pluripotency and self-renewal of pluripotent stem cells. The factors that inhibited reprogramming included those involved in differentiation, genomic integrity surveillance and cell cycle arrest (Kawamura *et al.,* 2009; Li *et al.,* 2009). Despite additional manipulations, the majority of the cells still failed to form hiPSCs, which clearly indicated the involvement of still unknown regulators of reprogramming. The complexity of the reprogramming process and limitations with the sensitivity of gene regulation studies are the major challenges in the Identification of novel factors involved in reprogramming.

Gene knockdown studies involving downregulation of specific gene of interest or RNAi screen have been particularly useful in the identification of novel factors in reprogramming (Mancarci *et al.,* 2012; Qin *et al.,* 2014; Sakurai *et al.,* 2014; Yang *et al.,* 2014). Since reprogramming is a multi-step process wherein the cells continuously change

their identity, the study design can influence the outcome of such studies. The effect of knockdown of genes is often evaluated based on its effect on reprogramming efficiency, which is estimated based on the number of cell or colonies expressing pluripotency markers like alkaline phosphatase, TRA-1-60, TRA-1-81 or NANOG. Since cells expressing these markers are known to vary in their level of pluripotency (Chan *et al.,* 2009; Valamehr *et al.,* 2012; Abujarour *et al.,* 2013), employing them individually to assess the effect on the reprogramming outcome will compromise the accuracy of screening tests. A more accurate assessment of the effect of gene regulation on reprogramming outcome would be possible by estimating the percentage of cells expressing a combination of markers, that specific stages in reprogramming. Moreover, most of the knockdown studies employ constitutive promoters in transfer vectors, and it is known that the efficiency of promoters to drive the expression of the genes of interest or small hairpin RNAs (shRNAs) is influenced by the type of target cells. Depending on the stages to which the reprogramming cells belong, the promoter activity and strength will differ, and this can interfere with the assessment of the role of genes tested. So it is important to choose appropriate promoters to achieve the gene regulation at reprogramming stages of interest.

Apart from the gene knockdown, an alternative to studying the role of specific factors in reprogramming and pluripotency is to derive iPSCs from patient cells with a genetic defect in the gene, which have led to impairment in associated molecular pathway. The somatic cells from patients can be stably transduced to carry inducible overexpression system with the functional gene. Through conditional complementation of genes at specific stages of reprogramming and in hiPSCs, the role of the gene and the associated pathway in pluripotency induction and maintenance, can be identified. Such patient-specific hiPSC lines would also be useful for elucidating the molecular basis of disease pathogenesis.

Thus, the novel factors influencing pluripotency induction and reprogramming can be identified by systematically following the reprogramming cells for the formation of intermediated cells and iPSCs under the influence of a robust gene knockdown or gene complementation system.

1.2. Objectives of the study

The principal objective of this thesis work is to identify the factors that determine pluripotency induction during reprogramming of human somatic cells to human induced pluripotent stem cells (hiPSCs). The study is described in the chapters 4 to 6, with each chapter dealing with specific sub-objectives.

Chapter 4: Establishment of reprogramming strategies for the generation of induced pluripotent stem cells (iPSCs) from human somatic cells

The chapter describes the establishment of reprogramming strategies for the generation of induced pluripotent stem cells from human adult dermal fibroblasts in the laboratory. The main objectives are as follows:

1. *Understanding the expression kinetics of molecular markers during the reprogramming process.*
2. *Assessment of the ability of the molecular markers to identify reprogramming intermediates and hiPSCs.*
3. *Derivation of hiPSC lines from human adult dermal fibroblasts by viral and non-viral vector mediated reprogramming and their characterization.*

Chapter 5: Derivation of Fanconi anaemia (FA) patient-specific hiPSC lines for elucidating the role of Fanconi anemia pathway in reprogramming and for disease modeling

The chapter describes the generation of Fanconi anaemia patient-specific hiPSCs, capable of the conditional restoration of functional FA pathway, for studying the role of this pathway in reprogramming and disease pathogenesis.

1. *Conditional complementation of FANCA gene in the fibroblasts derived from patients with defective FANCA genes (FA-A complementation group).*
2. *Evaluation of the effect of conditional complementation of FANCA on reprogramming of FA fibroblasts of FA-A complementation group.*
3. *Derivation and characterization of FA-A-hiPSC lines.*
4. *Evaluation of FA disease phenotype recapitulation in FA-A-hiPSC lines.*

Chapter 6: Identification of novel factors involved in somatic cell reprogramming by RNA interference (RNAi)

The aim of this chapter is to identify novel regulators that play a critical role in the reprogramming process. The main objectives are as follows.

1. *Identification of promoters that show robust activity in different stages of reprogramming.*
2. *Selection of potential genes that can influence pluripotency induction during hiPSC generation.*
3. *To study the effect of candidate gene knockdown on OSKM mediated reprogramming.*
4. *To identify the signaling pathways active in the intermediate and the late stages of reprogramming.*

Chapter 2

REVIEW OF LITERATURE

2.1. Pluripotent Stem Cells (PSCs)

Pluripotency refers to the developmental potential of cells to give rise to cells of all three embryonic germ layers (Weinberger *et al.,* 2016). Unlike totipotent cells, pluripotent stem cells (PSCs) cannot give rise to extra-embryonic tissues. Even though pluripotency represents a highly dynamic and transient state *in vivo*, the pluripotent cells derived from multiple sources can be maintained indefinitely in an induced self-renewal state under specific culture conditions *in vitro*.

Multiple types of PSCs are isolated from rodents and humans, and they are designated based on their donor cell of origin (**Figure 1**). Embryonic stem cells (ESCs) are derived from the inner cell mass (ICM) of developing pre-implantation mouse or human blastocysts. Epiblast stem cells (EpiSCs) are isolated from mouse post-implantation epiblasts. Due to ethical concerns, pluripotent cells are not derived from human post-implantation embryos. Embryonic germ cells are generated by reverting early migrating rodent primordial germ cells (PGCs) *in vitro* into pluripotent ESC-like cells. Similarly, male germ stem cells (GSCs) are obtained by converting mouse neonatal and adult spermatogonial stem cells to a pluripotent state. The GSCs retain only the male imprint signature, thereby increasing their tumorigenic potential. From primates, no validated GSCs have been derived yet.

PSCs exhibit characteristic cellular and molecular features that complement their pluripotency and self-renewal ability. These cells exist in a naïve and primed pluripotent state depending on the developmental stage of the source tissue. At the cellular level, PSCs exhibit specific features on their cell cycle kinetics, morphology, and energy metabolism (González and Huangfu, 2015). PSCs show a cell cycle profile with a reduced G0/G1 phase and a prominent S phase. This pattern of cell cycle kinetic give the pluripotent cells the advantage of shortening the exposure window to external cues that induce differentiation, which is likely to happen in prolonged G0/G1 phase. The pluripotent cells derived from inner cell mass of mouse and human embryos, like mouse ESCs (mESCs), mouse EpiESCs (mEpiSCs) and human ESCs (hESCs), show characteristic morphological features. mESCs

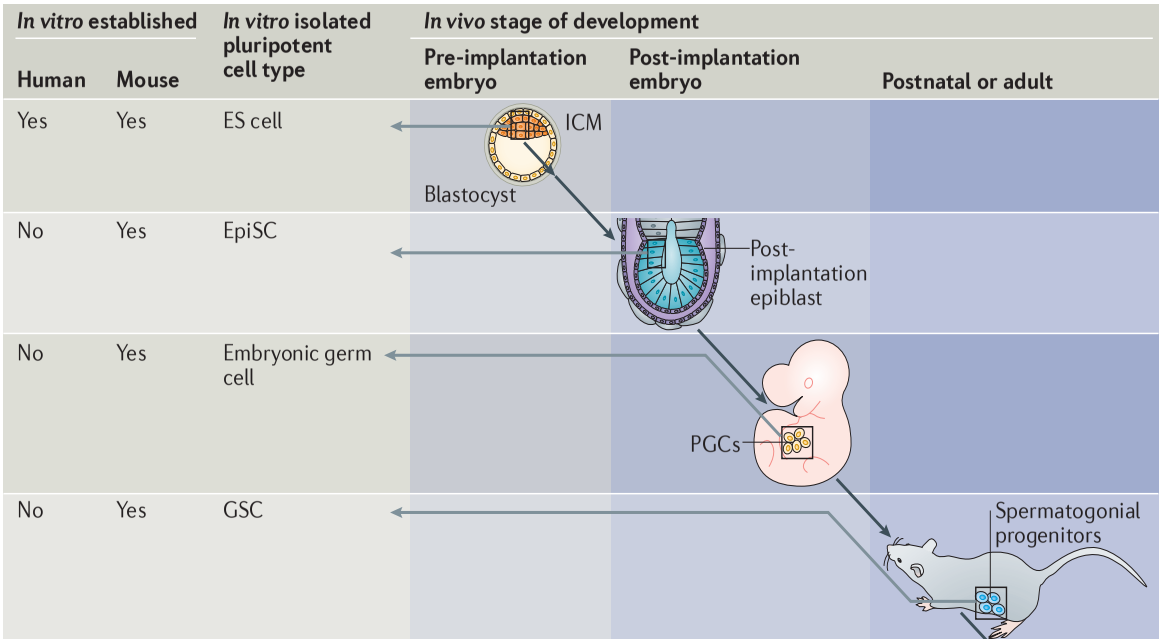


Figure 1: Types and sources of pluripotent stem cells (PSCs)
PSCs can be derived from different embryonic cells at various stages of mouse and human development. The PSCs derived during development in the chronological order are embryonic stem cells (ESCs) from inner cell mass (ICM) of pre-implantation embryo, epiblast stem cells (EpiSCs) from epiblast of post implantation embryo, embryonic germ cells (from primodial germ cells (PGCs) and germ stem cell (GSC) from spermatogonial progenitors. (Adapted from Weinberger *et al.*, 2016).

form dome shaped colonies with tightly packed cells. hESCs form flat colonies with defined boundary and closely packed cells. The individual cells in these colonies show prominent nuclei and high nuclei to cytoplasmic ratio. mEpi-ESCs resemble hESCs in their morphology. Pluripotent stem cells are dependent predominantly on glycolysis for their energy requirements and they exhibit reduced mitochondrial activity and elevated lactate production. This might be an adaptive feature of pluripotent cells to survive in low oxygen concentration ranging between 0.7-7% in the tissue *in vivo*.

At the molecular level, ESCs display distinctive chromatin features to support their special properties. In ESCs, the chromatin is in an "open" state, with the majority of the region existing in a euchromatin state. This is in contrast with differentiated cells were heterochromatin is highly prevalent. Consistent with this, repressive histone modifications are less prevalent, and active histone modifications are abundant genome-wide in ESCs compared to differentiated cells. The chromatin state of the cells dictates the transcription network and signaling pathways in cells to support the pluripotent state. Such a permissive chromatin supports the pluripotency of cells making them receptive to differentiation cues.

2.2. Applications of hESCs and their disadvantages

Even though mouse, rat and non-human primate models have enabled the dissection of disease mechanism in various cell types at different developmental stages *in vivo*, due to the genetic, developmental and physiological difference between these animal models and humans, the exact recapitulation of human disease phenotype was not achieved in some cases (Jucker 2010). Consistent with this, drugs that have been found to be effective in alleviating disease phenotype in animal models have sometimes failed to serve the purpose in humans. So it is preferable to conduct such studies on the human cells, which is limited to *in vitro* systems. Culturing patient-derived cells has been useful for studying the molecular and cellular features and helped in developing therapies for diseases like cancers. However, often disease-affected cells are in an inaccessible location *in vivo* and isolated cells cannot be maintained in culture for long. hESCs can serve as an unlimited source of

pluripotent and specialized cells and have found potential applications in disease modeling, drug screening and regenerative therapy as the discrepancies in the experimental outcome of studies involving animal models could be overcome by using hESCs. Also, hESCs offer the possibility of disease correction by gene editing techniques, which was difficult to achieve in patient-specific adult stem cells and affected cell types. Since multiple cell types can be derived from hESCs, the effect of cell autonomous and non-autonomous nature of disease can be explored. hESCs are the best suited for modeling monogenic disorder with an early onset, clear cellular phenotype and high penetrance. hESC-based monogenic disease models could be generated by mutagenesis of disease associated genes (Nakayama 2010; Ding *et al.,* 2013) in normal hESC lines or by deriving ESCs from affected blastocyst after PGD (Eiges *et al.,* 2007). Lesch-Nyhan syndrome disease model was developed by mutating the *HPRT* gene in normal hESCs (Urbach *et al.,* 2004). Early onset and developmental disorders like Turner and Down syndromes could be modeled using ESCs. Specialized cells derived from hESC are entering clinical trials for a few diseases that do not have any conventional treatment options around the globe (Trounson and Dewitt 2016). Before taking to the clinics, the therapeutic potential and safety of these cells are tested in mice and rat models. Once the cells clear this test, they are taken for clinical trials progressively from phase I to phase IV. Phase I clinical trial for Parkinson disease using human parthenogenetic embryonic stem cell-derived neural stem cells has been approved by Australian Therapeutic Goods Administration. hESC-derived retinal epithelial cells (RPE) will be soon used for the clinical trial of age-related macular degeneration (AMD) in UK and USA. In the phase I-II clinical trial of type I diabetes, β cell progenitor enclosed in a flat capsule is transplanted subcutaneously into patients. These cells on maturation release insulin and other islet cell hormones, thereby alleviating the disease phenotype.

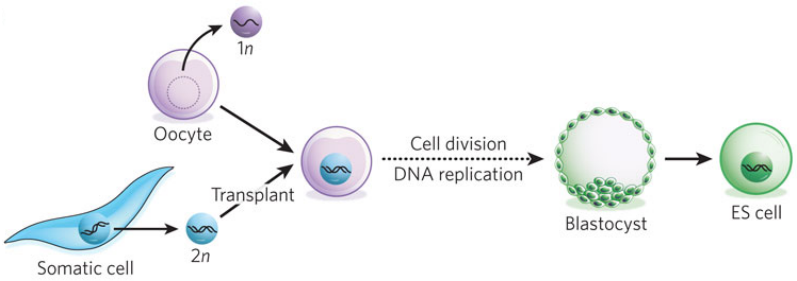
Due to the ethical concerns regarding the use of human embryos as a source of hESC, derivation of new hESC lines are restricted. Only a limited number of these lines are approved for research, and they lacked the diversity needed for downstream applications including disease modeling and drug screening. Moreover, their use in regenerative therapy is also limited by the fact that these cells may not match with patients immunologically. Even

though normal hESCs can be subjected to gene editing, generating an hESC model for complex disorders is nearly impossible since many unknown genes are involved in disease pathogenesis. Thus, there is a need for an alternative source of pluripotent stem cells that lacks at least some of the disadvantages of hESCs

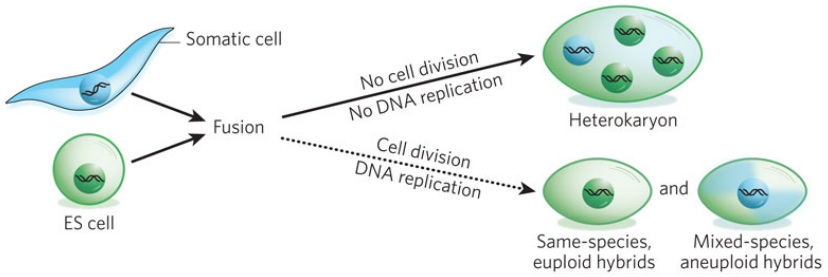
2.3. Somatic cell reprogramming

Somatic cell reprogramming refers to the techniques used for the generation of PSCs by directing a somatic nucleus to an early embryonic state. The three approaches used for somatic cell reprogramming are somatic cell nuclear transfer (SCNT), somatic cell nuclear fusion with embryonic stem cells and direct reprogramming with ectopic factor expression (Yamanaka and Blau 2010) (**Figure 2**). In SCNT, somatic nuclei are transferred into enucleated oocytes and upon oocyte activation, somatic genome gets rapidly remodeled to a pluripotent state by maternal factors. Zygotes generated by SCNT were found to be totipotent, and ES cells could be derived from ICM of the blastocyst stage of these embryos. In the second approach, somatic cells are fused with PSCs resulting in the formation of heterokaryons with two intact nuclei. In the heterokaryon, a certain pluripotency loci on somatic cell nucleus get activated in a replication-dependent manner. Following cell division, the two nuclei fuse and additional gene loci are reprogrammed to achieve a pluripotent state. Direct reprogramming of somatic cells by ectopic factor expression results in the formation of pluripotent cells called induced PSCs (iPSCs). The cells subjected to direct reprogramming go through multiple rounds of cell division in the presence of factor expression to become iPSCs. The slow kinetic and low efficiency is a major limitation of direct reprogramming method. The three approaches of somatic cells reprogramming differ in their pace and efficiency. The pace and efficiency of reprogramming is higher with cell fusion, whereas it is slow with SCNT and transcription factor mediated reprogramming. For disease modeling, drug screening and regenerative therapy, iPSCs have been found to be more advantageous over pluripotent cells derived by cell fusion and SCNT.

A.



B.



C.

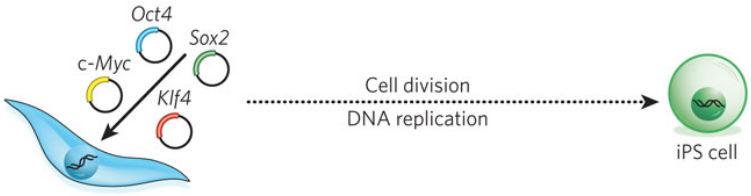


Figure 2: Somatic cell reprogramming approaches
(A) Somatic cell nuclear transfer (SCNT) – The diploid somatic cell nuclei (2*n*), transplanted into an enucleated oocyte, get reprogrammed to a pluripotent state on exposure to oocyte cytoplasmic contents. Blastocysts are generated from these oocytes from which diploid ESCs are derived. **(B)** Cell fusion – On fusing a somatic cell with an ESC, a heterokaryon or a hybrid is formed. If the fused cells fail to replicate, a short lived multinucleated heterokaryon is formed. If the fused cells udergo cell division, the nuclei fuse to form a tetraploid (4*n*) hybrid. Hybrids from the same species will be euploid and those from different species' will be aneuploid. **(C)** Direct reprogramming with ectopic expression of specific reprogramming factors like OCT4, SOX2, KLF4 and c-MYC that induce pluripotency in somatic cells resulting in the formation of induced pluripotent stem cells (iPSCs). (Adapted from Blau *et al.*, 2010.)

2.4. Induced Pluripotent Stem Cells (iPSCs)

Induced pluripotent stem cells are generated by inducing pluripotency in specialized somatic cells by resetting their epigenetic state with the aid of exogenous factors. In 2006, Takahashi *et al.,* first reported the generation of iPSCs from mouse embryonic fibroblasts (MEFs) and tail tip fibroblasts (TTF) by retroviral-mediated expression of four transcription factors, OCT4, SOX2, KLF4 and cMYC (OSKM) (Takahashi and Yamanaka 2006) (**Figure 3**). They identified the potential of these four factors, Yamanaka factors, to drive iPSC formation by systematically screening 24 factors that are associated with pluripotency in embryonic stem cells. Their factor screening reprogramming experiments were based on the hypothesis that factors that help in maintaining pluripotency might also be capable of inducing pluripotency in differentiated cells. The colonies generated were subjected to antibiotic selection facilitated by the puromycin resistance gene inserted into *Fbxo15* locus. The pluripotency of the miPSC lines generated was confirmed by their molecular and functional characterization. Subsequently in 2007, the same group reported the generation of hiPSCs from human dermal fibroblast using Yamanaka factors (Takahashi *et al.,* 2007). During the same time, Yu *et al* also succeeded in the generation of hiPSCs from fetal fibroblasts, newborn foreskin fibroblasts (BJ fibroblast), and primary synoviocytes (HFLS) by reprogramming with a different combination of factors, OCT4, SOX2, NANOG and LIN28 (OSNL) (Yu *et al.,* 2011).

2.5. Strategies for generation of iPSCs:

Derivation of iPSC lines is technically challenging, and the whole procedure is expensive and time-consuming. In the chronological order, the major steps associated with this process are isolation and expansion of donor cells, preparation of reprogramming factor delivery vectors, introduction of reprogramming factors into donor cells, expansion of these cells on feeder based or feeder-free culture system for generating iPSC clones, identification, isolation and expansion of iPSC clones, molecular and functional

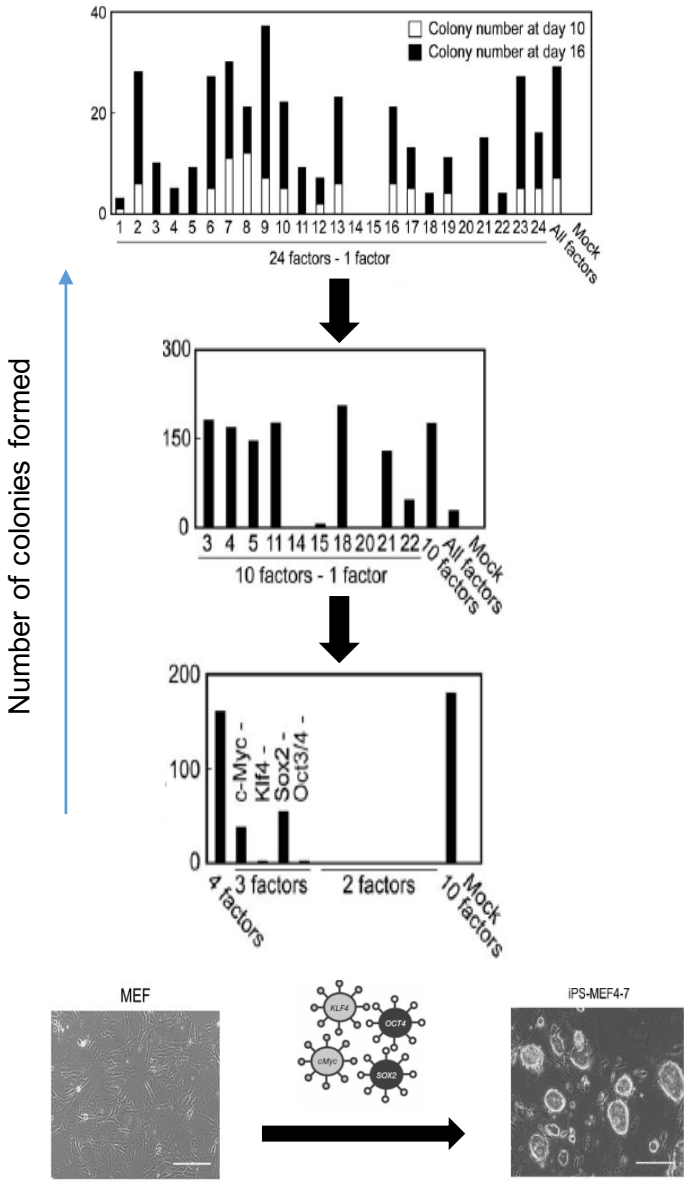


Figure 3: Discovery of direct reprogramming with ectopic factor expression in somatic cells

For identifying factors that can induce pluripotency in mouse embryonic fibroblasts (MEFs), 24 factors candidate factors were selected based on their role in maintaining pluripotency in ESCs. Various combinations of 23 factors, omitting one of the 24 factors in each set, were overexpressed in MEFs to identify the factors that can induce the formation of mESC-like colonies. The absence of ten factors adversely affected colony formation and they were further tested in the same way in combinations of 9 factors, omitting one of the 10 factors in each set. Among these ten factors, in the absence of four factors, namely, *Oct4* (O), *Sox2* (S), *Klf4* (K) and *cMyc* (M), colony formation was adversely affected. The overexpression of OSKM was found to be enough to induce pluripotency in MEFs resulting in the formation of clones with characteristic features of mESCs, which is termed as mouse induced pluripotent stem cells (miPSCs). (Adapted from Takahashi *et al.*, 2006.)

characterization to confirm their pluripotency and preparation of cryopreserved stocks (Stadtfeld *et al.,* 2010a; González *et al.,* 2011) (**Figure 4**).

2.5.1. Types of donor cells

Following the derivation of iPSCs from mouse and human, iPSC lines have been generated from a wide variety of organisms like a rodent, canine, equine, avian and bovine species (Stadtfeld *et al.,* 2010a). Almost all cell types could be reprogrammed to iPSCs (**Figure 4**). The efficiency of reprogramming a specific cell type has been found to be dependent on multiple factors like passage number, the extent of genetic aberration they carry, reprogramming cocktail and type of factor delivery mode employed.

Fibroblasts are the most frequently used cell type for reprogramming. Keratinocytes, the keratin dense epithelial cells, have shown relatively higher reprogramming efficiency compared to fibroblasts (Aasen *et al.,* 2008). The heterogeneous multipotent adipose stem cells (ASCs), which can be isolated in large quantities from lipoaspirated adult adipose tissue, could be reprogrammed to iPSCs (Sun *et al.,* 2009). Due to the high endogenous levels of SOX2 and cMYC, the human dermal papilla (DP) cells could be reprogrammed more efficiently than embryonic and dermal fibroblasts (Tsai *et al.,* 2010). Human urine-derived cells (HUCs) which include the cells drained out of the urinary system alone with urine have been reprogrammed successfully with a higher efficiency compared to any other cell type (Zhou *et al.,* 2012; Wang *et al.,* 2013) and these cells showed a very high efficiency of reprogramming, ranging from 0.2% to 4%, depending the methods used for reprogramming. Due to the non-invasive mode of isolation, reduced susceptibility to accumulate somatic mutation and high reprogramming efficiency, these cells are considered to be promising cell source for iPSC derivation.

Blood cells are the most favored source for iPSC induction as they can be obtained from multiple sources like peripheral blood (PB), cord blood (CB) and bone marrow (BM). These cells include hematopoietic stem cells (HSCs), progenitor cells and terminally differentiated cells. A major advantage of using blood cells is that even the frozen blood stored in blood banks could be reprogrammed, thereby providing the opportunity to investigate rare diseases retrospectively. PB is an easily accessible source of cells for

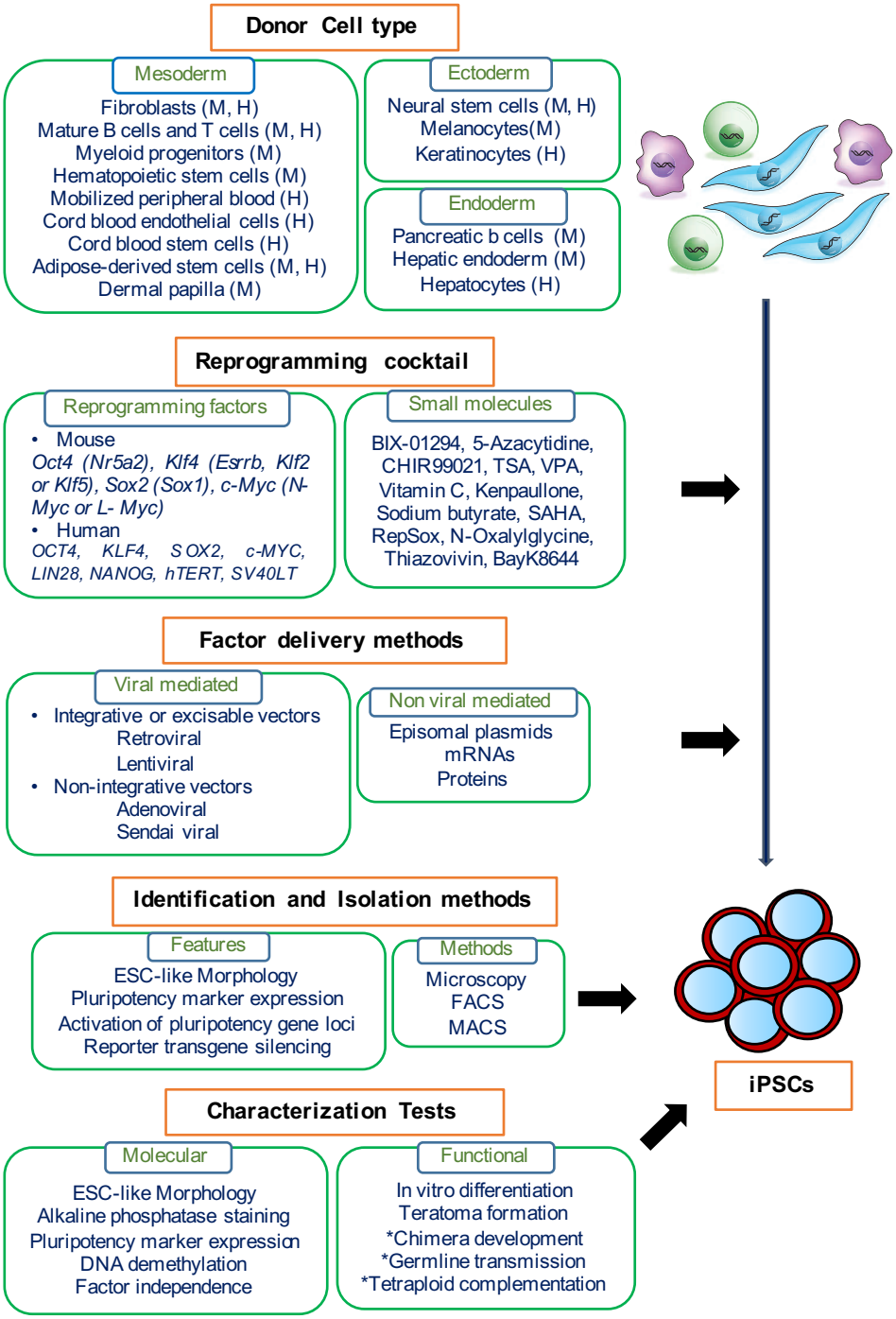


Figure 4: Strategies for the generation of iPSCs

iPSCs have been generated from almost all types of somatic cells using various combinations of reprogramming factors and small molecules. The factor delivery is carried out using integrating or non-integrating vectors. The iPSCs are identified and isolated based on morphology, pluripotency marker expression and reporter transgene silencing by microscopy, FACS and MACS. The pluripotency of the iPSC lines generated are confirmed by molecular and functional characterization tests. M-mouse, H-human. *not used for testing human iPSCs.

reprogramming and imposes a minimum risk to donors while harvesting. The major disadvantage associated with using PB is that reprogramming efficiency is extremely low at 0.001%. Moreover, mature T cells are found to get reprogrammed preferentially, and since they have undergone T cell receptor (TCR) rearrangement, such hiPSC lines might not be suitable for application in regenerative therapy (Loh *et al.,* 2010; Staerk *et al.,* 2010). Since HSCs and progenitor cells are in an early developmental stage, they are immunologically immature and carry minimum somatic mutations, unlike adult donor cells like fibroblasts and keratinocytes. These cells are rare in PB, but rich in CB, placenta and BM. Isolation of $CD34^+$ from BM is invasive and those mobilized into PB by (granulocyte colony stimulating factor) G-CSF administration is time-consuming and cumbersome (Loh *et al.,* 2009; Okabe *et al.,* 2009). As an alternate source, placenta and CB, rich in cells with multilineage differentiation potential, is promising (Haase *et al.,* 2009). Their use as a routine source of donor cells is limited by the availability.

Cancer cell lines and cells from patients with hematological malignancies and solid tumors were used as a source material for generating cancer-specific hiPSC lines (Fernandez *et al.,* 2013). Certain human cell types are rarely used as donor cells for reprogramming due to the inaccessibility *in vivo* location. These include neural stem cells (NSCs), pancreatic B cells, melanocytes, hepatocytes, meningiocytes and gastric epithelial cells.

2.5.2. Reprogramming factors

For iPSC induction, the combination of OCT4, SOX2, KLF4 and cMYC (OSKM) is the most widely used reprogramming factor combination and has been used in viral and non-viral mediated reprogramming of variety of cells from different organisms. OCT4, SOX2, NANOG and LIN28 (OSNL) is the second combination of factors that was initially employed to drive human cells to pluripotency (Yu *et al.,* 2007a). Individual factors in OSKM or OSNL combination could be omitted, replaced or complemented with other factors or small molecules to induce pluripotency (**Figure 4**). However, excluding one or more of the factors from OSKM or OSNL group without any additional manipulation, often resulted in a reduction in reprogramming efficiency. In the absence of cMYC in OSKM, MEFs showed a drastic

reduction in iPSC colonies formed. Also, a striking decrease in the rapidly proliferating cell was formed, thereby reducing the heterogeneity within the culture. The cells that show the high endogenous level of reprogramming factors has been found to require fewer factors to drive them to a pluripotent state. Adult mouse neural stem cells that express SOX2 and c-MYC at levels higher than ESCs could be reprogrammed with only O or OK or OM (Kim *et al.,* 2008; Kim *et al.,* 2009b). At the same time, some terminally differentiated cells required additional manipulation to achieve pluripotency. Mature murine B cells required C/EBPα overexpression or Pax5 knockdown along with OSKM overexpression to generate miPSCs (Hanna *et al.,* 2008).

The individual factors in the reprogramming cocktail could be replaced with other factors which are either homologs, regulators or effectors of the replaced factor (Ho *et al.,* 2011). NR5A1 and NR5A2 that regulate the expression of OCT4 by binding to its enhancer and promoter, and have been found to replace OCT4 in reprogramming. The homologs of KLF4 (KLF1, KLF2 or KLF5), SOX2 (SOX1, SOX3, SOX15 or SOX18) and c-MYC (n-MYC or L-MYC) were also found to be capable of inducing pluripotency. cMYC, being a target of WNT signaling, could be replaced by WNT3A for reprogramming mouse fibroblasts. The reprogramming efficiency was 20 fold more than with OSK alone. cMYC could also be substituted by its target, the miR90 cluster. ESSRB, capable of forming a protein complex with OCT4 and regulate KLF4 expression, could replace KLF4 and reprogram mouse fibroblasts with OSM. SV40 Large T antigen was found to reprogram human fibroblast with OS with 9 fold higher efficiency than OSKM. Together with hTERT and SLV40T, OSKM mediated reprogramming of human fibroblasts could be improved by three folds. UTF1 improved the reprogramming efficiency of human fibroblast with OSKM and OSK by 100 folds in the presence of p53 knockdown.

Small molecules additives in reprogramming culture medium were found to improve reprogramming efficiency or replace individual reprogramming factors, raising the possibility of deriving iPSCs using only small molecule cocktails (Feng *et al.,* 2009b; Ma *et al.,* 2013). Most of these molecules are associated with epigenetic modifiers and signaling pathways. DNA methyltransferase inhibitor, 5-aza-cytidine (AZA), promoted the conversion of partially

reprogrammed miPSCs to a fully reprogrammed state, thereby increasing the number of bonafide miPSCs derived from MEFs with OSKM by four folds. The histone deacetylase inhibitors like valproic acid (VPA), trichostatin A (TSA) and suberoylanilide hydroxamic acid (SAHA) improved the efficiency of OSKM mediated reprogramming of MEFs by 100, 15 and 2 folds, respectively. VPA could replace KM in human fibroblast reprogramming. The G9a methyltransferase inhibitor, BIX, could improve the OK mediated reprogramming of neural precursor cells that express high levels of endogenous SOX2, by 8 folds. The 2i inhibitors of MEK and GSK3 pathways, along with LIF, facilitated the transition of pre-iPSCs to fully reprogrammed miPSCs during OK mediated reprogramming of mouse neural stem cells.

2.5.3. Strategies for delivering reprogramming factors

For introducing reprogramming factors to the target cells, viral and non-viral approaches are used (González *et al.,* 2011; Burridge and Zambidis 2013) (**Figure 4**). Either a pool of vectors carrying individual factors or a single cassette reprogramming vector with each of the factors separated by a self-cleaving peptide signal is introduced into cells that are to be reprogrammed. Single cassette reprogramming vector ensures the delivery of all the factors in each of the infected cells at similar stoichiometry.

Viral vectors used for reprogramming include integrative, excisable or non-integrative vectors. Integrative retroviral and lentiviral vectors are highly efficient in bringing about the stable expression of reprogramming factors in the infected cells. Retroviral vectors offer the advantage that they get silenced once pluripotency is achieved whereas the lentiviral vectors continue to show variegated expression even after iPSC clones are formed (Hotta and Ellis 2008). Since lentiviruses can infect non-dividing cells efficiently, unlike retroviruses, they are preferred vehicles for expressing reprogramming factors in a wide variety of somatic cells. To address the safety concerns associated with integration of viral vectors on host genome, excisable lentiviral vectors carrying *loxP* site have been developed and these vectors can be excised out of host genome of iPSC clones by overexpression of Cre-recombinase (Chang *et al.,* 2009; Soldner *et al.,* 2009; Sommer *et al.,* 2009). The excisable lentiviral vector termed STEMCCA is widely used and ensures a reprogramming efficiency of 0.1-1.5% in human somatic cells (Somers *et al.,* 2010). The non-integrative viral

vectors employed for factor delivery are based on adenoviruses and sendai viruses. Using adenoviral vectors, only a low reprogramming efficiency of 0.001–0.0001% in mouse (Stadtfeld *et al.,* 2008b) and 0.0002% in human cells (Zhou and Freed 2009) could be achieved. Sendai viruses, being an RNA virus, remain in the cytoplasm without entering the nucleus, ensure a high level of expression of the protein and get diluted out of the cells in 10 passages after infection. Sendai-virus mediated reprogramming take about 25 days to generate iPSC colonies at an efficiency of 0.1% from blood and 1% from fibroblasts (Fusaki *et al.,* 2009; Seki *et al.,* 2010; Ban *et al.,* 2011). Unlike with retroviruses and lentiviruses, preparation of sendai viruses is tedious, but the availability of ready to use sendai viral preparation for Yamanaka factors from commercial sources have promoted usage of this factor delivery mode.

In non-viral approaches, the reprogramming factors are introduced as expression plasmids, mRNAs or proteins. These methods allow the generation of footprint-free iPSCs, overcoming the safety concerns associated with integrating vectors. By transient transfection with expression plasmids like episomal plasmids, minicircle vector, and *PiggyBac* vector, the reprogramming factors could be expressed in somatic cells for iPSC generation (González *et al.,* 2011). Episomal plasmids remain as an episome inside the cell without integrating into the host genome and will be lost as cells go through multiple rounds of replication. However, standard episomal plasmid vectors cannot ensure reprogramming factor expression long enough to bring about pluripotency induction and required repeated transfection even to achieve reprogramming with low efficiency (Okita *et al.,* 2009). Due to their ability to driving stable expression of factors for a prolonged period, OriP/EBNA based plasmids could be successful used for iPSC generation by single transfection. With a pool of three OriP/EBNA plasmid containing *OCT4, SOX2, KLF4, NANOG; OCT4, SOX2, SV40 T antigen*; and *c-MYC* and *LIN28*, human foreskin fibroblasts could be reprogrammed with a low efficiency of 0.0003-0.0006% by about day 20 (Yu *et al.,* 2009). The efficiency of OriP/EBNA plasmid vector-mediated reprogramming could be increased by ten fold by the addition of thiazovivin to cord blood cells and by passage 15, all iPSC clones lost the plasmid (Hu *et al.,* 2011). The efficiency could be further increased to 0.006-0.1% by co-expressing EBNA mRNA along

with reprogramming factors in fibroblasts under hypoxic condition in E8 medium. A new combination of oriP/EBNA vectors with *OSKML* in one vector and SV40T antigen on another could reprogram $CD34^+$ CB cells, and PB and BW mononuclear cells in the presence of sodium butyrate at an efficiency of 0.02%, 0.009% and 0.005% in about 2 weeks (Chou *et al.,* 2011). The bone marrow-derived iPSCs were found to have lost the episomal plasmid by passage 12 after colony isolation (Cheng *et al.,* 2012). Even though episomal plasmids mediated reprogramming is promising for the generation of footprint-free iPSCs, the need for modifying cell culture medium by addition of small molecules for achieving acceptable reprogramming efficiency has been a major limitation. Minicircle vector that contains only cDNAs and eukaryotic promoter to drive them has been used in reprogramming. The human adipose stromal cell could be reprogrammed with an efficiency of 0.005% in about 28 days using a minicircle vector expressing *OCT4, SOX2, NANOG, LIN28* and GFP (Narsinh *et al.,* 2011). Even though minicircle vector could also be used for reprogramming neonatal fibroblasts, the suitability of this vector in delivering reprogramming factors to other cell types is yet to be validated. Like the *piggyBac* transposons, the reprogramming vectors based on the transposon get the integrated into the host genome at the TTAA sites in the presence of transposase and get excised from there on re-expression of the enzyme, facilitating the generation of footprint-free iPSCs. The *piggyBac* vector containing Yamanaka factor could reprogram MEF with an efficiency of 0.02-0.05% in 2-4 weeks (Kaji *et al.,* 2009; Woltjen *et al.,* 2009) and vector could be excised the iPSC clones derived. A reprogramming efficiency of 0.02% could be achieved with *piggyBac* vector in human mesenchymal stromal cells (hMSCs) in the presence of sodium butyrate (Mali *et al.,* 2010). It is yet to be shown whether *piggyBac* vector can be excised from human iPSCs and can reprogramme other human cell types. mRNA and protein-based reprogramming factor delivery modes have also been employed in the generation of iPSCs from fibroblasts. Warren *et al.,* could achieve a reprogramming efficiency of 4.4% by transfecting mRNAs of LIN28 and Yamanaka factors into human fibroblasts in the presence of valproic acid at 5% O_2 (Warren *et al.,* 2010). The need for daily transfection for 7 days to achieve reprogramming is a major limitation associated with this process. Using bioactive proteins synthesized in *E.coli*, mouse (Zhou *et*

al., 2009) and human (Kim *et al.,* 2009a) fibroblasts could be reprogrammed with a low efficiency of 0.006% and 0.001% respectively. The low reprogramming efficiency and technical difficulties associated with synthesizing large quantities of bioactive proteins capable of penetrating plasma membrane are the major challenges in using proteins mediated reprogramming methods.

An ideal factor delivery approach would include vectors that can be easily prepared in the lab, omit the need for multiple rounds of factor delivery, ensure expression of reprogramming factors at the optimal stoichiometry and facilitate the formation of footprint-free iPSC. Such factor delivery methods will contribute immensely to the generation of clinical-grade iPSCs for translational research.

2.5.4. Identification and isolation of iPSCs

Due to the low efficiency of reprogramming and high heterogeneity in reprogramming cells, identification of iPSC colonies that attained all attributes of fully reprogrammed state has been technically challenging. For the identification and isolation of iPSCs from reprogramming cells, various methods based on microscopy, fluorescence activated cell sorting (FACS) and magnetic activated cell sorting (MACS) are being used (Chan *et al.,* 2009; Dick *et al.,* 2011; Abujarour *et al.,* 2013) (**Figure 4**). These methods are based on the acquisition of features similar to ESCs in morphology, activation of endogenous pluripotency genes and transgene silencing.

The most reliable method for the identification of hiPSCs is based on the morphology. The emerging hiPSC colonies attain hESC-like morphology, and the change becomes evident as the colonies increase in their cellularity. The colonies with hESC colony morphological features are identified by observing under the microscope and isolated manually or using automated systems. Even though this method has been widely used in established iPSC labs for deriving hiPSC lines, the correlation between morphology and pluripotency acquisition is yet to be explored. The need for having technical expertise in identifying the colonies exclusively based on microscopic observation is a major limitation in employing this technique in labs venturing into hiPSC research. Unlike with hiPSCs, the miPSCs in pre-iPSC and fully reprogrammed state possess morphology similar to mESCs

and hence morphology based isolation might not be right choice for deriving stable miPSC lines.

For mouse iPSCs, fluorescent reporter genes knocked into the endogenous loci of the *Nanog* or *Oct-4* pluripotency genes have been found to aid in the identification of reprogrammed colonies (Takahashi and Yamanaka 2006; Okita *et al.,* 2007; Wernig *et al.,* 2007). This identification strategy is based on the assumption that, when the reprogramming cells achieve a molecular state that activates endogenous *Oct* or *Nanog*, they express reporter proteins, thereby facilitating their identification. Similar strategies have also been used for identifying true human iPSC clones (Yu *et al.,* 2007b). Lentiviral constructs, which express a fluorescent protein under the control of early transposon promoter and *OCT-4* and *SOX2* (EOS) enhancers, aided in the identification and the isolation of both miPSCs and hiPSCs (Hotta *et al.,* 2009). The ability of PSCs to silence retroviral transgene expression were also used as an indicator of pluripotency acquisition (Nakagawa *et al.,* 2008; Chan *et al.,* 2009; Hotta *et al.,* 2009). However, this method has not been used extensively as a marker for identification and isolation of clones undergoing successful reprogramming. Though these methods aided in the derivation of pluripotency marker expressing iPSC clones, they required additional genetic manipulation of the donor cells and the ability of all the clones isolated by these methods to undergo *in-vitro* and *in-vivo* differentiation has not been investigated.

FACS or MACS of cells based on the expression of pluripotency-related surface markers is also being used for enriching the pluripotent cell population (Dick *et al.,* 2011; Valamehr *et al.,* 2012; Abujarour *et al.,* 2013; Kahler *et al.,* 2013; Quintanilla *et al.,* 2014). This omits the need for any additional genetic modifications of donor cells to express the reporter genes. Combining positive and negative surface markers also allows the selection and enrichment of iPSCs that can relatively reduce the efforts in culturing partially reprogrammed iPSC clones (Kahler *et al.,* 2013; Quintanilla *et al.,* 2014). A recent study showed that there is a high correlation between the expression of a surface marker, CD30, and endogenous NANOG in the reprogramming cells, and combining CD30 to SSEA-4 and TRA-1-81 in FACS significantly improved the specificity and efficiency of hiPSC selection

and derivation (Abujarour *et al.,* 2013). However, such enrichment methods generate heterogeneous colonies that lack clonal identity as they consist of cells originated from different donor cells. Additionally, these protocols involve single cell culture of hiPSCs that require modified culture conditions, and the clones generated from single cells often have increased the risk of karyotypic abnormalities (Valamehr *et al.,* 2012).

Live cell staining of reprogramming cells with surface markers like TRA-1-60 and TRA-1-81 could be used for isolation of intact hiPSC clones (Lowry *et al.,* 2008; Chan *et al.,* 2009; Polak *et al.,* 2012). Since it has been found that even the transgene-dependent partially reprogrammed cells express these markers (Chan *et al.,* 2009; Abujarour *et al.,* 2013; Tanaka *et al.,* 2015b), combining live cell imaging with morphology-based identification would be more reliable identification criterion.

2.5.5. Characterization of iPSCs

For the derivation of iPSC lines, the isolated clones are expanded and subjected to a series of characterization test to confirm their pluripotency (Stadtfeld *et al.,* 2010a) (**Figure 4**).

As an initial step, iPSC clones are subjected to molecular characterization by analyzing the expression of key pluripotency genes at levels comparable with hESC lines. The mRNA transcript levels of pluripotency markers like *NANOG, OCT4, SOX2, DNMT3A, ABCG2* and *REX1* are assessed by real-time PCR in hiPSCs. Also, by immunofluorescence assay, the expression of pluripotency markers like SSEA4, TRA-1-60, TRA-1-81, NANOG, OCT4, SOX2 and LIN28 at the protein level is also confirmed. The transcriptionally active state of endogenous pluripotency genes like *NANOG, OCT4* and *REX* is indicated by the hypomethylated state of corresponding promoters, and this is confirmed by sequencing PCR-based assessment of genomic DNA that is subjected to bisulfite conversion.

After confirming the transcriptionally active state of key endogenous pluripotency genes, the iPSCs are subjected to functional characterization. The developmental potential of iPSCs in terms of their ability to undergo trilineage differentiation and contribute to embryonic development are tested. The trilineage differentiation potential is tested *in vitro* by embryoid body (EB) mediated or directed differentiation and *in vivo* by teratoma assay. For

EB-mediated differentiation, the iPSC colonies are grown as suspension culture till they form cystic bodies that resemble gastrulating embryo termed as embryoid bodies. The EBs are subsequently grown on attachment culture with or without specialized matrix or feeder layer to support their further differentiation. Using specialized media containing growth factors and other nutrient supplements, iPSC or EBs can be directed to differentiate into cells belonging to specific lineages. With this regard, much progress has been made in the generation of neuronal and hematopoietic cells from iPSCs *in vitro*. The expression of lineage specific markers in the differentiated cells is confirmed by immunofluorescence, real-time PCR and/or Western blotting. In teratoma assay, the iPSC cells are injected into an immunocompromised mouse and monitored for the formation of teratoma, a benign growth that consists of cells belonging to three lineages. The teratoma is dissected and processed for histologic analysis. In addition to the pluripotency of iPSC clone, other technical factors like route and site of administration of cells and strain of mouse models used, have been found to influence the outcome of this assay. Nevertheless, teratoma assay is the gold standard test for confirming pluripotency of hiPSC lines. In the case of miPSCs, the cells are further characterized by their ability to contribute to normal development of an embryo *in vivo* by chimera formation, germline transmission, and tetraploid complementation. For chimera formation, the iPSC cells are injected into the blastocyst, and as the embryo develops, they differentiate into different cell types and get incorporated into the host system. If these iPSC chimeric mice are capable of given rise to all-iPSC cell offsprings, the potential of these iPSCs for germline transmission is confirmed. Tetraploid complementation assay is the most stringent pluripotency test for miPSCs. Here the iPSCs are injected into a tetraploid blastocyst and allowed to develop *in vivo*. Since tetraploid cells are incapable of contributing to developing embryo, the offspring borne will be entirely from iPSC introduced into a blastocyst.

2.6. Comparison between hESCs and hiPSCs

Although hESCs and hiPSCs are similar in their morphology and pluripotency properties, several studies have reported the substantial difference in epigenetic profile, gene expression, differentiation potential *in vitro* and teratoma-forming propensity (Bilic and Belmonte 2012). It is not clear whether these differences reflect inherent differences between hESCs and hiPSCs or a mere consequence of the reprogramming process. It has also been suggested the differences between ESCs and iPSCs is similar to the difference between multiple ESC lines or iPSC lines. Several factors like passage number, genetic background and reprogramming strategy, have been found to have a significant effect on gene expression and function of hiPSCs. Also, hiPSC are also found to carry copy number variations and point mutations, some of which pre-existed in somatic donor cells, whereas some were acquired during reprogramming or prolonged culture (Mayshar *et al.,* 2010). Differentiation bias is prominent in hiPSCs which preferentially take up the lineage of the donor cells from which they were derived (Kim *et al.,* 2011).

2.7. Applications of iPSCs

As a functional equivalent of ESCs, iPSCs are a potential replacement of ESCs and have tremendous applications in basic science, disease modeling, drug screening and regenerative medicine (Robinton and Daley 2012; Avior *et al.,* 2016; Trounson and Dewitt 2016) (**Figure 5**). iPSC has the advantage over ESC that it can be generated from live or frozen tissue sample of almost any individual, healthy or with disease and live or dead. Using disease corrected iPSC line as an isogenic control can nullify the effect in experimental outcome brought about by line to line specific variations.

2.7.1. Applications in basic research

The discovery of iPSCs by Yamanaka gave the ultimate proof for plasticity of cell identity. The ability of key pluripotency genes and associated factors in reverting cell fate from a differentiated state to pluripotent state confirmed their critical role in maintaining pluripotency (González and Huangfu 2015). Molecular dissection of reprogramming process

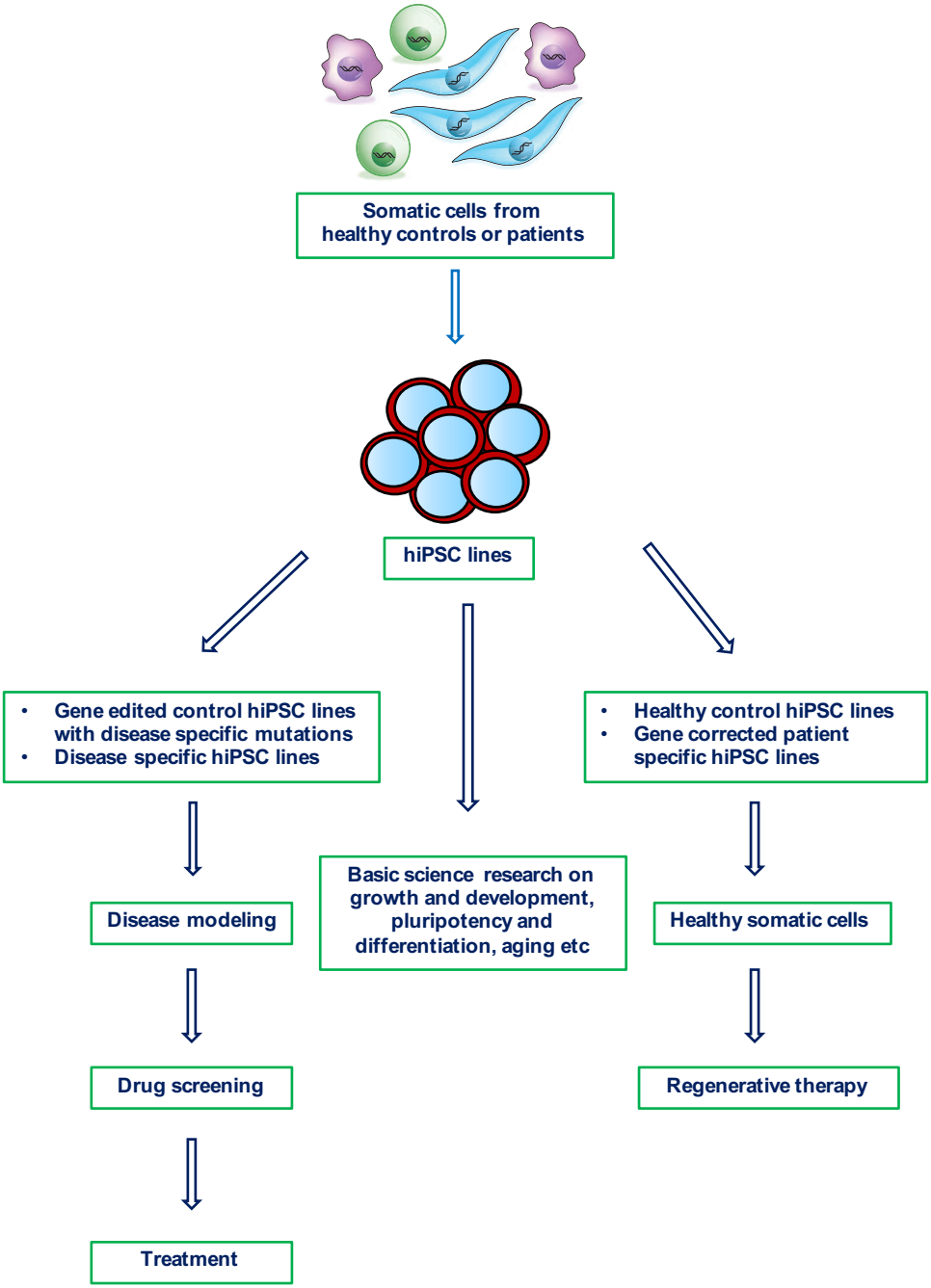


Figure 5: Applications of iPSCs
hiPSCs have potential applications in basic science research, disease modeling and drug screening. Healthy somatic cells derived from healthy control and gene corrected patient specific hiPSC lines can be used in regenerative therapy. Control hiPSC lines subjected to gene editing to carry specific genetic defects associated with the disease and unrepaired patient specific hiPSC lines can be used in modeling disease *in vitro* to understand disease pathogenesis and drug development.

and iPSCs has provided valuable insights into the relationship between epigenetic modification and cellular identity (Ho *et al.,* 2011; Liang and Zhang 2012; Smith *et al.,* 2016). It has been found that reprogramming shares common molecular events with cancer development, and hence iPSC technique can be used to understand transcriptional and epigenetic changes in cancer progression (Fernandez *et al.,* 2013). iPSC generation essentially involves rewinding the developmental clock and hence it is a promising tool for *in vitro* studies on aging.

2.7.2. Disease modeling and drug screening

Disease-specific iPSC lines have been generated from patients with neurological, cardiac, hepatic, and hematological disorders (Park *et al.,* 2008; Sterneckert *et al.,* 2014). Many of these iPSC lines could be to recapitulate certain aspects of disease pathogenesis *in vitro*. They could also be used for high throughput screening of drugs for developing therapeutics to alleviate or treat the disease. One of the first iPSC-based disease models was generated for spinal muscular atrophy (SMA) caused by loss of function of *SMN1* gene leading to motor neuron degeneration (Ebert *et al.,* 2009). The motor neurons derived from the SMA-iPSC line showed SMA characteristic nuclear germ formation and this phenotype could be alleviated by treatment with tobramycin and valproic acid. Similarly, long QT syndrome – iPSC-derived cardiomyocytes could mimic disease phenotype with longer action potential and identify novel drugs, β-adrenergic receptor inhibitors, nifedipine (a calcium channel blocker) and pinacidil (a K_{ATP} channel opener) that can potentially be used for treating the disease (Itzhaki *et al.,* 2011). iPSC-based disease modeling and drug discovery could also be achieved for some diseases like Wilson's disease, Adrenoleukodystrophy, Alzheimer disease, Parkinson's disease, Familial Dysautonomia, Niemann-Pick disease, bipolar disorders, Schizophrenia and Rett syndrome (Avior *et al.,* 2016).

iPSC-based disease modeling has several advantages over other hESC-derived disease models. iPSC lines derived from multiple patients will enable the analysis of the similar mutation in diverse genetic backgrounds. For studying the mechanism of complex disorders, individual specific iPSC disease model is the best choice, as there are several unknown genes involved. Drug screening on patient-specific iPSC lines would be helpful in

developing personalized medicine. There are a few limitations associated with hiPSC based disease models. As in the case with Fanconi anemia, disease-associated defects make the patient-derived cells refractory to reprogramming (Raya *et al.,* 2009). The differentiated cells derived from hiPSC are found to exhibit fetal features, and this may interfere with modeling diseases like β-Thalassaemia in which adult proteins are involved (Chang *et al.,* 2011). For modeling late on set diseases like Alzheimer disease and Parkinson disease, addition manipulations in culture are required to promote cellular aging *in vitro* and accelerate disease pathogenesis (Avior *et al.,* 2016).

2.7.3. Regenerative Medicine

iPSC lines generated from healthy donors can provide an unlimited supply of cells for replacement therapy. Specialized cells derived from iPSCs have been tested in vitro and *in vivo* in animal models to check their suitability for disease treatment. In one of the first proof of principle experiment, iPSC derived from tail tip fibroblast of humanized mice with sickle cell anemia was gene corrected by homologous recombination and differentiated to hematopoietic progenitor cells *in vitro* (Hanna *et al.,* 2007). On transplantation of these cells to the affected donor mice, the production of functional haemoglobin was restored leading to recovery from the disease. Retinal epithelial cells (RPE) derived from hiPSCs has been approved for a clinical trial of age-related macular degeneration (Trounson and Dewitt, 2016).

2.8. Major molecular events in successful reprogramming

Reprogramming is a gradual multi-step process that involves the loss of specialized function of somatic cells and gain of pluripotency and self-renewal properties, similar to ESCs. In the chronological order, the major molecular events associated with successful reprogramming process are down-regulation of somatic-specific programme, mesenchymal to epithelial transition (MET), activation of endogenous pluripotency network, X chromosome reactivation, erasure of epigenetic memory and exogenous reprogramming factor (or

transgene) independence (Ho *et al.,* 2011; Liang and Zhang 2012; David and Polo 2014; González and Huangfu 2015) (**Figure 6**).

2.8.1. Downregulation of somatic cell specific programme and mesenchymal to epithelial transition (MET)

Soon after the overexpression of reprogramming factors, downregulation of somatic cell specific genes is observed in the majority of the cells during miPSC and hiPSC generation (Mikkelsen *et al.,* 2008a; Stadtfeld *et al.,* 2008a; Chan *et al.,* 2009). This process is closely associated with mesenchymal to epithelial transition, which is one of the initial events in the successful reprogramming process. MET related events are observed before the expression of pluripotency markers like SSEA-1, Nanog, Sall4 and Oct3/4 (David and Polo 2014). MET associated gene expression changes include suppression of key mesenchymal genes, such as *Snail, Slug, Zeb1, Zeb2, Twist1, Twist2, N-cad* and *Fn* and the upregulation of key epithelial genes, such as *E-cadherin, Ocln, Cldn3, Dsp, Pkp1, Pkp3, Ep-CAM, Krt8* and *Krt19* (David and Polo 2014). The promoting the epithelial state, such as TGF-β inhibitors, BMPs, microRNAs, miR200s and miR302/367, and Cdh1 enhance iPSC generation, and in some cases, they can substitute for reprogramming factors (Ichida *et al.,* 2009; Samavarchi-tehrani *et al.,* 2010; Liao *et al.,* 2011; Liang and Zhang 2012). In contrast, factors that suppress the epithelial state (e.g., TGF-β) or depletion of key epithelial adhesion molecules (e.g., Cdh1, Epcam) are able to inhibit iPSC generation (Chen *et al.,* 2010; Li *et al.,* 2010; Samavarchi-tehrani *et al.,* 2010; Redmer *et al.,* 2011). Morphologically, reprograming cells that gain epithelial properties show reduced cell size and compact cell-cell interaction (Li *et al.,* 2010; Samavarchi-tehrani *et al.,* 2010; Smith *et al.,* 2010). They have been shown to escape cell cycle arrest and secure a faster division rate (Smith *et al.,* 2010; Ruiz *et al.,* 2011). Along with altered proliferation rate, a shift in energy metabolism from oxidative phosphorylation to glycolysis is initiated during this stage (Panopoulos *et al.,* 2012).

2.8.2. Activation of endogenous pluripotency circuitry

Activation of endogenous pluripotency circuitry is the central event in acquiring and establishing a pluripotent state in somatic cells subjected to reprogramming. This occurs

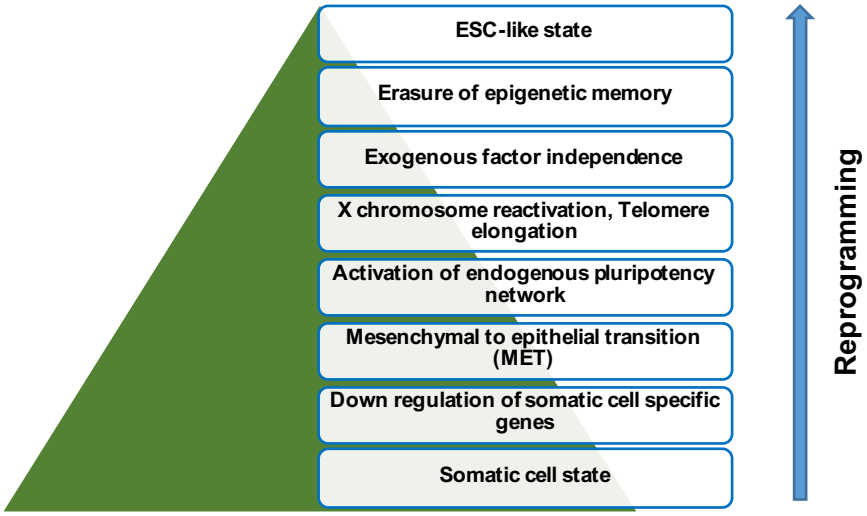


Figure 6: Molecular events in successful reprogramming to iPSCs
Reprogramming cells transit through multiple stages to achieve pluripotency. In the chronologic order, they achieve downregulation of somatic cell specific genes, mesenchymal to epithelial transition, activation of endogenous pluripotency network, X chromosome reactivation, telomere elongation, exogenous factor independence and erasure of epigenetic memory. During each stage of reprogramming, cells undergo multiple barriers and as a result only a subset of cells enter the subsequent step towards pluripotency acquisition. As majority of cells fail to overcome these barriers, the reprogramming efficiency is extremely low.

gradually with initial activation of certain pluripotency-related genes and the rest, over the course of time, to finally establish the endogenous pluripotency circuitry, thereby suggesting an activation hierarchy of pluripotency factors (Buganim *et al.,* 2012). The hierarchical activation of pluripotency genes suggests that pluripotency can be induced and established through alternative pathways, and this supports the ability of novel combinations of reprogramming factors in inducing pluripotency (Feng *et al.,* 2009b).

The pluripotency genes that are initially activated may serve as pioneer factors in establishing pluripotency network (Ichida *et al.,* 2009; Samavarchi-tehrani *et al.,* 2010; Liang *et al.,* 2012; Onder *et al.,* 2012). These include pluripotency genes like *Sox2, Esrrb, Utf1, Lin28, Dppa2, Fbxo15, Fgf4* and *Oct4.* The most upstream event in this hierarchy is the activation of *Sox2* gene which is suggested to be due to permissive chromatin at the corresponding locus (Buganim *et al.,* 2012). The strong correlation of *Esrrb, Utf1, Lin28* and *Dppa2* upregulation, even before the activation of pluripotency circuitry, with successful reprogramming outcome, suggests that they are actively involved in some rate limiting steps in reprogramming process (Samavarchi-tehrani *et al.,* 2010; Buganim *et al.,* 2012). However, activation of *Fbxo15, Fgf4* and *Oct4* does not guarantee the derivation of stable iPSCs and are found to be expressed by even partially reprogrammed colonies. Certain studies reported the activation of *Nanog* soon after MET gene activation. In line with this, NANOG, even though is not an inevitable factor in reprogramming cocktail, has been found to be indispensable for entry to pluripotency (Mitsui *et al.,* 2003; Silva *et al.,* 2009), increase reprogramming efficiency on overexpression and convert partially reprogrammed cells to a pluripotent state. Unlike *Sox2*, activation of *Nanog* occurs after chromatin reorganization induced by reprogramming factors. After the pioneer gene activation and subsequent activation of other pluripotency genes, the activation of pluripotency circuitry occurs at the later stages of iPSC formation. The exact sequence of events in this process is not clearly understood. The activation of pioneer, pluripotency genes is so critical to reprogramming that the presence of non-permissive chromatin at their loci poses a major epigenetic barrier for the subsequent activation of pluripotency circuitry. Upon activation of pluripotency circuitry, reprogramming cells achieve self-renewal ability.

2.8.3. X chromosome reactivation

The somatic cells from females show random inactivation of one of the X chromosomes (XCI) and as a result, only single X chromosome remains active. During reprogramming of mouse somatic cells, silenced X chromosome get reactivated so that miPSCs possess two active X chromosomes like mESCs (Maherali *et al.,* 2007). This event occurs in late stages of reprogramming alongside with the activation of endogenous *Nanog* and *Oct4* (Stadtfeld *et al.,* 2008a). On the contrary, hiPSCs retain the inactive X chromosome from the somatic state (Tchieu *et al.,* 2010) and on extended culture, some hiPSCs clones had been found to show dosage erosion wherein the inactive X chromosome get derepressed (Mekhoubad *et al.,* 2012). This is consistent with XCI observed in different female hESC lines, wherein many lines show complete XCI in an undifferentiated state and a few lines partially carry two active X chromosomes (Silva *et al.,* 2008b).

2.8.4. Exogenous factor independence

On successful reprogramming, the iPSCs retain their pluripotent state with the aid of established endogenous pluripotency network independently of exogenous reprogramming factors, confirming stable conversion of cell fate (Brambrink *et al.,* 2008; Stadtfeld *et al.,* 2008a). Exogenous factor independence is achieved in the late stages of reprogramming, and premature depletion of the exogenous factors revert the reprogramming cells to their initial differentiated phenotype or partially reprogrammed state. This need not necessarily be associated with transgene silencing. In retroviral mediated reprogramming, transgene silencing is very efficient (Maherali *et al.,* 2007) probably through the actions of Trim28, Zfp806, and histone and DNA methyltransferases (Lei *et al.,* 1996; Wolf and Goff, 2009; Matsui *et al.,* 2010; Rowe *et al.*, 2010), whereas constitutively active lentiviral vectors, continued to express transgenes in established iPSC lines. Even though transgene silencing is not an absolute requirement for establishing a self-renewing pluripotent ESC-like state, persistent expression of exogenous factors after establishing endogenous pluripotency network in iPSCs impairs the differential potential of these cells (Brambrink *et al.,* 2008; Sommer *et al.,* 2010).

2.8.5. Erasure of epigenetic memory

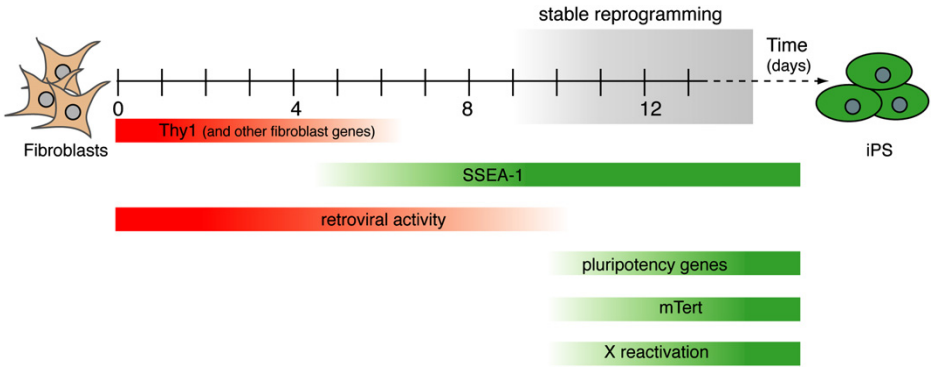
Despite being pluripotent, the isolated iPSC in the initial passages have been found to retain certain epigenetic traits characteristic of the donor cells, and this epigenetic memory has been found to influence their differentiation potential. These clones display an increased propensity to differentiate into their original lineage compared to other lineages, and this trend might be favoured by the gene expression profile and chromatin state resembling that of starting cells (Kim *et al.,* 2010; Ohi *et al.,* 2011; Kim *et al.,* 2012). The epigenetic memory in the iPSCs also contributes to their molecular identity distinct from ESCs (Ghosh *et al.,* 2010). Upon expansion in culture over several passages, the epigenetic memory is gradually lost thereby diminishing the disparities between ESCs and iPSCs in molecular and functional properties (Chin *et al.,* 2010; Polo *et al.,* 2010). The erasure of epigenetic memory could be accelerated by exposure to drugs that modulate the activity of epigenetic factors (Kim *et al.,* 2010).

2.9. Expression kinetics markers used for defining stages of reprogramming

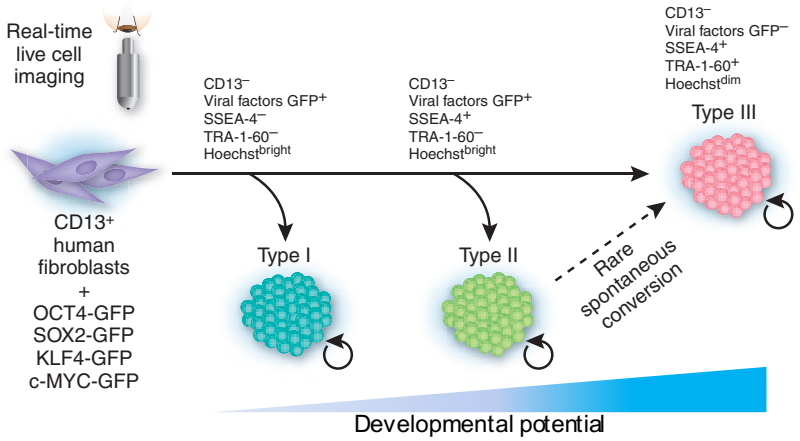
For the identification of markers that can define stages of reprogramming, the reprogramming cells were monitored by time-lapse fluorescent imaging and FACS analysis. In miPSC reprogramming, Thy1, CD44, ICAM, SSEA-1, retroviral fluorescent reporter and endogenous NANOG, SOX2 or OCT4 coupled with a fluorescent reporter have been used to define stages of reprogramming (Stadtfeld *et al.,* 2008a; O'Malley *et al.,* 2013). In hiPSC reprogramming, to serve the same purpose, the fibroblast marker CD13, pluripotency markers SSEA-4, TRA-1-60, TRA-1-81 and CD30 and retroviral fluorescent reporter have been used (Takahashi *et al.,* 2014; Tanaka *et al.,* 2015a).

In one of the first studies on expression kinetics of markers during reprogramming, a doxycycline (DOX)-inducible secondary reprogramming system in MEF was used (Stadtfeld *et al.,* 2008a) (**Figure 7**). When the expression kinetics of Thy1 and SSEA-1 was followed, it was found that 30%, 50% and 70% of cells achieved $Thy1^-$ state by day 3, day 5 and day 7

A.



B.



C.

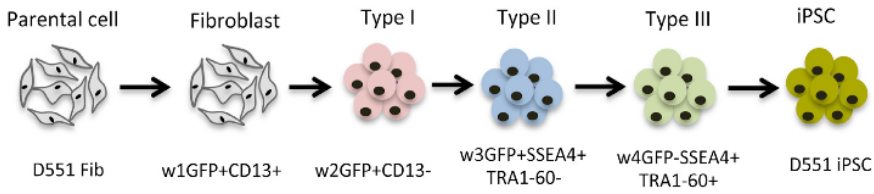


Figure 7: Expression kinetics of markers used for defining stages of reprogramming

Schematics of the expression kinetics of markers during **(A)** miPSC reprogramming and **(B-C)** hiPSC reprogramming. (Adapted from **(A)** Stadtfeld *et al.*, 2008, **(B)** Chan *et al.*, 2009 and **(C)** Tanaka *et al.*, 2015).

following DOX addition, respectively. About 1-5% of Thy1$^-$ cells were found to initiate SSEA-1 expression by day 5, and by day 7, 5%– 10% of total cells showed an SSEA-1$^+$Thy1$^-$ phenotype. A week after DOX withdrawal, SSEA-1$^+$Thy1$^-$ miPSC colonies were formed. Thus, it was found that Thy1 down-regulation is an early event during the reprogramming process that occurs in the majority of cells, and SSEA-1 expression is initiated in a small fraction of Thy1$^-$ cells, first at low and eventually at high levels. On reprogramming *mTert-GFP*, *X-GFP* and *Sox2-GFP*, it was found that in a small fraction of SSEA-1$^+$ cells, the reactivation of the pluripotency transcription factor, Sox2, the silenced X chromosome, and telomerase occurs. The occurrence of these late events correlated with the period when cells start to become independent of DOX, and thus exogenous factor expression. Using *Sox-GFP* MEF labelled with retroviral tdTomato, it was identified that retroviral transgene silencing is a gradual process that gets completed in established clone. In a similar study on MEF reprogramming with the DOX-inducible secondary system, it was found that SSEA-1 cannot accurately delineate reprogramming process Nanog-eGFP+ cells were found to originate from both SSEA-1$^+$ and SSEA-1$^-$ fractions (O'Malley *et al.,* 2013). It was found that the appearance of CD44$^-$ and ICAM1$^+$ cells at later time points closely correlated with Nanog–eGFP expression. By coupling the markers, CD44 and ICAM1, with NANOG-GFP, the intermediate reprogramming subpopulations could be isolated successfully.

By serial live cell imaging of human fibroblast subjected to reprogramming, three distinct types of expandable hESC-like colonies formed during reprogramming were identified via expression patterns of virus-derived GFP, fibroblast marker CD13 (ANPEP), and two pluripotent markers SSEA-4 and TRA-1-60 (Chan *et al.,* 2009) (**Figure 7**). Type I cells are defined by continuous expression reprogramming genes (CD13$^-$GFP$^+$SSEA-4$^-$TRA-1-60$^-$). Type II cells express pluripotency marker SSEA-4 and continue expressing reprogramming factors (CD13$^-$GFP$^+$SSEA-4$^+$TRA-1-60$^-$). Type III cells show expression of TRA160 as well as SSEA4 (CD13$^-$GFP$^-$SSEA-4$^+$TRA-1-60$^+$). Among these types of colonies, only type III has similar molecular phenotypes with hESCs and become bona fide hiPSCs. Type I and type II cells are partially reprogrammed cells and display negative nuclear NANOG staining, low expression of several pluripotency related proteins (e.g., DNMT3B and

REX1), and a distinct epigenetic state from type III cells and hESCs. Type I cells remain in their incomplete reprogramed state while a small population of type II cells may still convert to type III cells and complete hiPSC reprogramming. Using the same marker combinations, the primary human fibroblasts undergoing reprogramming with pMSCV-IRES-GFP-based retroviral vectors expressing OSKM were analyzed by FACS to follow the expression kinetics of markers (Tanaka *et al.,* 2015a) (**Figure 7**). Based on the expression profile of pluripotency genes in the different cell fractions analysed, it was found that $CD13^{-}GFP^{-}$ $SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{+}$ represents the type III cells.

2.10. Methods used to dissect the molecular mechanism of reprogramming

For unraveling the molecular mechanism of reprogramming, different approaches are currently being used. The study methods used, their advantages and limitations are described below.

2.10.1. Analysis of molecular changes in reprogramming cells on factor overexpression

In the most straight forward approach, transcriptional and epigenetic changes in total cell populations at different time points after factor induction have been analyzed (Koche *et al.,* 2010). Such studies have helped to understand the global effect of reprogramming factor overexpression in somatic cells. By following the reprogramming cells using time-lapse imaging, insights into the morphological changes and surface marker expression kinetics during reprogramming could be obtained (Chan *et al.,* 2009; Smith *et al.,* 2010). The information gained from such studies have been very useful in identification and isolation of reprogramming intermediates and iPSC colonies. Molecular dissection of partially reprogrammed colonies, which have achieved only a few features of pluripotent cells, has served in the identification of molecular barriers in reprogramming (Mikkelsen *et al.,* 2008a; Silva *et al.,* 2008a). By testing the ability to convert partially reprogrammed colonies to a fully reprogrammed state, the role of novel factors in reprogramming was identified.

A more precise understanding of the transcriptional changes that occur at the

different stages of reprogramming could be achieved by analyzing the intermediate cell populations isolated based on specific surface marker expression (Hansson *et al.,* 2012; O'Malley *et al.,* 2013; Takahashi *et al.,* 2014; Tanaka *et al.,* 2015a). For molecular analysis, cells expressing different combinations of these markers or a specific pluripotency marker at a specific time point or different time points during the course of reprogramming are used (Stadtfeld *et al.,* 2008a; Polo *et al.,* 2012; O'Malley *et al.,* 2013; Tanabe *et al.,* 2013; Takahashi *et al.,* 2014; Tanaka *et al.,* 2015a).

Since reprogramming is an extremely inefficient process, the cells undergoing successful reprogramming constitute less than 1% of the total population. In an attempt to overcome the problem of cell heterogeneity interfering the dissection of the molecular basis of reprogramming, the reprogramming process has also been studied at single-cell resolution. Two single-cell techniques, Fluidigm BioMark and a single-molecule-mRNA fluorescent *in situ* hybridization (sm-mRNA-FISH), have been used to quantify gene expression in the rare cells that undergo reprogramming (Raj *et al.,* 2008; Polo *et al.,* 2012; O'Malley *et al.,* 2013).

2.10.2. Manipulation of specific factors by molecular intervention

Molecular interventions of pathways associated with reprogramming events, by overexpression or knockdown of specific genes, or activation or inhibition of specific proteins, have identified several novel factors involved in reprogramming (Li and Rana 2012; Mancarci *et al.,* 2012; Zhang *et al.,* 2012; Qin *et al.,* 2014). By manipulating the expression of these genes or activity of proteins, their stage-specific role in pluripotency induction could be identified and reprogramming efficiency could be improved.

One of the major considerations in gene regulation studies is the extent of exogenous transgene expression or endogenous gene knockdown achieved in the target cells. This, in turn, is found to be mainly dependent on several functional elements in the transfer vectors. One of the major factors that determine the efficiency is the type of the promoters used. However, it has been reported that these promoters varying in their strength and susceptibility to silencing in somatic cells, ESCs and ESC-derived cells. The high, stable and persistent transcriptional activity of EF1*α* promoters during propagation and

differentiation of mammalian ESCs have been reported by multiple studies (Kim *et al.,* 2007; Wang *et al.,* 2008). EF1*α* showed high level of transgene expression in mouse tail fibroblasts (129TF), MEF, mouse myoblasts (C2C12), rat mesenchymal stem cells (MSC), human fibroblasts (MRC5), human fibrosarcoma cells (HT1080), human embryonic kidney cells (293T), and rhesus macaque mammary tumor cells (CMMT) and this was comparable with CAGG and higher than SV40, PGK and UbC (Qin *et al.,* 2010). In this study, it was observed that UbC was consistently the weakest promoter in all the cell types while PGK was also consistently weak, though typically stronger than UBC. CMV promoter is the most variable, being very strong in some cell types (e.g., 293T and CMMT) and rather weak in others (e.g., MRC5). When multiple promoters with single copy integration to *ROSA* locus in mESCs were compared, CAG was found to yield the highest levels of expression at approximately 9–10 fold the level of the endogenous *ROSA26* promoter (Chen *et al.,* 2011). This was followed by EF1*α* (4.5 fold), UbC (3-5 fold), CBA, PGK, MC1 (1 fold) and CMV (0.1 - 0.3 fold). Even though CMV promoter strongly drives transgene expression in several mammalian cells, CMV promoter drove GFP expression at very low levels in mouse, human and monkey ESCs (Kim *et al.,* 2007). In human $CD34^+$ hematopoietic stem/progenitor cells, the highest level of GFP expression was driven from the MSCV LTR promoter, followed by the PGK, GALV LTR, EF1*α*, CMV promoter, and CAG promoters (Ramezani *et al.,* 2000).

Since reprogramming involves the continuous transition of cell states to reach pluripotency, depending on the type of promoters and the reprogramming stage to which the cells belong, the extent of gene regulation brought about by these vectors is likely to vary. This alteration in promoter activity during reprogramming will potentially influence the reprogramming outcome and interfere with the assessment of the role of novel factors in reprogramming. Thus, it is important to choose the promoter in overexpression and knockdown vectors used for elucidating the molecular mechanism of reprogramming. However, so far, there are no studies on evaluating the promoter activity in cells undergoing reprogramming.

2.10.3 RNAi screening

A few studies have reported RNAi screening using shRNA libraries to identify novel factors in reprogramming (Qin *et al.,* 2014; Sakurai *et al.,* 2014; Yang *et al.,* 2014). This method involves transducing the somatic cells with shRNA libraries and then subjecting the cells to factor mediated reprogramming. The reprogramming cells are isolated based on the expression of specific cell surface markers at multiple time points. From the sorted cell fraction, DNA is isolated and subjected to high throughput analysis to identify the shRNAs that are enriched or depleted in isolated cell fractions. The enrichment of a specific shRNA in the pluripotent fraction indicates that the target gene of that particular shRNA act as a barrier to reprogramming. By a kinome-wide RNAi based analysis, 59 kinases that act as barriers in MEF reprogramming were identified (Sakurai *et al.,* 2014). Using a genome-wide RNAi screen, specific genes that are associated with transcription, chromatin regulation, ubiquitination, dephosphorylation, vesicular transport, and cell adhesion were identified as barriers in hiPSC reprogramming (Qin *et al.,* 2014). The potential of RNAi screen is yet to explored to identify stage specific role of novel factors in reprogramming.

2.11. Mechanistic models of reprogramming

The stochastic and deterministic models have been proposed to explain the mechanism of reprogramming (Yamanaka 2009) (**Figure 8**). In the “stochastic” mode, iPSCs appear with variable latencies, and it cannot be predicted whether or when a given cell would become an iPSC. The stochastic model is strongly supported by the single-cell cloning experiments which demonstrated that sister cells from an early colony generate iPSCs with variable latency and with some sister cells never giving rise to iPSCs (Hanna *et al.,* 2009a). This stochasticity may reflect differences in cell cycle stage, cell-intrinsic fluctuation in gene expression and epigenetic status. Acceleration of reprogramming kinetics in a cell division rate-dependent manner further supports the stochastic nature of reprogramming. In “deterministic” mode, the reprogrammed cells would be generated with a fixed latency. This mode of reprogramming is displayed by granulocyte monocyte progenitor

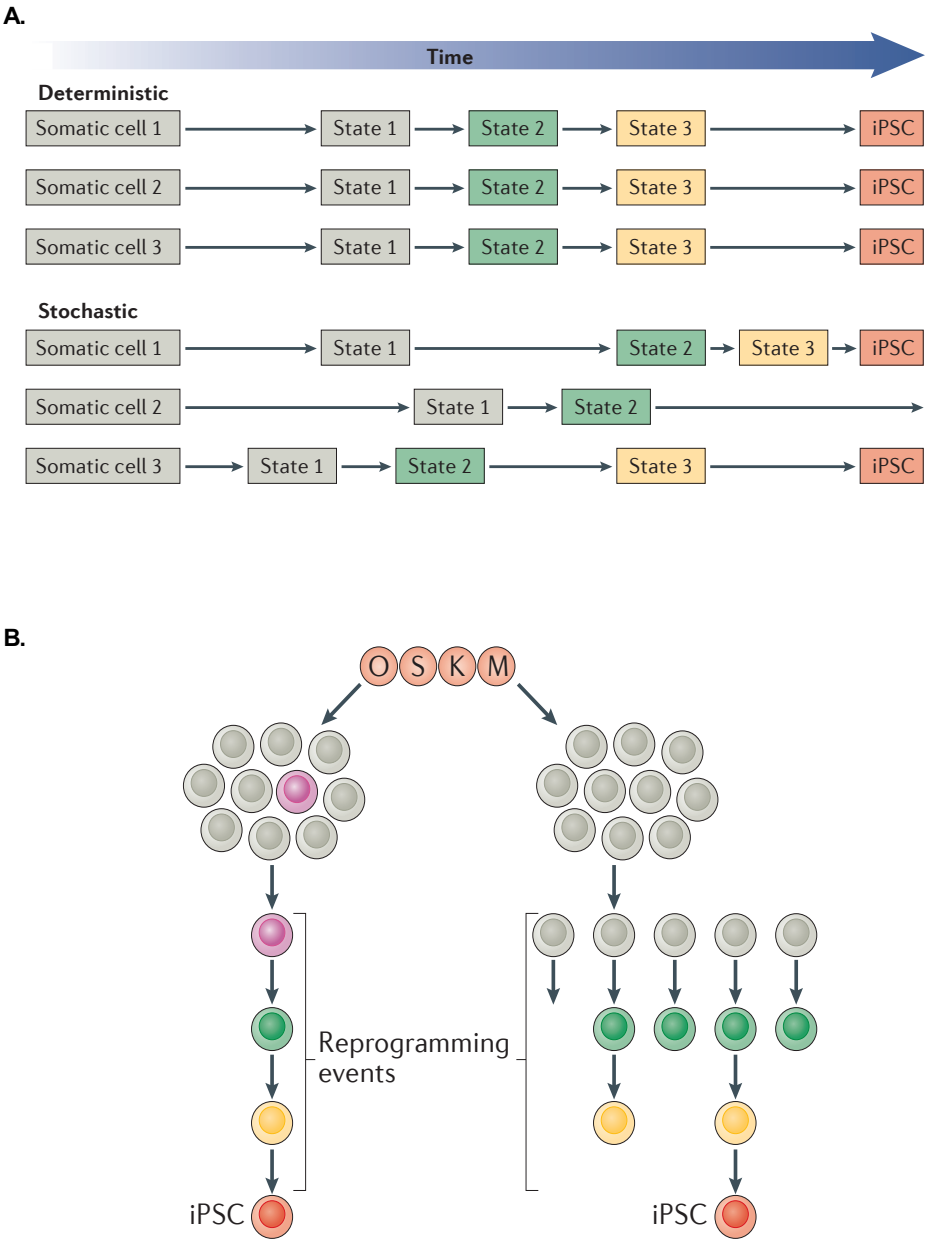


Figure 8: Mechanistic models of reprogramming
(A) In deterministic model, reprogramming cells follow a same route to pluripotency and become iPSCs with a fixed latency. In stochastic model, reprogramming cells take alternate routes to pluripotency and achieve iPSC state with different latencies. **(B)** In elite model, only a small number of cells present within the starting cell population (shown in magenta) has the potential to become iPSCs under the influence of reprogramming factors. An alternate model suggests that reprogramming is initiated in many cells, but majority of the cells fail to transverse through the multi-step process, and only a few cells become iPSCs. (Adapted from Takahashi *et al.*, 2016.)

(GMP) cells with an ultrafast cell cycle of 8 hours, giving rise to iPSC colonies almost synchronously in only four to five cell divisions upon OSKM expression (Guo *et al.,* 2014). Similarly, chromatin derepression through Mbd3 depletion allows rapid, deterministic reprogramming of MEFs (Rais *et al.,* 2013).

Both of these models are further divided into subcategories where 'all' or only a subset of 'elite' cells are permissive to reprogramming (**Figure 8**). The evidence for all cells permissive for reprogramming comes from the observation that >92% of the pre-B cell can give rise to reprogrammed colonies if the reprogramming factors are expressed for enough time (Hanna *et al.,* 2009a). The extremely low reprogramming efficiency led to the suggestion that only the elite cells in donor population were capable of reprogramming. The elite cells may be either predetermined or induced upon viral delivery. The existence of predetermined elite cells capable of reprogramming comes from the fact that hematopoietic stem and progenitor cells generate up to 300 times more iPSC colonies than terminally differentiated B and T cells (Eminli *et al.,* 2009). The idea of elite cells getting induced upon viral delivery is based on the fact that random transgene integration cause heterogenous transgene expression and only the rare cells that achieve optimal expression, and stoichiometry of the reprogramming factors get reprogrammed. However, about >90% of the secondary reprogrammable MEFs (Wernig *et al.,* 2008) or transgenic reprogrammable MEFs with a doxycycline (dox)-inducible polycistronic OSKM expression cassette integrated into the *Collagen* (*Col1a1*) locus (Carey *et al.,* 2010; Stadtfeld *et al.,* 2010b), failing to reprogram argue against the concept of an elite model of reprogramming.

2.12. Molecular mechanism of reprogramming somatic cells to iPSCs

Reprogramming of somatic cells to iPSCs is a complex process that involves multiple genetic-epigenetic signaling pathways to induce pluripotency, self-renewal and loss of specialized cellular functions. This involves the activity of a large set of transcription factors, chromatin- and DNA-modifying enzymes, non-coding RNAs, and signal transduction pathways (**Figure 9**).

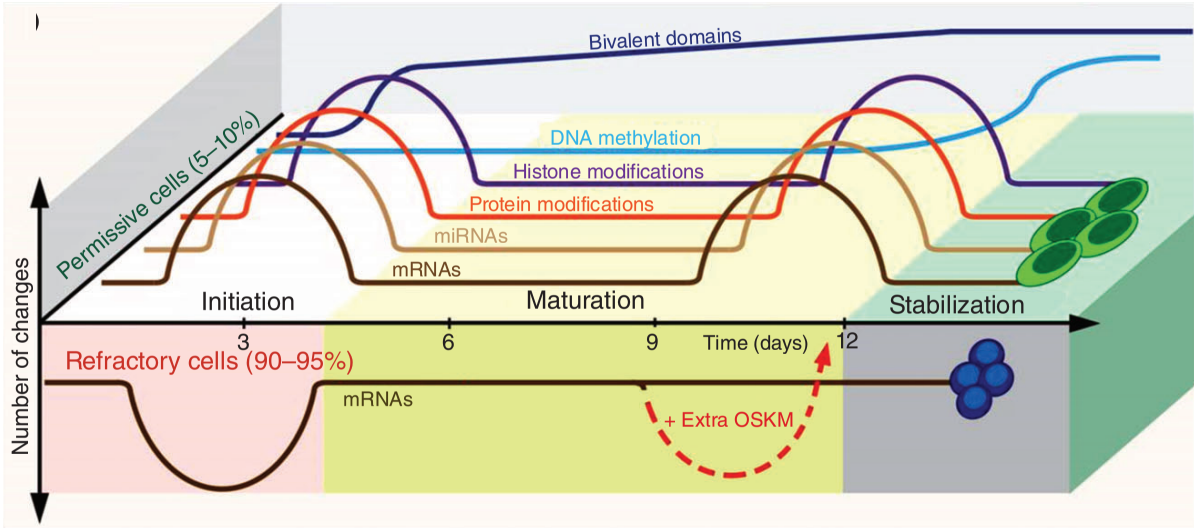


Figure 9: Dynamics of molecular changes in pluripotency induction during reprogramming

A schematic diagram indicating the dynamics of molecular changes during the initiation, maturation and stabilization phases of reprogramming. Permissive reprogramming cells (positive y-axis) show a biphasic pattern of protein, mRNA, and microRNA (miRNA) expression changes following the kinetics of individual histone modifications throughout the initiation phase and during the late maturation phase of reprogramming. Establishment of bivalent domains occur gradually throughout the three phases, with an initial burst in initiation phase. DNA methylation changes occur predominantly at the late maturation phase and stabilization phase. Refractory cells (negative y-axis) undergo a similar wave of expression changes during the initiation phase but remain relatively stable thereafter. Forced expression of OKSM in refractory cells can rescue their ability to form iPSCs. (Adapted from González *et al.*, 2016.)

2.12.1. Role of transcription factors

OCT4, SOX2 and NANOG form core transcriptional regulatory network in ESCs and they co-occupy promoters of key pluripotency genes that are up-regulated in ESCs and also the developmental genes that are silenced in pluripotent cells (Jaenisch and Young 2008). They also form an autoregulatory network binding to each other's and their own promoters to maintain transcription regulatory activities. cMYC is associated with cell cycle regulation and it promotes cell cycle progression by downregulating cell cycle checkpoint genes by inhibiting cyclin-dependent kinase inhibitors and promoting DNA synthesis by associating with prereplication complexes (Meyer and Penn 2008). During reprogramming, a drastic change in the promoters occupied by OSK occurs with changes in the expression profile of the respective genes (Sridharan *et al.,* 2009; Koche *et al.,* 2010; Soufi *et al.,* 2012). In the initiation phase, OSK bind to promoters and enhancers of active and repressed genes, of which half of the enhancers are specific for PSCs (PSCs), suggesting the critical role of OSK in opening chromatin at the pluripotency loci. Through the interaction with Mediator, Cohesin complexes or RNA Pol II elongation factor ELL3, OSK might be recruiting these factors to non-canonical enhancers and promoters bringing about chromatin opening (Kagey *et al.,* 2010; Buganim *et al.,* 2013; Lin *et al.,* 2013). During the stabilization phase, gene promoter occupancy of OSK in reprogramming cells looks similar to that in PSCs, confirming their role in establishing pluripotency in iPSCs. KLF4 is involved in somatic gene repression initially and in pluripotency gene activation in later stages (Polo *et al.,* 2012). The role of cMYC in favouring reprogramming is thought to be mainly through regulating cell proliferation. cMYC activity has been found to be restricted to initiation phase of reprogramming and all cMYC binding sites get occupied during this phase. The increase in proliferation rate brought about by cMYC may help the cells to reset the genome with the help of OSKM. It has also been found that cMYC brings about global amplification of gene expression at all active promoters, and they may be enhancing transcription of genes activated by OSK (Lin *et al.,* 2012; Nie *et al.,* 2012; Soufi *et al.,* 2012).

In addition to OSKM, several other pluripotency-related genes have been found to be capable of inducing reprogramming. *Nanog, Lin28, Glis1, Sall4,* and the orphan nuclear

receptor *Esrrb* are some of such factors (Yu *et al.,* 2007b; Tsubooka *et al.,* 2009; Maekawa *et al.,* 2011). NANOG homeodomain transcription factor is part of the core pluripotency transcriptional network and is required for the maintenance of pluripotency in epiblasts and ESCs (Pan and Thomson 2007). NANOG is involved in the transition from maturation phase to stabilization phase during reprogramming of human mesenchymal cells by OCT4, SOX2, NANOG and LIN28 (OSNL) (Yu *et al.,* 2007b) and mouse neural stem cells (NSCs) by OSK (Silva *et al.,* 2009). NANOG accelerates OSKM-mediated reprogramming of mouse B cells in a cell-division-rate-independent manner (Hanna *et al.,* 2009a). LIN28 blocks maturation of *let-7* miRNAs, which are negative regulators of cycle-regulating miRNAs associated with ESC self-renewal. LIN28 has been shown to mediate post-transcriptional regulation of *Oct4* in hESCs and cell-cycle regulators in mESCs (Shyh-Chang and Daley 2013). So LIN28 in the OSNL set may be a functional analogous to that of cMYC in OSKM. Lin28 is not inevitable in OSNL combination, and its removal brings about a five-fold reduction in the number of ESC-like colonies formed, might be improving reprogramming through both miRNA- dependent and -independent mechanisms (Yu *et al.,* 2007b; Viswanathan and Daley 2010). GLIS1, a transcription factor enriched in unfertilized oocytes and one-cell stage embryos, was found to promote OSK and OSKM mediated reprogramming of mouse and human fibroblasts (Maekawa *et al.,* 2011). The ability of GLIS1 to activate pluripotency genes like *Foxa2, Essrb, Lin28a* and *Nanog*, and *Myc* family genes like *MycN* and *Mycl1* but downregulation of cMyc expression might be responsible for its effect on reprogramming. SALL4, with overlapping binding sites with OSN, is required for pluripotency maintenance in ESCs and its knockdown lead to differentiation of ESCs to trophectoderm related cells. Sall4 overexpression gave two-fold more $NANOG^{+}$ iPSC colonies on OSK mediated reprogramming of MEFs and *Sall4* knockdown gave twofold reduction (Tsubooka *et al.,* 2009). ESRRB targets in ESCs include self-renewal and pluripotency genes (Ivanova *et al.,* 2006; Loh *et al.,* 2006), and it localizes with KLF4 (Feng *et al.,* 2009a). It was found to be capable of replacing KLF4 in OSKM mediated reprogramming of MEF though OSEM exhibited a twofold decrease in efficiency (Feng *et al.,* 2009a).

To maintain pluripotency, in addition to repressing key developmental regulators

(Young 2011), the pluripotency factors may act as early lineage specifiers and the balance of their antagonistic activities results in the maintenance of the undifferentiated state (Loh and Lim 2011). Supporting this observation, overexpression of GATA3 (a mesendoderm-lineage specifier) and GMNN (an ectoderm-lineage specifier) could replace Oct4 and Sox2 in OSKM mediated reprogramming of MEF (Shu *et al.,* 2013). A similar mode of action has been identified for other lineage transcription factors in miPSC and hiPSC reprogramming indicating a more complex mechanism of pluripotency induction.

2.12.2. Role of epigenetic remodeling factors

Reprogramming aims at switching the molecular, cellular and functional identity of somatic cells to a pluripotent state by global epigenetic resetting (**Figure 10**). Epigenetic modifications associated with reprogramming involve chromatin remodeling, post-translational modification of histones and DNA methylation/ demethylation. These changes are essential for gaining access to the host genome for altering its transcriptional activity thereby inducing pluripotency. The major epigenetic events in reprogramming include silencing of somatic cell specific gene expression profile, activation of endogenous pluripotency network, maintenance of imprinting status, reactivation of X chromosome and silencing of exogenous factors. Coordinated action of the following epigenetic modifiers brings about these changes resulting in induction of pluripotency in terminally differentiated cells.

Histone modifiers

Histone modifiers like histone methyltransferases, demethylases and acetyltransferases have been found to be associated with reprogramming events. Through addition and removal of methyl or acetyl groups on specific amino acid residues like H3K4, H3K9, H3K27 and H3K79, these histone modifiers regulate gene expressions associated with reprogramming.

Knockdown of *Wdr4*, an H3K4me3 methyl transferases in Trithorax group (TrxG) complexes, brought about a decrease in occupancy of WDR5 on chromatin, the global reduction of H3K4me3, a locus-specific decrease of H3K4me3 at pluripotency-associated gene promoters including those of *Oct4* and *Nanog*, and a decrease in the efficiency of

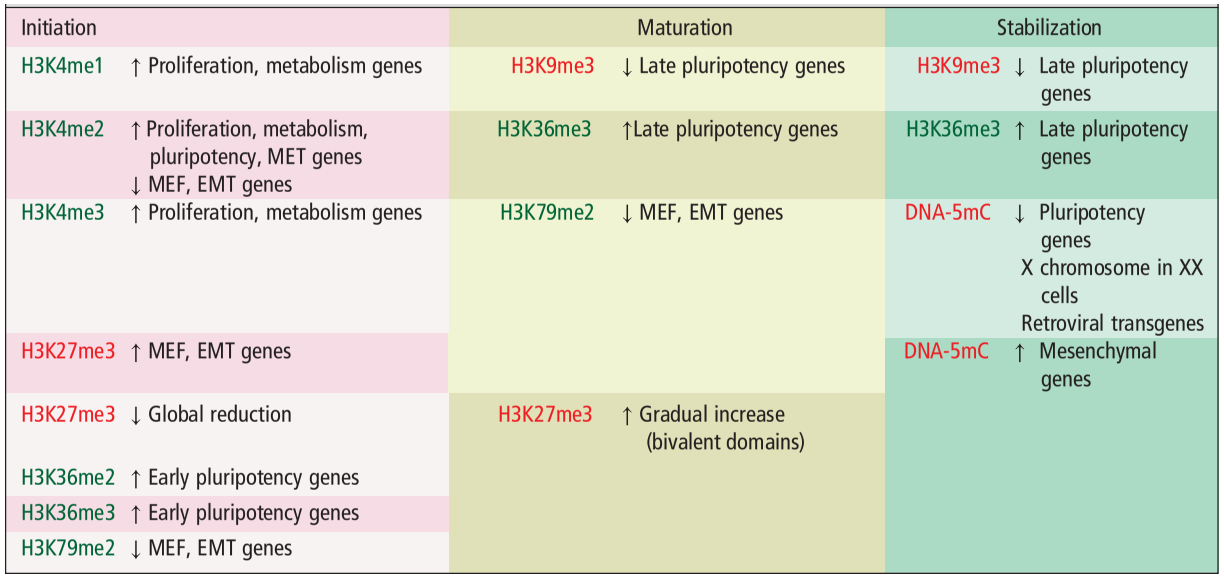


Figure 10: Epigenetic modifications during reprogramming

H3K4me1/2/3, H3K36me2/3 and H3K79me2 (in green) are generally associated with transcriptionally active loci and H3K9me3, H3K27me3 and DNA-5me (in red) are generally associated transcriptionally repressed loci. In the initiation phase, the epigenetic changes bring about activation of proliferation, metabolism and early pluripotency associated genes, and repression of MEF and EMT genes. In the maturation phase, late pluripotency genes get activated through increase in H3K36me3 and decrease in H3K9me3 marks. With a gradual increase in H3K27me3, bivalent domains get established on developmental genes. In the stabilization phase, the epigenetic changes bring about an activation of late pluripotency genes through enrichment of active marks and depletion of repressive marks. A decrease in DNA methylation was associated with X chromosome reactivation and an increase was associated with retroviral transgene silencing and downregulation of mesenchymal genes. Arrows pointing up or down indicate an increase or a decrease in the corresponding epigenetic mark, respectively. (Adapted from González *et al.*, 2016).

reprogramming in MEFs. During reprogramming Wrd5 is upregulated through direct binding of OCT4 to its promoter and OCT4 facilitate binding of Wdr5 at pluripotency loci to reestablish H3K4me3 active marks leading to robust transcriptional activation (Ang *et al.,* 2011).

The addition of H3K9me3 mark which is usually associated with repressed heterochromatin, bring about gene repression. Depletion of H3K9 methyltransferases, *Ehmt1, Ehmt2* or *Setdb1* and the H3K9me3 binding heterochromatin protein, CBX3 has been found to increase miPSC formation (Sridharan *et al.,* 2013). This is in line with the fact that H3K9me3-containing regions prevent OSKM binding in human fibroblasts during reprogramming (Soufi *et al.,* 2012), and their presence at pluripotency loci represents a major roadblock in the transition from pre- to fully reprogrammed mouse iPSCs (Chen *et al.,* 2012; Sridharan *et al.,* 2013). H3K9 methyltransferases seem to act downstream of bone morphogenetic proteins (BMPs), and together with their corresponding demethylases, they constitute a switch controlling the transition from pre-iPSC to iPSC by regulating H3K9 methylation levels at core pluripotency loci (Chen *et al.,* 2012).

Knockdown of H3K27 methyltransferases belonging to the PRC1 (BMI1 and RING1) and PRC2 (EZH2, EED, and SUZ12) PcG complexes substantially reduces the number of iPSC colonies obtained from human fibroblasts (Onder *et al.,* 2012). This is consistent with the role of PcG complex in transcription repression of the developmental genes regulated by OSN in the pluripotent stem cell. Further studies have shown that EZH2 negatively regulates transforming growth factor - β (TGF-β) signaling components and pro-EMT miRNAs such as the miR-23a cluster through association with cMyc during the initiation phase of reprogramming in human fibroblasts (Rao *et al.,* 2015). Inhibition of the H3K27 demethylase UTX (also known as KDM6A) and the H3K36 demethylases JHDM1A and JHDM1B (also known as KDM2A and KDM2B) also decreases reprogramming efficiency. Utx has been shown to promote reprogramming through physically interacting with OSK and removing H3K27me3 from early pluripotency genes such as *Fgf4*, *Sall4*, or *Sall1* (Mansour *et al.,* 2012). JHDM1B plays a prominent role during the initiation phase of reprogramming when it enhances the activation of early responsive genes through binding and demethylating their

promoters (Liang *et al.,* 2012). Moreover, JHDM1A and JHDM1B were also shown to mediate the enhancing effect of vitamin C on reprogramming, by repressing the *Ink4/Arf* locus, increasing proliferation, and suppressing senescence (Wang *et al.,* 2011).

Inhibition of the H3K79me2 methyltransferase DOT1L promotes reprogramming and can replace the reprogramming factors KLF4 and cMYC in OKSM (Onder *et al.,* 2012). These effects are likely mediated through early upregulation of NANOG and LIN28 and loss of H3K79me2 at EMT genes (*SNAI1, SNAI2, ZEB1, and TGFB2*).

Histone variants

Manipulation of expression levels of histone variants associated with DNA damage response (H2A.X) and heterochromatin (MacroH2A) and enriched in oocytes (TH2A, and TH2B) has been found to influence reprogramming efficiency.

It has been shown that H2A.X depletion leads to decreased reprogramming efficiency, and that aberrant accumulation of H2A.X marks poor-quality iPSCs (Buganim *et al.,* 2014; Wu *et al.,* 2014). This effect on reprogramming outcome might be due to their role in DNA damage response brought about by ectopic expression of the reprogramming factors, wherein an increase in H2A.X phosphorylation (γ-H2A.X) is observed during the initiation phase of reprogramming (Strati *et al.,* 2009; Müller *et al.,* 2012a; González *et al.,* 2013).

More efficient reprogramming has been observed in macroH2A1- and macroH2A2-depleted mouse NSCs (Pasque *et al.,* 2012) and double knockout mouse fibroblasts (Gaspar-Maia *et al.,* 2013). This is consistent with the observation in mouse fibroblasts that macroH2A1 and macroH2A2, together with H3K27me3, co-occupy repressed pluripotency genes (Pasque *et al.,* 2012; Gaspar-Maia *et al.,* 2013) and are highly enriched at UTX target genes, which are reactivated early during reprogramming (Gaspar-Maia *et al.,* 2013). In human keratinocytes, macroH2A1 occupies pluripotency and bivalent genes and in line with this, macroH2A1 knockdown leads to increased reprogramming efficiency, while macroH2A1 overexpression inhibits reprogramming (Barrero *et al.,* 2013).

TH2A and TH2B, involved in activation of the paternal genome after fertilization in oocytes, stimulate iPSC generation when overexpressed with OSKM in MEFs (Shinagawa *et*

al., 2014). Similarly, knockdown and overexpression studies of SF1A, a histone-remodeling chaperone enriched in human oocytes, showed that this histone chaperone promotes reprogramming of HDFs (Gonzalez-Munoz *et al.,* 2014).

Nucleosome remodelers

BRG1 and BAF155, two components of the BAF (SWI/SNF) chromatin remodeling complex, increase OSKM-mediated reprogramming of MEFs by promoting DNA demethylation of the promoters of pluripotency genes such as *Oct4, Nanog,* and *Rex1* (Singhal *et al.,* 2010). This is consistent with the role of components of this complex in maintaining ESC self-renewal and differentiation (Ho and Crabtree 2010).

Other chromatin-remodeling proteins including CHD1 and INO80 have been shown to be required for pluripotency and reprogramming. CHD1 associates with euchromatin in ESCs and preferentially targets genes involved in chromatin organization and transcription (Gaspar-Maia *et al.,* 2009). INO80 co-occupies pluripotency gene promoters in an OCT4- and WDR5-dependent manner, maintaining open chromatin and facilitating the recruitment of Mediator and RNA Pol II for gene activation (Wang *et al.,* 2014). Both CHD1 and INO80 are required for efficient reprogramming (Gaspar-Maia *et al.,* 2009; Wang *et al.,* 2014).

MBD3, a core member of the nucleosome remodeling and deacetylation (NuRD) repressor complex, is found to have a contradictory effect on reprogramming. In MEFs, *Mbd3* overexpression significantly reduces reprogramming and its knockdown using shRNAs leads to a 10-fold increase in reprogramming efficiency (Luo *et al.,* 2013). Using optimized OSKM transgene delivery and 2i/LIF (leukemia inhibitory factor) ground-state pluripotency culture conditions, *Mbd3*-/- MEFs can be reprogrammed with close to 95% efficiency (Rais *et al.,* 2013). During OSKM-mediated reprogramming of MEFs, the genes bound by MBDD3 and CHD4 (another NuRD component) are enriched for *Klf4*, *Oct4*, *Sox2*, and *Esrrb* targets. In contrast to these observations, *Mbd3* ablation in NSCs reduces iPSC formation (Dos Santos *et al.,* 2014), and its knockdown in human fibroblasts differentiated from hESCs has a negative effect on iPSC induction (Mancarci *et al.,* 2012).

DNA methyltransferases / demethylases

Silencing of somatic genes and activation of pluripotent genes during

reprogramming involve DNA methylation and demethylation, respectively. DNA methyl transferases (DNMT1, DNMT3a, and DNMT3b), methylcytosine dioxygenases (TET1 and TET2), PARP1 and AID, have been found to be associated with these process (Liang and Zhang 2012; González and Huangfu 2015).

Knockdown of maintenance methyl transferase, *Dnmt1,* brought about replication-dependent passive DNA demethylation and promoted the transition of reprogramming cells from maturation phase to stabilization phase (Mikkelsen *et al.,* 2008b). Downregulation of *DNMT3a* and *Dnmt3b* in human and mouse fibroblasts did not have a significant effect on reprogramming (Pawlak *et al.,* 2011).

TET1 and TET2 methylcytosine dioxygenases bring about active DNA demethylation by hydroxylation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and the subsequent replacement of 5hmC with an unmethylated cytosine through TDG-mediated base excision repair (Kohli and Zhang 2013). The inability of Triple *Tet1, 2, 3*-null and TDG-null MEFs to get reprogrammed, confirmed that oxidative demethylation in an indispensable for reprogramming of mesenchymal cells (Hu *et al.,* 2014). However, keratinocytes and neural progenitor cells (NPCs) can be successfully reprogrammed without any *Tet* genes. In miPSC reprogramming, *Oct4* can be replaced by *Tet1* and the overexpression of latter increased reprogramming efficiency (Chen *et al.,* 2013; Gao *et al.,* 2013). The close association of *Tet1* and *Tet2* with *Nanog* is evident from observations that NANOG targets *Tet1* and *Tet2* to many pluripotency loci (Costa *et al.,* 2013), overexpression of *Tet1* or *Tet2* with Nanog enhance reprogramming, and inhibition of *Nanog* or *Tet2* compromises iPSC formation (Silva *et al.,* 2009; Costa *et al.,* 2013). During hiPSC reprogramming, an elevation in the levels of 5hmC in correlation with TET1 activation observed might be associated with active DNA demethylation events.

PARP1 has been found to regulate 5mC modification to bring about the establishment of early epigenetic marks during somatic cell reprogramming (Doege *et al.,* 2012). PARP1 is also involved in relieving the inhibitory effect of excessive SOX2 on FGF4 expression both in PSCs and during reprogramming by PARP1-mediated poly(ADP-ribosyl)ation of SOX2 leading to the dissociation of SOX2 from the *Fgf4* enhancer (Gao *et*

al., 2009; Weber *et al.,* 2013).

The role of activation-induced cytidine deaminase (AID) in pluripotency and reprogramming is yet to be understood thoroughly. A four to five-fold decrease in iPSC colony formation was observed on transient knockdown of AID within 2 days of reprogramming. Consistent with this, AID overexpression lead to a two-fold increase in OSKM-mediated reprogramming (Bhutani *et al.,* 2013), but no effects have been observed in subsequent studies (Kumar *et al.,* 2013; Habib *et al.,* 2014; Shimamoto *et al.,* 2014). Contradictory to this observation, AID-null MEFs, which were initially hypersensitive to OSKM-mediated reprogramming, gave rise to $NANOG^+$ iPSC colonies with six-fold higher efficiency compared to control MEF and majority of AID-null iPSC colonies failed underwent differentiation by 4 weeks.

2.12.3. Role of signal transduction pathways

The reprogramming process involves changes in activities of signal transduction pathways linked to development, stress, and metabolism (**Figure 11**). Some of these pathways support, whereas the others inhibit the transition of cellular identity. The pathways favoring the reprogramming involve the ones that play a central role in maintaining pluripotent state of cells. The main pathway that has an inhibitory effect of reprogramming is stress induced signaling pathways. Specific inhibitors and stimulators of signaling pathways had been able to improve reprogramming efficiency and even been able to replace key reprogramming factors.

Developmental signaling pathways

Pathways associated with pluripotency and MET/EMT play important in reprogramming. The addition of cytokines and inhibitors associated these pathways like Wnt, Activin/Nodal, bFGF, LIF, and 2i have been found to influence the pluripotent induction and pluripotent state of cells (Zhang *et al.,* 2012). The effect of signaling molecules was found to be dependent on cell type and the stage of reprogramming. BMP promotes MET in the early initiation phase (Samavarchi-tehrani *et al.,* 2010) but blocks the late transition of pre-iPSC to iPSCs (Chen *et al.,* 2012). MEK inhibition limits the growth of non-iPSC colonies and promotes the growth of reprogrammed iPSCs from NPCs (Shi *et al.,* 2008), and GSK3

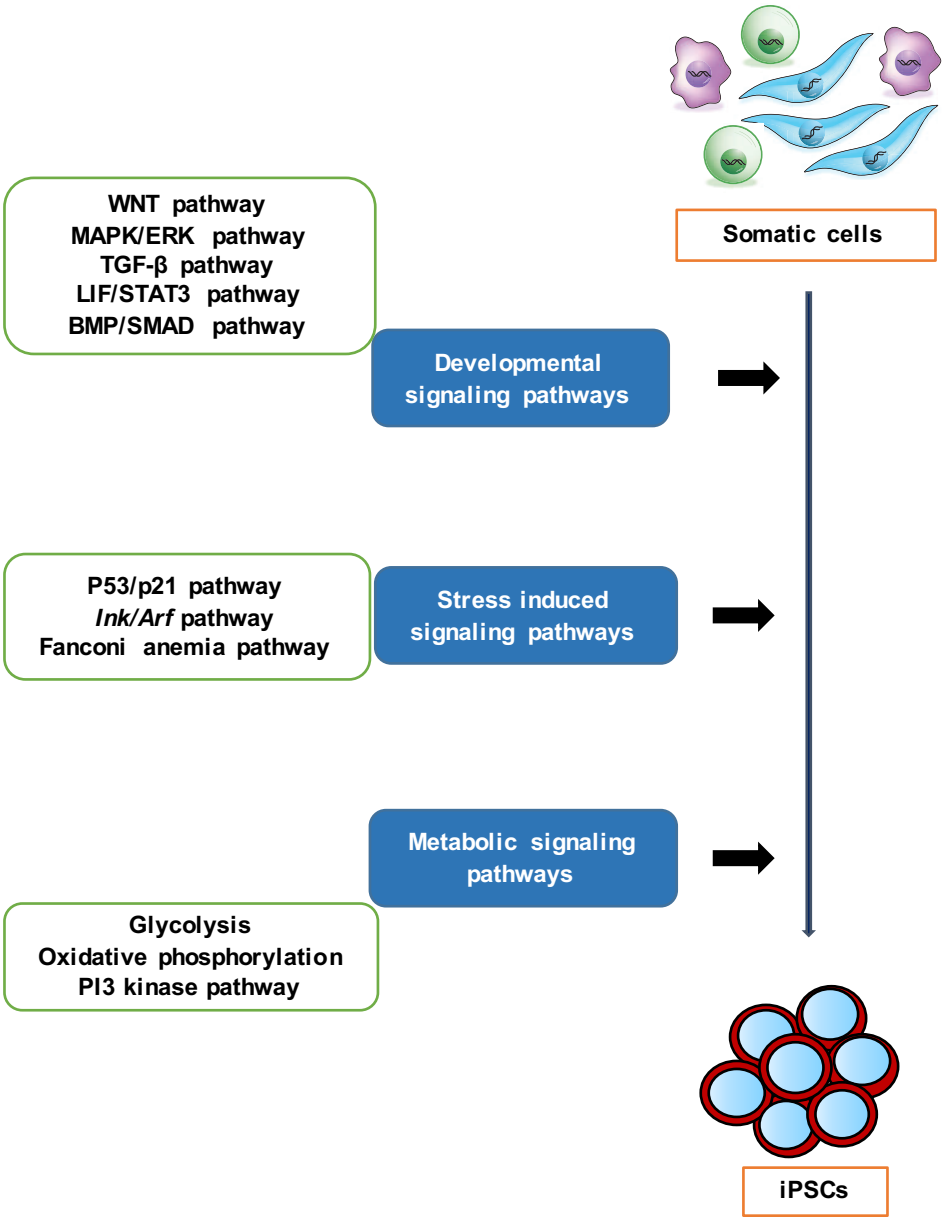


Figure 11: Role of signal transduction pathways in reprogramming
Several molecular pathways get activated or inhibited during the reprogramming process. Developmental signaling pathways get switched from those active in somatic cell to those associated with pluripotency, and the failure to do so adversely affect the reprogramming efficiency. The genotoxic stress brought about by the integration or/and overexpression of reprogramming factors activates DNA damage response pathways, leading to cell cycle arrest and senescence, thereby reducing the reprogramming efficiency. Similar to their *in vivo* counterparts, iPSCs are dependent more on glycolytic pathway for energy metabolism and a gradual switch from oxidative phosphorylation to glycolysis is observed as somatic cells attain iPSC state.

inhibition facilitates reprogramming with OS alone. Even though Wnt3a conditioning in early days of reprogramming had an inhibitory effect on MEF reprogramming, it had a stimulatory role late in reprogramming (Hou *et al.,* 2013). Wnt3a addition could improve OSK mediated MEF reprogramming (Marson *et al.,* 2008) through repression of *Tcf3*, which is involved in repressing *Oct4*, *Nanog*, and *Sox2* in mouse ESCs (Cole *et al.,* 2008). β-Catenin is indirectly regulated by Nanog and is necessary for the conversion of pre-iPSCs into iPSCs (Marucci *et al.,* 2014). Combining 2i and LIF promotes the transition of partially reprogrammed cells toward bona fide iPSCs (Silva *et al.,* 2008a). Similarly, activation of the Jak-Stat3 pathway by LIF promotes reprogramming by facilitating the transition from partially reprogrammed cells toward fully reprogrammed iPSCs (Yang *et al.,* 2010). Inhibition of TGF-β signaling allowed reprogramming with OK, probably through promoting MET (Li *et al.,* 2010; Zhang *et al.,* 2012). Cell type specific influence of reprogramming pathway on reprogramming came from the observation that 80% of MEFs reprogram through OSKM expression, TGF-β inhibition, Wnt activation, and ascorbic acid treatment, whereas hepatoblast or blood progenitors only require TGF-β inhibition or Wnt activation alone, respectively, to achieve similar reprogramming efficiencies with OSKM.

The conventional hESCs which share similarities with primed mouse EpiSCs than the naïve mESCs, could be taken to the naïve state by treating with different combinations of chemical compounds and cytokines (Silva *et al.,* 2008a; Hanna *et al.,* 2010; Theunissen *et al.,* 2014; Ware *et al.,* 2014). Modulation of reprogramming efficiency and replacement of reprogramming factors could be achieved by the addition of cyclic adenosine monophosphate (cAMP) (Wang and Adjaye 2011), phosphoinositide 3-kinase (PI3K) (Liao *et al.,* 2013), Hippo/Yap (Lian *et al.,* 2010), Src (Lian *et al.,* 2010) and fibroblast growth factor (FGF) (Weber *et al.,* 2013). Hou *et al.,* performed an extensive small-molecules screen and identified a combination of seven chemical compounds sufficient to reprogram mouse somatic cells with up to 0.2% efficiency (Hou *et al.,* 2013).

Stress-Induced Signaling Pathways

Due to genotoxic stress resulting from overexpression of reprogramming factors, the stress induced signaling pathways get activated to bring about the elimination of genetically

abnormal cells. The active involvement of DNA damage surveillance machinery in resisting the transformation of cells by the reprogramming factors is critical for generating iPSC clones without genetic abnormalities.

The p53 mediated DNA damage response pathway resolve intracellular and extracellular signals into DNA repair, cell cycle arrest, and cell death. Overexpression of reprogramming factor in MEFs brings about an increase in p53 activity and upregulation of p19, p21, and MDM2, leading increase in cell death (Banito *et al.,* 2009; Hong *et al.,* 2009; Kawamura *et al.,* 2009a). p53 mediated cell death during reprogramming was prominent in cells with prior DNA damage like short telomere confirming the critical role of p53 in resisting the formation of genetically abnormal iPSC colonies. This protective role of p53 was supported by studies that showed a decrease in reprogramming efficiency on knockdown ataxia-telangiectasia mutated (ATM), which is a key protein involved in cellular response to double-stranded break that involves elevation of p53 activity. Consistent with the inhibitory effect of p53 on reprogramming efficiency, downregulation of *p53, p19, p21* and *Mdm2* in MEFs was found to bring about an increase in the number of iPSC colonies (Hong *et al.,* 2009; Kawamura *et al.,* 2009a; Strati *et al.,* 2009; Utikal *et al.,* 2009). The effect p53 deletion or depletion was found to be proliferation dependent (Hanna *et al.,* 2009b).

The depletion of tumor suppressor proteins, $p16^{Ink4a}$ and $p15^{Ink4b}$ coded from *Ink/Arf* locus and RB was found to promote reprogramming (Banito *et al.,* 2009; Li *et al.,* 2009; Utikal *et al.,* 2009). Since these three proteins are negative regulators of cell cycle, the adverse effect of these proteins on reprogramming might be through limiting proliferation in reprogramming cells. However, it was found that RB acted through repressing pluripotency network in somatic cells and not in a proliferation-dependent way.

The genes involved in DNA double-strand break (DSB) repair are also associated with the reprogramming process. These include DNA double-strand break repair genes such as *Brca1*, *Brca2*, and *Rad51* (González *et al.,* 2013; Soyombo *et al.,* 2013; Navarro *et al.,* 2014), DNA interstrand cross-link repair genes like Fanconi anemia pathway genes (Rodrı *et al.,* 2009; Müller *et al.,* 2012a; Yung *et al.,* 2013; Liu *et al.,* 2014; Rio *et al.,* 2014) and nonhomologous end joining DSB repair genes (Molina-Estevez *et al.,* 2013).

Metabolic Signaling Pathways

ICM cell exists in a low oxygen niche *in vivo* (Mohyeldin *et al.,* 2010) and probably as an adaptation to such condition; pluripotent cells are dependent on glycolysis for energy metabolism rather than on oxidative phosphorylation. Consistent with this, hypoxic conditions had been found to enhance reprogramming efficiency of mouse and human somatic cells (Yoshida *et al.,* 2009). Also, activation of glycolysis or blockade of mitochondrial OXPHOS using fructose 2,6-bisphosphate (PFK1 activator), 2,4-dinitrophenol (mitochondria decoupler), quercetin (HIF activator), or PS48 (PDK1 activator) (Zhu *et al.,* 2010) were shown to enhance reprogramming efficiency.

Oxygen-sensitive transcription factors, hypoxia-inducible factors (HIFs), which facilitate cellular adaptation to low oxygen concentration, have been found to have addition role in pluripotent cells. HIF1α bring about activation of Notch signaling and maintain the undifferentiated state of various stem and progenitor cell populations (Gustafsson *et al.,* 2005) and HIF2α regulate *Oct4* expression by binding to its promoter in PSCs (Covello *et al.,* 2006). Their role in reprogramming became evident with the reduction in the number of iPSC colonies formed on knockdown of HIF1α or HIF2α and significant improvement reprogramming efficiency and acceleration in the switch toward glycolytic metabolism by overexpression of non-degradable forms of either protein during early phases of reprogramming (Mathieu *et al.,* 2014).

The addition of antioxidants like ascorbic acid to reprogramming cells altered their properties that improved the efficiency of OSK- and OSKM- mediated reprogramming. Vc treatment brought about (i) a significant reduction in p53 and p21 levels leading to decrease in cell senescence (Esteban *et al.,* 2010) among reprogramming cells, (ii) improvement in iPSC quality through preventing aberrant silencing of the imprinted *Dlk1-Dio3* locus (Stadtfeld *et al.,* 2012), (iii) promotion of Vc- dependent demethylating dioxygenases JHDM1A/1B activity and (iv) TET1 mediated 5hmC formation at loci critical for MET during reprogramming (Chen *et al.,* 2013), all of which favoured reprogramming. In addition to H3K36me2/3 demethylation, Jhdm1b accelerated cell cycle progression, suppressed cell senescence by repressing the *Ink4/Arf* locus, and cooperated with OCT4 to activate the miR-

302-367 cluster (Wang *et al.,* 2011).

2.13. Molecular barriers in reprogramming

The extremely low efficiency in reprogramming is caused by the molecular barriers that resist the transition of cellular identity from the differentiated state to the pluripotent state (**Figure 12**). These barriers are associated with chromatin state, nuclear architecture, apoptosis and senescence in the reprogramming cells. They either abort the reprogramming process or result in the formation of partially reprogrammed colonies, the pre-iPSCs that fail to achieve full features of the pluripotent state. The ability of a fraction of pre-iPSCs to get fully reprogrammed on exposure to different small molecules confirmed that these cells represent intermediate cell types that have encountered specific barriers during pluripotency induction, and they are not mere byproducts of aborted reprogramming process (Zhang *et al.,* 2012). The requirement for specific additional manipulations to achieve pluripotent state and the low efficiency of this conversion indicated that pre-iPSCs are highly heterogeneous and are formed as a result of roadblocks at different stages. At the same time, similarities in the pre-iPSC state formed from various somatic cell types suggested that reprogramming cells proceed through similar routes wherein they encounter similar molecular barriers (Mikkelsen *et al.,* 2008a). Molecular dissection of pre-iPSCs has helped to identify barriers in reprogramming and develop reprogramming strategies that can increase the kinetics and efficiency of reprogramming.

Inability to reset the chromatin state to ESC-like state results in the formation of pre-iPSCs. The major obstacles associated with this step include the inaccessibility to the binding sites of exogenous reprogramming factors and their inability to bind to the targets in the absence of their functional partners. This can lead to inappropriate DNA methylation and histone modifications resulting in aberrant silencing and activation of somatic cell specific genes and pluripotency genes respectively. Consistent with this, in pre-iPSCs, endogenous *Oct4* and *Nanog* genes, and somatically silenced X chromosome genes remained inactive (Maherali *et al.,* 2007; Silva *et al.,* 2008a; Sridharan *et al.,* 2009). They have been found to

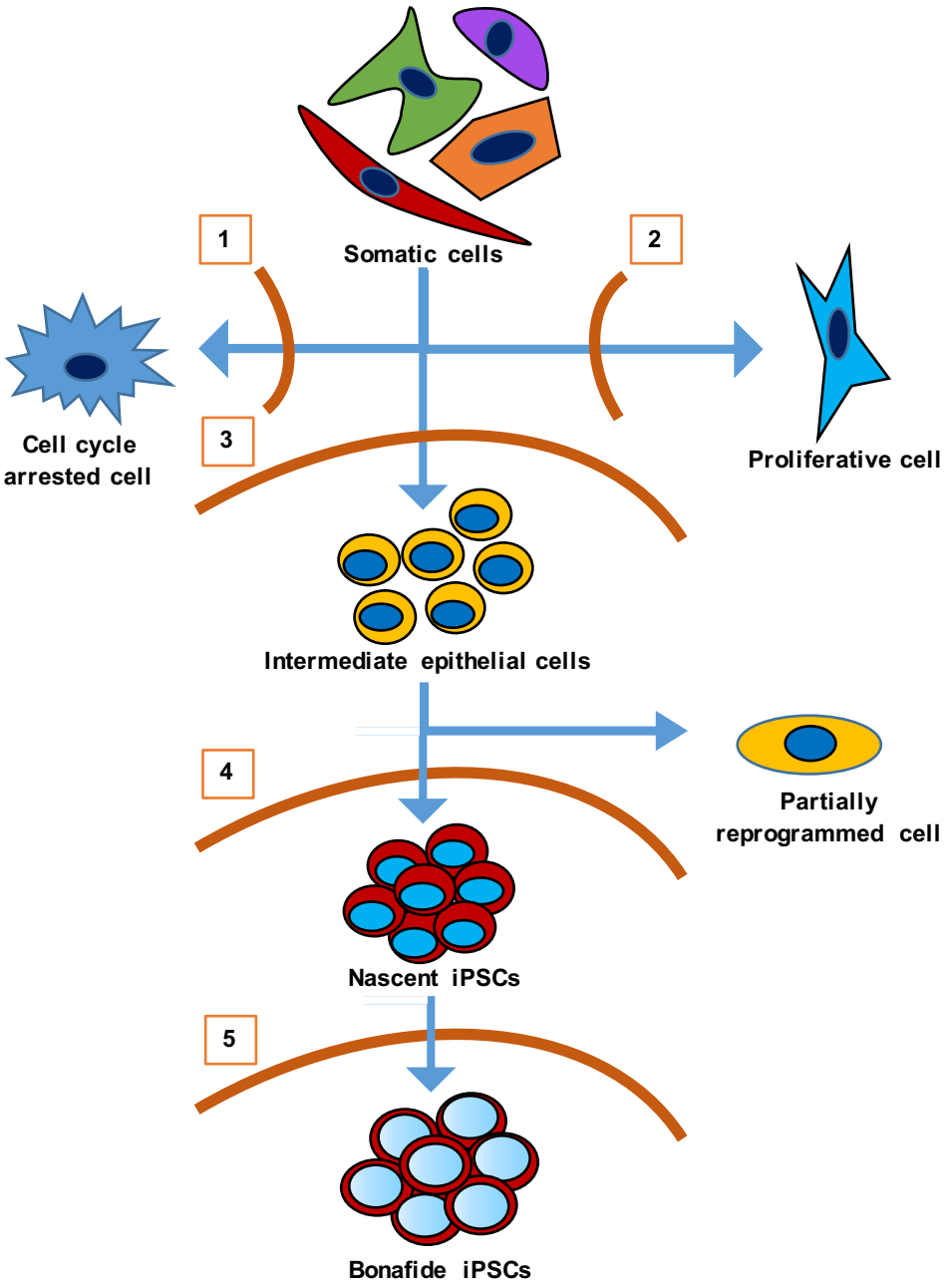


Figure 12: Barriers in reprogramming
Due to the genotoxic stress resulting from the overexpression of reprogramming factors, DNA damage response pathway get activated and majority of the cells undergo cell cycle arrest or senescence (barrier 1). As an initial response, cells acquire a high proliferation rate and fraction of them remain with accelerated proliferation without entering the next stage of reprogramming (barrier 2). Downregulation of somatic cell gene expression along with acquisition of epithelial properties through MET is hindered by epigenetic state of somatic cell (barrier 3). Epigenetic barriers also result in failure to activate pluripotency network in intermediate epithelial cells leading to formation of partially reprogrammed clones (barrier 4). Nascent iPSCs formed have been found to retain somatic cell features and are yet to achieve pluripotency equivalent to ESCs (barrier 5).

express genes that are neither active in ESCs nor iPSCs. By manipulating the activities of epigenetic modifiers that can make the chromatin more permissive, pre-iPSCs could be converted to iPSCs and reprogramming kinetics could be improved. Treatment with DNA demethylating agents like 5'-azacytidine or depletion of *DNMT1* improved the conversion of pre-iPSCs to iPSCs (Mikkelsen *et al.,* 2008a). Addition of histone deacetylase inhibitors including valproic acid (VPA), trichostatin A (TSA), suberoylanilide hydroxamic acid (SAHA) and butyrate (Huangfu *et al.,* 2008a; Liang *et al.,* 2010; Zhang *et al.,* 2012), and H3K9 methyltransferase inhibitors like BIX (Shi *et al.,* 2008) increased the reprogramming efficiency. The positive effect of these components was found to be stage specific and was even found to replace reprogramming factors. NaB promotes the reprogramming only in the presence of exogenously expressed c-Myc at the early stage of reprogramming (Liang *et al.,* 2010). VPA can replace *KLF4* and *c-MYC* to induce iPS cells from human neonatal fibroblasts (Huangfu *et al.,* 2008b). Similarly, overexpression of ESC-specific chromatin remodelers including SWI/SNF-type BAF complex were found to enhance reprogramming efficiency and kinetics (Singhal *et al.,* 2010). Nanog expression has been found to lower the barriers during the final steps in reprogramming. Since NANOG co-binds many of the OCT4, SOX2 and KLF4 targets in ESCs (Wang *et al.,* 2006) its absence might be hindering the binding of OCT4, SOX2 and KLF4 on their target sites. In line with this idea, the conversion of pre-iPSCs to iPSCs by TGF-β inhibition or ERK/GSK inhibition have been associated with upregulation of Nanog expression (Ichida *et al.,* 2009). In addition to chromatin state, the inability to reorganize the chromatin in the nucleus also hinders the reprogramming process. Consistent with this, chromatin organization of pre-iPSCS was found different from ESCs, and it was somatic cell like with an accumulation of compacted chromatin at the periphery of the nucleus. In the pre-iPSCs, the *Oct4* gene was found to localize in periphery away from transcriptional factories, thereby remaining transcriptional inactive in these cells.

Overcoming the barrier imposed by senescence and apoptosis triggered by cellular stress associated with reprogramming is a major efficiency limiting step in reprogramming. MEFs at late passages with higher senescence rate showed lower reprogramming efficiency compared to those at early passages (Utikal *et al.,* 2009). It has been found that the

improvement in reprogramming efficiency achieved by Vitamin C supplementation is through alleviating p53 induced cell senescence along with modulating epigenetic regulators (Wang *et al.,* 2011). Inhibition of components of DNA damage surveillance machinery like p53, p21, p16 and p19 in fibroblasts increased the number of iPSC colonies formed as well as kinetics of the process (Banito *et al.,* 2009; Hong *et al.,* 2009; Kawamura *et al.,* 2009a; Strati *et al.,* 2009; Utikal *et al.,* 2009). Even though on the addition of SV40 T antigen iPSC colonies were formed within 8 days, the majority of them carried chromosomal abnormalities. However, achieving rapid and highly efficient reprogramming at the cost of genetic and chromosomal abnormalities is unacceptable. So, while manipulating apoptosis and senescence pathways for enhancing reprogramming outcomes, care should be taken not to interfere with their role in safeguarding the genomic integrity of the cells. Inherent defects in somatic cells that make them incapable of maintaining genomic integrity have also been found to offer a barrier to reprogramming. Cancer and malignant cells show the low efficiency of reprogramming due to a high degree of chromosomal abnormalities in these cells. Similarly, Fanconi anemia patient-derived fibroblasts had been found to be refractory to reprogramming and give rise to stable iPSC clones only when defective FA gene is complemented. Valuable insights into the role of DNA damage response pathway in pluripotency can be obtained by analyzing reprogramming cells and iPSCs derived from the defective somatic cells from the patients.

2.14. Fanconi anemia pathway in reprogramming and disease pathogenesis

Fanconi anaemia (FA) is a hereditary disorder clinically characterized by developmental malformations, progressive bone marrow failure (BMF), and increased the incidence of leukemia and solid tumors (Auerbach, 2009). The genetic defects causing FA phenotype in patients has been associated with members of FA / BRCA interstrand cross-link repair (FA/BRCA ICR) pathway (Su and Huang, 2011; Kee and Andrea, 2012; Kottemann and Smogorzewska, 2013; Walden and Deans, 2014; Ishiai *et al.,* 2016). Aberrant functioning of this pathway makes the cells highly sensitive to DNA cross-linking

agents, resulting in accumulation of genetic aberrations and leading to cell cycle arrest (Kee and Andrea 2012; Garaycoechea and Patel 2013; Walden and Deans 2014; Ishiai *et al.,* 2016).

Fibroblasts derived from FA patients have been found to be refractory to reprogramming unless the defective gene is complemented. The increase in reprogramming efficiency of FA cells with p53 knockdown or under hypoxic condition indicated that the resistance offered by FA cells is due to their increased sensitivity to DNA damage induced by the reprogramming factors. Moreover, the colonies generated with such additional manipulations are mostly found to generate iPSC colonies with karyotypic abnormalities. This indicates that functional FA pathway is critical for generating genetically stable iPSCs. Analyzing the FA reprogramming cells, FA-hiPSCs and the differentiated cells from FA-hiPSCs may provide valuable insights into the role of FA pathway in reprogramming, pluripotency and differentiation. Even though the molecular interactions between FA / BRCA ICR pathway genes are well established, the exact mechanism by which they contribute to highly variable FA disease phenotypes is not known. Due to the inherent limitations of conventional cellular and animal models of FA, they fail to recapitulate the key phenotypes that are observed in humans. The iPSC-based disease models have been useful for modelling several human diseases and is a promising alternative for understanding FA pathogenesis in humans.

2.14.1. Clinical features of Fanconi anaemia

Bone marrow failure is the major cause of mortality in FA patients. In addition to BMF, most FA patients develop haematological malignancies, myelodysplastic syndrome (MDS), acute lymphocytic leukemia (ALL) or acute myelogenous leukemia (AML) (Butturini *et al.,* 1994). Hypoplastic anaemia due to BMF in these patients is detected at the median age of seven years. The median age of onset of AML, the common haematological malignancy observed in FA, was found to be 14 years and the risk of AML is 800-fold higher than that of the general population (Kee and Andrea, 2012). Specific chromosomal abnormalities have been reported in FA patients with MDS or AML (e.g. gain of 1q23-32, 3q26) (Tönnies *et al.,* 2003).

The congenital abnormalities characteristic of FA are present in approximately 70% of the patients, and these include café au lait spots or hypopigmentation, radial ray defects, microphthalmia, malformations of kidneys, gastrointestinal tract and heart, short stature, mental retardation and hearing defects (Auerbach, 2009; Kottemann and Smogorzewska, 2013). In addition to hematological malignancies, one-third of FA patients develops squamous cell cancers (SCCs) of the head and neck and cervical/gynecological cancers, by the fourth decade of life (Kutler *et al.,* 2003; Rosenberg *et al.,* 2003). Many FA individuals display endocrine abnormalities, which is indicated by short stature among approximately half of FA individuals, correlating with insufficient growth hormone production and hypothyroidism (Sherafat-Kazemzadeh *et al.,* 2007; Eyal *et al.,* 2008). Abnormal glucose or insulin metabolism is also found to be associated with FA (Elder *et al.,* 2008).

Spontaneous chromosomal instability, hypersensitivity to DNA crosslinking agents, G2/M cell cycle arrest, and reduced cell survival are major defects in FA cells (Bogliolo and Surrallés, 2015). FA cells display much higher levels of chromosomal aberrations upon exposure to the chemotherapeutic drugs such as mitomycin C (MMC) that induce DNA interstrand cross-links (ICLs) (Sasaki and Tonomura 1973; Deans and West, 2011). This observation proposed that FA is a DNA repair disorder that is specifically defective in ICL repair (Sasaki and Tonomura, 1973). Hypersensitivity to ICLs in FA is associated with the accumulation of G2/M-arrested cells and the cell cycle arrest at G2/M is due to increased activity of checkpoint kinase CHK1 (Guervilly *et al.,* 2008).

FA patients display similar cellular and clinical phenotypes and the diagnostic tests for FA are based on cellular phenotypes that FA cells manifest. The most widely used diagnostic test for FA is the measurement of *in vitro* hypersensitivity of FA cells to DNA interstrand crosslinking agents (ICLs), such as diepoxy butane (DEB) and mitomycin C (MMC) (Talmoudi *et al.,* 2013; Auerbach, 2015). As FA cells display a marked increase of cells in G2/M phase (4N DNA content), examining the cell cycle profile of FA cells also has been used for the diagnosis (Seyschab *et al.,* 1995; Schindler *et al.,* 2007). Following the discovery of the protein complexes involved in FA pathway, analyzing FANCD2 ubiquitination status and complementation test also has been used for the diagnosis of FA

(Chandra *et al.,* 2005; Casado *et al.,* 2007; Pinto *et al.,* 2009). The complementation test for FA subtyping involves transducing FA cells with viral vectors to express cDNAs complementing the different FA subtypes. Transduction with the appropriate complementation group cDNA corrects the cellular FA phenotypes, such as hypersensitivity to DNA ICLs and cell cycle arrest.

2.14.2. DNA interstrand cross-links (ICLs)

ICLs are the most toxic lesions incurred during normal metabolism or cancer chemotherapy. ICLs covalently tether both strands of a DNA duplex, thereby blocking DNA strand separation and preventing transcription and replication-fork progression (Thompson *et al.,* 2005; Kee and D'Andrea, 2010). Chemotherapeutic agents like MMC and Cisplatin and metabolic byproducts like aldehydes are potent DNA interstrand cross-linking agents. The removal of ICLs is critical for restoring normal DNA metabolism and maintaining genomic integrity. The mechanism of ICL repair is dependent upon the cell cycle. During S phase, when a homologous sister chromatid is present, the FA / BRCA interstrand cross-link repair (FA/BRCA ICR) pathway coordinates ICL excision (Raschle *et al.,* 2008; Knipscheer *et al.,* 2009; Shen *et al.,* 2009; Long *et al.,* 2011). In the G1 phase of the cell cycle, when a sister chromatid is not present to act as a template for homologous recombination (HR), the nucleotide exchange repair (NER) machinery is likely responsible for coordinating ICL repair (Wang *et al.,* 2001; Enoiu *et al.,* 2012).

The major cause of FA is the inability to repair ICLs in S phase of the cell cycle due to genetic defects in the members of FA/BRCA ICR pathway that involves the proteins in the FA pathway, translesion synthesis, homologous recombination and nucleotide exchange repair (Kottemann and Smogorzewska 2013; Walden and Deans 2014; Ishiai *et al.,* 2016).

2.14.3. Fanconi anaemia pathway

The FA DNA repair pathway is the key pathway that coordinates reactions that remove ICL damage during S phase to restore genome integrity. DNA damage and stalled replication forks due to ICLs activate the FA pathway. In addition to their role in ICL repair, FA proteins provide a surveillance mechanism for monitoring unrepaired DNA during cytokinesis, and the absence of FA proteins in these fragile sites was associated with

increased chromosome instability and binucleated cells. FA pathway has also been found to be associated with alternative lengthening of the telomeres (ALT) pathway that helps in maintaining normal telomere function (Spardy *et al.,* 2008; Fan *et al.,* 2009; Rhee *et al.,* 2010).

Nineteen genes have been identified to function in FA pathway. These include *FANCA*, *FANCB*, *FANCC*, *FANCD1*, *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *FANCI*, *FANCJ*, *FANCL*, *FANCM*, *FANCN*, *FANCO*, *FANCP, FANCQ, FANCR, FANCS* and *FANCT* (Bogliolo and Surrallés, 2015). However, only 15 are classified as bona fide FA genes (*FANCA, B, C, D1, D2, E, F, G, I, J, L, N, P, Q* and *T*) (Bogliolo and Surrallés, 2015). *FANCO, FANCR* and *FANCS* are FA-like genes, as they were found to cause a chromosome fragility syndrome with FA- related malformations but without BMF (Vaz *et al.,* 2010; Sawyer *et al.,* 2015; Wang *et al.,* 2015). Since the loss of function variants of *FANCM* do not result in any haematological abnormalities, it has been recommended to consider *FANCM* as an FA-associated gene (Lim *et al.,* 2014). The majority of FA genes are located on autosomes, except *FANCB* that is on the X chromosome.

Depending on the functional role in FA pathway, FA genes can be classified into four groups namely FA core complex members, ID complex members, endonucleases and homologous recombination (HR) repair related genes (**Figure 13**). In addition to DNA repair, FA proteins are also involved in modulating cytokine response and interactions with REDOX sensing proteins.

FA core complex: FA core complex consists of eight FA proteins (FANCA, B, C, E, F, G, L, and M) which interact with each other constitutively. Only FANCL and FANCM in the core complex have enzymatic domains and all others function to form a scaffold. FANCL has a PHD-type RING finger domain at the C-terminus (Meetei *et al.,* 2003) and FANCM has a DEAH helicase domain at the N-terminus and inactive ERCC1 nuclease domain in the C-terminus (Meetei *et al.,* 2005). FANCL possesses intrinsic E3 ubiquitin ligase activity, and the primary function of the complex is to monoubiquitinate two FA proteins, FANCD2 and FANCI (Kim and D'Andrea 2012; Kottemann and Smogorzewska, 2013). The PHD-type RING finger domain of FANCL facilitates its interaction with an E2 ubiquitin–conjugating

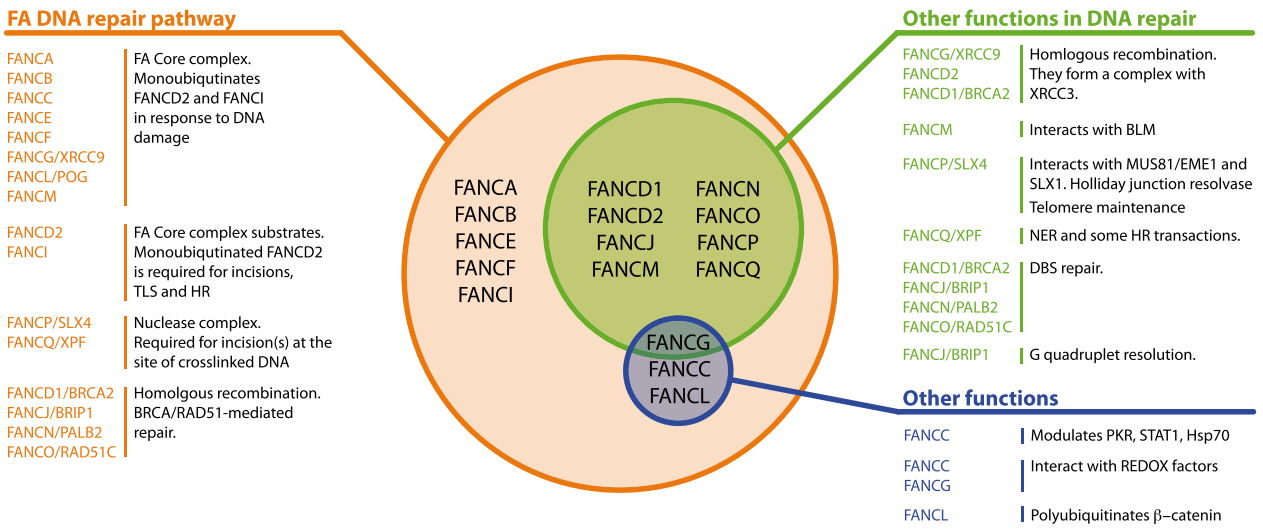


Figure 13: Multifunctionality of members of Fanconi anaemia pathway

Depending on the functional role in FA pathway, FA genes are classified in to four groups namely FA core complex members, ID complex members, endonucleases and homologous recombination (HR) repair related genes. FA core complex consists of eight FA proteins (FANCA, B, C, E, F, G, L, and M) which interact with each other constitutively. FA ID complex consists of FANCI and FANCD2. Endonucleases consist of FANCP and FANCQ. HR related genes include FANCD1, FANCJ, FANCN and FANCO. In addition to their role in FA pathway, some of the FA proteins have other functions in DNA repair and cytokine response. (Adapted from Garaycoechea et al., 2014.)

enzyme, UBE2T during this process (Meetei *et al.,* 2003; Machida *et al.,* 2006). The FA core complex also contains non-FA proteins such as FA-associated proteins (FAAPs), FAAP100, FAAP24 and FAAP20, and a heterodimeric complex consisting of MHF1 and MHF2 proteins. They are involved in partnerships with the core complex members, resulting in the subcomplexes like FANCA-G-FAAP20, FANCL-B-FAAP100 and FANCM-FAAP24-MHF1-MHF2 (Medhurst *et al.,* 2006), and the mode action of these subcomplexes is still unknown.

FA ID-complex: FANCD2 and FANCI proteins are expressed as a constitutive heterodimer that is referred to as the ID-complex. Monoubiquitination of FANCD2 and FANCI by FA core complex results in the activation of ID complex (Smogorzewska *et al.,* 2007). After activation, ID complex translocates and assembles into DNA repair foci. The exact mechanisms through which monoubiquitination affects the activity or conformation of the heterodimer is not yet understood. Monoubiquitinated FANCD2 serves as a signal for recruiting downstream effector proteins that have an affinity for the ubiquitin. In DNA repair foci, the ID complex interacts with the FA pathway downstream proteins like FANCD1, FANCN, FANCJ, FANCP and FANCO (Hussain *et al.,* 2004; Wang *et al.,* 2004; Zhang *et al.,* 2009) and other DNA repair pathways like RAD51, PCNA and REV1 (Howlett *et al.,* 2009; Geng *et al.,* 2010; Long *et al.,* 2011).

Endonucleases: FANCP/SLX4, FANCQ/XPF/ERCC4, and FAN1 are involved in the endonucleolytic activity in FA pathway, and they function to make a nucleolytic incision to remove the ICLs. FANCP/SLX4 functions as a scaffold DNA repair protein that can interact with several endonucleases including XPF/ERCC1 complex (Fekairi *et al.,* 2009). FANCP bind to monoubiquitinated FANCD2, and is thereby delivered to DNA damage sites where it forms foci. FANCQ/XPF/ERCC4 associates with ERCC1 and ERCC4/ERCC1 endonuclease is responsible for the incision on the 5' side of DNA lesions (Bhagwat *et al.,* 2009). The Fanconi anaemia-associated nuclease 1 (FAN1), a structure- specific nuclease that is recruited to damaged DNA with monoubiquitinated FANCD2 (Liu *et al.,* 2010; MacKay *et al.,* 2010).

Homologous recombination (HR) related genes: FA proteins FANCD1/BRCA2, FANCJ/BRIP1 and FANCN/PALB2 and FA-related proteins FANCO/RAD51C,

FANCR/RAD51 and FANCS/BRCA1 are involved in the homologous recombination repair (HRR) step of the FA/BRCA ICR pathway (Vaz *et al.,* 2010). These genes are linked to familial breast and ovarian cancer (FBOC) highlighting the link between FA and FBOC (Venkitaraman 2004; Wang, 2007).

2.14.4. Mechanism of FA/BRCA DNA Interstrand cross-link repair pathway

The FA/BRCA pathway coordinates ICL excision during S phase when a homologous sister chromatid is present as a template for genomic correction. The major events in this pathway are damage recognition, ID complex activation, unhooking, translesion synthesis, homologous recombination repair and deubiquitination of ID complex (**Figure 14**). The FA proteins play a critical role in coordinating all the steps in this pathway.

Damage recognition: When DNA containing an ICL undergoes replication, the leading strands of 2 converging replication forks stop at the lesion. Upon collision of a replication fork with an ICL, the FA core is loaded onto single-stranded (ss) DNA at the site of the lesion via the activity of the FANCM/FAAP24 helicase/translocase subcomplex. The core complex also stabilizes the localization of the ATR-ATRIP kinase thereby providing linkage between ATR and ID complex. Proteins involved in ATR signaling are required for the efficient induction of FANCD2/ FANCI monoubiquitylation (Ho *et al.,* 2006; Friedel *et al.,* 2009).

ID complex activation: DNA damage-activated PI3K-family kinases ATR-ATRIP kinase trigger FANCI phosphorylation (Ho *et al.,* 2006) which promotes FANCD2-FANCI interaction that triggers monoubiquitination of FANCD2 and FANCI (Ishiai *et al.,* 2008) by FANCL E3 ubiquitin ligase present in the activated FANC core complex (Kim and D'Andrea 2012; Kottemann and Smogorzewska, 2013). UBE2T/FANCT supplies the E2 conjugase activity necessary for this process to occur. Monoubiquitination of the ID complex is the most critical regulatory step in the activation of the FA pathway. Monoubiquitinated ID complex accumulates at the damaged chromatin resulting in foci formation and the chromatin-bound ID complex acts as a scaffold to recruit endonucleases such as FAN1, FANCP and FANCQ (Bhagwat *et al.,* 2009; Fekairi *et al.,* 2009; Liu *et al.,* 2010).

***Nucleolytic incision or Unhooking*:** FAN1, FANCP and FANCQ are involved in the nucleolytic cleavage of 3' and 5' site of DNA to unhook the ICLs causing a double-stranded

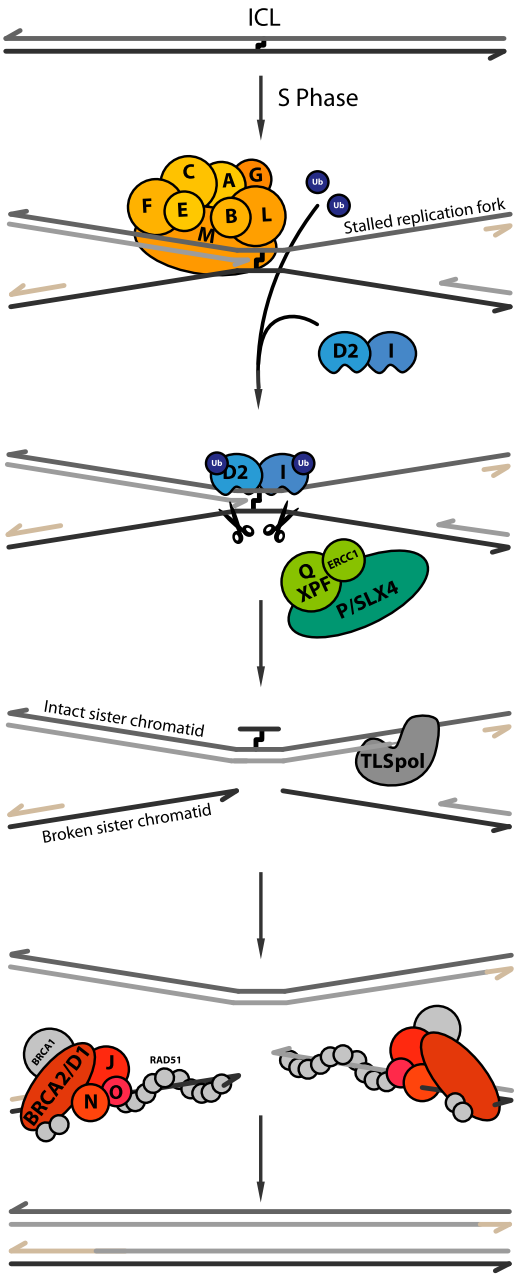


Figure 14: Mechanism of FA/BRCA DNA interstrand cross-link repair pathway
During the replication of DNA containing an ICL, the leading strands of 2 converging replication forks stop at the lesion. The FA core complex gets recruited to the chromatin and brings about monoubiquitination of ID complex. Through dual incision on either side of the ICL, mediated by SLX4-XPF-ERCC1, the 2 sister chromatids get uncoupled. The nascent strand gets extended beyond the ICL by a translesion DNA polymerase (TLS pol). Through the action of nucleotide excision repair (NER) on the top duplex and homologous recombination (HR) on the bottom duplex with the help of HR related FA proteins, 2 fully repaired DNA duplexes are generated . (Adapted from Garaycoechea *et al.*, 2014.)

break (DSB). DSB is an intermediate lesion during ICL repair, and this eventually leads to uncoupling of one sister chromatids from another.

***DNA Translesion synthesis (TLS)*:** Incision of ICL is followed by TLS, a damage tolerance mechanism in which specialized low-fidelity polymerases bypass bulky damaged lesions. The ssDNA gap opposite to the unhooked crosslink are filled by translesion DNA polymerases such as REV1, polymerase ζ and polymerase η (Waters *et al.,* 2009; De Groote *et al.,* 2011). The bypassed, unhooked crosslink resembles single nucleotide lesions and can be repaired by NER proteins (Evans *et al.,* 1997; Kuraoka *et al.,* 2000) or by the DNA glycosylase NEIL1 (Bandaru *et al.,* 2002). The leading strand thus extended serves as a substrate that is processed by successive HRR reactions.

Homologous recombination: The FA proteins play a critical role in making the choice between error-free HR and error-prone NHEJ to repair DSB breaks by enhancing the former and inhibiting the latter. The extensive end resection of a DSB by nucleolytic degradation of the 5' strand is required for the initiation of homologous recombination. This is brought about by MRE11-RAD50-NBS1 (MRN) complex and exonuclease 1 (Exo1) singly or in combination with the Bloom's syndrome RecQ helicase-like protein (BLM) and DNA replication helicase 2 (DNA2). FANCM is involved in the recruitment of MRN complex, and FANCD2 is involved BLM and DNA2. The ability of FAND2 to interact with FANCG of the core complex might be favouring the localization of FANCD2 and its interacting partners at the site of lesion (Hussain *et al.,* 2003). The long 3' overhang created by end resection is stabilized by binding of replication protein A (RPA) and FANCM is involved in the recruitment of RPA (Huang *et al.,* 2010). The recombination mediators, FANCD1, and FANCN, remove RPA and load 3' overhang with RAD51 to form an RAD51 presynaptic filament. FAND1 and FANCN increase the intrinsic stability of the RAD51 presynaptic filament and protect RAD51 from removal. Once the presynaptic filament is assembled, the homology search is initiated. It aligns with homologous DNA segment on sister chromatid and forms the synaptic complex. This is followed by strand invasion wherein the complementary strand of the sister chromatid is displaced leading to the formation of a transient structure known as the D-loop structure. The 3′ overhang of the invading strand is then freed from RAD51 and used as a primer for

elongation. At this time, it is unclear which replication machinery is used for this elongation though it has been shown that TLS polymerases, Pol η in particular, demonstrate an affinity for D-loop elongation After elongation, two pathways can be utilized to resolve the D-loop. In the first pathway, referred to as the synthesis-dependent strand-annealing (SDSA) pathway, D-loop extension continues for a short distance and is disassembled by regulator of telomere elongation helicase 1 (RTEL1) and reannealed with the ssDNA associated with the other DSB end. DSB repair is then completed by gap filling and ligation. The second pathway involves BLM helicase, an ATP- dependent 3′-5′ DNA helicase that can unwind D-loops.

Deubiquitination of ID complex: Deubiquitination of FANCD2 is required for the completion of FA pathway and efficient ICL repair. FANCD2 is deubiquitinated by USP1, an ubiquitin-specific cysteine protease, which localizes to chromatin and interacts with FANCD2. The stability and enzymatic activity of USP1 are strongly stimulated by UAF1, the USP1 associated factor 1 (Cohn *et al.,* 2007). The precise mechanism regulating FANCD2 deubiquitination is not known. It is proposed that deubiquitination of FANCD2 may be required to release FANCD2 from DNA repair complex(es), allowing subsequent repair steps to take place in order to complete the ICL repair process and persistence of FANCD2 monoubiquitination may be toxic to cells causing a significant increase in MMC sensitivity (Nijman *et al.,* 2005; Oestergaard *et al.,* 2007; Kim *et al.,* 2009c).

2.14.5. Molecular pathogenesis of Fanconi Anaemia in Humans

Mutations in the genes that encode FA core complex proteins account for 90% of FA cases (Wang and Smogorzewska, 2015). FANCA, FANCC and FANCG contribute to 64%, 12% and 8%, respectively. Mutations in FANCD2 and FANCI account for 4% and 1% of cases, respectively and FANCL, FANCM, FANCN, FANCO, FANCP, FANCQ and FANCS were each seen at a frequency of less than 1% in FA patients.

FA patients with mutations in any of the FA pathway genes show common phenotypes like growth retardation, thumb and radial ray defects, developmental abnormalities, hypogonadism, microcephaly, male infertility and female reduced fertility and hyperpigmentation (Neveling *et al.,* 2009; Kottemann and Smogorzewska, 2013). The susceptibility to developing BMF, haematological malignancies and cancer have been found

to be dependent on the FA gene which is defective. Except the patients belonging to complementation group FANCO, FANCR and FANCS, the rest developed BMF. Patients with mutations in FANCD1 and FANCN showed an early onset and higher rate of malignancies, developing a wider spectrum of cancers including AML, ALL, medulloblastoma, neuroblastoma and Wilms tumor. Even the carriers show increased susceptibility to breast, ovarian and prostate cancer. On the contrary, no malignancies have been reported in patients and carriers with FANCL and FANCM mutations. Patients belonging to rest of the complementation groups, exhibited susceptibility to AML, head and neck squamous cell carcinoma (HNSCC), esophageal, gynecological, and liver cancers, with no report of cancer in carriers, except FANCC defect carriers who showed increased risk of developing breast cancer.

2.14.6. Disease manifestation in FA mouse models

Although most mouse models exhibit certain features of the disease at the phenotypic and molecular level, they do not fully recapitulate bone marrow failure (Parmar *et al.,* 2009; Bakker *et al.,* 2013). *Fancp* mutant mice have lower white blood cell counts and very low levels of platelets, and hypomorphic *Fancd1* mice show a proliferation defect in haematopoietic progenitors (Navarro *et al.,* 2006; Crossan *et al.,* 2011). These two mouse models are suitable for studying the molecular basis of haematopoietic failure in FA. The HSCs obtained from *Fancc*, *Fancd2* and *Fancg mutants* were found to have a reduced repopulating ability (Haneline *et al.,* 1999; Zhang *et al.,* 2008; Barroca *et al.,* 2012) and hence, they have been used to study the role of FA pathway in haematopoietic homeostasis. However, BMF could be induced in *Fancc*-mutant mice by treating them with very high dose of mitomycin C (MMC) or by prolonged exposure to MMC (Carreau *et al.,* 1998). This suggested that although HSCs from *Fancc*-deficient mice maintain haematopoietic homeostasis under normal conditions and they fail to do so under conditions of stress when HSCs need to be mobilized and replicated. *Fancg*-deficient HSCs also show defects in mobilization and homing to the bone marrow (Barroca *et al.,* 2012). These findings suggest that the absence of anaemia in most FA mouse models might be because FA mutant mice housed in laboratories are protected from stressful environmental cues.

The *Fancd2* mutant mouse show poor performance in competitive repopulation and they have a reduced number of HSCs in the early age (3 weeks), which indicate that an early defect in HSC development and its maintenance (Zhang *et al.,* 2010). HSCs from *Fancd2*-deficient mice exhibited increased cell cycle entry losing their normal quiescent state. The phenotype of *Fancd2*-mutant HSCs makes this model suitable for preclinical studies to test compounds that could reverse the cell cycle state (Zhang *et al.,* 2010). The lower number of HSCs in young *Fancd2*-mutant mice suggested that the development or maintenance of HSCs is defective in the absence of an active FA pathway.

Homozygous disruption of *Fancd1*, *Fancn* and *Fanco,* causes embryonic lethality. *Fancl* mice are embryonic lethal on a pure 129/Sv or C57BL/6 background but are viable on a mixed C57BL/6×FVB background (Agoulnik *et al.,* 2002). Perinatal lethality has been described for several FA mouse mutants, including *Fancc* (Carreau, 2004), *Fancd2* (Houghtaling *et al.,* 2003; Van de Vrugt *et al.,* 2011), *Fancl* (Agoulnik *et al.,* 2002), *Fancm* (Bakker *et al.,* 2009) and *Fancp* (Crossan *et al.,* 2011) and perinatal lethality has been found to depend strongly on genetic background (Van de Vrugt *et al.,* 2011). These data suggest that viability of FA mice is influenced by the presence of modifier loci in certain genetic backgrounds. The phenotypic variations among patients with same FA gene defect might be due to a similar contribution from individual-specific genetic variations in associated genetic loci. Although embryonic and perinatal lethality has not been reported for humans with FA, those with hypomorphic mutations in *FANCD1* or *FANCN* genes die at a young age. Complete loss of function of *FANCD1* or *FANCN* is probably embryonic lethal in humans.

Most FA mouse models do not show any FA-specific developmental abnormalities. However, microphthalmia has been observed in Fanca (Wong *et al.,* 2003), Fancc (Carreau, 2004), Fancd2 (Houghtaling *et al.,* 2003) and Fancp (Crossan *et al.,* 2011). A proportion of Fancp-deficient mice were also reported to have an abnormally shaped and enlarged skull (Crossan *et al.,* 2011). Gonadal abnormality and associated infertility, which is common in FA patients, has been observed in all FA mouse models described thus far (Parmar *et al.,* 2009; Tischkowitz and Winqvist, 2011). This has been attributed to the reduction in the

number of primordial germ cells (PGCs) in the mutant mice. This common phenotype of FA mouse models suggests that the FA pathway is important in early embryogenesis, probably for repairing replication-associated stress.

FA mouse models do not recapitulate the types of tumors observed in FA patients. A statistically significant increase in the incidence of both adenomas and carcinomas of epithelial origin was reported in *Fancd2*-deficient mice (Houghtaling *et al.,* 2003). *Fancf*-deficient mice have been reported to develop late-onset ovarian tumours (Bakker *et al.,* 2012), and *Fancm*-deficient mice have increased cancer incidence, without skewing to a particular tumour type (Bakker *et al.,* 2009).

Using double knock out mice epistatic and non-epistatic relationships between FA genes have been established. *Fanca*$^{-/-}$ *Fancc*$^{-/-}$ *did not* show any evidence of an additive phenotype, suggesting an epistatic relationship between the two genes (Noll *et al.,* 2002). However, *Fancc*$^{-/-}$ *Fancg*$^{-/-}$ mice showed a more severe phenotype than that observed in the single knock out mice as they developed BMF, myelodysplasia, and complex cytogenetic abnormalities, which are not observed in the single knockouts (Pulliam-Leath *et al.,* 2010). However, such additive phenotype was not observed in *Fanca*$^{-/-}$ *Fancc*$^{-/-}$ and *Fanca*$^{-/-}$ and *Fancg*$^{-/-}$ mice. The difference in the outcomes of epistasis analysis might be explained by the difference in the FA pathway modifiers in different mouse strains.

In addition to the manifestation of the disease phenotypes by mutant FA mouse models, they provided valuable insights into the molecular mechanism of the disease. The role of aldehydes in aggravating the phenotypes of FA has been extensively studied in FA mouse models (Garaycoechea and Patel, 2013). Aldehydes which are produced endogenously during normal metabolism induce DNA-DNA and DNA-protein crosslinks (Barker *et al.,* 2005; Deans and West, 2011). It was found that Fancd2 mutant mice were not viable in the absence of aldehyde dehydrogenase 2 (ALDH2) as they died of acute lymphoblastic leukemia (ALL) within 3–6 months (Langevin *et al.,* 2011). Young Fancd2$^{-/-}$ Aldh2$^{-/-}$ mice had a severe reduction in bone marrow cellularity with γH2AX induction, indicating DNA damage. Thus, mouse models have helped in our understanding of the necessity of the FA pathway for acetaldehyde detoxification. Similarly, the role of

reactive oxygen species in causing FA phenotype was evident from mice carrying mutations in *Fancc* and superoxide dismutase 1 (*Sod1*) genes (Hadjur *et al.,* 2012). In the absence of reactive oxygen species (ROS) metabolism by SOD1, these double mutant mice showed a decrease in the red blood cell counts and the colony-forming capacity of bone marrow cells, thereby suggesting the role of FANCC in counteracting superoxide-mediated genotoxicity.

2.14.7. Basis of haematological complications in FA patients

Molecular and functional analysis of patient-derived FA cells, FA deficient hESCs and HSCs and FA mutant mouse models have helped in exploring the basis of haematological complications in FA patients.

The bone marrow failure and other haematological complications in FA patient are attributed to inherent stem cell defect. The evidence for this come from the findings like deficiency of all the blood lineages in the FA patients, depletion of $CD34^+$ HSCs in the BM of young FA patients, impaired hematopoiesis by hESCs with FA gene knockdown and the reduced number of HSCs and the inability of HSCs from FA mutant mouse to reconstitute to blood production in irradiated mice after transplantation (Haneline *et al.,* 1999; Tulpule *et al.,* 2010; Zhang *et al.,* 2010; Ceccaldi *et al.,* 2012). Although it has been well established that the FA pathway gene products are involved in repairing DNA cross-links, their role in preserving bone marrow function has not been proven. The facts that FA patients are prone to solid tumours and haematological malignancies and they often harbour gross chromosomal abnormalities, and both MDS and AML are thought to originate at the stem or progenitor cell levels suggest that the defects in the FA pathway be directly involved in the haematological phenotypes (Butturini *et al.,* 1994; Bonnet and Dick 1997; Alter *et al.,* 2000; Nilsson *et al.,* 2000; Nilsson *et al.,* 2007; Chen *et al.,* 2008; Welch *et al.,* 2012).

The role of DNA damage as an initiator of BMF came from studies evaluating the role p53 in FA-deficient cells. It was found that deletion or knockdown of p53 rescued the defects of FA human and mouse hematopoietic progenitors, but at the enormous cost of enhanced genomic instability and tumor formation (Ceccaldi *et al.,* 2012). This indicates that that unresolved DNA damage in FA cells induces p53, which might then lead to HSC depletion and hence progressive BMF in FA, and the genesis of neoplastic clones. The DNA

damage is the cause of BMF is supported by the fact that the first organ system that fails following total body irradiation is bone marrow, and DNA damage of HSCs limits the regeneration of haematopoiesis (Milyavsky *et al.,* 2010). This is also supported by the fact that the mouse models bearing mutations in the DNA damage response genes show severe defects in the quantity and quality of their HSC pools (Ruzankina *et al.,* 2007; Niedernhofer 2008; Zhang *et al.,* 2011).

Investigation on the source of DNA damage in bone marrow identified metabolic intermediates like ROS and aldehydes as molecules that might contribute to the endogenous burden of DNA damage driving the FA phenotype (Langevin *et al.,* 2011; Hadjur *et al.,* 2012). ROS arise in cells as a consequence of normal cellular metabolism. Superoxide dismutases (SODs) that provide a defense against oxygen toxicity. Improvement in growth feature of FA cells under hypoxia compared to normoxia, correlation in frequency of chromosomal aberration in FA cells with oxygen tension, alleviation in reprogramming resistance of FA deficient cell to iPSCs in hypoxia, and bone marrow hypocellularity and decreased numbers of colony-forming units in *Sod1*$^{-/-}$*Fanc*$^{-/-}$ double mutant mice supported the probable role of ROS in inducing DNA damage in bone marrow (Joenje *et al.,* 1981; Schindler and Hoehn 1988; Hadjur *et al.,* 2012; Müller *et al.,* 2012a). Formaldehyde and acetaldehyde, two highly reactive small aldehydes that are ubiquitously found in the environment and formed as by-products of cellular metabolism, are capable of forming DNA adducts *in vitro* and *in vivo* and cause DNA damage (Wang *et al.,* 2000; Garcia *et al.,* 2011). A family of 18 aldehyde dehydrogenase enzymes oxidizes acetaldehyde to acetate, thereby preventing the accumulation of these genotoxic agents (Marchitti *et al.,* 2008). Mouse HSCs have been found to possess high levels of Aldh2 activity. It has been found that the exposure to aldehyde leads to activation of the FA pathway in wild-type cells and accumulation of double-strand breaks and chromosomal aberrations in FA-deficient cells (Ridpath *et al.,* 2007; Mechilli *et al.,* 2008; Langevin *et al.,* 2011; Rosado *et al.,* 2011). The direct association of aldehyde induced DNA damage and BMF in FA patients originated from the studies that showed a profound defect in the stem and progenitor cell pool in *Aldh2*$^{-/-}$ *Fancd2*$^{-/-}$ mice. Accelerated BMF and increase in the frequency of malformation were

observed in the patients with ALDH2 deficiency (Langevin *et al.,* 2011). Thus, it can be concluded that BMF in FA patients could result from endogenous ROS or/and aldehyde-induced toxicity, which then leads to the depletion of HSCs.

Hypersensitivity of FA cells to cytokines has also been suggested to contribute to BMF in FA patients. Exposure to tumor necrosis factor a (TNF-α) and interferon γ (IFN-γ) induced an apoptotic response in FA cells *in vitro* and bone marrow dysfunction in FA mice *in vivo* (Rathbun *et al.,* 1997; Haneline *et al.,* 1998; Wang *et al.,* 1998; Li *et al.,* 2004). Also, cytokines are detected at elevated levels in FA cells and overproduced by FA-deficient cells *in vitro* (Rosselli *et al.,* 1992; Sarkies *et al.,* 2012). The molecular basis of cytokine overproduction and sensitivity in FA cells is yet to be understood fully.

Thus, it can be concluded that the inability of FA deficient HSCs to maintain genomic integrity on exposure to genotoxic metabolic byproducts makes them accumulate chromosomal abnormalities, activate DNA damage response pathways and undergo cell cycle arrest or cell death leading to bone marrow failure. The exact nature of the type of lesion introduced by ROS and aldehydes, mode of detection of these lesions and mechanism of repair of these lesions by FA proteins need to be explored further to develop drugs for alleviating the disease phenotypes.

2.14.8. Challenges in the derivation of FA patient specific hiPSC lines

FA specific hiPSCs can serve as an excellent tool to study the role of FA pathway in pluripotency, differentiation and disease pathogenesis. However, the resistance to reprogramming posed by FA cells is a major challenge in the derivation of FA patient specific hiPSC line (Raya *et al.,* 2009; Müller *et al.,* 2012a; Yung *et al.,* 2013). Raya *et al.,* demonstrated that the fibroblasts from patients with mutations in FANCA and FANCD2 genes could not be reprogrammed unless the cells were complemented with functional genes. The iPSC lines thus generated could be cultured for 20-40 passages without loss of pluripotency and signs of replicative crisis while maintaining normal karyotype (Raya *et al.,* 2009). The early iPSC-like colonies derived from uncorrected fibroblasts of an FA-D2 patient with a hypomorphic mutation that confers residual expression of normal FANCD2 protein failed to survive in culture beyond three passages (Raya *et al.,* 2009). Consistent with this

study, in attempt to generate iPSCs from fibroblasts obtained from FA-A, FA-C, and FA-D2 patients, Yung *et al.,* could derive FA-A, FA-C, and FA-D2 deficient hiPSC clones, out of which only FA-C deficient hiPSC lines could be maintained in culture for an extended number of passages (Yung *et al.,* 2013). However, the generated hiPSC clone failed to generate teratomas. The two FA-C iPSC lines retained the abnormal karyotype found in the FA fibroblasts along with additional karyotypic abnormalities, which were most probably generated due to the DNA damage associated with the reprogramming process. However, Mueller *et al.,* reported that iPSCs could be derived from fibroblasts of FA-A and FA-C patients without complementation, under hypoxic conditions (Müller *et al.,* 2012a). The established hiPSC lines maintained pluripotency and normal karyotypes without replicative crisis. They showed the features characteristics of FA cells including the inability to make FANCD2 foci formation after treatment with hydroxyurea while this phenotype could be rescued by complementation.

The low reprogramming efficiency of FA cells has been linked with increased susceptibility to genotoxic stress (Müller *et al.,* 2012a). Reprogramming itself is associated with cellular stress, potentially increasing DNA damage and activating cell cycle checkpoint pathways such as *TP53* (Hong *et al.,* 2009; Kawamura *et al.,* 2009b). It was shown that reprogramming-induced stress correlated with an accumulation of reactive oxygen species, leading to increased DNA DSBs, senescence and activation of the FA complex-associated DNA repair pathway (Müller *et al.,* 2012a). These features were elevated in *Fanca*$^{-/-}$ mouse tail tip fibroblast resulting in decreased reprogramming efficiency and *Fanca* complementation alleviated these phenotypes and rescued reprogramming (Müller *et al.,* 2012a). Knocking down the levels of p53 in the FA-A fibroblasts facilitated their successful reprogramming under normoxic conditions (Liu *et al.,* 2014) and the established iPSC lines maintained normal karyotype till passage 13. However, on extended culture, these FA-A deficient clones showed a reduction in clonogenicity and increased sensitivity to DNA cross-linking agents and susceptibility to chromosome fragility.

The outcomes of various reprogramming strategies employed for FA-iPSC generation indicate a clonal selection of mutated cells with an improved ability to

compensate for reprograming stress. These experiments suggested that a functional FA pathway is crucial for successful reprogramming or iPSCs once derived is particularly susceptible to DNA damage and dependent on DNA repair pathways.

2.14.9. iPSC-based disease modeling of Fanconi Anaemia

Novel insights into the mechanism of disease pathogenesis have been obtained using iPSC-based disease modeling of neurological, cardiac, lung, haematological, lysosomal storage and developmental disorders (Colman and Dreesen 2009; Tiscornia *et al.,* 2011; Liu *et al.,* 2012; Zhang 2014; Cao *et al.,* 2015). Disease-specific hiPSCs can be generated from the cells obtained from patients or by genome editing of normal iPSCs, and these mutant iPSC lines can serve as an unlimited source for affected cell types of interest. iPSC-based disease models are invaluable tools for studying the mechanisms of diseases like FA for which cellular and animal models do not faithfully recapitulate the disease phenotypes.

Despite the difficulties in obtaining FA patient-specific iPSC in the first place, current FA-iPSC based haematopoietic differentiation models have been shown to recapitulate aspects of the disease phenotype. FA-A and FA-C deficient iPSC lines were found to be capable of undergoing hematopoietic differentiation without gene complementation (Müller *et al.,* 2012b; Yung *et al.,* 2013). However, a significant reduction in ability to form hematopoietic colonies and increase in apoptosis of haematoendothelial progenitors ($CD34^{+}CD31^{+}CD45^{-}$) was reported for FANCC-deficient iPSC lines (Yung *et al.,* 2013). It has to be noted that these FA-C deficient iPSC clones were hypomorphic which had additional chromosomal abnormalities when compared to patient fibroblasts. These acquired abnormalities might have also contributed to defect in haematopoietic differentiation described in this study and hence such FA deficient iPSC lines with secondary genetic changes is less suited for elucidating the underlying mechanism of FA. The *FANCA* complemented iPSC lines showed a robust multilineage hematopoietic differentiation potential, resulting in erythroid and myeloid hematopoietic colony forming units (CFUs) to a similar degree as compared to normal donor controls (Müller *et al.,* 2012a). In line with these studies, Liu *et al.,* reported a significantly lower yield of haematopoietic progenitor cells

(HPCs) from FA deficient iPSCs compared to wild type control iPSCs and the deficit in generating HSCs was significantly rescued by FANCA gene correction (Liu *et al.,* 2014). Moreover, HPCs derived from FA deficient iPSCs were restricted to the colony-forming unit-granulocyte macrophage (CFU-GM) and were not able to generate colonies containing erythroblasts and/or megakaryocytes, whereas HPCs derived from gene corrected FA–iPSCs gave rise to all typical hematopoietic colonies.

The manifestations of cellular, molecular and functional phenotypes like sensitivity to genotoxic stress and cross-linking agents, reduced FANCD2 monoubiquitination and loci formation at the site of DNA damage, increased chromosome fragility and G0/G1 cell cycle arrest, and defective haematopoietic differentiation reassures the suitability of FA-iPSC line for studying the molecular mechanism of FA disease. Due to the clonal variability among hiPSC lines that may influence their differentiation potential and interfere with the assessment of disease-associated defects, it is challenging to choose the right hiPSC clone suitable for disease modeling. An ideal model system should include patient-specific iPSC that has been conditionally gene-corrected, enabling a comparison of isogenic iPSC lines that are distinct only with regard to the expression of a specific FA protein. Careful analysis of hematopoietic differentiation of such iPSCs holds the promise of uncovering new insights into BMF and may enable high-throughput screening with the goal of identifying compounds that ameliorate hematopoietic failure. While elucidating the disease mechanism using hiPSC based disease models, a few aspects like inter-patient variability with regard to the age of onset, rapidity of progression and severity of the disease despite the lack of genotypic differences, are to be considered.

Chapter 3

MATERIALS & METHODS

3.1. Culture of human dermal fibroblasts

Human adult dermal fibroblasts (hADFs) obtained from commercial sources (Cascade Biologicals) were cultured in two types of media, Amniomax II complete medium (Life Technologies) and αMEM20 (**Table 1**). hADFs from FA patients were generated from 10mm^2 skin biopsies, after the informed consents approved by the institutional review board, Christian Medical College, Vellore. Skin biopsies were cut into small pieces, adhered to a well of a 6 well plate containing Amniomax medium till fibroblasts outgrowths appeared. Subsequently, fibroblasts were passaged using 0.05% trypsin and maintained in αMEM20.

3.2. Preparation of SNL feeder cells

SNL feeder cells were obtained as a gift from Dr. Allan Bradely. Gelatin coated tissue culture (TC) dishes were prepared by adding 0.1% gelatin on dishes followed by incubation for 3 hours in 37^{0}C / 5% CO_2 incubator. Following incubation, the gelatin solution was removed, and SNL cells were seeded in D10 medium (**Table 1**) at a confluency of around 40%. When the cells reached 80-90% confluency, the cells were split at 1:5 ration onto gelatin-coated dishes in D10 medium. For making inactivated SNL feeder cells, the cells at 80-90% confluency were treated with mitomycin C at a concentration of 10µg/ml for 3hrs in 37^{0}C / 5% CO_2 incubator. After incubation, the medium was removed, and the cells were washed twice with 1X PBS and then harvested using 0.05% trypsin. The cells were centrifuged at 1000rpm for 5 minutes. The cell pellet was resuspended in freezing medium (**Table 1**), aliquoted in cryovials and stored in liquid nitrogen for up to 1 year. For making feeder dishes, the frozen mitomycin C-treated feeder cells were revived and seeded on 0.1% gelatin coated TC dishes at a count of 1.8 x 10^6 cells per 10cm dish or 6 well plate in D10 medium and incubated at 37^{0}C/5% CO_2 for 24hrs. Before seeding of reprogramming cells or hiPSC colonies on feeder dishes, the spent medium from the feeder plate was removed, washed with 1X PBS, incubated in hiPSC medium for conditioning the medium and used ideally with 6hrs.

Table 1: Cell culture media used in this study

Application	Media	Components	Concentration	Volume
Culture of Fibroblasts	αMEM20 medium	MEM, α modification	-	make up to 100ml
		FBS	20%	40ml
		200mMGlutamine	2mM	1ml
		100X Penicillin Streptomycin solution	1X	1ml
Culture of Plate E, HEK293T and SNL cells	D10 medium	DMEM	-	make up to 100ml
		FBS	10%	10ml
		200mMGlutamine	2mM	1ml
		100X Penicillin Streptomycin solution	1X	1ml
Reprogramming and culture of hiPSCs	hiPSC medium	DMEM : F12 nutrient mix	-	make up to 100ml
		Knockout Serum Replacement (KOSR)	20%	20ml
		100X MEM NEAA	1X	1ml
		55mM β-Mercaptoethanol	0.1mM	200µl
		100X Penicillin Streptomycin solution	0.5X	500µl
		1µg/ml bFGF	6ng/ml	60µl
Freezing of hiPSCs	2X hiPSC freezing medium	FBS, ESC grade	60%	1.2ml
		hiPSC medium	20%	400µl
		DMSO	20%	400µl
In-vitro differentiation of hiPSCs	EB suspension culture medium	DMEM : F12 nutrient mix	-	make up to 100ml
		Knockout Serum Replacement (KOSR)	20%	20ml
		100X MEM NEAA	1X	1ml
		55mM β-Mercaptoethanol	0.1mM	200µl
		100X Penicillin Streptomycin solution	0.5X	500µl
	EB attached culture medium	DMEM : F12 nutrient mix	-	make up to 100ml
		FBS, ESC grade	20%	20ml
		100X MEM NEAA	1X	1ml
		55mM β-Mercaptoethanol	0.1mM	200µl
		100X Penicillin Streptomycin solution	0.5X	500µl

Differentiation of hiPSCs to HSCs	iPSC-HSC differentiation medium (for days 0-1)	RPMI	-	make up to 10 ml
		200mM Glutamine	2mM	100µl
		100X Penicillin Streptomycin solution	0.5X	50µl
		50mg/ml Ascorbic Acid	50µg/ml	10µl
		100mM Monothioglycerol	0.4mM	40µl
		50ng/µl BMP4	5ng/ml	1µl
		100ng/µl VEGF	50ng/ml	5µl
		40ng/µl WNT3a	25ng/ml	6.25µl
	iPSC-HSC differentiation medium (for day 2)	RPMI	-	make up to 10 ml
		200mM Glutamine	2mM	100µl
		100X Penicillin Streptomycin solution	0.5X	50µl
		50mg/ml Ascorbic Acid	50µg/ml	10µl
		100mM Monothioglycerol	0.4mM	40µl
		50ng/µl BMP4	5ng/ml	1µl
		100ng/µl VEGF	50ng/ml	5µl
		10ng/µl bFGF	20ng/ml	20µl
	iPSC-HSC differentiation medium (for day 3)	Stem Pro	-	make up to 10 ml
		200mM Glutamine	2mM	100µl
		100X Penicillin Streptomycin solution	0.5X	50µl
		50mg/ml Ascorbic Acid	50µg/ml	10µl
		100mM Monothioglycerol	0.4mM	40µl
		50ng/µl BMP4	5ng/ml	1µl
		100ng/µl VEGF	50ng/ml	5µl
		10ng/µl bFGF	20ng/ml	20µl
	iPSC-HSC differentiation medium (for day 4-5)	Stem Pro	-	make up to 10 ml
		200mM Glutamine	2mM	100µl
		100X Penicillin Streptomycin solution	0.5X	50µl
		50mg/ml Ascorbic Acid	50µg/ml	10µl
		100mM Monothioglycerol	0.4mM	40µl
		100ng/µl VEGF	15ng/ml	1.5µl
		10ng/µl bFGF	5ng/ml	5µl

Differentiation of hiPSCs to HSCs	iPSC-HSC differentiation medium (for day 6)	SFD medium	-	make up to 10 ml
		200mM Glutamine	2mM	100µl
		100X Penicillin Streptomycin solution	0.5X	50µl
		50mg/ml Ascorbic Acid	50µg/ml	10µl
		100mM Monothioglycerol	0.4mM	40µl
		100ng/µl VEGF	50ng/ml	5µl
		10ng/µl bFGF	50ng/ml	50µl
		50ng/µl SCF	50ng/ml	10µl
		50ng/µl Flt3L	5ng/ml	1µl
	iPSC-HSC differentiation medium (for day 7-10)	SFD medium	-	make up to 10 ml
		200mM Glutamine	2mM	100µl
		100X Penicillin Streptomycin solution	0.5X	50µl
		50mg/ml Ascorbic Acid	50µg/ml	10µl
		100mM Monothioglycerol	0.4mM	40µl
		100ng/µl VEGF	50ng/ml	5µl
		10ng/µl bFGF	50ng/ml	50µl
		50ng/µl SCF	50ng/ml	10µl
		50ng/µl Flt3L	5ng/ml	1µl
		40ng/µl TPO	50ng/ml	12.5µl
		10ng/µl IL6	10ng/ml	10µl
	SFD medium - 10ml	IMDM	75%	7.5ml
		Ham's F12	25%	2.5ml
		100X N-2 supplement	0.5%	50µl
		50X B-27 supplement	1%	100µl
		1% BSA in 1XPBS	0.05%	100µl

3.3. Plasmids obtained from public repositories, other investigators and commercial sources

The complete list of the plasmids used in this study is given in **Table 2**.

3.4. Generation of doxycycline-inducible overexpression plasmids

pINDUCER20-FANCA was constructed by Gateway cloning. The destination vectors, pINDUCER20 were obtained from Stephen Elledge, HHMI. The entry vector pENT223-FANCA was obtained from DNASU. Gateway cloning was carried out using Gateway LR Clonase II Enzyme mix (Life Technologies) as per the manufacturer's instructions. Briefly, 2μl of LR clonase II enzyme mix was added to 10μl of TE buffer containing 150ng of destination vector and 100ng of entry vector. The reaction mix was mixed well by vortexing briefly twice, followed by a brief spin to settle the contents. The mix was incubated at 25^0C for 90 minutes, and the reaction was terminated by incubating with 1μl of proteinase K at 37^0C for 10 minutes. From the reaction mix, 1μl was used for transformation of 50μl of XL1 Blue competent cells. After 30 minutes incubation on ice, the competent cells were subjected to heat shock at 42^0C for 45 seconds. The cells were immediately transferred to ice and cooled for 2 minutes. The transformed cells were expanded in 250μl of SOC medium in a shaker incubator at 37^0C/800rpm for 90 minutes. Following incubation, 50-100μl of culture was plated on LB plates with 100ng/ml Ampicillin. The plates were incubated for 16hrs at 37^0C bacterial incubator for colony formation. Suspension culture was initiated by inoculating five individual colonies in 5ml LB broth with ampicillin that was incubated for 16hrs in 37^0C incubator shaker. Plasmid was isolated from the culture using MiniPrep kit following the manufacturer's protocol (BioBasics). The final gateway expression plasmids were digested with specific restriction enzymes (**Table 3**) and sequenced to confirm the insert.

Table 2: Plasmids used in this study

Plasmids	Source	Reference
Lentiviral Plasmids		
pLenti6/UbC/Slc7a1	Addgene 17224	Takahashi *et al*., 2007
pHAGE2-hSTEMCCA	gift from Gustavo Mostoslavsky, MIT	Sommer *et al*., 2010
pINDUCER20	gift from Stephen Elledge, HHMI	Meerbrey, KL *et al*., 2011
pINDUCER21		
pENT223-FANCA	Transomic-TOH6003	-
pENT223-GRHL1	DNASU HsCD00515977	-
pENT223-GRHL2	DNASU HsCD00514302	-
pENT223-GRHL3	DNASU HsCD00513349	-
pINDUCER20 - FANCA	constructed in house	-
pZIP-mEF1a	Transomics TLN0005, Promoter selection kit	-
pZIP-hEF1a		-
pZIP-mCMV		-
pZIP-hCMV		-
pZIP-SFFV		-
pZIP-UbC	constructed in house	-
pZIP-CBA	constructed in house	-
Retroviral Plasmids		
pMXs-mRFP1	Addgene 21315	Hotta *et al*., 2009
pMXs-hOCT3/4	Addgene 17217	Hotta *et al*., 2009; Takahashi *et al*., 2007
pMXs-hSOX2	Addgene 17218	
pMXs-hKLF4	Addgene 17219	
pMXs-c-MYC	Addgene 17220	
Pooled shRNA library module		
DECIPHER Pooled shRNA Library - Human Module 1 - signaling pathways	Addgene 28285	-

Table 3: Restriction digestion patterns of doxycycline inducible overexpression plasmids and constitutive shRNA knockdown plasmid

Plasmid	Restriction enzyme	Band sizes
pINDUCER20-FANCA	*BstXI*	1177bp, 10694bp
pINDUCER20-GRHL1		635bp, 1503bp, 10326bp
pINDUCER20-GRHL2		2159bp, 10326bp
pINDUCER20-GRHL3		449bp, 1483bp, 10326bp
pINDUCER21-GRHL1		635bp, 1503bp, 10694bp
pINDUCER21-GRHL2		2159bp, 10694bp
pINDUCER21-GRHL3		449bp, 1483bp, 10694bp
pZIP-hEF1a-shRNA	*HpaI / MluI*	242bp, 8713bp

3.5. Generation of shRNA lentiviral plasmids

For cloning of shRNAs the protocol described by Knott *et al* was used (Knott *et al.* 2014). An oligo with the Ultramir scaffold sequence, GTTAACTGAATACCTTGCTATCTCTT TGATACAATTTTTACAAAGCTGAATTAAAATGGTATAAATTAAATCACTTTACGCGT was synthesized commercially and amplified with the primers, Ultramir-F and Ultramir-R (**Table 4**). The amplified product was digested with *HpaI* and *MluI* and then it was cloned into pZIP-EF1α-shRNA plasmid to generate the plasmid containing ultramiR scaffold. shRNA sequences were obtained from SHERWOOD (http://sherwood.cshl.edu) (**Table 5**) and the oligos were amplified with primers, Ultramir-Hpa-I-F and Ultramir-Hpa-I-R (**Table 4**). Subsequently, pZIP-EF1α-shRNA plasmid contaning ultramiR scaffold was digested with *Hpa*I and the the amplified shRNAs were cloned by Gibson assembly following the manufacturer's protocol. Plasmids obtained after transformation were verified by digestion with *Hpa*I and *Mlu*I and by DNA sequencing (**Table 3**).

3.6. Generation of lentiviruses

HEK293T (CRL-1126, ATCC) cells were cultured in D10 medium. For preparing lentiviruses, HEK 293T cells were seeded at a count of 4×10^6 cells on a 10 cm dish. About 12-16 hrs later, the cells were transfected with 7 µg of lentiviral expression or shRNA plasmid along with 3.5 µg of pMD2.G envelope plasmid (Addgene 12259) and 3.5 µg of psPAX2 packaging plasmid (Addgene 12260) (gifts from Didier Trono), using X-tremeGENE HP (Roche Life Science) following the manufacturer's protocols. After 24 hrs, transfection complex containing medium was replaced with fresh medium. The viral supernatant was collected at 48, 60 and 72 hrs, pooled and filtered through a 0.45 um filter. The lentiviruses were concentrated using Lenti-X Concentrator (Clontech Laboratories, CA) following the manufacturer's protocols. Briefly, Lenti-X Concentrator was added at a volume equal to $1/3^{rd}$ of viral supernatant, mixed well by inverting tube 4-5 times and left undisturbed at 4^0C overnight for precipitation of viral particles. The viral supernatant was centrifuged at 1500g

Table 4: Primers used in this study

Real Time PCR for analyzing the expression of pluripotency genes	
hβACTIN Sense	GACGACATGGAGAAAATCTG
hβACTIN Antisense	ATGATCTGGGTCATCTTCTC
hNANOG Sense	GCTTGCCTTGCTTTGAAGCA
hNANOG Antisense	TTCTTGACTGGGACCTTGTC
hOCT3/4 Sense	CCTCACTTCACTGCACTGTA
hOCT3/4 Antisense	CAGGTTTTCTTTCCCTAGCT
hSOX2 Sense	CCCAGCAGACTTCACATGT
hSOX2 Antisense	CCTCCCATTTCCCTCGTTTT
hGDF3 Sense	AAATGTTTGTGTTGCGGTCA
hGDF3 Antisense	TCTGGCACAGGTGTCTTCAG
hDNMTB Sense	ATAAGTCGAAGGTGCGTCGT
hDNMTB Antisense	GGCAACATCTGAAGCCATTT
hZFP42 Sense	GAATTCAGACCTAACCATCG
hZFP42 Antisense	TGAGCACTACTAGAGTGAAG
Real Time PCR for measuring the transgene expression	
Exo hOCT3/4 Sense	TTAAGGATCCCAGTGTGGTG
Exo hOCT3/4 Antisense	TAGCCAGGTCCGAGGATCAA
Exo hSOX2 Sense	AATTAAGGATCCCAGTGTGGTG
Exo hSOX2 Antisense	CTTCAGCTCCGTCTCCATCAT
Exo hKLF4 Sense	AATTAAGGATCCCAGTGTGGTG
Exo hKLF4 Antisense	GTGGAGAAAGATGGGAGCAG
Exo hc-MYC Sense	AATTAAGGATCCCAGTGTGGTG
Exo hc-MYC Antisense	CGAGGTCATAGTTCCTGTTGGT
Exo mRFP1 Sense	CCCGCCGACATCCCCGACTA
Exo mRFP1 Antisense	CTGGGTCACGGTCACCACGC
Amplification of bisulfite converted DNA for DNA methylation analysis	
Me hOCT3/4 Sense	GAGGTAATATAAGAGTGGGGGAAAA
Me hOCT3/4 Antisense	CCCTAAACTCCCCTTCAAAA
Me hNANOG Sense	GGGGGTTTTGTTATGTTGGTTA
Me hNANOG Antisense	ATCCCATCTACCAATCTCACCA
Real Time PCR for calculating knockdown efficiency	
hPRDM14 Sense	GAGCCTTCAGGTCACAGAGC
hPRDM14 Antisense	GTCCACACAGGGGGTGTACT
hPRDM1 Sense	AACGGCCTTTCAAATGTCAG
hPRDM1 Antisense	TCTTGAGATTGCTGGTGCTG
hMECOM Sense	ACAGCAGTGTGAAGCCCTTT
hMECOM Antisense	ATTTGGGTTCTGCAATCAGC

Real Time PCR for calculating knockdown efficiency	
hSETD7 Sense	GGGAGGCCTCACTTTGAACT
hSETD7 Antisense	GATCTGGAAGAAGAGCATTGG
hPRDM12 Sense	CTCGGCATCTTCTCCAAGAC
hPRDM12 Antisense	CGTGCCATCCTCATTGAAC
hKDM1B Sense	CCCGGGTACTCGGTGATAAT
hKDM1B Antisense	GGGAGGAACGTGACCAAAA
hKDM6B Sense	TATCGCCTCTGAGGTGGAAG
hKDM6B Antisense	GGGGAAAGGTAGGACTCTCG
hKDM2B Sense	CATGCAGCAGAAAAGCAAAA
hKDM2B Antisense	TCCGACAAGTCCTCGTTCTC
hDNMT3A Sense	AGACCCCTGGAACTGCTACA
hDNMT3A Antisense	GGGACAGGTGGGTAAACCTT
hTET1Sense	GGTGAATGCACTTACTGCAAGA
hTET1 Antiense	TTTGGGCTTCTTTTCCCTCT
hJMJD1C Sense	AGGCTCAACCTTACGGGATT
hJMJD1C Antisense	TGTTAGGCATAGTCCGTCCA
hJARID2 Sense	TCCAAAGTGTGCTGTGGGTA
hJARID2 Antisense	TGCTTCTGCTTGTGCAATCT
hKDM1A Sense	CGGCATCTATAAGAGGATAAAACC
hKDM1A Antisense	CCTGGCTTCCAAAAGTGTGA
hKDM2A Sense	ACTGCATAACCAACCGTTCC
hKDM2A Antisense	TTCTCGATCCACTGCTTCCT
hKDM4A Sense	GGGCAGTCTGGCCTCTTTAC
hKDM4A Antisense	CGCTCGAGCTCTTCAAACTC
hTET2 Sense	TGGATACACCTGTCAAGACTCAA
hTET2 Antisense	GGACCTGCTCCTAGATGGGTA
hBMI1 Sense	ATGCAGCTCATCCTTCTGCT
hBMI1 Antiense	TCTGGTCAAAGAATTCAATGGATA
hGRHL1 Sense	CTGAGCCAGATCACAGCAAA
hGRHL1 Antisense	GCACTCTGTTTTCTCCAGCA
hGRHL2 Sense	AGTTGGGGCTGAGGAGTACA
hGRHL2 Antisense	CCTGCTTCTGACGGAGAGAT
hGRHL3 Sense	GTTCCGGAGGAAGGTCAAGT
hGRHL3 Antisense	CAGTCTCTGGCCGAAGGTAG
hSUV420H2 Sense	TCGATACGGCAATGTGTCTC
hSUV420H2 Antisense	AAAGTCAGCTCTTCTCCAGCA

Primers used for pZIP-hEF1a-shRNA vector construction by Gibson assembly	
Ultramir-F	CTTCTTCAGGTTAACTGAATACCTTGCTATCTC
Ultramir-R	GTCCAGACGCGTAAAGTGATTTTAATTTATAC
Ultramir-Hpa-I-F	CTGGGATTACTTCTTCAGGTTAACCCAACAGAAGGCTAAAGAAGGTA TATTGCTGTTGACAGTGAGCG
Ultramir-Hpa-I-R	AGAGATAGCAAGGTATTCAGTTTTAGTAAACAAGATAATTGCTCCTAA AGTAGCCCCTTGAAG
Primer used for sequencing shRNA in the pZIP-hEF1α--shRNA vector	
pZIP-Seq-For	GGTGCCCGAAGGACCG
Primers used for high throughput sequencing of shRNA barcodes (Ref: Cellecta Decipher RNAi screen manual)	
FwdHTS	TTCTCTGGCAAGCAAAAGACGGCATA
RevHTS	TGCCATTTGTCTCGAGGTCGAGAA
FwdGex	CAAGCAGAAGACGGCATACGAGA
IND1AS	AATGATACGGCGACCACCGAGAGGTTCAGAGACGTACAGTCCGAAA
IND2AS	AATGATACGGCGACCACCGAGAGGTTCAGATTATTACAGTCCGAAA
IND3AS	AATGATACGGCGACCACCGAGAGGTTCAGAAGTATACAGTCCGAAA
IND4AS	AATGATACGGCGACCACCGAGAGGTTCAGACCGCTACAGTCCGAAA
GexSeqN	ACAGTCCGAAACCCCAAACGCACGAA

Table 5: Sequence of shRNAs used for knocking down the expression of target genes

Target Gene	shRNA Sequences
hBMI1	TGCTGTTGACAGTGAGCGAATGAAGATAAGAGAATTATAATAGTGAAGCCACAGATGTATTATAATTCTCTTATCTTCATCTGCCTACTGCCTCGGA
hDNMT3A	TGCTGTTGACAGTGAGCGATCCAGATGTTCTTCGCTAATATAGTGAAGCCACAGATGTATATTAGCGAAGAACATCTGGAGTGCCTACTGCCTCGGA
hGRHL1	TGCTGTTGACAGTGAGCGCCAGAGTGCAAGTACTGAAAAATAGTGAAGCCACAGATGTATTTTTCAGTACTTGCACTCTGTTGCCTACTGCCTCGGA
hGRHL2	TGCTGTTGACAGTGAGCGCCAGCGAGACCGGAGACAACAATAGTGAAGCCACAGATGTATTGTTGTCTCCGGTCTCGCTGATGCCTACTGCCTCGGA
hGRHL3	TGCTGTTGACAGTGAGCGCCAGGGGCAATGAGACGACCTATAGTGAAGCCACAGATGTATAGGTCGTCTCATTGCCCCTGATGCCTACTGCCTCGGA
hJARID2	TGCTGTTGACAGTGAGCGCCTCCATGGAGAAGTTACTCTATAGTGAAGCCACAGATGTATAGAGTAACTTCTCCATGGAGATGCCTACTGCCTCGGA
hJMJD1C	TGCTGTTGACAGTGAGCGCAACCTCCAAGATAAATGTAAATAGTGAAGCCACAGATGTATTTACATTTATCTTGGAGGTTTTGCCTACTGCCTCGGA
hKDM1A	TGCTGTTGACAGTGAGCGACAGCAGCTCGACAGTTACAAATAGTGAAGCCACAGATGTATTTGTAACTGTCGAGCTGCTGCTGCCTACTGCCTCGGA
hKDM1B	TGCTGTTGACAGTGAGCGCCTGGAGATGAAGTGCAGGTTATAGTGAAGCCACAGATGTATAACCTGCACTTCATCTCCAGATGCCTACTGCCTCGGA
hKDM2A	TGCTGTTGACAGTGAGCGCCAGCATGGATTTGGAGTTAAATAGTGAAGCCACAGATGTATTTAACTCCAAATCCATGCTGATGCCTACTGCCTCGGA
hKDM2B	TGCTGTTGACAGTGAGCGACCCGATTGACCGCCAGCGATATAGTGAAGCCACAGATGTATATCGCTGGCGGTCAATCGGGCTGCCTACTGCCTCGGA
hKDM4A	TGCTGTTGACAGTGAGCGCCTGGAAAAATCTTACATTCAATAGTGAAGCCACAGATGTATTGAATGTAAGATTTTTCCAGTTGCCTACTGCCTCGGA
hKDM6B	TGCTGTTGACAGTGAGCGACAGCAGGAATGCCAAGGTGAATAGTGAAGCCACAGATGTATTCACCTTGGCATTCCTGCTGGTGCCTACTGCCTCGGA
hMECOM	TGCTGTTGACAGTGAGCGAAAAGTGATAAAGAGAAATTTATAGTGAAGCCACAGATGTATAAATTTCTCTTTATCACTTTCTGCCTACTGCCTCGGA
hPRDM1	TGCTGTTGACAGTGAGCGCCACCCAGTTTGTGCACCTGAATAGTGAAGCCACAGATGTATTCAGGTGCACAAACTGGGTGATGCCTACTGCCTCGGA
hPRDM12	TGCTGTTGACAGTGAGCGAATCAAGTGTGCACGTAACGAATAGTGAAGCCACAGATGTATTCGTTACGTGCACACTTGATGTGCCTACTGCCTCGGA
hPRDM14	TGCTGTTGACAGTGAGCGACAGAAGGCTACAGATGTGAAATAGTGAAGCCACAGATGTATTTCACATCTGTAGCCTTCTGCTGCCTACTGCCTCGGA
hSETD7	TGCTGTTGACAGTGAGCGCTGGAGAGATGATAGAAGGCAATAGTGAAGCCACAGATGTATTGCCTTCTATCATCTCTCCATTGCCTACTGCCTCGGA
hSUV420H2	TGCTGTTGACAGTGAGCGCCTGGACTATGAGTCTGATGAATAGTGAAGCCACAGATGTATTCATCAGACTCATAGTCCAGATGCCTACTGCCTCGGA
hTET1	TGCTGTTGACAGTGAGCGCCAGGAAGGAAGATGTAAACAATAGTGAAGCCACAGATGTATTGTTTACATCTTCCTTCCTGATGCCTACTGCCTCGGA
hTET2	TGCTGTTGACAGTGAGCGAATGGTGAAAATCAGTATTCAATAGTGAAGCCACAGATGTATTGAATACTGATTTTCACCATGTGCCTACTGCCTCGGA

Note: The sequence corresponding to hairpin stem is shown in violet colour and loop in red colour. The sequences in black colour are part of the miR30 backbone used for binding of primers during PCR amplification. The shRNA sequences were obtained from http://sherwood.cshl.edu.

for 1hr at 4^0C. The viral pellet was dissolved in cold DMEM to get 50X concentrated viruses. Concentrated viruses were aliquoted and stored in -80 for future use.

3.7. Generation of retroviruses

PlatE cells (RV-101, Cell Biolabs) were cultured in D10 medium. For preparing retroviruses, Plate-E cells were seeded at a count of 3.6 x 10^6 on a 10 cm dish and, after 12-16hrs, they were transfected separately with 14 µg of retroviral expression plasmids. After 24 hrs, transfection complex containing medium was replaced with fresh medium. The viral supernatant was collected at 48, 60 and 72 hrs, pooled and filtered through a 0.45 um filter. The freshly made retroviral supernatant was directly used for infecting the cells.

3.8. Doxycycline-inducible expression of transgenes

For the conditional complementation of FANCA, the FA-A fibroblast cells were transduced with pINDUNCER20-FANCA lentiviruses and the transduced cells were enriched by G418 antibiotic selection. The pINDUCER20 - FANC A - fibroblast were treated with 100-1000ng/µl of doxycycline (DOX) to induce the expression of FANCA transgene.

3.9. Validation of the knockdown efficiency of shRNA plasmids

hAHDF or HELA cells were seeded at a count of 8x10^5 cells on a 12 well plate in αMEM20 or D10, and 12-14hrs later, the cells were subjected to lentiviral transduction. The cells in each well were infected with 20µl of 50X concentrated shRNA lentiviruses in 1ml of medium with 0.5 µl of 1X polybrene to obtain ~40% of transduction efficiency. The cells were incubated at 37^0C/5% CO_2 incubator for 24 hours. Following incubation, the virus-containing medium was replaced with fresh medium. On day 2, the cells were split at 1:3 ratio. From day 3 or 4, the cells were subjected to puromycin selection to enrich stably transduced cells. Following antibiotic selection, the cells were checked under a fluorescent microscope to confirm that all the cells were GFP^+. The cells were grown for one or two more passages and

then harvested for RNA extraction. Real-time PCR was performed using specific primers (**Table 4**) to measure the knockdown efficiency.

3.10. Human somatic cell reprogramming strategy

Three different protocols for reprogramming were used in this study. The details of these protocols are given below.

3.10.1. Using lentiviral and retroviral vectors

For lentiviral-mediated reprogramming, hADFs were seeded at a count of 8×10^5 cells on a 12 well plate. About 12 – 16 hrs later, the cells were transduced with concentrated hSTEMCCA lentiviruses to get a transduction efficiency of about 70-80% as measured by immunofluorescence analysis of OCT4 expression in the transduced fibroblasts. Six days after transduction, the cells were seeded on mitomycin C-treated SNL feeders at a count of 5×10^5 on a 6 well plate in the hiPSC medium. To facilitate ecotopic retroviral-mediated reprogramming of hADFs, they were transduced with lentiviruses to express mouse Slc7a1, and the transduced cells were selected with Blasticidin S (Life Technologies, CA). $Slc7a1^+$ hADFs were transduced with pMXs-mRFP1 retroviruses, and RFP^+ cells were sorted by FACS. For retroviral-mediated reprogramming, about 12 - 16 hrs before transduction, RFP^+ or RFP^- $Slc7a1^+$ hADFs were seeded at a count of $8x10^5$ cells in fibroblast medium on a 10 cm dish. The cells were subjected to two rounds of transduction with pools of freshly prepared OSKM retroviral supernatants at 1:1:1:1 ratio in an interval of 48 hrs between the first and the second transductions. On day 4, $5x10^5$ transduced hADFs were seeded on a 6 well plate containing mitomycin C-treated SNL feeder cells in fibroblast medium. Two days later, the medium was changed to hiPSC medium. The reprogramming cells were fed daily with fresh medium and were observed for the emergence of hiPSC colonies.

3.10.2. Using episomal vectors

For episomal mediated reprogramming, hADFs were transfected with Y4 combination of plasmids using Neon transfection system (Life Technologies, CA) based on a protocol previously described (Okita *et al.* 2011). Briefly, 6×10^5 hADFs loaded in 100µl Neon

tip were electroporated with 1µg each of the plasmids, pCXLE-hOCT3/4-shp53-F (Addgene 27077), pCXLE-hSK (Addgene 27078) and pCXLE-hUL (Addgene 27080) (gifts from Shinya Yamanaka) using the condition 1650V / 10 mS / 3 pulses. Transfected cells were seeded on one well of a 6 well plate in Amniomax (Life Technologies, CA) and fed with fresh medium on alternate days. After 6 days, transfected cells were seeded on mitomycin C-treated SNL feeder cells at a count of 3×10^5 cells per 10 cm dish in Amniomax and the medium was changed to hiPSC medium on the next day. The cells were fed daily with hiPSC medium and observed for the emergence of colonies.

3.10.3. Using Sendai virus (SEV) based vectors

For SeV based reprogramming, pIND20-FANCA - patient fibroblasts were transduced with CytoTune 2.0 Sendai reprogramming vectors (A16517, Life Technologies) according to the manufacturer's protocol. Briefly, 70 - 80% confluent fibroblasts were transduced with 20µl each of CytoTune 2.0 KOS (hKLF4-OCT4-SOX2), CytoTune 2.0 hc-MYC and CytoTune 2.0 hKLF4 vectors in 1ml of fibroblast culture medium per well of a 6 well plate. Following overnight incubation with the reprogramming cocktail, the medium was replaced with fresh fibroblast medium and cells were fed daily with fresh medium. On day 7, the transduced cells were seeded on mitomycin C-treated SNL feeder layer in fibroblast medium at a count of 2×10^5 cells per 10cm dish. From next day, cells were fed daily with hiPSC medium and were monitored for the emergence of Fanconi anemia (FA)-A hiPSC colonies. The expression of *FANCA* gene was induced according to experimental requirements by the addition of doxycycline into the culture medium.

3.11. Derivation and maintenance of hiPSC lines

For the derivation of hiPSC lines by retroviral method (retro hiPSC), the RFP^- colonies with hESC-like morphology were manually picked up in the 3^{rd} – 4^{th} weeks of reprogramming. hiPSC clones were isolated solely based on morphology for deriving hiPSC lines by episomal and Sendai virus-mediated reprogramming. The isolated clones were mechanically broken into small clumps and seeded on mitomycin C-treated SNL feeder layer

in the hiPSC medium. When the confluency reached 60-70% (4-5 days), the clones were subsequently passaged onto feeders by enzymatic treatment with 1mg/ml Collagenase IV (Life Technologies) for 45 minutes in 5% CO_2 incubator. For establishing a feeder-free culture of hiPSC lines, the colonies maintained on feeders were harvested using 1mg/ml Collagenase IV and seeded on vitronectin-coated TC dishes in the hiPSC medium. The next day, the hiPSC medium is replaced partially with E8 medium (Life Technologies). From the next day onwards, the complete medium is replaced with the fresh E8 medium. For subsequent passaging of cells, the cells were harvested using EDTA solution and grown in E8 medium. For maintaining FA-A hiPSC lines, doxycycline at a concentration of 100-1000ng/ml was added to the medium.

3.12. Cryopreservation of hiPSC lines

For making hiPSC line stocks, 4-5 days old undifferentiated hiPSC culture maintained on mitomycin C-treated SNL feeder layers were taken for cryopreservation. The colonies at 70% confluency on a 10cm dish or a 6 well plate were harvested in hiPSC medium with Collagenase IV as mentioned above. After a spin at 800rpm for 5 minutes, colony pellets were resuspended in 2ml of spent hiPSC medium by tapping gently. The colony suspension was aliquoted into 10 cryovials, each with 200 μl. Then, 200μl of cold 2X hiPSC freezing medium (**Table 1**) was added to each vial drop by drop, by tapping the vial after the addition of each drop. The vials are immediately transferred to Mr. Frosty (Nalgene) maintained at 4^0C and shifted to -80^0C for slow freezing. The frozen cell stocks were transferred to liquid nitrogen vapour phase the following day.

3.13. Revival of cryopreserved hiPSC lines

Before reviving hiPSCs, D10 medium in the mitomycin C-treated feeder plates was substituted with hiPSC medium containing 10mM Rock Inhibitor (Y-27632, Calbiochem) and incubated at 37^0C/5%CO_2 for at least 3hrs. For thawing frozen hiPSC lines, 1ml of warm hiPSC medium was gently added to the frozen cells in the cryovials and the thawed contents

were immediately transferred to a 15ml tube. This process was repeated till the entire content in the cryovial was moved to the 15ml tube. The thawed colony suspension was pipetted gently once and centrifuged at 800rpm for 5 minutes. The pelleted hiPSC colonies were resuspended in 1ml warm hiPSC medium with 10mM rock inhibitor by gentle tapping and seeded onto the feeder plates. The colonies were uniformly distributed on feeder layer and left undisturbed in 37^0C/5% CO_2 incubator for 36hrs. Following this, the hiPSC lines were fed daily with hiPSC medium and maintained on feeder layers as described earlier.

3.14. Immunofluorescence assay

For detecting surface marker expression, the cells maintained in culture were washed with phosphate-buffered saline (PBS), fixed with 4% paraformaldehyde at room temperature for 20 minutes and then blocked with a buffer containing 1% bovine serum albumin (BSA) and 5% FBS for 45 minutes at room temperature. For intracellular protein analysis, the cells were permeabilized with blocking buffer containing 0.1% TritonX-100. The fixed and permeabilized cells were incubated with primary antibodies in the blocking buffer for 4 hours at 4^0C and 1 hour at room temperature followed by incubation with suitable labeled secondary antibodies for 1 hour at room temperature. Imaging was done under a fluorescent microscope (Leica Microsystems) with appropriate filters. The pluripotency of hiPSC lines was confirmed by analysis of expression of SSEA-4, TRA-1-60, TRA-1-81, NANOG, OCT4, SOX2, and LIN28. For confirming the trilineage differentiation potential of hiPSC lines, the EB-derived cells were analyzed for expression of α-fetoprotein, α-smooth muscle actin, and βIII-Tubulin. For checking FA pathway functionality in FA-hiPSCs, the primary antibodies used were FANCA and FANCD2. The complete list of primary and secondary antibodies used in this experiment is given in **Table 6**.

3.15. Flow cytometry

For analyzing expression of pluripotency marker by hiPSC clones, the colonies on feeder were treated briefly with 1mg/ml Collagenase IV to remove feeders, and the attached

Table 6: Antibodies used in this study

Purpose	Antibody-Flurochrome	Cat no	Source
a. Flow cytometry			
Analysis and sorting of reprogramming cells	CD13-PE	555394	BD Pharmingen, San Diego, CA
	SSEA-4-PE	560128	
	SSEA4-AF647	560796	
	TRA-1-60-BV421	562711	
	TRA-160-PE	560193	
	TRA-1-85-AF700	FAB3195N	R&D Systems, Minneapolis, MN
	TRA-1-85-FITC	FAB3195F	
Assessment of haematopoietic differentiation of hiPSCs	CD31-AF647	561654	BD Pharmingen, San Diego, CA
	CD235a-APC	561775	
b. Immunofluorescence Assay			
Analysis of pluripotency marker expression	SSEA-4	SCR001	Millipore, MA
	TRA-1-60	sc-21705	Santa Cruz Biotechnology, Texas
	TRA-1-81	sc-21706	
	NANOG	4903	Cell Signaling Technology, MA
	Oct-04	2840	
	SOX2	3579	
	LIN28	3695	
Analysis of trilineage differentiation of hiPSCs	α-Fetoprotein	MAB1368	R&D Systems, MN
	α-Smooth Muscle Actin	ab5694	Abcam, Cambridge
	βIII-Tubulin	ab5568	Cell Signaling Technology, MA
Analysis of FA pathway functionality	FANCA	PA5-22276	ThermoFisher Pierce, MA
	FANCD2	SC-20022	Santa Cruz Biotechnology, Texas
Secondary antibodies raised in goat	anti-mouse IgGs (H+L) - AF488	11029	Invitrogen, IL
	anti-mouse IgGs (H+L) - AF594	11032	
	anti-rabbit IgGs (H+L) - AF488	11034	
	anti-rabbit IgGs (H+L) - AF594	11037	
Abbreviations*: APC - Allophycocyanin, PE - phycoerythrin, AF488 - Alexaflour488, AF594 - Alexaflour 594, AF647 - Alexaflour 647, BV421 - Brilliant Violet 421, AF700 - Alexaflour 700, FITC - Fluorescein*			

colonies were then harvested using TryLE Express (12604-021, Life Technologies). For analyzing the expression kinetics of surface markers in the cells undergoing reprogramming and for sorting of the cells based on marker expression profile, the cells were treated with the TrpLE express or 1mg/ml Collagenase IV, followed by Accutase (Chemicon Millipore, CA). To the single cell suspension in dissociation solution, an equal volume of warm D10 was added, mixed by pipetting gently twice, filtered through a 70 µm cell strainer (BD Biosciences) and centrifuged at 800rpm for 5 minutes. The harvested cells were resuspended in hiPSC medium to get a count of about $1x10^6$ cells per 100µl of the medium. Optionally, feeder cells in the sample were depleted using feeder removal microbeads and LD column on Quadro MACS following the manufacturer's protocol (Miltenyi Biotec, Bergisch Gladbach, Germany). The harvested cells were then stained with labelled primary antibodies at recommended dilution in the hiPSC medium for 30 minutes in the dark at 4°C with the intermittent tapping of cells. Following incubation, the cells were washed with cold D10 once and centrifuged at 800rpm for 5 minutes. The cell pellet was resuspended in cold hiPSC medium at a volume 3-4 times the initial volume and taken for analysis or sorting using FACS machine. FACS Aria III flow cytometer (BD Biosciences, San Jose, CA), FACS Calibur flow cytometer (BD Biosciences, San Jose, CA) and CytoFLEX flow cytometer (BC, Pasadena, CA) were the FACS machines used. The complete list of labelled antibodies used in this study is given in **Table 6**.

3.16. Cell cycle analysis

Cell cycle analysis of FA-A hiPSC clones was carried out using ClickIT EdU proliferation assay kit (Life Technologies). Briefly, iPSC clones at actively growing state were treated with EdU at 20mM concentration for 1 hr. Cells were harvested as single cell suspension using TryLE Express. After giving a PBS wash, the cells were fixed with 4% paraformaldehyde for 15 minutes at room temperature. Fixed cells were washed with wash buffer containing 1X PBS and 1% BSA and then permeabilized using wash buffer containing saponin solution for 30 minutes at room temperature. The permeabilized cells were washed

with wash buffer containing saponin wash buffer followed by treatment with ClickIT reaction buffer containing Azide-Alexaflour488 for 30 minutes at room temperature. The cells were washed with wash buffer and then stained with DAPI for minimum 2 hours at 4^0C. The cells were finally washed with wash buffer and the cell pellet was resuspended in PBS. The cell cycle profile of the cells was analyzed using FACS Aria III flow cytometer based on fluorescence signal from Alexaflour488 and DAPI.

3.17. Quantitative Real-Time PCR

RNA was extracted from fibroblasts, cell lines, flow-sorted reprogramming cells and hiPSCs using Tri-reagent (Sigma-Aldrich). 1 µg of total RNA was used for reverse transcription reaction using high capacity cDNA reverse transcription kit (Life Technologies) according to the manufacturer's instructions. Quantitative RT-PCR was set up with SYBR Premix Ex Taq II (Takara Bio) using specific primers (**Table 4**) and analyzed with ABI 7500 (Life Technologies) or QuantStudio12K Flex (Life Technologies) real-time PCR systems.

3.18. Bisulfite sequencing

Genomic DNA extracted from established hiPSC lines was subjected to bisulfite conversion using EpiTect Bisulfite Kit (Qiagen) according to the manufacturer's recommendations. Using specific primers (**Table 4**) that can bind bisulfite converted DNA, the promoter regions of human *OCT4* and *NANOG* genes were amplified. The PCR products were cloned into a TA cloning vector pCR2.1 (Life Technologies, CA), and plasmids were extracted from five to ten transformed bacterial clones and were screened by DNA sequencing.

3.19. Western Blot analysis

hiPSCs were harvested as described in the previous section and lysed using RIPA buffer (150mM sodium chloride, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS and 50mM Tris, pH 8) containing phenyl methane sulfonyl fluoride (PMSF) and protease

inhibitor for 30 minutes on ice. The cell lysate was centrifuged for 30 minutes at 12000rpm. The supernatant was aliquoted into multiple tubes and stored in -70^{0}C for future use. Protein concentration was estimated using Bradford's method. For performing WB, 30µg of protein was denatured by boiling for 5minutes at 90^{0}C in appropriate volumes of 4X Lamelli buffer. The sample were cooled on ice for 5 minutes and centrifuged at 12000rpm for 20 minutes at 4^{0}C. The supernatant was loaded on to 5% of stacking gel in parallel with 3-colour broad range prestained protein ladder (Puregene). The bands were resolved on a 7.5% resolving gel. For detection of FANCA and FANCD2, the electrophoresis was carried out at 100V for 3hrs and 7hrs, respectively. Separated protein bands were transferred onto 0.45µ polyvinylidene (PVDF) membrane filter (Millipore) at 30V for 16hrs. After transfer, the membrane was blocked using 1X PBS with 0.01% Tween 200 (PBST) containing 5% non-fat milk powder (Bio-Rad). Blocked membrane was incubated with primary antibodies against FANCA or FANCD2 (**Table 6**) for 4hrs at room temperature. Following PBST wash, the membrane was incubated with secondary antibodies conjugated with horse raddish peroxidase (HRP) for 2hrs at room temperature. The membrane was washed with PBST for 30 minutes to remove unbound antibodies. The membrane was treated with Westar supernova chemiluminescent substrate (Cyanogen). The protein bands were visualized under FluorChemE gel documentation system (Protein Simple). FANCA, FANCD2, and ubiquitinated FANCD2 give bands of size 163kDa, 143kDa and 150kDa repectively.

3.20. Alkaline phosphatase staining

The FA-A hiPSC colonies maintained on feeder-free or mitomycin C-treated SNL feeder layers, in the presence or absence of DOX, were taken for alkaline phosphatase staining on fourth-fifth day after passaging. The staining was performed using Leukocyte alkaline phosphatase kit (86R, Sigma) following the manufacturer's protocol. Briefly, the colonies were given a 1X PBS rinse and fixed using fixative containing citrate solution, acetone and 37% paraformaldehyde (25:65:10) for 30 seconds. The fixed colonies were rinsed with distilled water and then incubated in dark with dye mix for

30 minutes in dark. Following incubation, the whole well of the culture plate was imaged.

3.21. Embryoid body (EB)-mediated differentiation

To initiate EB formation, 4 - 5 days old hiPSCs were harvested as large clumps using 1 mg/ml Collagenase IV and were transferred to a low attachment plate in EB suspension culture medium (**Table 1**). After 8 days of floating culture, EBs formed were seeded on 0.1% gelatin-coated culture dish in EB attached culture medium (**Table 1**) to induce further differentiation. The suspension and the attached culture were fed with the fresh medium on alternate days. The attached culture was monitored for differentiation into different cell types for 1 - 2 weeks. The differentiated cells were analyzed by immunofluorescence assay for the expression of α-fetoprotein (endoderm marker), α-smooth muscle actin (mesoderm marker) and βIII-tubulin (ectoderm marker) to confirm *in-vitro* trilineage differentiation potential of hiPSC clones.

3.22. Haematopoietic differentiation of hiPSCs

Differentiation of hiPSCs to HSCs was performed based on previously described protocol Mills *et al* (Mills *et al.* 2014). The hiPSC colonies maintained on vitronectin dishes were harvested using EDTA solution and seeded on growth factor reduced Matrigel matrix basement membrane (354230, Corning, Bedford, MA) coated dishes as small colonies in E8 medium to get a count of 1-2x10^5 cells / well. The culture was incubated in hypoxic condition in 37^0C / 5% CO_2 / 5% O_2 / 90% N_2 incubator for 24-48hrs till the colonies reached about 70% confluency. The culture was fed daily with the specific haematopoietic differentiation medium as indicated in **Table 1**. Exogenous FANCA expression was induced in the culture by doxycycline addition at 0.1-1µg/ml. The HSC differentiation was tracked by FACS by looking for expression of haematoendothelial and haematopoietic markers (**Table 6**).

3.23. Teratoma formation

Teratoma assay was performed based on a previously described protocol (Park *et al.* 2008). Briefly, feeder cells were removed from a 4 - 5 days old 60-70% confluent hiPSC culture by a brief treatment with 1 mg/ml Collagenase IV, and the cells were harvested using a cell scraper in DMEM-F12 medium. The cell pellet was resuspended in 40 µl of cold DMEM-F12, 80 µl of Collagen I (Life Technologies) and 120µl of hESC-qualified Matrigel (BD Biosciences) and 100 µl each of this colony suspension was injected intramuscularly into hind limbs of immunocompromised B6.CB17-Prkdcscid/SzJ (The Jackson Laboratory) or CB17/Icr-Prkdcscid/IcrIcoCrl (Charles River).

3.24. Karyotyping

Karyotyping was carried out at Unit of Cytogenetics, Christian Medical College, Vellore.

3.25. Experiments to identify the suitable lentiviral promoters for the expression of shRNAs in reprogramming experiments

Lentiviral shRNA plasmids that co-express shRNA and GFP under the control of 7 different promoters, mCMV, hCMV, mEF1α, hEFL1α, SFFV, CBA and UbC were used for generation of lentiviruses using the protocol mentioned in section 3.6. hADFs were transduced with lentiviruses, the transduced cells were enriched by puromycin selection, and the enrichment was confirmed by fluorescent microscopic and by FACS analysis of GFP+ cells. These cells were then subjected to hSTEMCAA infection using the protocol mentioned in section 3.10.1 to get an infection efficiency of 20-30%. On day 6 after hSTEMCAA infection, 1/10th of the cells were seeded on mitomycin C-treated SNL feeders in hiPSC medium with 8ng/ml bFGF (Life Technologies). On day 15 and day18, the reprogramming cells were analyzed by FACS to measure the GFP expression status in cells expressing different combinations of CD13, SSEA-4, and TRA-1-60 markers. Based on GFP status, the promoter activity was assessed in the cells at different stages of reprogramming.

3.26. Candidate gene knockdown reprogramming experiment

shRNA vectors generated as described in section 3.5 was used for the generation of lentiviruses as described in section 3.6. hADFs stably transduced with shRNA lentiviruses described in section 3.9, were taken for reprogramming on day 9 of shRNA knockdown. The cells were subjected to reprogramming as described in the section 3.10.1 and FACS analysis as described in section 3.15. The percentage of the cells expressing different combinations of CD13, SSEA-4 and TRA-1-60 were estimated to assess the effect of the gene knockdown on reprogramming efficiency.

3.27. shRNA library screening reprogramming experiment

For signaling library shRNA screening experiment, Cellecta shRNA library module 2 with shRNAs that target 5000 signalling pathway genes was used. The hADFs were first transduced with concentrated signaling library shRNA viruses at an infection efficiency of 40% to get a low copy number integration of individual shRNA. The infected cells were selected using puromycin, and the enrichment of transduced cells was confirmed by FACS analysis. The cells were subjected to reprogramming using the protocol described in section 3.25. On days 15 to 17 of reprogramming, the $SSEA4^{-}TRA\text{-}160^{-}$, $SSEA4^{+}TRA\text{-}160^{-}$ and $SSEA4^{+}TRA\text{-}160^{+}$ cells were sorted. Using Gentra DNA extraction Kit (Qiagen), DNA was isolated from from puromycin selected transduced hAFDs before reprogramming (sample 1) and the three flow sorted fractions, $SSEA4^{-}TRA\text{-}160^{-}$ (sample 2), $SSEA4^{+}TRA\text{-}160^{-}$ (sample 3) and $SSEA4^{+}TRA\text{-}160^{+}$ (sample 4). The DNA yield from sample 1, 2, 3 and 4 were 200 µg, 40 µg, 80 µg and 50 µg respectively. The shRNA specific barcodes were amplified from the genomic DNA by two rounds of PCRs. For sample 1, 200 µg of DNA and for the other samples, the entire DNA isolated was used for the 1^{st} round PCR in multiple 50 µl reactions (**Table 7**). The amplified products from each sample were pooled, and 2 µl of the 1^{st} round PCR product of each sample was used for setting up the 2^{nd} round PCR in two 100 µl reactions (**Table 7**). The second round PCR was performed for 13 cycles after the initial optimization of the number of cycles to obtain an equal amount of PCR products from all the

Table 7: shRNA specific barcode amplification PCR

i. First round of PCR

Reaction Components	50ul reaction
Water	17µl
Buffer	5µl
dNTP (10mM)	4µl
Forward Primer	1.5µl
Reverse Primer	1.5µl
Titanium Taq Polymerase	1µl
DNA template	20µl
Note: PCRs for samples 1, 2, 3 and 4 were set up with 25µg, 20µg, 20µg and 12.5µg each in 8, 2, 4 and 4 reactions.	

ii. Second round of PCR

Reaction Components	100ul reaction
water	73µl
Buffer	10µl
dNTP	4µl
Forward Primer	5µl
Reverse Primer	5µl
Titanium Taq polymerase	1µl
1^{st} round PCR product	2µl
Note: For each sample, two 100µl reaction were set up.	

DNA samples. To facilitate multiplex sequencing, specific indexing primers (**Table 4**) were used for amplification of each sample in the 2nd round PCR, and the high throughput (HT) sequencing was carried out on HiSeq 2500 using GexSeq primer (**Table 4**) at Medgenome, Begaluru. The barcode deconvolution and enumeration were carried out using software available on the Cellecta Decipher Project website (www.decipherproject.net/software/). T-test was performed to identify differentially present shRNAs. Gene ontology analysis was performed using String.db tool (https://string-db.org) available online.

Chapter 4

ESTABLISHMENT OF REPROGRAMMING STRATEGIES FOR THE GENERATION OF INDUCED PLURIPOTENT STEM CELLS FROM HUMAN SOMATIC CELLS

4.1. Introduction

Reprogramming of somatic cells to iPSCs is an extremely inefficient and slow process, with an average success rate of >1%. Derivation of hiPSC lines from human somatic cells involves multiple steps and establishing these protocols is technically challenging. Among the different factor delivery modes available, viral-based methods are easy to perform and give relatively better reprogramming efficiency. Fibroblast is one of the most commonly used somatic cells for reprogramming. So, for establishing reprogramming techniques in our laboratory, viral vector-mediated reprogramming of human fibroblasts was opted. The technical experience gained through this experiment laid the base for deriving integration-free hiPSC lines (this chapter) and Fanconi anaemia patient specific hiPSC lines (**Chapter 5**), and setting up gene knockdown reprogramming experiment (**Chapter 6**).

One of the major challenges in deriving hiPSC lines is the identification of fully reprogrammed hiPSC colonies from a highly heterogeneous background of reprogramming cells and colonies. The heterogeneity results from the extensive changes in gene and protein expression patterns brought about by ectopic expression of reprogramming factors (Chan *et al.,* 2009; Ruiz *et al.,* 2011). The currently used hiPSC identification methods based on pluripotency marker expression and morphology have been met with varying degree of success. Even though there are individual studies on analysing the expression patterns of fibroblast markers and pluripotency markers such as SSEA-4, TRA-1-60 and CD44 in the various stages of human somatic cell reprogramming (Chan *et al.,* 2009; Hotta *et al.,* 2009; Abujarour *et al.,* 2013; Kahler *et al.,* 2013; Tanaka *et al.,* 2015), a systematic analysis of their temporal expression has been lacking. Silencing of the expression of retroviral transgenes (RV-Tg) driven by long terminal repeats (LTRs) is an inherent property of pluripotent cells (Hotta and Ellis, 2008). In reprogramming using retroviral vectors, the transgene silencing marks, in addition to the acquisition of pluripotency, the exogenous factor independence of hiPSCs for the maintenance of pluripotency (Okita *et al.,* 2007; Takahashi *et al.,* 2007; Aasen *et al.,* 2008; Stadtfeld *et al.,* 2008). Even though silencing of retrovirally expressed fluorescent proteins has been reported to facilitate the identification of iPSC colonies (Chan

et al., 2009; Hotta *et al.,* 2009) this has not been correlated systematically with the expression of other pluripotency markers, NANOG, SSEA-4, TRA-1-60 and TRA-1-81 to understand the kinetics of transgene silencing and pluripotency induction in successful reprogramming events. In this study, human dermal fibroblast undergoing reprogramming under the influence of OSKM were systematically monitored for pluripotency marker expression, morphology acquisition and retroviral transgene silencing to understand their correlation with each other and with the pluripotent state.

4.2. Objectives

This chapter describes the establishment of reprogramming strategies for the generation of induced pluripotent stem cells from human adult dermal fibroblasts in the laboratory. The main objectives are as follows:

1. *Understanding the expression kinetics of molecular markers during the reprogramming process.*
2. *Assessment of the ability of the molecular markers to identify reprogramming intermediates and hiPSCs.*
3. *Derivation of hiPSC lines from human adult dermal fibroblasts by viral and non-viral vector mediated reprogramming and their characterization.*

4.3. Results

4.3.1. *Generation of hiPSCs from adult dermal fibroblast using retroviral vectors to express OSKM*

For establishing the methods to derive iPSCs using retroviral methods to express OSKM, the protocol described by Takahashi *et al.,* was used (Takahashi *et al.,* 2007) (**Figure 15A**). Briefly, the human adult dermal fibroblasts (hADFs) were modified to express mouse solute receptor, SLC7A1, to aid the entry of retroviral vectors carrying OCT4, SOX2, KLF4, cMYC and RFP (OSKMR). Red fluorescent protein (RFP) was additionally included in

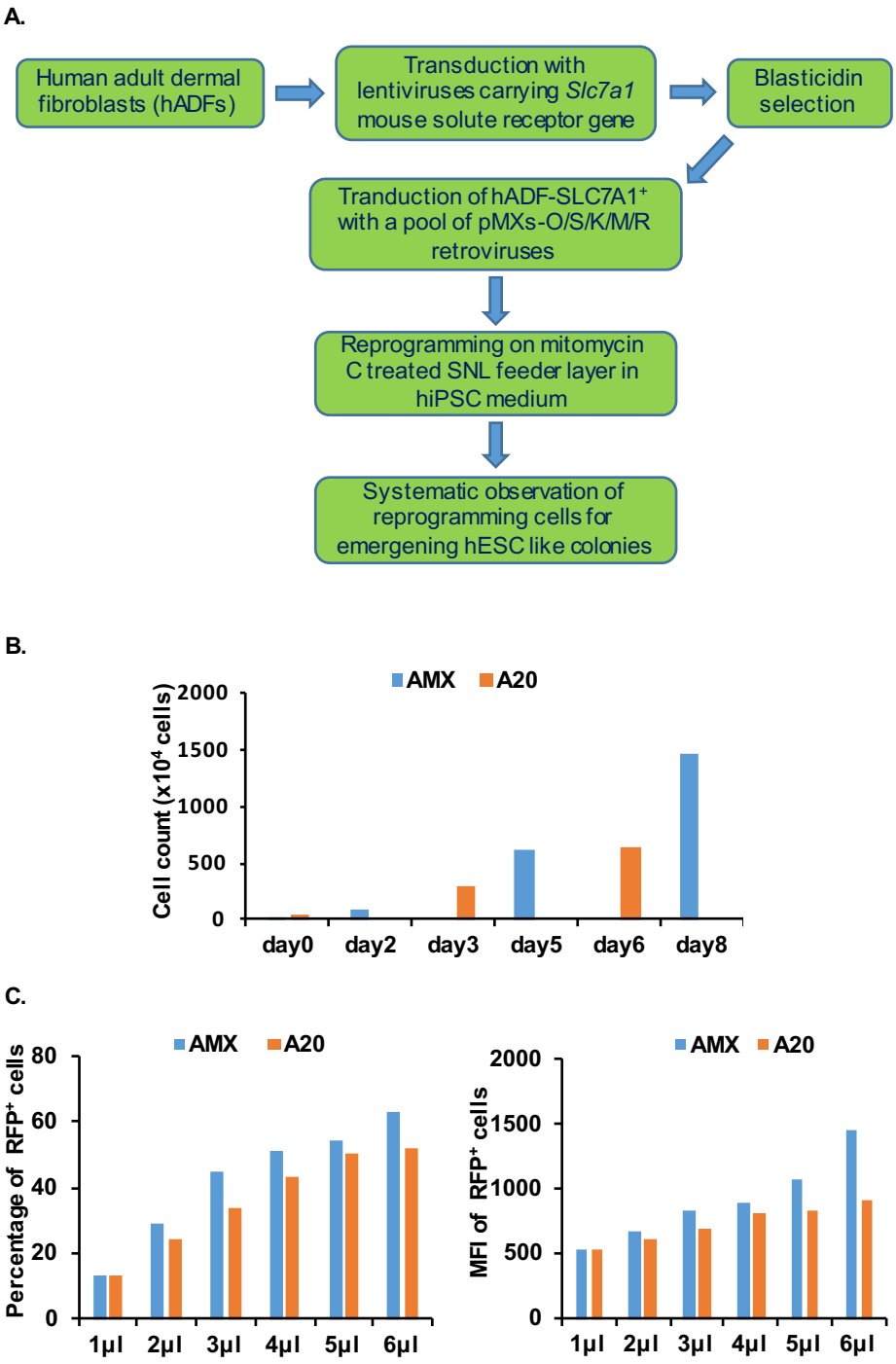


Figure 15: Strategy used for retroviral mediated reprogramming of human adult dermal fibroblasts

(A) Schematic of hiPSC generation from hADFs by retroviral mediated reprogramming with OSKM. **(B)** Cell count of hADFs maintained in Amniomax and A20 at the end each passage. **(C)** Percentage of RFP$^+$ cells and mean fluorescent intensity (MFI) of RFP in hADFs transduced with indicated volumes of concentrated retroviral RFP. hADFs - human adult dermal fibroblasts. OSKM – OCT4, SOX2, KLF4 and cMYC.

the reprogramming cocktail to monitor transgene silencing and aid in the identification and isolation of exogenous factor independent hiPSC clones as described in section 4.3.4.

It was observed that hADFs maintained in Amniomax medium (AMX) showed rapid proliferation compared to conventional A20 medium (see Materials and Methods) (**Figure 15B**). Since retroviral vectors have higher infection efficiency on rapidly proliferating cells, the fibroblasts cells preconditioned in Amniomax medium were used. As expected, hADF-AMX showed high transduction efficiency and high transgene expression level when infected with retroviral RFP vector (**Figure 15C**). Following infection with OSKMR, the reprogramming was carried on a mitomycin treated SNL feeder layer and maintained in conventional KOSR based hiPSC medium. The hiPSC colonies were identified as described in section 4.3.4 and they gained the cellularity required for isolation by the end of 2nd week of reprogramming. The reprogramming efficiency of 0.05%-0.1%, calculated based on the number of TRA-1-60^{+} hES-like colonies formed from the total number of cells initially infected with OSKM retroviruses, was achieved. For the derivation of hiPSC lines, the clones were isolated mechanically, expanded on feeder based culture system and passaged by enzymatic treatment with Collagenase IV. The established clones were characterized for their pluripotency at molecular and functional levels as described in Sections 4.3.4 and 4.3.6. The characterized hiPSC lines were cryopreserved for future applications and could be successfully revived back to culture using the protocols standardized in this thesis.

4.3.2. Generation of hiPSCs from adult dermal fibroblast using lentiviral vectors to express OSKM

For understanding the molecular mechanism of reprogramming, it would be advantageous to use a method that ensures high reprogramming efficiency. Therefore, a polycistronic lentiviral vector, hSTEMCCA, which has been reported to successful reprogram fibroblasts even with single copy number integration (Somers *et al.,* 2010), was chosen for this application in this study. Also, unlike in the case of using individual vectors for expressing each reprogramming factors, hSTEMCCA vector will help to avoid the heterogeneity brought about by differences in factor stoichiometry in reprogramming cells. Using the expertise developed in hiPSC generation using retroviral method, human adult

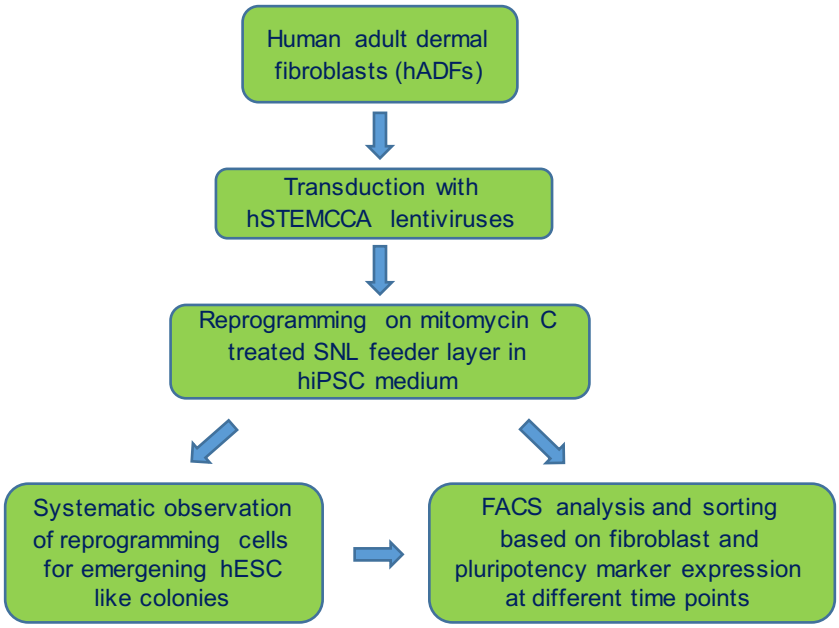
dermal fibroblasts were subjected to hSTEMCAA mediated reprogramming (**Figure 16A**). The hiPSC like colonies appeared by the end of the second week of reprogramming and they were ready for isolation from mid of the third week (**Figure 16B**). A reprogramming efficiency of about 0.5%-1% was achieved, which was calculated as mentioned in section 4.3.1. The reprogramming cells were isolated based on expression of a different combination of fibroblast and pluripotency markers and used for understanding their expression kinetics in section 4.3.3. This reprogramming strategy was also used for understanding the molecular mechanism of reprogramming by knockdown and overexpression studies in **Chapter 6**.

4.3.3. SSEA-4 and TRA-1-60 are expressed in both fully and partially reprogrammed cells

For determining the specific markers that can be used for the isolation of pluripotent clones, we monitored the expression kinetics of surface markers in the reprogramming cells before and after the colonies emerged. The hADFs undergoing reprogramming with polycistronic hSTEMCAA lentiviral vectors were analyzed on days 6, 8, 12, 16, 17 and 20 by flow cytometry for the expression of a fibroblast marker, CD13, and the pluripotency markers, SSEA-4 and TRA-1-60. The analysis was carried out within the TRA-1-85^{+} population to obtain an accurate percentage of marker expressing cells, without the interference of SNL feeder cells in the data analysis. The emerging colonies were analyzed by immunofluorescence assay for the expression of pluripotency markers, SSEA-4, TRA-1-60 and NANOG on different days of reprogramming with retroviral OSKM vectors.

As expected, the percentage of CD13^{+} cells reduced with a concurrent increase in SSEA-4^{+} and TRA-1-60^{+} cells as the reprogramming progressed (**Figures 17A and 17B**). CD13^{+} cells reduced from 100% on day 0 to 71% on day 8 and to 43% on day 20, and SSEA-4^{+} cells were 50% on day 8, which increased up to 80% on day 12 (**Figure 17A**). The percentages of CD13^{+}SSEA-4^{-}, CD13^{+}SSEA-4^{+} and CD13^{-}SSEA-4^{+} cells were at the ratio 65:25:10 on day 6, 17:50:33 on day 12, 23:23:54 on day 16 and 25:20:55 on day 20 (**Figures 17C and 17D**). These results indicated that SSEA-4 expression is initiated before CD13 expression was completely shutdown, and this transition continued until the end of the

A.



B.

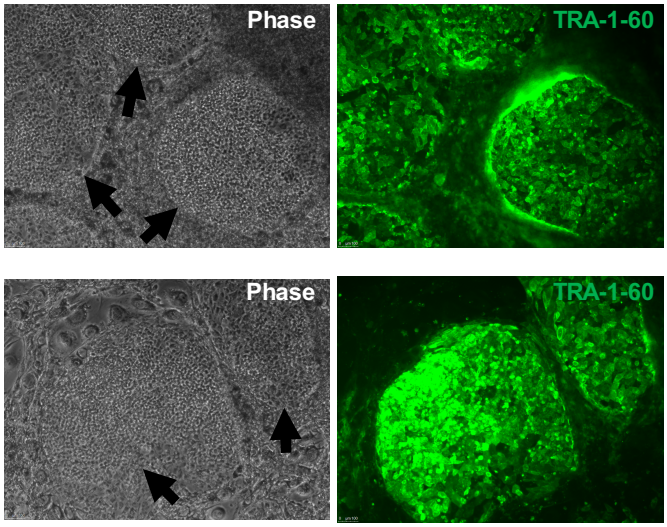
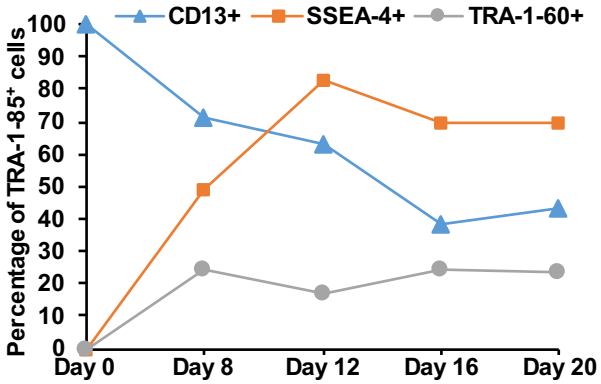


Figure 16: Strategy used for lentiviral mediated reprogramming of human adult dermal fibroblasts

(A) Schematic of hiPSC generation from hADFs by lentiviral mediated reprogramming with OSKM. **(B)** hESC-like TRA-1-60^{+} colonies emerging from reprogramming cells. Arrows show the hiPSC colonies. Images are at 10X magnification. hADFs - human adult dermal fibroblasts. OSKM – OCT4, SOX2, KLF4 and cMYC.

A.



B.

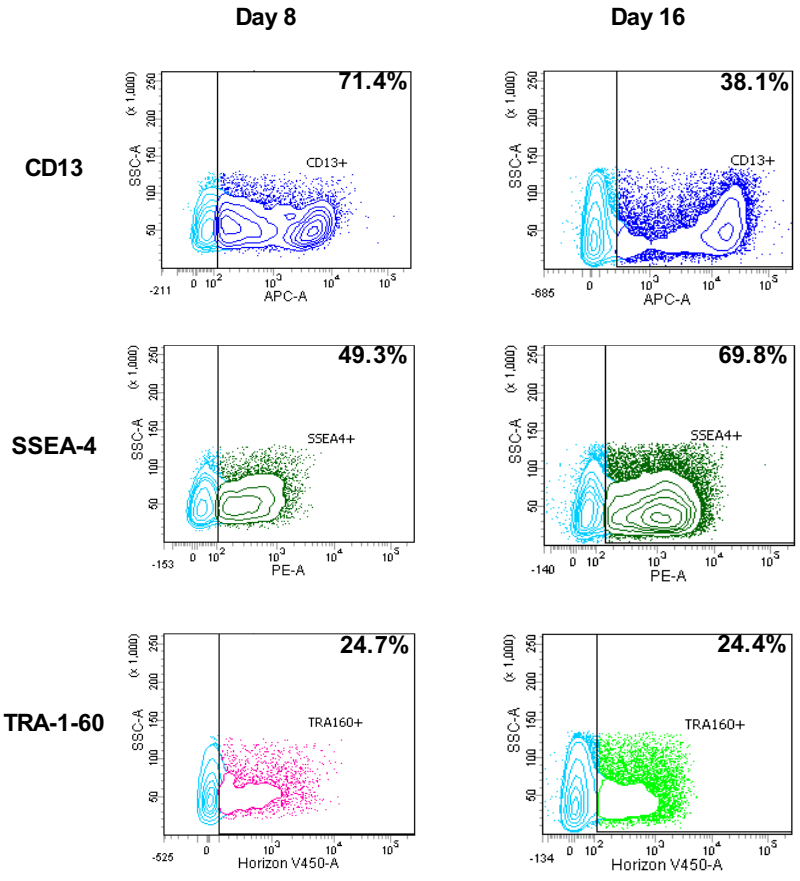
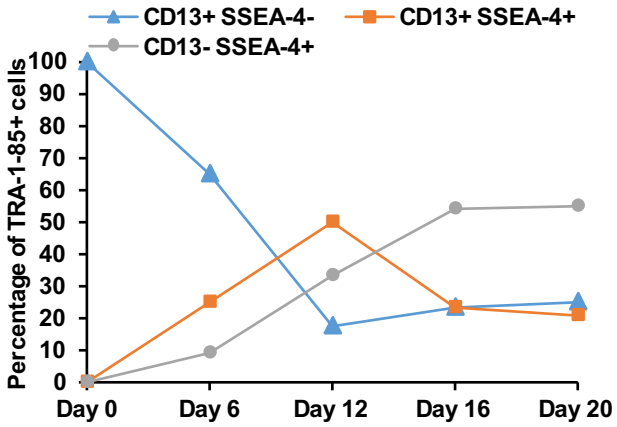
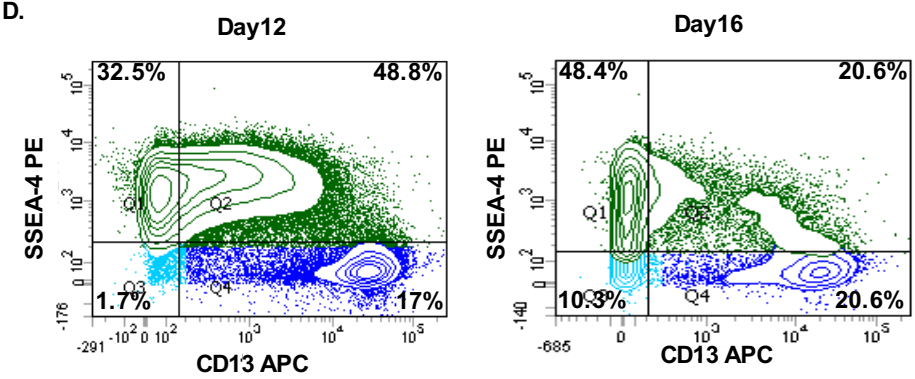


Figure 17: Analysis of fibroblast (CD13) and pluripotency (SSEA-4 and TRA-1-60) markers during reprogramming of hADFs by flow cytometry
(A) Percentages of $CD13^+$, $SSEA\text{-}4^+$ and $TRA\text{-}1\text{-}60^+$ cells on different days of reprogramming in the $TRA\text{-}1\text{-}85^+$ population of cells estimated by FACS and **(B)** representative FACS contour plots.

C.



D.



E.

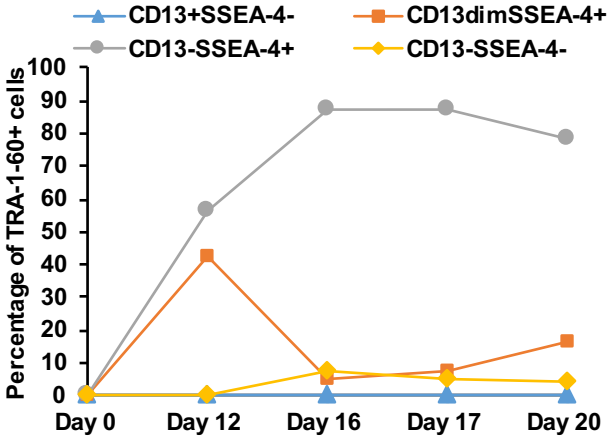
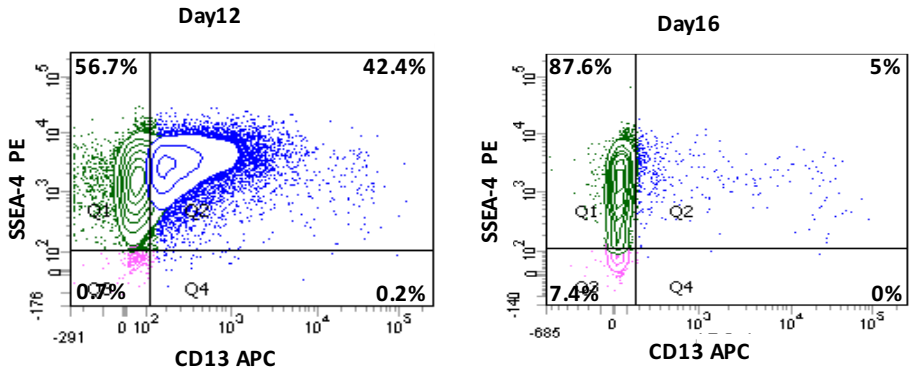
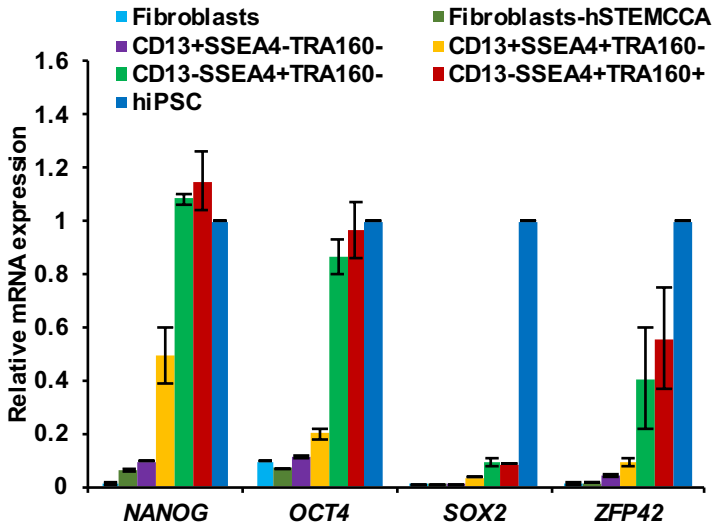


Figure 17 continued: **(C)** Percentages of the cells that co-express CD13 and SSEA-4 in the TRA-1-85$^+$ population of cells estimated by FACS. **(D)** their representative FACS contour plots. **(E)** Percentages of the cells that co-express CD13 and SSEA-4 in the TRA-1-60$^+$ fraction of cells.

F.



G.



H.

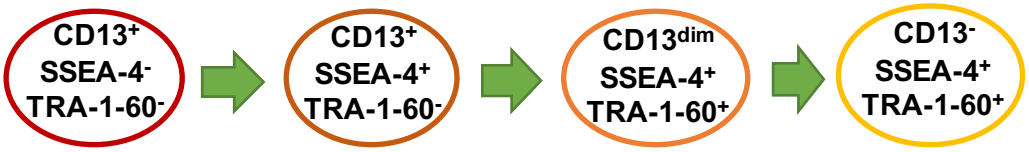


Figure 17 continued: (F) Representative FACS contour plots of the cells that co-express CD13 and SSEA-4 in the TRA-1-60$^+$ fraction of cells. **(G)** Real-time PCR analysis of mRNA levels of pluripotency markers, *NANOG, OCT4, SOX2* and *ZFP42* in hADFs, in the cells 6 days after transducing with hSTEMCCA lentiviruses and the cell fractions sorted by flow cytometry based on the expression of CD13, SSEA-4 and TRA-1-60 (n=2). The fold changes were calculated relative to expression levels in hiPSCs. **(H)** Sequence of molecular states, defined by coexpression pattern of CD13, SSEA-4 and TRA-160, through which fibroblasts transit during successful reprogramming.

2nd week of reprogramming resulting in a large fraction of reprogramming cells achieving SSEA-4$^+$ state (50% on day 8 to 80% on day 12) (**Figure 17A**).

TRA-1-60, a late-stage marker of human somatic cell reprogramming, was expressed in ~20% of the reprogramming cells on day 8, and its level remained nearly constant till day 20 (**Figure 17A**). Almost all (>95%) TRA-1-60$^+$ cells were SSEA-4$^+$, and they were CD13^{dim} initially and CD13$^-$ later, which indicated that, in the successfully reprogramming cells, TRA-1-60 expression followed SSEA-4 expression after the somatic cell gene silencing is completed (**Figures 17E and 17F**)**.** This data also highlighted the significant difference in the pluripotency levels in the SSEA-4$^+$ and TRA-1-60$^+$ cells. The presence of CD13^{dim}SSEA-4$^+$ and CD13$^-$SSEA4$^-$ cells along with pluripotent CD13$^-$SSEA4$^+$ cells within the TRA-1-60$^+$ population clearly showed the heterogeneity among the TRA-1-60$^+$ cells (**Figure 17E**). Analysis of the late stage pluripotency genes, *NANOG, OCT-4, SOX2* and *ZFP42*, in the cell fractions sorted based on the expression of CD13, SSEA-4 and TRA-1-60 showed that their levels increased steadily from CD13$^+$SSEA-4$^+$TRA-1-60$^-$ to CD13$^-$SSEA-4$^+$TRA-1-60$^-$ to CD13$^-$SSEA-4$^+$TRA-1-60$^+$ cells (**Figure 17G**). Based on the change in the relative percentage of cells expressing different combination of markers during the reprogramming process and the expression levels of pluripotency markers, it could be proposed that in successful reprogramming the sequence of states that CD13$^+$SSEA-4$^-$TRA-1-60$^-$ fibroblasts transit through are CD13$^+$SSEA-4$^+$TRA-1-60$^-$, CD13^{dim}SSEA-4$^+$TRA-1-60$^+$ and CD13$^-$SSEA-4$^+$TRA-1-60$^+$ and the cells that fail to enter this path in the first week of reprogramming may remain incompletely reprogrammed as CD13$^+$SSEA-4$^-$TRA-1-60$^-$, CD13$^+$SSEA-4$^+$TRA-1-60$^-$, CD13$^-$SSEA-4$^+$TRA-1-60$^-$, CD13$^-$SSEA-4$^-$TRA-1-60$^-$ and CD13$^-$SSEA-4$^-$TRA-1-60$^+$ (**Figure 17H**). These results showed that with the co-expression profile of CD13, SSEA-4 and TRA-1-60, cells belonging to the different stages of reprogramming can be demarcated.

When the emerging colonies were analyzed, a similar trend in the expression kinetics of pluripotency markers was observed. At the beginning of the second week of reprogramming (day 9), small cell clusters or colonies made up of 20 - 60 tightly packed cells were visible. Although they lacked hESC morphology, most of them expressed SSEA-4

(**Figure 18A** upper panel), but only about 10% of the SSEA-4$^+$ cell clusters expressed TRA-1-60. All the TRA-1-60$^+$ colonies were NANOG$^+$ at this stage itself (**Figure 18A** middle panel), and this strong correlation suggested that TRA-1-60 can be used as a surrogate marker for identifying NANOG expressing cells. By the third week of reprogramming (day 16), as the colony size increased their morphologies became evident, and many large colonies expressed SSEA-4, TRA-1-60, and NANOG (**Figure 18B**). The reprogramming efficiency estimated based on the number of TRA-1-60$^+$ colonies remained constant at about 0.1% in the 2nd and the 3rd weeks of reprogramming with retroviral vectors.

It was interesting to find that the reprogramming cells achieve a CD13$^-$SSEA-4$^+$TRA-1-60$^+$ pluripotent state in the very early days of reprogramming (less than 10 days) even before hiPSC colonies appear. Even though 10% on day 6 and 33% on day 12 of the reprogramming cells become CD13$^-$SSEA-4$^+$ before the hESC-like colonies appear and ~20% of the cells attain CD13$^-$SSEA-4$^+$TRA-1-60$^+$ state during the course of reprogramming, the reprogramming efficiency in the generation of TRA-1-60$^+$NANOG$^+$ colonies with viral vectors was very low (<0.1%), as reported earlier (González *et al.,* 2011; Bayart and Cohen-Haguenauer, 2013). This clearly indicates the further heterogeneity that exists in the TRA-1-60$^+$ cells and the limitation in the use of TRA-1-60 as a sole marker for the enrichment and isolation of true hiPSCs.

4.3.4. Strong correlation between retroviral transgene silencing and acquisition of pluripotency

To understand the correlation between RV-Tg silencing, morphological changes and expression of pluripotency markers in the emerging clones, the fibroblasts subjected to retroviral OSKM mediated reprogramming were co-transduced with retroviruses to express RFP. The reprogramming cells were monitored for assessing the kinetics of acquisition of these features. On day 11, 65-70% of the SSEA-4$^+$ colonies were TRA-1-60$^+$, but only ~35% of TRA-1-60$^+$ colonies showed RV-RFP silencing (RV-RFP$^-$) and in the 3rd week of reprogramming, about 85% of the TRA-1-60$^+$ colonies were RFP$^-$ suggesting that though a large number of colonies become TRA-1-60$^+$ in the early stages of reprogramming they do not achieve complete reprogramming. Routine monitoring of the emerging clones throughout

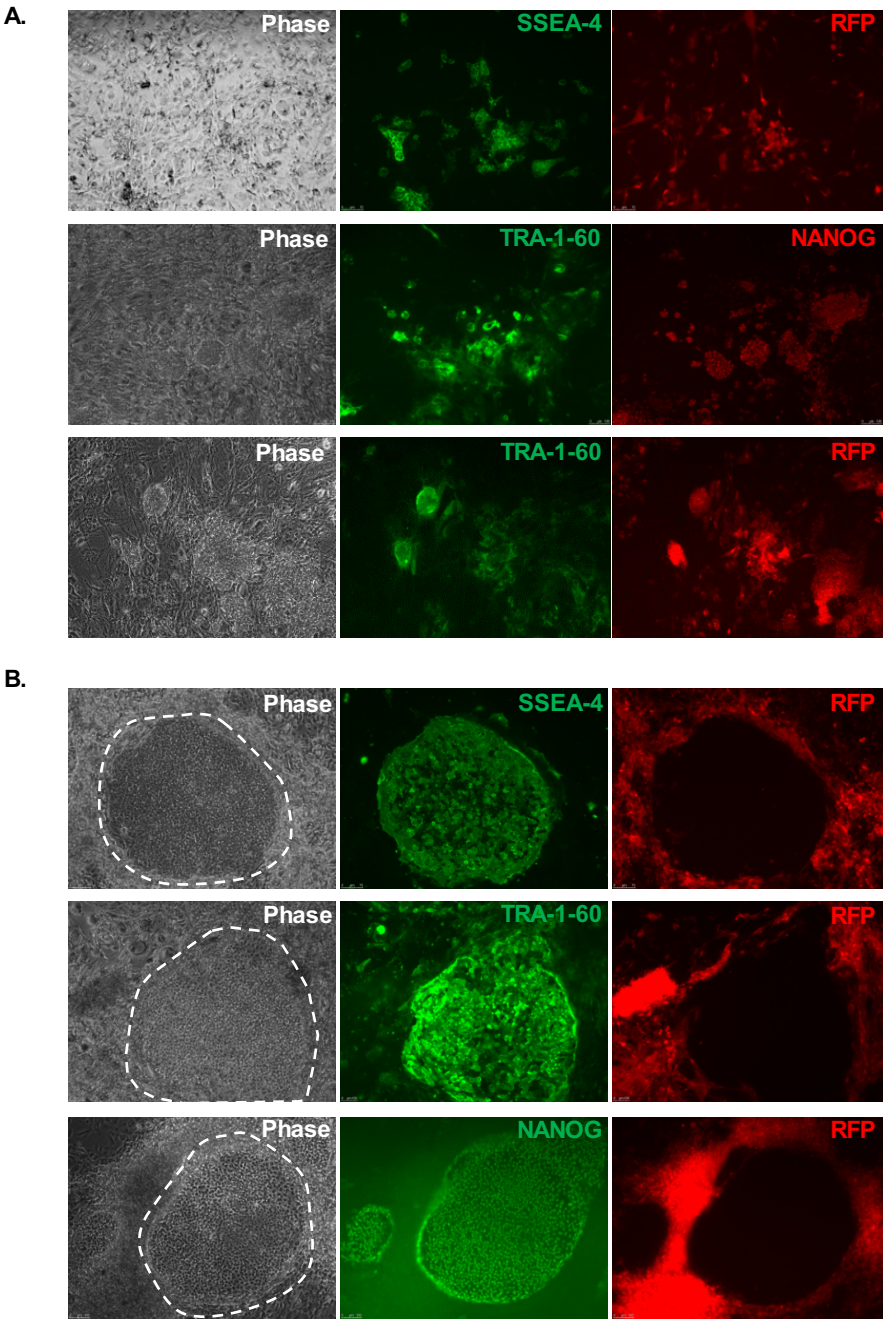
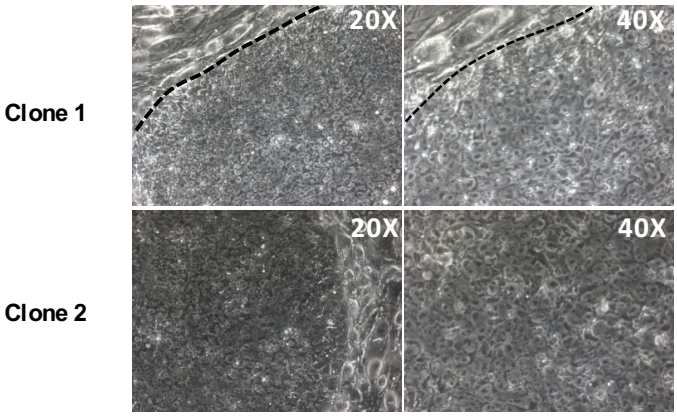


Figure 18: Analysis of morphology, pluripotency marker expression and transgene silencing in the colonies emerging during reprogramming
(A) Expression of SSEA-4 and RV-RFP (upper panel), TRA-1-60 and NANOG (middle panel) and TRA-1-60 and RV-RFP (lower panel) in the cell clusters / colonies on day 9 of reprogramming showing the initiation of pluripotency marker expression before the cells achieve hESC-like morphology and transgene silencing. **(B)** Expression of the pluripotency markers (SSEA-4, TRA-1-60 and NANOG) and RV-RFP silencing in the colonies on day 16 of reprogramming. The emerging hiPSC colonies showed characteristic hESC-like morphology and retroviral transgene silencing allowing their easy identification.

C.



D.

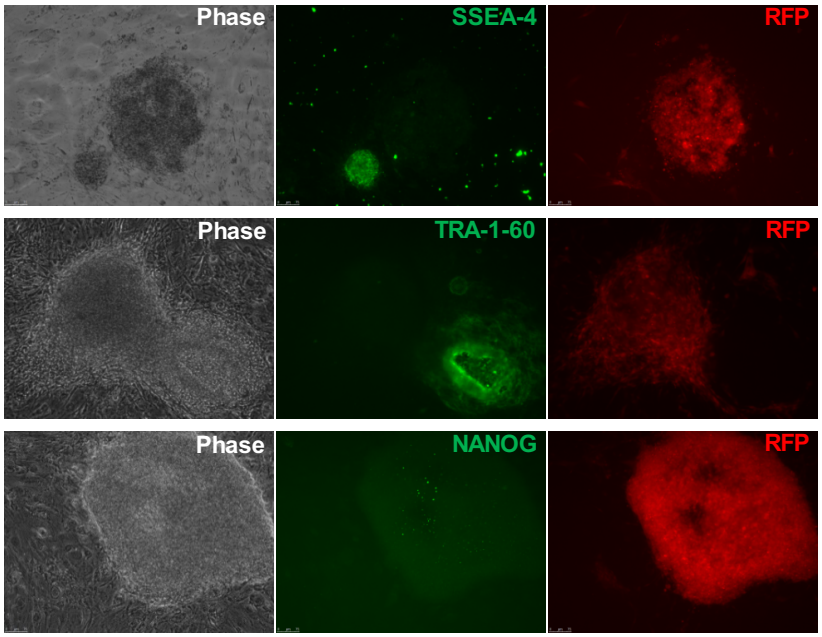


Figure 18 continued: (C) Higher magnification images of RV-RFP$^-$ hiPSC colonies showing their hESC-like morphology- flat appearance, defined boundary and high nuclear to cytoplasmic ratio. **(D)** Non-hESC like RFP$^+$ colonies which lacked the expression of pluripotency markers on day 16. All images are at 10X magnification unless otherwise indicated. The broken lines show the characteristic boundaries of the emerging hiPSC colonies on the feeder cells.

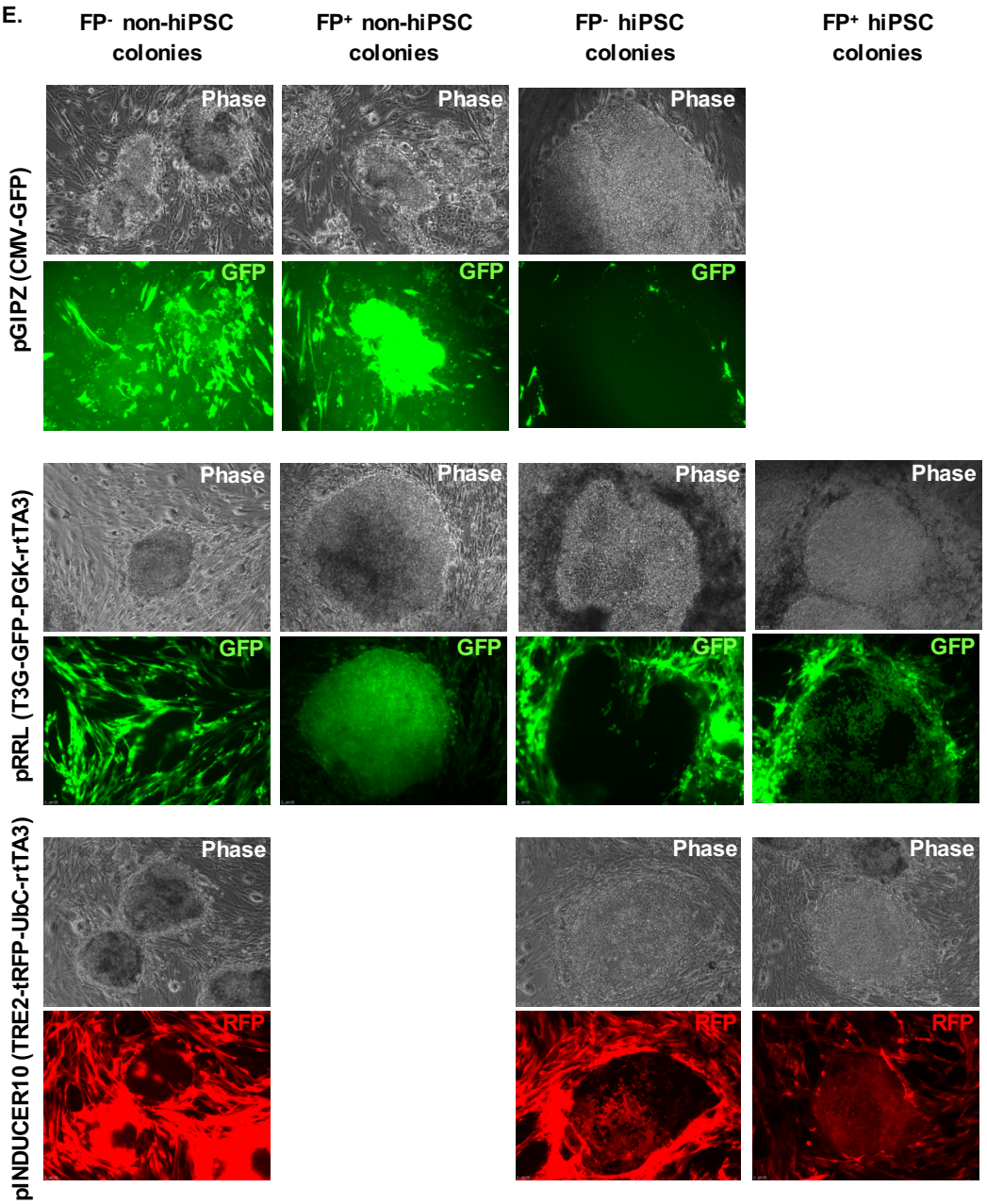


Figure 18 continued: (E) Representative images of hiPSC and non-hiPSC colonies exhibiting variegated expression of fluorescent protein (FP) (GFP or RFP) from lentiviral vectors with different promoters, pGIPZ (CMV-GFP) lentiviral vector with CMV promoter to express GFP (Ansaloni et al., 2010), pRRL (T3G-GFP-PGK-rtTA3) inducible lentiviral vector with PGK promoter to express rtTA3 and T3G promoter to express GFP (Fellmann et al., 2013) and pINDUCER10 (TRE2-tRFP-UbC-rtTA3) inducible lentiviral vector with UbC promoter to express rtTA3 and TRE2 promoter to express tRFP (Meerbrey et al., 2011). All images are at 10X magnifications. CMV-human cytomegalovirus intermediate early enhancer and promoter, rtTA3-reverse tetracycline transactivator-3 gene, TRE2- Tet responsive element-2 promoter and T3G- optimized third generation Tet-responsive element promoter.

the course of reprogramming with the aid of RFP reporter protein expression revealed characteristic morphological changes in them. The TRA-1-60^{+}RFP^{-} colonies emerged from highly proliferative transformed cells, and the colonies appeared as symmetric RFP^{-} clusters of cells among the RFP^{+} transformed cells (**Figure 18B**). Characteristically, they emerged dislodging the feeder cells around them radially, thereby forming flat colonies with an almost symmetrical shape and defined boundaries, and they expressed NANOG, in addition to SSEA-4 and TRA-1-60 (**Figure 18B**). The close microscopic observation of RFP^{-} colonies showed that they possessed typical hESC- like morphology - flat colonies containing closely packed cells with increased nuclear to cytoplasmic ratio (**Figure 18C**). This morphology was similar to a well-established hiPSC line that was maintained in our laboratory (BC1-hiPSC line; a gift from Linzhao Cheng) (data not shown). In the third week of reprogramming, almost all (>95%) the colonies without hESC morphology were RFP^{+}, and they lacked the expression of SSEA-4, TRA-1-60 and NANOG (**Figure 18D**). Experiments using lentiviral vectors with different promoters driving the expression of fluorescent proteins did not show a correlation between transgene silencing, pluripotency and morphology of the colonies wherein the colonies exhibited variegated expression of fluorescent proteins (**Figure 18E**). Thus, the specificity and reliability of retroviral transgene silencing as an indicator of pluripotency of hESC-like hiPSC colonies were confirmed.

Nine stable hiPSC lines were established from the RFP^{-} colonies which were isolated based on the acquisition of hiPSC morphological features described above. All the established retrovirally generated lines (RV-hiPSC lines) maintained high-level expression of *NANOG, OCT4, SOX2, GDF3* and *DNMT3B* mRNA transcripts as measured by real-time PCR (**Figure 19A**) and SSEA-4, TRA-1-60, TRA-1-81, NANOG, LIN28 and OCT4 by immunofluorescence, up to 20 passages (**Figure 19B**). Integration of the retroviral transgenes in these clones was confirmed by genomic PCR (data not shown), but their expression was significantly reduced in all the clones due to transgene silencing (**Figure 19C**). Three of these clones were expanded in long-term culture (>50 passages) without the loss of morphology and expression of the pluripotency markers. They also had

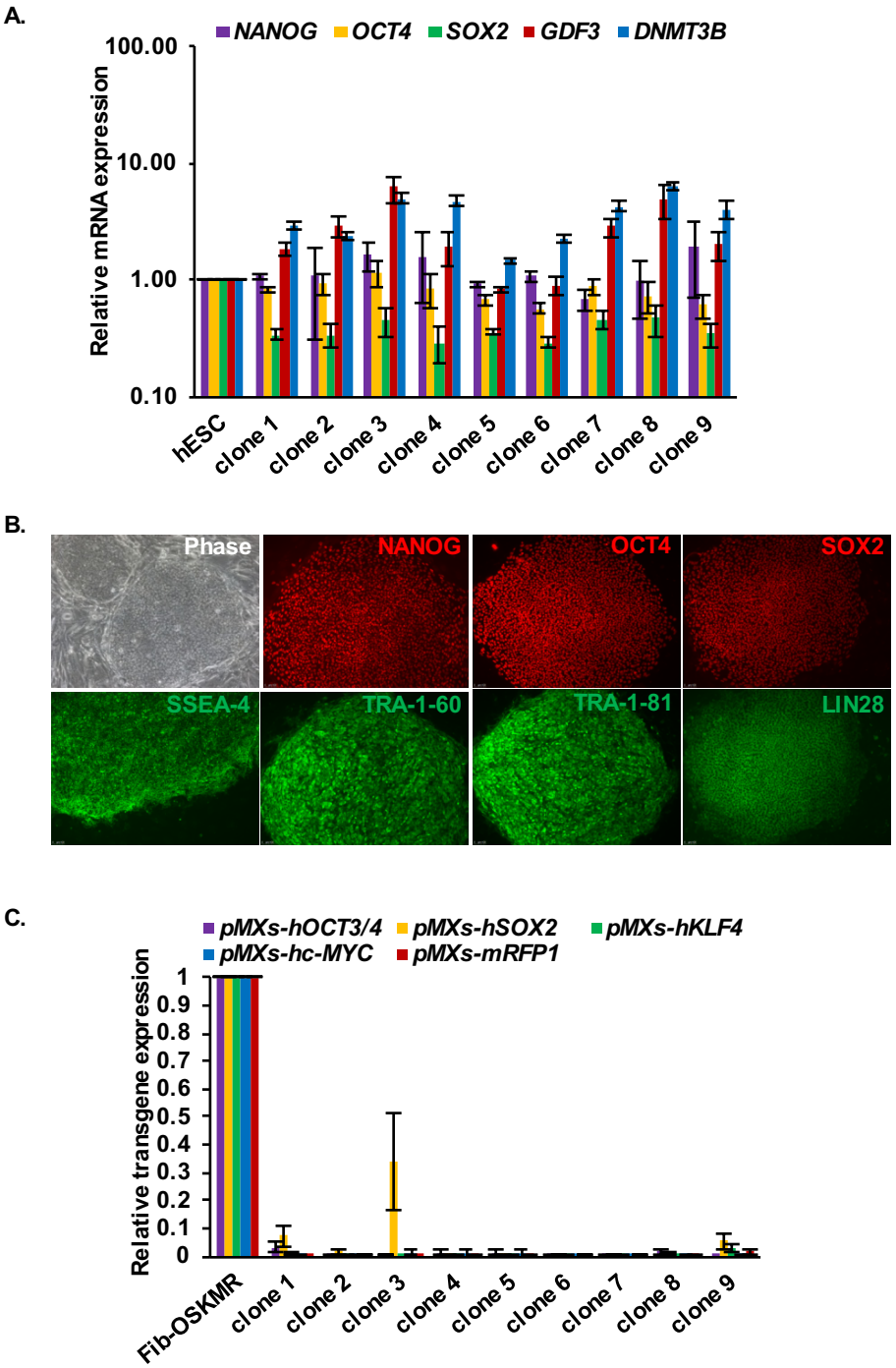
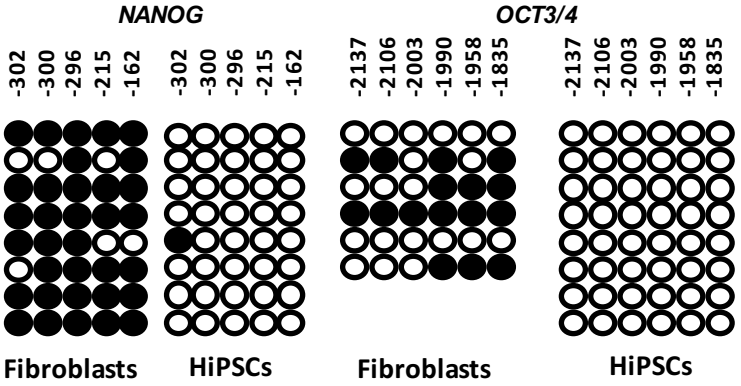


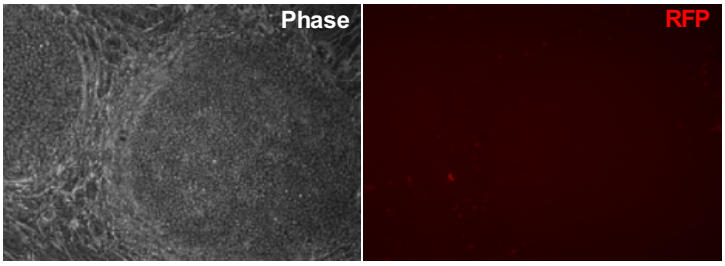
Figure 19: Characterization of RV-hiPSC clones isolated based RV-Tg silencing and hESC-like morphology

(A) Real time PCR analysis of pluripotency markers in the isolated RV-hiPSC clones. The fold change was calculated relative to the expression levels in hESCs (n=2). **(B)** Immunofluorescence analysis of pluripotency markers in the isolated clones. **(C)** Real time PCR analysis of expression of retroviral transgenes in the clones (n=2). The fold changes were calculated relative to the expression levels in fibroblast transduced with OSKMR (Fib-OSKMR).

D.



E.



F.

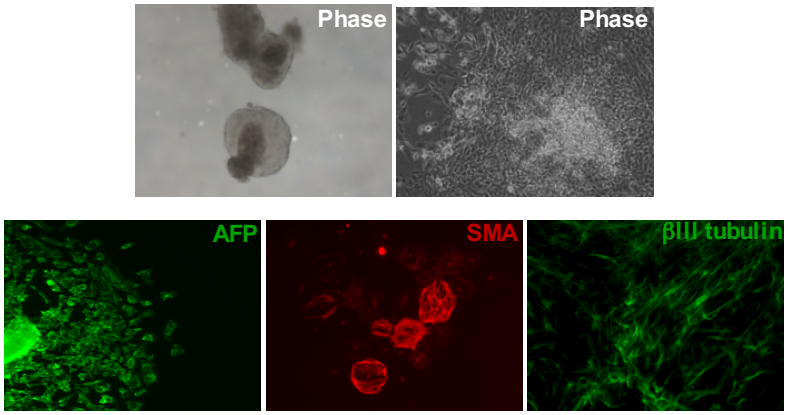


Figure 19 continued: (D) Bisulfite sequencing results of *OCT4* and *NANOG* promoters in fibroblasts and established RV-hiPSC clones showing hypomethylation of these regions. **(E)** Microscopic image of an established RV-hiPSC clone confirming the stable silencing of transgenes throughout the culture. **(F)** *In vitro* differentiation of established hiPSC clones. Data from a representative clone is shown. hiPSCs formed cystic embryoid bodies (EBs) in suspension culture and these EBs were differentiated further in adherent culture into the cells expressing markers characteristic of three germ layers - endoderm (α-feto protein, AFP), mesoderm (α-smooth muscle actin, SMA) and ectoderm (βIII-Tubulin). All images are at 10X magnification.

hypomethylated *OCT4* and *NANOG* promoters (**Figure 19D**). They remained RFP$^-$ throughout the culture (**Figure 19E**) indicating stable silencing of retroviral transgenes and their transgene independence for the pluripotency maintenance. Pluripotency of these clones was further confirmed by their ability to undergo the trilineage differentiation *in-vitro* (**Figure 19F**).

Taken together, these results showed that there is a good correlation between RV-Tg reporter silencing, the hESC-like morphology and the pluripotency of hiPSCs, and RV-Tg reporter silencing can facilitate easier identification of true hiPSC clones from the heterogeneous colonies that appear in the reprogramming dish. By combining to hESC-like morphology, it can serve as a more reliable and convenient pluripotency marker than SSEA-4 and TRA-1-60, which are also expressed in a very large number of reprogramming cells that do not form hiPSC colonies and even before hESC-like morphology and RV-Tg silencing are attained in the reprogramming cells. The reprogramming efficiency estimated based on the total number of TRA–1-60$^+$ colonies remained the same in the 2nd and 3rd weeks of reprogramming, but as the majority of TRA-1-60$^+$ cells in the initial stages of reprogramming do not show RV-Tg silencing, they are molecularly distinct from the late stage TRA-1-60$^+$ cells. Hence it is proposed that combining RV-Tg reporter silencing to TRA-1-60 expression will enrich highly pluripotent cells especially during the initial days of reprogramming. Moreover, the characteristic temporal morphology changes that successfully reprogramming cells undergo while they give rise to RV-RFP$^-$ hiPSC colonies on feeders was identified that facilitate easier identification of true hiPSC clones.

4.3.5. Morphology-based isolation yields bona fide hiPSC clones

For confirming the observation that morphology serves as the most reliable marker for identifying bona fide iPSC clones, this identification criteria was used to isolate hiPSC clones generated from fibroblasts by non-integrative vector-mediated reprogramming methods. For the generation of integration-free hiPSCs from human fibroblasts, oriP / EBNA1 based episomal plasmids (Okita *et al.,* 2011) and Sendai viruses (Ban *et al.,* 2011) were employed. The expertise in the identification of the right morphology of hiPSC clones

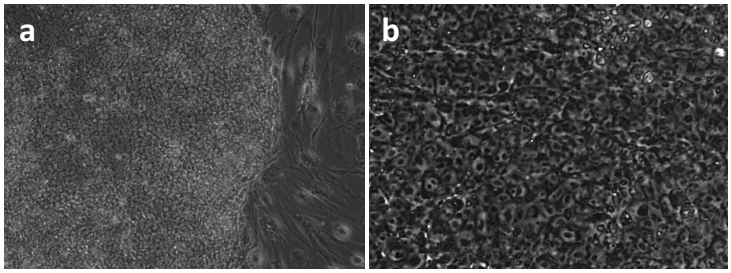
that was established based on correlation of morphological changes and RV-RFP silencing in our previous experiment was used to derive integration-free hiPSCs.

Using a combination of polycistronic oriP / EBNA1 based episomal plasmids carrying OCT4, SOX2, KLF-4, L-MYC, LIN28 and p53-shRNA, a reprogramming efficiency of 0.01% from hADFs was obtained, that was estimated based on TRA-1-60 expression and acquisition of hESC-like morphology in the colonies. Six clones were isolated based on their resemblance in morphology with the clones identified based on retroviral silencing and morphology (**Figure 20A**) They maintained hESC-like morphology (**Figure 20A**) and consistent high-level expression of pluripotency markers over 10 passages (**Figures 20B and 20C**). Compared to retroviral vectors, the episomal vectors showed a slower pace of reprogramming; the hiPSC colonies emerged only by the third week, and they were ready for isolation only by the 4^{th} week of reprogramming. This method generated a relatively lower number of rapidly proliferating transformed cells and non-iPSC colonies.

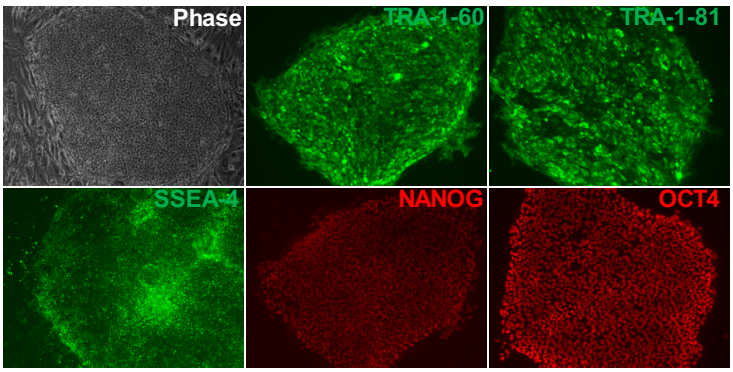
With SeV vectors to express OSKM, a reprogramming efficiency of 0.2%, estimated based on TRA-1-60 expression and hESC-like morphology in colonies generated from patient-derived fibroblasts was achieved (refer section 5.3.3). The colonies with emerging hiPSC morphology were visible from the beginning of the 3^{rd} week (day 16) after transduction and, in the fourth week of reprogramming, 9 colonies were isolated based on their hESC-like morphology. The details regarding the characterization of these clones are mentioned in chapter 5 in section 5.3.3.

Though episomal and SeV based vectors showed slower pace in reprogramming, undesirable colonies were fewer by these methods and hence identification of hiPSC colonies based on morphology was found to be relatively easier compared to the methods using retroviral and lentiviral vectors. These results showed that the morphology is a reliable marker for identification of true hiPSCs, and the laboratories which gained experience in the correct morphology of hiPSC clones can derive hiPSCs lines with a high level of pluripotency without the use of additional markers.

A.



B.



C.

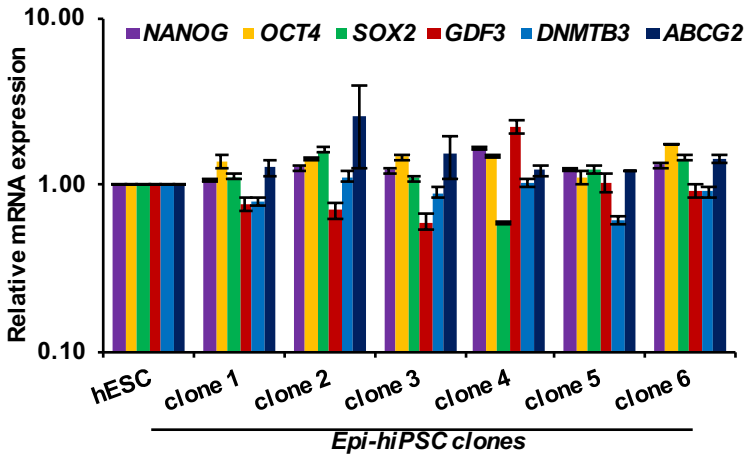


Figure 20: Confirmation of pluripotency of hiPSC lines generated using using episomal (Epi) plasmids

(A) Morphology of a representative emerging hiPSC colony. Images in (a) lower (10X) and (b) higher (20X) magnifications are shown to represent the defined boundaries and the increased nuclear to cytoplasmic ratio of the emerging colonies. **(B)** Morphology and immunofluorescence analysis of pluripotency markers in an established Epi-hiPSC line. Images are at 10X magnification. **(C)** Real time PCR analysis of pluripotency markers in Epi-hiPSC lines (n=2). The fold changes were calculated relative to the expression levels in hESCs.

4.3.6. HiPSCs isolated based on morphology showed consistent in vivo differentiation potential

To confirm the pluripotency of hiPSC clones isolated based on the morphology criterion, their *in vivo* differentiation potential was tested by their ability to form teratomas. Since the efficiency of teratoma formation depends on multiple factors like pluripotency level of the injected cells, the sites of injection and the host immunophenotype (Ozolek and Castro, 2011), teratoma assay of individual clones was performed in two strains of SCID mice, B6.CB17-Prkdcscid/SzJ (black SCID) and CB17/Icr-Prkdcscid/ IcrIcoCrl (white SCID). To favour teratoma formation, the cells were injected with Matrigel and Collagen I intramuscularly into the hind limbs of 4 - 7 weeks old mice, using a previously described protocol (Park *et al.,* 2008a). Four hiPSC lines (three hiPSC lines generated using retroviral vectors and one hiPSC line generated using episomal expression plasmids), which were isolated based on the morphology and subsequently confirmed to have a high-level expression of pluripotency markers were taken for teratoma assay (**Figure 21A**). All the four lines formed teratomas constituting of cells representing endoderm, mesoderm and ectoderm in 8 to 14 weeks (**Figure 21B**). Both black and white SCID mice showed similar efficiencies in teratoma formation; 92% (22/24 sites) and 100% (17/17 sites), respectively. It was noteworthy that when hiPSCs suspended in hiPSC basal culture medium were injected for teratoma generation (Takahashi *et al.,* 2007), without the use of Matrigel and Collagen I, the efficiency of teratoma formation was very low (<10%) (data not shown).

4.4. Discussion

Though significant progress has been made in the methods to reprogramme somatic cells from different genetic backgrounds to hiPSCs, their enrichment and isolation criteria widely employed are still based on observable features like hESC-like morphology and expression of pluripotency markers (Takahashi *et al.,* 2007; Park *et al.,* 2008b; Chan *et al.,* 2009). However, there is significant heterogeneity in the pluripotency levels among the clones generated during reprogramming, and this makes the derivation of bona fide hiPSC

A.

hiPSC clones	Number of sites (mice) generated teratoma / Number of sites (mice) injected	
	Black SCID	White SCID
R8	6 (3) / 6 (3)	6 (3) / 6 (3)
R13	6 (3) / 6 (3)	6 (3) / 6 (3)
R48	8 (5) / 9 (5)	4 (2) / 4 (2)
N27	2 (2) / 3 (2)	1 (1) / 1 (1)
Success rate of teratoma formation	92%	100%

B.

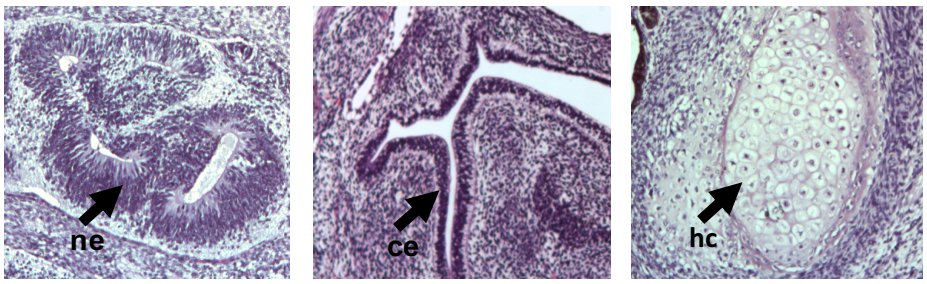


Figure 21: hiPSC clones isolated based on morphology with and without using retroviral silencing as a marker generate high grade teratomas
(A) Table showing the outcome of teratoma assays performed in black SCID and white SCID mice. R8, R13 and R48 are RV-hiPSC lines and N27 is an Epi-hiPSC line. **(B)** Haematoxylin and eosin staining of formalin fixed teratoma sections showing tissues of all the three germ layers. Representative images are shown. ne-neuroepithelium (ectoderm), he-hyaline cartilage (mesoderm) and ce-columnar epithelium (endoderm). Images are at 20X magnification.

lines suitable for downstream applications tedious, expensive and time-consuming. This heterogeneity has been attributed to cell-to-cell genetic variations in the donor cells (Liang and Zhang, 2013), Persistent donor cell gene expression (Kim *et al.,* 2011), the residual expression of the reprogramming factors in the isolated clones (Okada and Yoneda, 2011) and the low efficiency of exogenous factors to bring about complete reprogramming (Raya *et al.,* 2009). Flow cytometry studies on reprogramming fibroblasts had suggested that the enrichment of cells expressing specific combinations of fibroblast and pluripotency markers can improve the derivation efficiency of true hiPSCs (Valamehr *et al.,* 2012; Abujarour *et al.,* 2013; Kahler *et al.,* 2013). However, all these methods require additional tedious and expensive steps and yield hiPSC lines with significant heterogeneity and without clonal identity. As the reprogramming cells, on attaining pluripotency, achieve complete silencing of retroviral vectors this property has also been used to identify hiPSCs (Hotta and Ellis 2008; Chan *et al.,* 2009). Serial live cell imaging experiments had shown that hiPSC colonies that resemble hECs morphologically and molecularly exhibit silencing of retrovirally expressed exogenous factors and fibroblast markers, and expression of pluripotency markers, SSEA-4 and TRA-1-60 (Chan *et al.,* 2009; Tanaka *et al.,* 2015). However, the majority of the cell clusters with this marker expression profile observed in the initial days of reprogramming failed to grow, and many colonies morphologically similar to hESCs were found to be partially reprogrammed without achieving TRA-1-60 expression and transgene silencing (Chan *et al.,* 2009; Subramanyam and Blelloch, 2009). A systematic comparison of the currently used methods for their efficiency to identify high-quality hiPSCs in the early and late stages of reprogramming has been lacking.

The reprogramming cells, as well as the emerging colonies, were observed throughout the reprogramming process in this study. Consistent with the previous reports (Stadtfeld *et al.,* 2008; Tanaka *et al.,* 2015), it was found that shutdown of fibroblast markers is one of the early events in hiPSC reprogramming, and this happens in the majority of the starting population of cells. It was observed that the SSEA-4^{+} state is achieved by the reprogramming cells soon after the initiation of the downregulation of the fibroblast programme whereas the TRA-1-60^{+} state is achieved by SSEA-4 expressing cells towards

the completion of the shutdown process. The continuous induction of SSEA-4 till the beginning of the third week of reprogramming that results in a large percentage of cells expressing SSEA-4 confirms that the reprogramming cells can achieve SSEA4^{+} state with relative ease. However, TRA-1-60 induction is completed as early as in the 1st week of reprogramming, and only about 25% of the SSEA-4^{+} cells attain TRA-1-60^{+} state. These observations clearly suggested that TRA-1-60 can be used as a more reliable pluripotency marker, than SSEA-4, to isolate pluripotent cells even in the early stages of reprogramming before colonies emerge. These findings also suggested that there is a major barrier in reprogramming when cells transit from SSEA-4^{+}TRA-1-60^{-} to SSEA4^{+}TRA-1-60^{+} state. Analysis of the cells belonging to these two reprogramming states will help in understanding the molecular basis of this major barrier, thereby leading to the development of more efficient strategies to improve the reprogramming efficiency. As the cells that were amenable to successful reprogramming achieved CD13$^{dim/-}$SSEA-4^{+}TRA-1-60^{+} state by day 8 after OSKM expression, it is suggested that the protocols that involve enrichment of CD13$^{dim/-}$ SSEA-4^{+}TRA-1-60^{+} cells within the 2nd week of reprogramming can reduce the formation of multiple subclones, thereby decreasing the number of clones that need to be screened for deriving hiPSC lines. However, despite 20% of the cells being TRA-1-60^{+}, the reprogramming efficiency achieved was only less than 0.1%, which indicated that TRA-1-60^{+} cells even in the later days are significantly heterogeneous at the molecular level, and TRA-1-60^{+} is not a false proof marker for the identification of true hiPSCs.

A striking correlation between retroviral RFP silencing (RV-RFP^{-}) and hESC-like morphology was observed in the emerging colonies. These RV-RFP^{-} hESC-like colonies expressed NANOG, a definitive marker of pluripotency, in addition to SSEA-4 and TRA-1-60. A detailed description of the characteristic morphological changes in these emerging hiPSC colonies that can be easily observed by routine monitoring of reprogramming cells has been given, and this can aid in the morphology-based identification and isolation of hiPSC colonies. The isolated RV-RFP^{-} colonies generated stable hiPSC lines and they showed high expression of pluripotency genes for more than 20 passages and could be successfully differentiated *in vitro* and *in vivo*. Thus it was observed that silencing of retroviral LTR-driven

fluorescent marker will help in the identification of the successfully reprogrammed hiPSC colonies as they appear as a defined patch of cells lacking the expression of the fluorescent protein among the unreprogrammed cells/colonies that express the fluorescent protein as early as in the 2nd week after OSKM delivery in fibroblasts.

An extremely high efficiency of teratoma formation was achieved in both black and white SCID mice after intramuscular injection of hiPSCs using the cell processing protocol described by Park *et al.,* (Park *et al.,* 2008a). Previous studies showed that the efficiency of teratoma formation of hESCs was 100% with kidney capsule injections, and it was 60% for intratesticular, 33% for subcutaneous and 12.5% for intramuscular injections (Prokhorova *et al.,* 2009). For hiPSCs, the subcutaneous injection was reported to give very low efficiencies in teratoma formation, but intratesticular injection showed up to 80% efficiency (Ohnuki *et al.,* 2009). The high efficiency of teratoma formation by hiPSC clones derived in our laboratory confirmed that the clones isolated based on characteristic morphological features we described guarantees isolation of bona fide hiPSC clones.

Although integration-free hiPSCs are preferred for their potential use in regenerative medicine, retroviral vectors to express OSKM are still being used for generation of hiPSC lines for disease modelling (Colman and Dreesen 2009; Tulpule *et al.,* 2013; Djuric *et al.,* 2015). Based on this study, it is suggested that the use of retroviral vectors to express fluorescent proteins in the initial experiments of somatic cell reprogramming will be helpful for the laboratories, which lack previous experience in hESC culture and reprogramming, for easy identification of bona fide hiPSCs. It is also proposed that combining silencing of retroviral fluorescent proteins with traditional markers SSEA-4 and TRA-1-60 can significantly increase the specificity and enhance the efficiency of the FACS-based selection protocols compared to the previously described strategies.

Chapter 5

DERIVATION OF FANCONI ANAEMIA PATIENT-SPECIFIC HiPSC LINES FOR ELUCIDATING THE ROLE OF FANCONI ANAEMIA PATHWAY IN REPROGRAMMING AND FOR DISEASE MODELING

5.1. Introduction

Fanconi anaemia (FA) is the most common inherited bone marrow failure disease. The patients manifest a wide range of developmental abnormalities, and 90% of them progressively develop bone marrow failure (BMF), and a significant number of the patients develop additional complications including leukemia and solid tumors that increase the risk of lethality (Kutler *et al.,* 2003). The disease is caused by an inherent defect in genes associated with DNA interstrand cross-link repair (ICR) pathways. On exposure to DNA cross-linking agents, due to the inability of the cells to repair interstrand cross-links, the cells accumulate genetic aberrations, which in turn trigger the DNA surveillance system and induce cell cycle arrest and the cells with genetic aberrations that manage to bypass the DNA damage checkpoint increase the risk of developing cancer.

The molecular interactions of the FA-pathway have been extensively studied and characterized, but the physiological consequences resulting in the clinical and highly variable phenotypes, in particular leading to BMF, are less clear. BMF is identified to be caused by gradual exhaustion of defective haematopoietic stem cells (HSCs) in the bone marrow (Garaycoechea and Patel, 2013). Studying FA in the cells obtained from patients is often impractical due to the rarity of the disease, the low cellularity of patients' bone marrow and inaccessibility to the tissues. Transformed FA cell lines, due to transformation-related artifacts, do not faithfully recapitulate FA disease phenotypes. Although fibroblasts of FA patients are useful in studying DNA damage repair in FA, this cellular model is not suited for studying haematopoietic stem cell defects. Due to the difference in metabolic profile and exposure to environmental stress between mouse and human, the majority of the mouse models carrying genetic defects in FA pathway genes failed to develop haematological phenotypes observed in the patients with FA. One of the current concepts on BMF in FA patients is that the HSCs get depleted due to continuous exposure to cross-linking agents like the endogenous aldehydes in the HSC niche (Garaycoechea and Patel, 2013). The exact mechanism by which this happens and the role of an individual gene is FA pathway has not been completely understood.

Induced pluripotent stem cells (iPSCs), because of their unlimited replicative capacity, clonability and the ability to differentiate into specific cell types, have been extensively used for cellular and molecular studies to understand the molecular basis of genetic diseases (Colman and Dreesen 2009; Meng *et al.,* 2010; Tiscornia *et al.,* 2011; Liu *et al.,* 2012; Zhang 2014; Cao *et al.,* 2015; Jung *et al.,* 2015). The generation of FA iPSCs is found to be challenging due to the increased sensitivity of cells to reprogramming associated genotoxic stress. The low reprogramming efficiency of FA cells, accumulation of chromosomal abnormalities in FA-deficient hiPSC lines and their progressive exhaustion in culture, and a prerequisite for gene complementation for deriving stable hiPSC lines with normal karyotype are the challenges faced in FA cell reprogramming. Presently employed reprogramming strategies derive either FA deficient or FA gene complemented hiPSCs lines (Raya *et al.,* 2009; Müller *et al.,* 2012a; Müller *et al.,* 2012b; Yung *et al.,* 2013; Liu *et al.,* 2014). The exact role of FA pathway in reprogramming has not been explored yet. By comparing haematopoietic differentiation potential of FA deficient hiPSC lines with FA gene corrected hiPSC lines and normal hiPSC lines, it was seen that the former showed impaired production of different cell types of the haematopoietic compartment (Raya *et al.,* 2009; Müller *et al.,* 2012a; Müller *et al.,* 2012b; Yung *et al.,* 2013; Liu *et al.,* 2014). Due to clonal variations among the hiPSC lines, qualitative and quantitative assessment of defects in haematopoiesis by such comparisons will be less accurate. Using isotypic mutant and gene corrected hiPSC lines would be ideal to elucidate the effect of defective FA pathway on haematopoietic differentiation. This is possible by conditional complementation of normal FANC gene during haematopoietic differentiation and studying the cells at different stages of differentiation. A recent study has reported the generation of FA-hiPSC lines from FA fibroblasts manipulated with a doxycycline (DOX)-inducible construct for driving the expression FANCA gene (Cusulin *et al.,* 2015).

In the study carried out in this thesis, FA-A hiPSC lines were generated from the fibroblasts derived from a patient with FANCA gene deficiency (FA-A fibroblasts) by conditionally expressing FANCA gene in a DOX-inducible manner. A non-integrative Sendai virus-mediated method was used for generating FA-A hiPSC lines. This DOX-inducible

FANCA lentiviral system facilitated the investigation of the role of FA pathway in the reprogramming process and the maintenance of pluripotency. The FA-A-hiPSC lines complemented with FANCA were evaluated for their ability to undergo to haematopoietic differentiation. Compared to all other models reported in the literature, these FA hiPSC lines can be maintained and differentiated with the conditional expression of FANCA and used to mimic the cellular phenotypes of FA by withdrawing DOX in the culture medium.

5.2. Objectives

The objectives of the study described in this chapter are:

1. *Conditional complementation of FANCA gene in the fibroblasts derived from patients with defective FANCA genes (FA-A complementation group).*
2. *Evaluation of the effect of conditional complementation of FANCA on reprogramming of FA fibroblasts of FA-A complementation group.*
3. *Derivation and characterization of FA-A-hiPSC lines*
4. *Evaluation of FA disease phenotype recapitulation in FA-A-hiPSC lines.*

5.3. Results

5.3.1. Generation of inducible lentiviral vector for conditional restoration of functional Fanconi anaemia pathway in FA-A fibroblasts

The fibroblasts obtained from three patients with FA clinical phenotype, who were also diagnosed by chromosome breakage analysis and FANCA complementation test, were used in this study. For complementation analysis, FANCA gene was cloned into the pINDUCER20 vector by Gateway cloning to generate a DOX-inducible lentiviral system to express FANCA gene (**Figure 22A**). pINDUCER20 has a multicistronic design in which promoter drives the constitutive expression of rtTA and neomycin as a bicistronic transcript, and tetracycline response element 2, TRE2, facilitates DOX-inducible expression of the gene of interest. Upon DOX addition, DOX binds to rtTA3 and prevent it from occupying the TRE2,

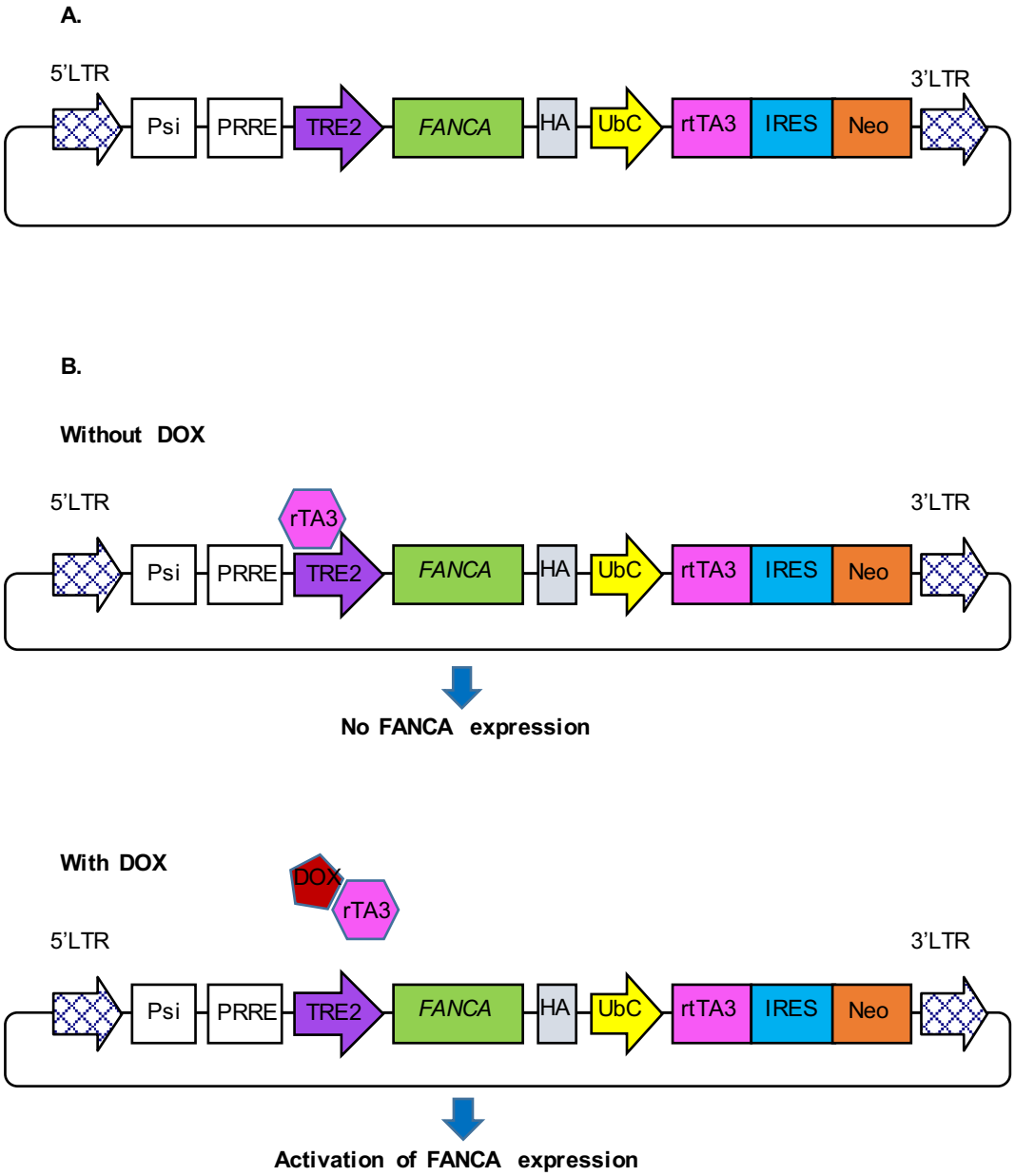


Figure 22: Conditional restoration of functional FA pathway in FA-A patient derived fibroblasts

(A) Schematic representation of pINDUCER20-FANCA vector. **(B)** The mechanism of activation of FANCA expression from pINDUCER20-FANCA vector. In the absence of doxycycline (DOX^-), the constitutively expressed rTA3 bind to TRE2 and block the transcription of *FANCA*. On addition of doxycycline (DOX^+), DOX binds to rTA3 and prevent the latter from binding TRE2. Thus, under DOX^+ condition, TRE3 drive the transcription of FANCA.

C.

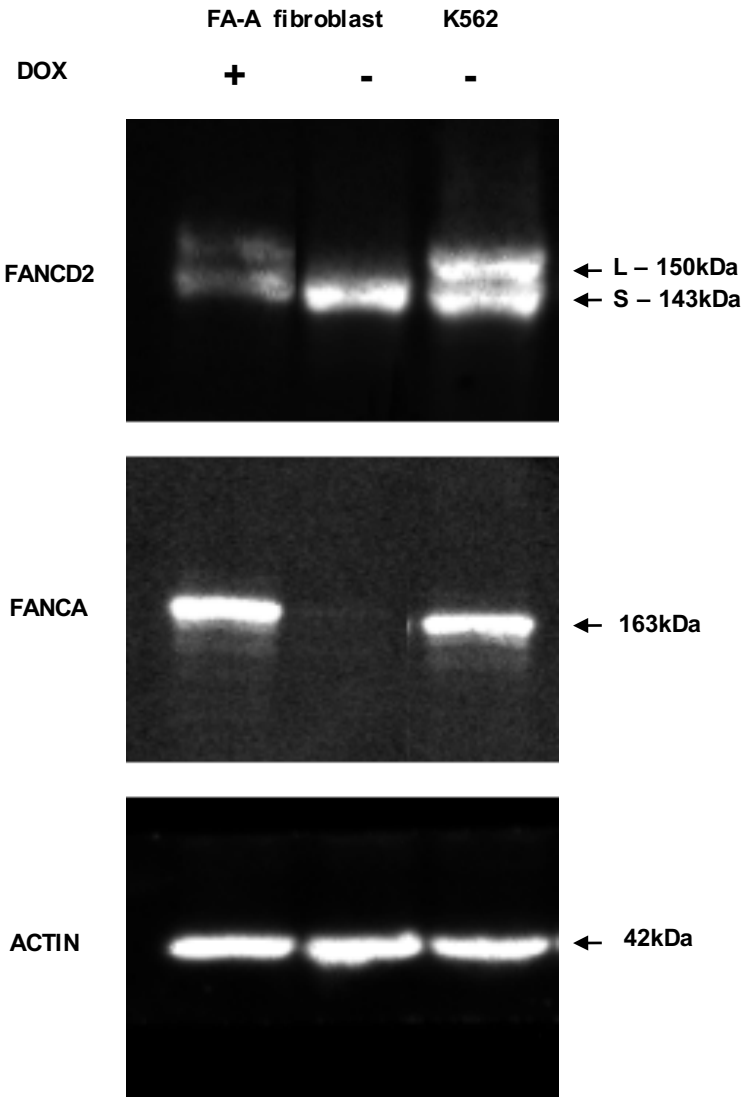


Figure 22 continued: (C) Western blot analysis of FANCD2 and FANCA expressions in the pINDUCER20-FANCA transduced FA-A fibroblasts in DOX^+ and DOX^-. Non-ubiquitinated and monoubiquitinated FANCD2 bands are indicated by S and L , respectively. K562 serves as a positive control.

and this activates TRE2 driven transcription of the gene of interest (**Figure 22B**). Neomycin acts as a selection marker to enrich the cells that are stably transduced with the vector. Since pINDUCER20 express both reverse tet-transactivator 3 (rtTA3) and the gene of interest, it offers the advantage of achieving uniform temporal, dose-dependent and reversible control of gene expression by exposure to DOX with a single vector. In the transduced fibroblasts, in the absence of DOX (DOX^-), there was a complete absence of exogenous FANCA and on the addition of DOX (DOX^+), FANCA expression was induced (**Figure 22C**). By analyzing FANCD2 ubiquitination status, it was found that with the complementation of FANCA, restoration of functional FA pathway could be achieved in these FA-A fibroblasts. In DOX^-, only non-ubiquitinated FANCD2 was present and in DOX^+, both ubiquitinated and non-ubiquitinated FANCD2 were present (**Figure 22C**). Thus, we generated FA-A fibroblasts capable of complementation of FANCA, and restoration of functional FA pathway with DOX addition in the culture medium.

5.3.2. Effect of conditional complementation of FANCA on reprogramming of FA-A fibroblasts.

Resistance to reprogramming exhibited by FA cells has been reported to be overcome by restoring functional FA pathway with constitutive gene complementation (Rodrı *et al.,* 2009; Müller *et al.,* 2012a). We analyzed the effect of constitutive and temporal complementation of FANCA on reprogramming of patient fibroblasts with hSTEMCCA polycistronic lentiviral vector (**Figure 23A**). When FANCA was constitutively expressed by continuous addition of DOX in the FA reprogramming cells TRA-1-60^+ hiPSC-like colonies were formed by the 4^{th} week of reprogramming (**Figure 23B**). However, the normal control fibroblasts formed colonies by the 3^{rd} week of reprogramming. This showed a significant delay in reprogramming in the FA fibroblasts even after FANCA complementation. Consistent with this, flow cytometric analysis on day 18 after reprogramming showed that the percentages of SSEA4^+ and TRA-1-81^+ cells were markedly lower in the complemented FA cells compared to those from control fibroblasts (**Figure 23C**).

Subsequently, the experiments were carried out in DOX^+ and DOX^- conditions to understand whether induced expression of FANCA affects lentiviral-mediated

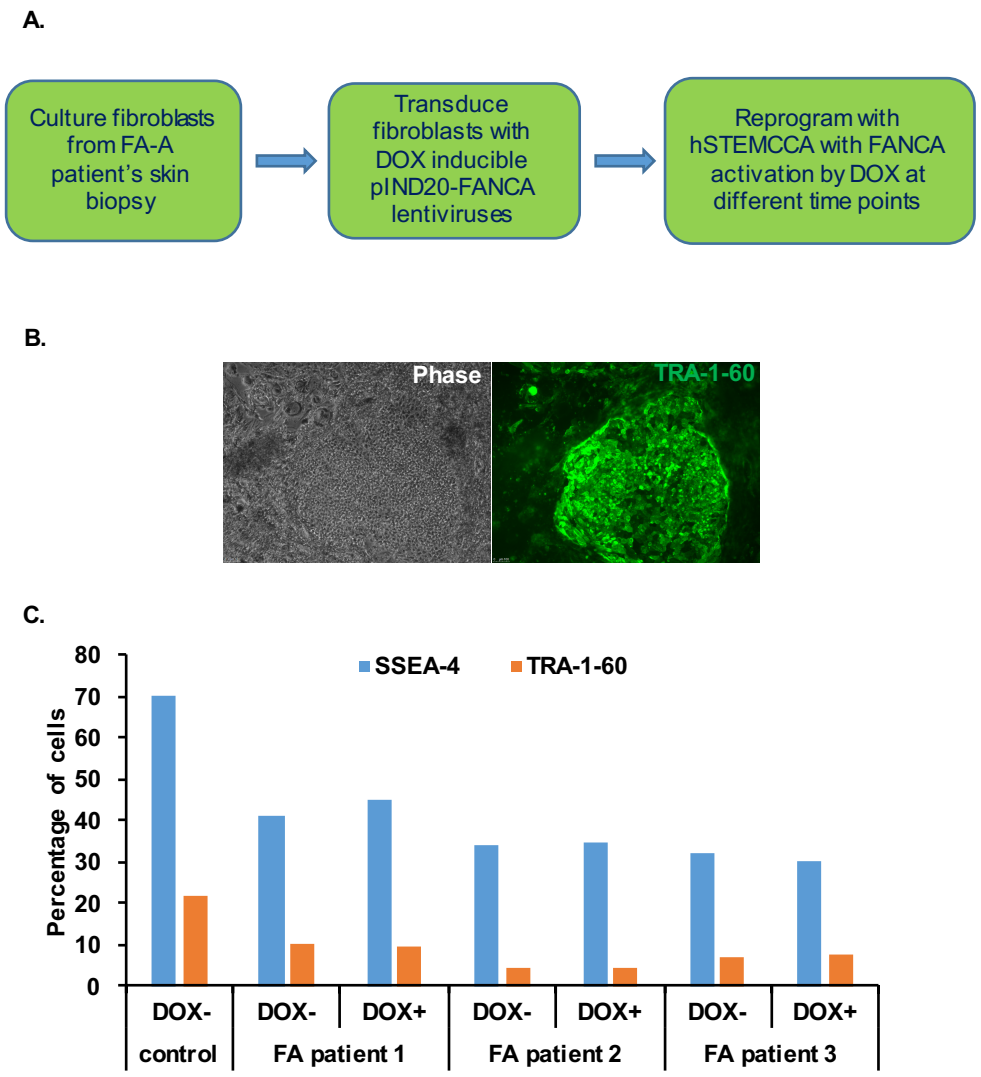
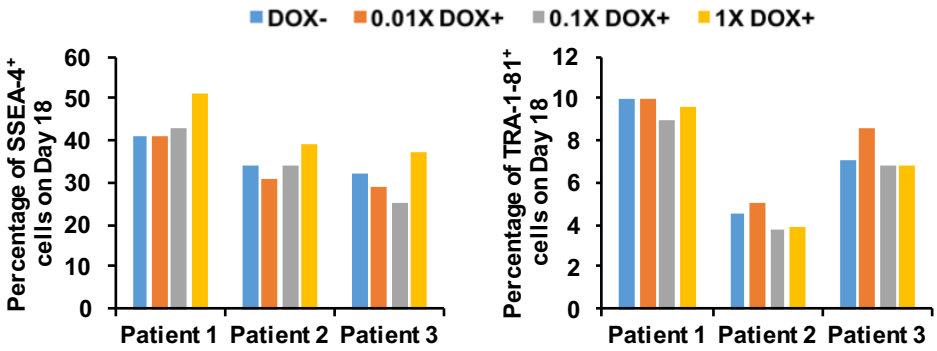
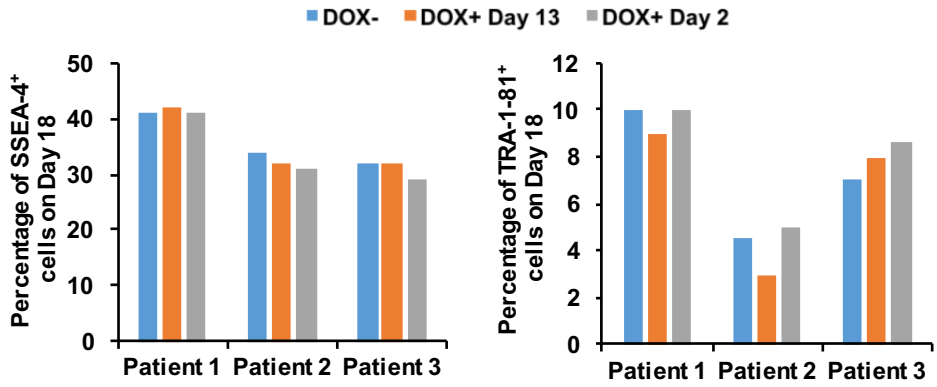


Figure 23: Effect of inducible complementation of FANCA on lentiviral vector (hSTEMCCA) mediated reprogramming of FA-A fibroblasts
(A) Schematic representation of hSTEMCCA mediated reprogramming of FA fibroblasts with conditional expression of FANCA. **(B)** Images of emerging TRA160^{+} hiPSC colonies at 10X magnifications. **(C)** Percentages of SSEA-4^{+} and TRA-1-81^{+} cells formed from FA fibroblasts, with and without complementation of FANCA, and control fibroblasts.

D.



E.



F.

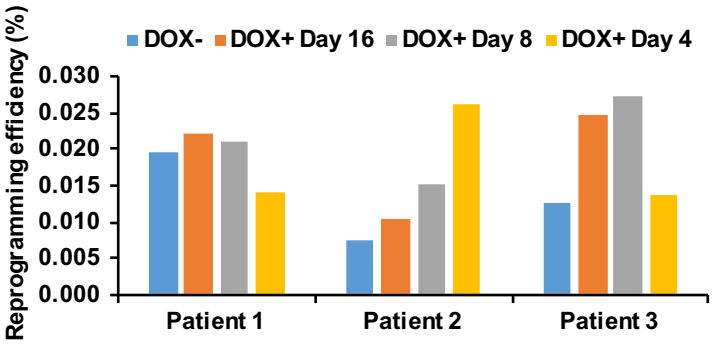


Figure 23 continued: **(D-E)** Percentages of SSEA-4$^+$ and TRA-1-81$^+$ cells formed from FA cells on day 18 with induction of FANCA expression by exposure to DOX **(D)** at different concentrations and **(E)** different time points. **(F)** Reprogramming efficiency of FA cells with FANCA induction at different time points estimated based on the number of TRA-1-60$^+$ colonies formed on day 29 of reprogramming.

G.

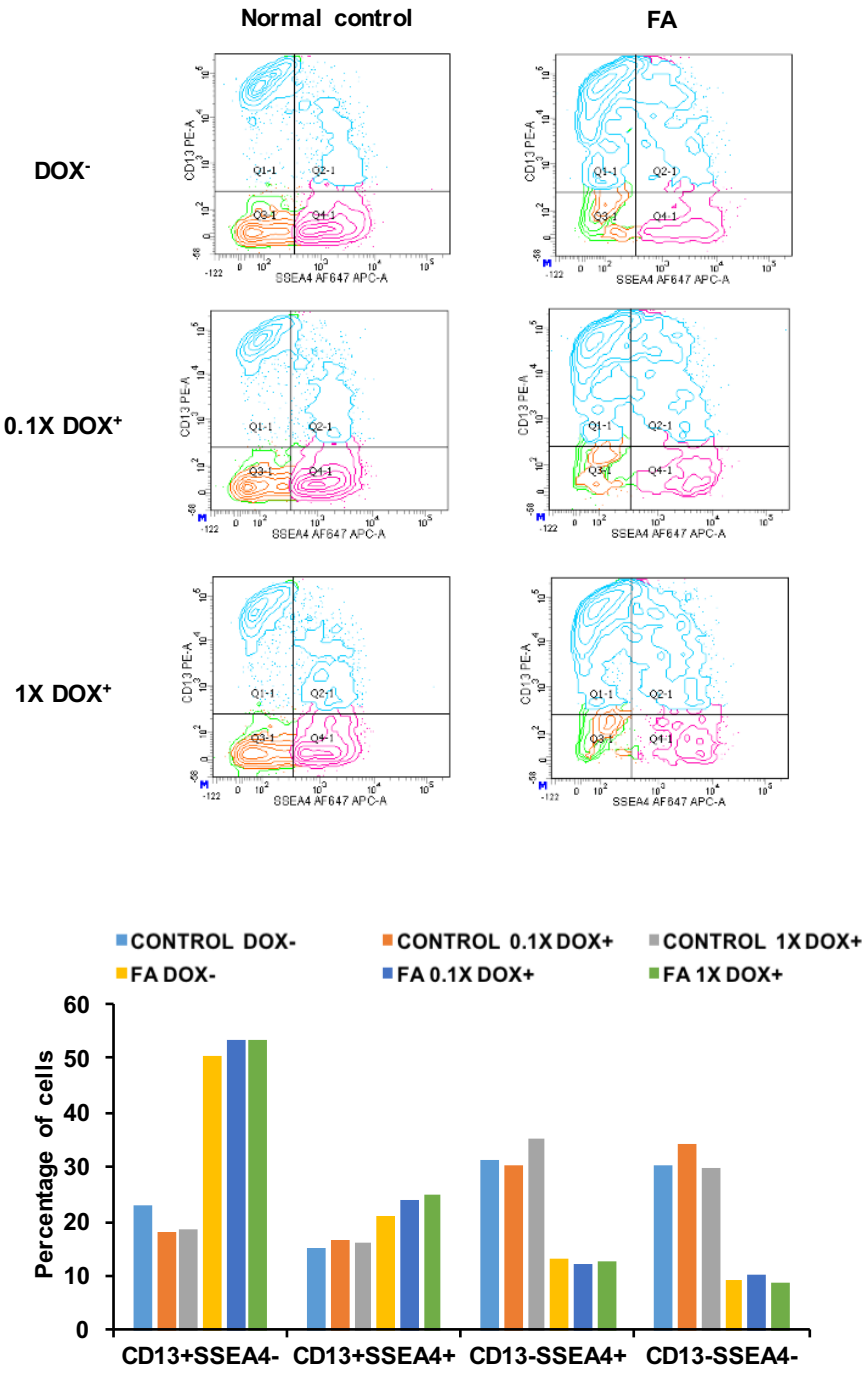


Figure 23 continued: **(G)** The co-expression pattern of CD13 and SSEA-4 in normal control and FA patient fibroblasts undergoing reprogramming, both with inducible FANCA expression, on day 15 of reprogramming. 0.01X, 0.1X and 1X DOX$^+$ indicate the concentrations of DOX at 1ng/ml, 10ng/ml and 100ng/ml, respectively.

reprogramming. DOX was added at different concentrations, 0.1ng/ml, 1ng/ml and 10ng/ml (**Figure 23D**) and at different time points, on days 2, 4, 8, 13 and 16 (**Figures 23E and 23F**). When the reprogramming cells were analysed on day 18, no significant difference was found in the percentages of SSEA-4$^+$ and TRA-1-60$^+$ cells in the absence and the presence of DOX at different concentrations (**Figure 23D**) and different time points (**Figure 23E**). Moreover, the colonies that were formed in any of the different DOX conditions did not have hiPSC morphology. On day 29, when the reprogramming efficiency of the cells exposed to DOX at different time points was estimated with respect to the number of TRA-1-60$^+$ hESC-like colonies formed, the timing of FANCA induction was not found to have any specific effect on hiPSC generation (**Figure 23F**). However, there was a patient to patient variation in the efficiency of SSEA-4 and TRA-1-60 induction (**Figures 23C, 23D, 23E and 23F**). Thus, it is seen that FA-A fibroblasts are unable to restore the pace and the efficiency of lentiviral-mediated reprogramming, even after the functional FA pathway is restored.

The slow pace and low efficiency of lentiviral-mediated reprogramming of FA cells were further confirmed by comparing the expression profile CD13 and SSEA-4 in the reprogramming cells from FANCA complemented fibroblasts and normal fibroblasts transduced with pINDUCER20-FANCA. When they were reprogrammed in the presence and absence of DOX, the percentage of CD13$^+$ cells was about 50 times more in the FA cells compared to the control, indicating that shutdown of somatic cell programme is impaired in the FA cells even after complementation (**Figure 23G**). Similarly, the induction of SSEA-4 was inefficient in FA reprogramming cells compared to the normal cells, with a decrease of about 25% (**Figure 23G**). The expression levels and co-expression pattern of CD13 and SSEA-4 in the reprogramming cells revealed very high levels of heterogeneity among the FA cells, compared to the normal control fibroblasts (**Figure 23G**). Among the SSEA-4$^+$ fraction, about 60% remained CD13$^+$ in FA sample compared to just 33% in control (**figure 23G**).

To understand the effect of non-integrating vectors in the reprogramming of FA-fibroblasts, experiments were carried out using Sendai viruses to express OSKM. Unlike with lentiviral-mediated reprogramming, no hiPSC colonies emerged in DOX$^-$ condition. The

conditional complementation FANCA at a later time point of day 16 was enough to generate colonies with hiPSC morphology.

Thus, it was found that FA cells are incapable of undergoing reprogramming efficiently by lentiviral approach despite the complementation with FANCA. This might be because of the inherent inability of the FA fibroblasts to tolerate DNA damage brought about during the integration of lentiviral vectors, in addition to the pre-existing chromosomal abnormalities in these cells. The restored FA pathway might be hindering the reprogramming of such cells to maintain genomic integrity. This hypothesis is further supported by the increased reprogramming efficiency observed with non-integrative sendai virus-mediated reprogramming.

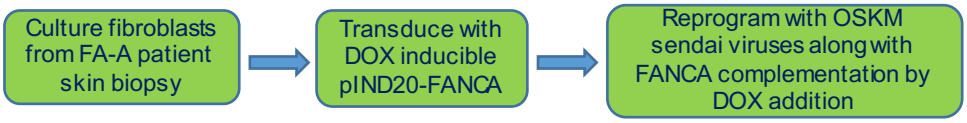
5.3.3. Generation and characterization of FA patient-specific iPSC lines

As the non-integrative Sendai virus-mediated approach for reprogramming was found to be more effective than lentiviruses in FA cells after FANCA complementation, the experiments to generate the hiPSC lines from FANCA deficient fibroblasts were carried out using Sendai viruses to express OSKM. FA fibroblasts transduced with pINDUCER20-FANCA lentiviruses were subjected to reprogramming (**Figure 24A**).

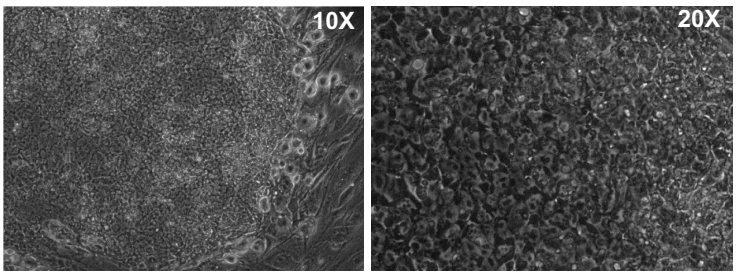
The reprogramming of FANCA transduced FA fibroblasts was carried out in the presence of DOX in the culture medium for the constitutive expression of FANCA. After 18-23 days, hiPSC colonies emerged, and ten colonies were isolated based on their hESC-like morphology (**Figure 24B)**. Out of these ten colonies, six could be maintained as stable hiPSC lines for more than 20 passages and all of them expressed pluripotency markers when they were analyzed at 5, 10, 15 and 20 passages (**Figures 24C and 24D**). These six FA-A hiPSC lines could be adapted to feeder free culture system (**Figure 24E**). Karyotypic analysis of four hiPSC lines did not show any chromosome number variations or chromosome structural abnormalities (**Figure 24F**).

One of the FA-A hiPSC clones, C6, was analyzed for its *in-vivo* differentiation potential by teratoma by injecting into SCID mice. The tumor was found after two months, and the histological analysis showed all the three germ layers in the tumor tissues confirming its pluripotency (**Figure 24G**).

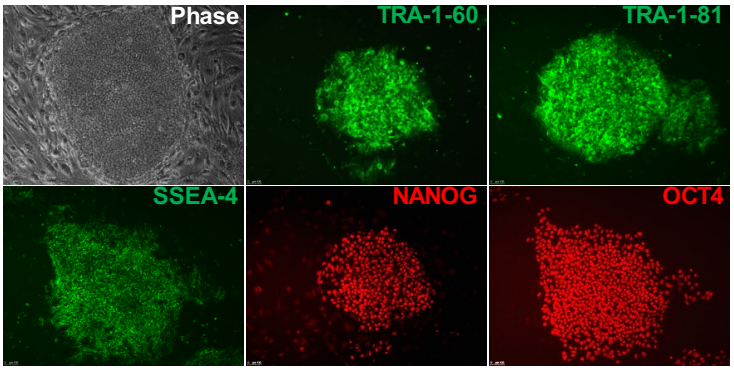
A.



B.



C.



D.

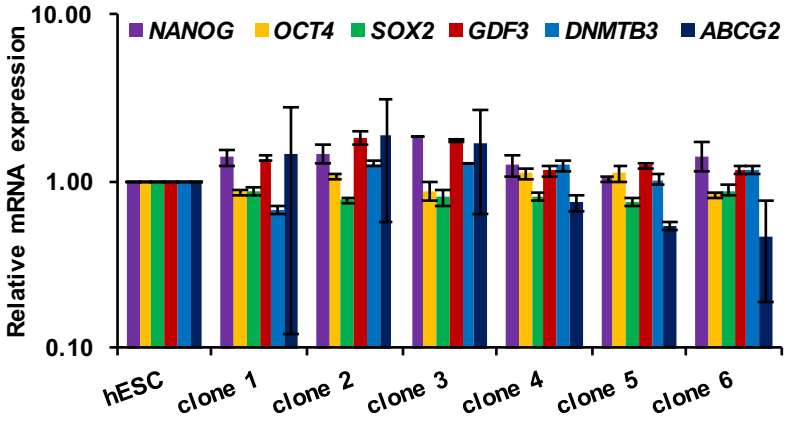


Figure 24: Generation and characterization of FA patient specific hiPSC lines **(A)** Schematic representation of derivation of FA-A hiPSC lines from FA-A fibroblasts by Sendai virus mediated reprogramming. **(B)** Morphology of emerging hiPSC colonies. **(C)** Immunofluorescence staining for TRA-1-60, TRA-1-81, SSEA-4, NANOG and OCT4. Images are at 10X magnification. **(D)** Real time PCR analysis for expression of *NANOG, OCT4, SOX2, GDF3, DNMT3B* and *ABCG2*. The fold changes were calculated relative to the expression levels in hESCs (n=2).

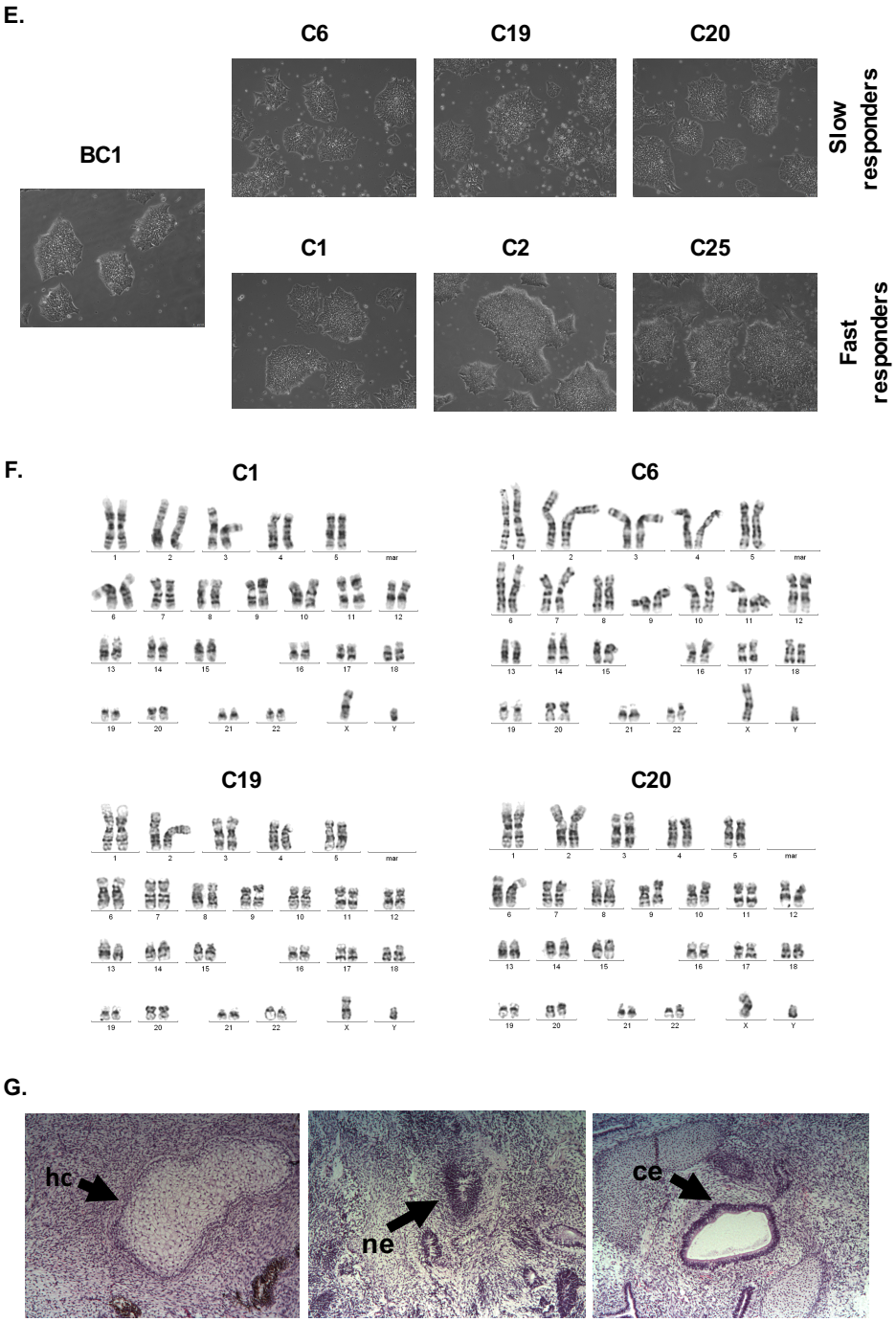


Figure 24 continued: (E) Representative colony images of BC1 control and six FA-A hiPSC lines adapted to feeder free system at passage P20-25. **(F)** Karyotype of four FA-A hiPSC lines at passage P12-15. **(G)** Haematoxylin and eosin staining of formalin fixed teratoma sections of C6 FA-A hiPSC line showing tissues of all the three germ layers. Representative images are shown. ne -neuroepithelium (ectoderm), he -hyaline cartilage (mesoderm) and ce-columnar epithelium (endoderm). Images are at 10X magnification.

Thus, by using a non-integrative approach, four pluripotent FA-A hiPSC lines could be generated which maintained a normal karyotype and could be adapted to feeder-free culture system.

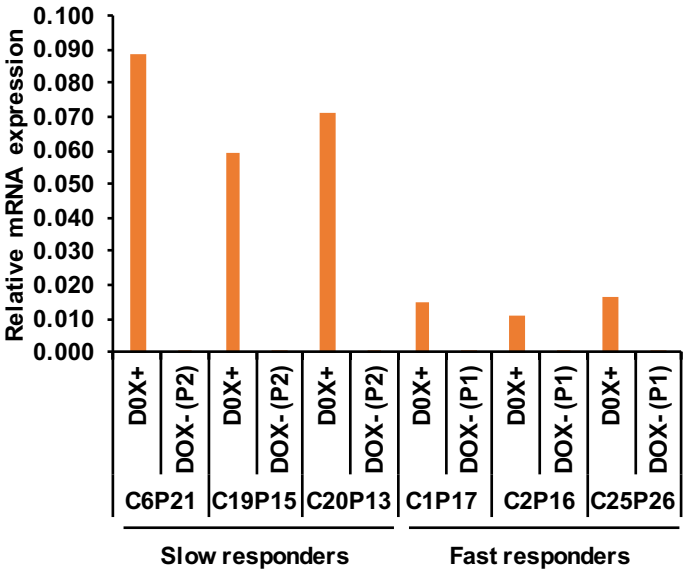
5.3.4. Confirmation of functional DOX-inducible FANCA system in the FA-A hiPSC lines

To confirm that the FA-A hiPSC lines have an active DOX-inducible FANCA system, the colonies cultured in the presence and absence of DOX were analyzed for expression of FANCA transgene. The mRNA transcript analysis showed a progressive decline in the expression of FANCA transgene on DOX withdrawal (**Figure 25A**), confirming that *FANCA* expression is regulated by DOX in the hiPSC lines. Among the six hiPSC lines that were analyzed for transgene expression, three lines had very high levels of FANCA, and three lines had lower levels of FANCA (**Figure 25A**). The lines that showed a low level of transgene expression diminished exogenous FANCA level faster by first passage on DOX withdrawal. Hence, these hiPSC lines were categorized as fast responders in this study. In the lines that had a high level of transgene expression, exogenous FANCA was observed for up to 10 days or till passage 2 even after DOX withdrawal. These hiPSC lines were classified as slow responders. The reduction in the transgene expression could be confirmed by western blot analysis of FANCA expression in FA-A hiPSC clones in DOX^+ and DOX^- conditions. (**Figure 25B**). Analysis of FANCD2 ubiquitination by western blot analysis showed that both ubiquitinated and non-ubiquitinated FANCD2 were present when the clones were cultured in the presence of DOX, and only non-ubiquitinated FANCD2 was present when DOX was withdrawn (**Figure 25B**). These data suggested that the FA-A hiPSCs, generated in the presence of DOX to induce FACNA expression, activate FANCD2 ubiquitination and the FA pathway and the withdrawal of DOX results in defective FA pathway.

5.3.5. Manifestation of FA disease phenotype in C6 hiPSC lines

One of the major cellular phenotypes of FA cells is G2/M cell cycle arrest. As a result, FA-iPSC lines capable of mimicking FA will undergo a progressive replicative crisis in culture following DOX withdrawal. To check whether the FA hiPSC lines generated in this

A.



B.

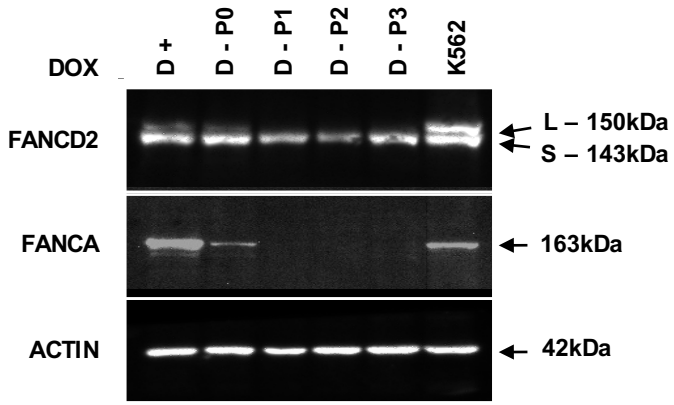


Figure 25: Confirmation of functional doxycycline inducible FANCA system in FA–A hiPSC lines

(A) Real time PCR for FANCA transgene expression in six FA-A hiPSC lines. **(B)** Western blot analysis of FANCD2 and FANCA expressions in slow responder C6 FA-hiPSC line during DOX withdrawal from culture. K562 serves as a positive control.

study can be used for disease modeling, the colonies in culture were analyzed for cell cycle kinetics and replication potential on DOX withdrawal from the medium. The slow responders FA hiPSC lines with high levels of transgene expression showed cell death after three passages and fast responder FA hiPSC lines with lower transgene expression showed cell death in two passages (**Figure 26A**).

Cell cycle analysis of slow responding, C6, at different time points of DOX withdrawal confirmed that the replicative crisis of this clone is due to G2/M cell cycle arrest (**Figure 26B**). The percentage of cells in S phase reduced from 60% to 38% and that in G2/M phase increased from 22% to 44% by day 8 of DOX withdrawal. These data showed that FA-hiPSC lines that were generated in this study could mimic the *in vivo* cellular phenotypes of FA cells, confirming its suitability to model FA *in vitro*.

Since haematological defects are the major FA disease-related phenotypes in humans, FA hiPSC lines generated in the lab were tested for their potential to give rise to cells of haematopoietic lineage. In parallel with a control hiPSC line (BC1-hiPSC line; a gift from Linzhao Cheng) the FA hiPSC line C6 was subjected to haematopoietic differentiation in vitro using adherent monolayer protocol (**Figure 26C**) described by Mills *et al.,* (Mills *et al.,* 2014). The differentiation of FA hiPSCs was carried out in the presence of DOX to complement FANCA expression. The differentiated cells were analysed for expression of hemato-endothelial marker, CD31 (PECAM) and pan-haematopoietic marker CD235/ glycophorin A. On day 6 of differentiation, when the cells are in the hemato-endothelial specification, 40.5% of the attached cells and 78.7% of the suspension cells derived from C6 were $CD31^{+}$, and this was similar to BC1 culture that showed 45% and 62%, respectively (**Figure 26D)**. In a successful haematopoietic differentiation, HSC progenitors start appearing as loosely attached non-adherent cells on the monolayer of differentiating cells by day 8 (Mills *et al.,* 2014). Even though such cells were observed in the BC1 culture (**Figure 26E**), they were not seen in C6 culture. On day 10, when 22.5% of suspension cells were $CD235a^{+}$ in BC1, this was only 7% for C6. Thus, it was found that the differentiation to haemato-endothelial cells in C6 FA-A hiPSC line was comparable to a control line, but the formation of HSC progenitors was impaired in the FA-A hiPSC line (**Figure 26D**). The actual

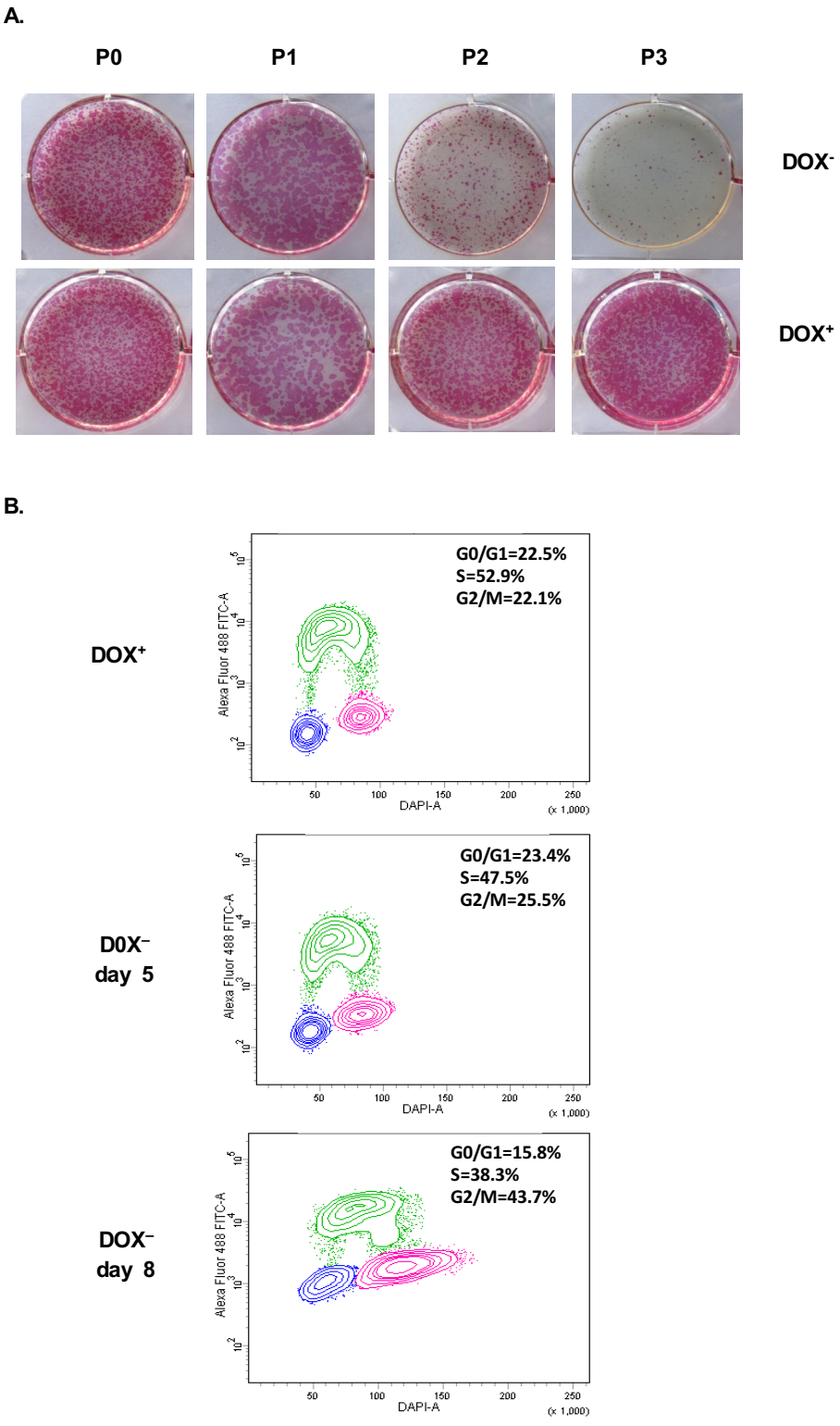
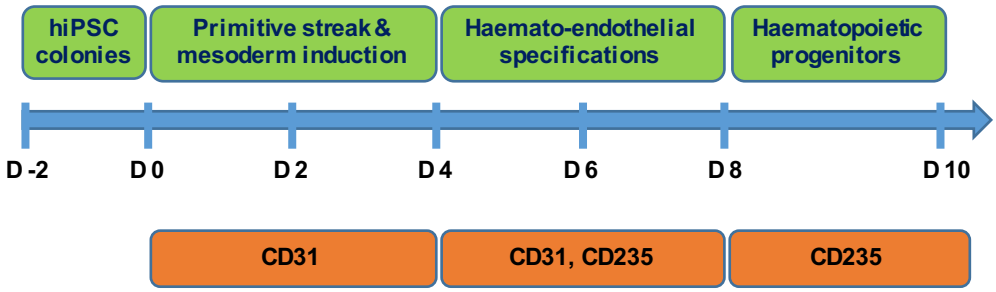


Figure 26: Recapitulation of FA by FA-A hiPSC lines
(A) Alkaline phosphatase staining showing replicative crisis of FA-A iPSCs in the abscece of DOX and FANCA expression and **(B)** the cell cycle analysis of slow responder C6 FA-A hiPSC cultured in DOX^+ and DOX^- conditions.

C



D.

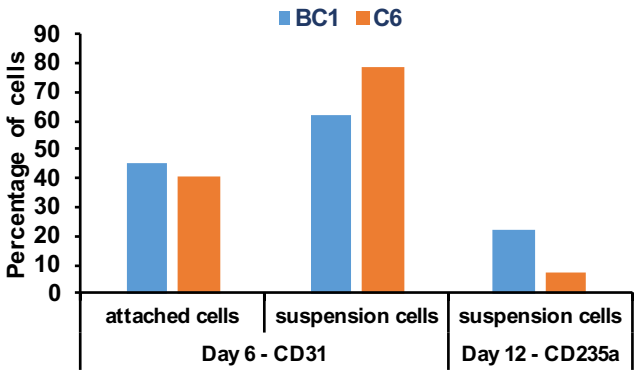


Figure 26 continued: (C) Schematic representation of the protocol for haematopoietic differentiation of hiPSCs. Upper part shows the stages of differentiation, middle part shows different days of differentiation and the lower part shows the markers analyzed for differentiation. **(D)** FACS analysis of hiPSC lines subjected to haematopoietic differentiation.

mechanism for this difference in the abilities to differentiate to HSCs could not be elucidated. As the FA cells were complemented with FANCA expression, this deficiency in differentiation in C6 is not FA pathway related. The other abnormalities present in the FA fibroblasts, related or unrelated to the FA disease, may be responsible for this difference. It is noteworthy that the BC1 control hiPSC line was derived from PBMNCs, and it has been shown that the iPSCs derived from haematopoietic cells have a high propensity to undergo haematopoietic differentiation (Kim *et al.,* 2011).

5.4. Discussion

FA is a genetic disease with defects in DNA double crosslink repair pathway, namely FA/BRCA pathway (Kottemann and Smogorzewska 2013; Ishiai *et al.,* 2016). Even though mouse models with defects in FA/BRCA pathway genes showed disease phenotypes affecting the functioning of various organs and organ systems, they failed to develop BMF, which is the major disease manifestation in humans (Parmar *et al.,* 2009; Bakker *et al.,* 2013). iPSC-based disease modeling offers an alternate approach for exploring the pathogenesis of FA. However, the progress in this field had been hindered by the refractiveness of FA-deficient cells to reprogramming and the pre-requisite for gene correction for generating stable hiPSC lines (Raya *et al.,* 2009; Müller *et al.,* 2012a; Yung *et al.,* 2013). However, there are a few reports on the generation of FA deficient FA-hiPSC lines under normoxic conditions (Raya *et al.,* 2009), but these lines either inherited or acquired chromosomal abnormalities making them unsuitable for any downstream application. The FA deficient hiPSC lines with normal karyotype and cell cycle profile could be generated without gene correction under hypoxic condition (Müller *et al.,* 2012a). However, in long-term culture, these clones have a higher risk of accumulating genetic abnormalities due to defective FA pathway.

In this study, the potential of the inducible lentiviral system to promote reprogramming by conditional complementation of FANCA was explored in FA patient-derived fibroblasts and FA-A patient specific hiPSC lines could be successfully generated. Conditional complementation of FANCA expression and restoration of FA pathway could be

achieved in the patient specific FA-A hiPSC line by controlling the exposure to DOX. Therefore, these hiPSC clones facilitated the side by side comparison of FANCA deficient and FANCA expressing iPSC lines. FANCA deficient lines could recapitulate *in-vitro* the molecular and cellular features of FA-deficient cells and FANCA expressing cells showed normal phenotypes. On DOX withdrawal from the hiPSC culture medium, FA pathway became non-functional indicated by non-ubiquitination of FAND2 and G2/M cell cycle arrest leading to replicative crisis in the hiPSCs. It was observed that the individual hiPSC clones differed in their response time to DOX withdrawal and the slow responding clones developed disease phenotype over a period of about 10 days while going through two to three passages. Such slow responding FA-hiPSC clones will offer the advantage of the time while evaluating FA associated defects in haematopoiesis. Thus, a system could be established for obtaining isotypic FA deficient and proficient patient specific hiPSC lines for investigating the role of FA pathway in reprogramming, pluripotency and FA pathogenesis. The same reprogramming strategy established in the lab can be extended to other complementation groups to derive patient-specific FA-hiPSC lines.

While this study was going on, a recent study reported the generation of patient-specific FA-hiPSC lines through complementation of FANCA using the same inducible lentiviral vector (Cusulin *et al.,* 2015). The reprogramming was performed on patient-derived keratinocytes and fibroblasts using polycistronic lentiviral vectors expressing OSKMR and episomal vectors expressing OSKM along with p53-shRNA respectively. These reprogramming strategies pose a risk of acquiring reprogramming associated genetic defects due to the integration of lentiviral vectors and p53 knockdown. The reprogramming strategy described in this thesis study is advantageous as the FA-hiPSC lines were generated using a non-integrative approach and hence the genome integrity of the clones was less challenged. This is of special mention because DNA damage repair pathways are involved in the pathogenesis of FA.

Previous studies analyzing the haematopoietic differentiation potential of FA pathway deficient hiPSC lines showed a marked deficiency in the generation of blood cells, suggesting the suitability of using FA-hiPSC lines in disease modeling (Müller *et al.,* 2012b;

Yung *et al.,* 2013; Liu *et al.,* 2014). Such studies were confined to analyzing the differentiation potential of FA-hiPSCs lines to different cell types of haematopoietic compartment. However, in such comparative studies, it is important to take into account the clone to clone variation that exists between hiPSC lines that may potentially influence their differentiation potential and interfere with the assessment of the disease-related defects. Moreover, FA defective hiPSC line has a risk of acquiring new genetic defects that may interfere with disease manifestations. So it is preferred to perform such comparative studies using isotypic hiPSC lines with a functional and defective FA pathway at the time of investigations. The hiPSC clones generated in this study in which FA pathway functionality could be regulated by DOX exposure serve this purpose. Analyzing the hiPSCs and cells undergoing haematopoietic differentiation in the presence and absence of DOX can provide valuable insights into disease mechanism.

The present understanding of the cause for BMF is that the hematopoietic stem cells and progenitor cells undergo progressive exhaustion from the bone marrow due to high sensitivity to DNA double strand cross-linking agents in the absence of functional FA pathway (Garaycoechea and Patel, 2013). The increased susceptibility of FA patients to BMF and mouse models to hematological complications in the background of genetic defects in aldehyde dehydrogenase indicated that the endogenous cross-linking agents like aldehydes are the major inducer of DNA double-stranded breaks in HSC niche in FA patients (Langevin *et al.,* 2011; Rosado *et al.,* 2011). The exact molecular basis of endogenous cross-linking agents induced BMF is not known. The addition of cross-linking agents to the culture medium of FA-hiPSCs capable of inducible complementation of FANCA offer the advantage of studying the mechanism by which defective FA pathway contribute to defects in haematopoietic differentiation. Moreover, FA-hiPSC can be subjected to additional genetic modification to regulate the expression of other genes that are associated with disease pathogenesis like those coding for proteins involved in aldehyde metabolism.

Chapter 6

IDENTIFICATION OF NOVEL FACTORS INVOLVED IN SOMATIC CELL REPROGRAMMING BY RNA INTERFERENCE (RNAi)

6.1. Introduction

Reprogramming of somatic cells to iPSCs involves the continuous transition of cells from one state to another under the influence of exogenous factors as well as endogenous factors regulated by them. Several molecular pathways are linked to the reprogramming process, some of which promote pluripotency induction, whereas the others inhibit the process. The complex interplay between multiple pathways during the transition of somatic cells to hiPSCs leads to the generation of colonies at different reprogramming states that have achieved only some features of pluripotent state (Ho *et al.,* 2011; Sanges and Cosma 2011; Liang and Zhang 2013; González and Huangfu 2015). This demands extensive characterization of clones for the derivation of hiPSC lines suited for downstream applications. Understanding the molecular mechanisms underlying somatic cell reprogramming to pluripotency is critical for the creation of high quality pluripotent cells and may be useful for therapeutic applications. Moreover, gaining insights from *in vitro* reprogramming approaches may yield relevant information for SCNT or cell fusion-mediated reprogramming and may broaden our understanding of fundamental questions regarding cell plasticity, cell identity and cell fate decisions.

A successful reprogramming event results in the generation of hiPSCs which show properties similar to hESC-like morphology, growth requirements, cell cycle profile, self-renewal ability and pluripotency. The inactivation of somatic cell specific gene network, mesenchymal to epithelial transition, activation of endogenous pluripotency network and silencing of exogenous reprogramming factors are some of the molecular events associated with such a process. Molecular interventions of pathways associated with these events by overexpression, knockdown or chemical inhibition of specific genes have identified several novel factors involved in reprogramming (Li and Rana 2012; Mancarci *et al.,* 2012; Zhang *et al.,* 2012; Qin *et al.,* 2014). By manipulating the expression of these genes, their stage-specific role in pluripotency induction could be identified and reprogramming efficiency could be improved. The factors that could favour reprogramming included the genes that are part of the pathways maintaining pluripotency and self-renewal of pluripotent stem cells. The

factors that inhibited reprogramming included those involved in differentiation, genomic integrity surveillance and cell cycle arrest. Despite these additional manipulations, the majority of the cells still fail to form hiPSCs, and this clearly indicates the involvement of still unknown regulators of reprogramming.

In the studies attempting to elucidate the molecular mechanisms of reprogramming, overexpression or knockdown vectors with different promoters are being used. It is known that the efficiency of eukaryotic promoters to drive the expression of the genes of interest or shRNAs is influenced by the type of target cells. Hence, alterations in the promoter activity are likely to happen during reprogramming as the cells transit through multiple stages. Moreover, most of the commonly used promoters are known to get silenced in pluripotent stem cells through endogenous epigenetic mechanisms. Such cell type-specific alterations in promoter activity can interfere with the assessment of the role of candidate factors in the reprogramming process by gene regulation approaches. To identify factors that play a critical role throughout the reprogramming process, a transfer vector with a promoter that remains active in these different stages should be used. To identify the promoter that can serve this purpose, we assessed the capability of the commonly used promoters like CBA, UbC, SFFV, mCMV, hCMV, mEF1α and hEF1α to remain active, based on the expression status of GFP driven by them in a reprogramming set up. For this, human adult dermal fibroblast transduced with transfer vectors carrying GFP were subjected to reprogramming and GFP expression status of cells at different stages of reprogramming, and hiPSC colonies were analyzed.

In our search for new factors that might potentially regulate pluripotency induction in fibroblasts, two approaches were adopted for knockdown studies namely candidate gene knockdown approach and shRNA library screening approach. In candidate gene approach, the target genes were chosen based on their known biological roles that are likely to make them capable of regulating different molecular steps in reprogramming and based on their variations in the expression levels in the somatic cells, reprogramming intermediates and pluripotent stem cells in the published datasets (www.stemformatics.com) and the data generated in our lab. The role of these candidate genes in reprogramming was assessed by

estimating the effect of their knockdown on the generation of cells belonging to different stages of reprogramming. In the library screen approach, a pool of lentiviral shRNAs that targets signaling pathway genes was used. The key signaling pathways playing a critical role in reprogramming were identified by comparing selective enrichment and depletion of specific shRNAs in the cells at different stages of reprogramming.

6.2. Objectives

The aim of this chapter is to identify novel regulators that play a critical role throughout the reprogramming process. The main objectives are as follows.

1. *Identification of promoters that show robust activity in different stages of reprogramming.*
2. *Selection of potential genes that can affect pluripotency induction during hiPSC generation.*
3. *To study the effect of candidate gene knockdown on OSKM mediated reprogramming.*
4. *To identify the signaling pathways active in the intermediate and the late stages of reprogramming.*

6.3. Results

6.3.1. Variation in transcriptional activity of promoter during reprogramming

To ensure consistent knockdown of genes throughout the reprogramming process, a promoter that remains active in all the stages of reprogramming is required. The experiments to identify the most efficient promoters that remain active throughout reprogramming and show very high transcriptional activity as the cells transit through multiple have not been systematically carried out so far. In this thesis, lentiviral vectors with 7 different promoters, CBA, UbC, SFFV, mCMV, hCMV mEF1α and hEF1α, were analyzed in the reprogramming cells (**Figure 27A**). For this, individual lentiviral vectors with each of these promoters driving the expression of GFP were transduced into hADFs and subjected

to reprogramming with hSTEMCCA lentiviruses. The promoter activity was measured by two means: (a) the strength of the promoters estimated based on the mean fluorescence intensity (MFI) of GFP and (b) the resistance to silencing measured by the total percentage of GFP^+ cells.

Transcriptional silencing of these promoters in the hADFs showed that SFFV, mCMV, hCMV mEF1α and hEF1α promoters remained active in >99% of the transduced fibroblasts **(Figure 27B**). However, GFP silencing was observed in a small fraction of the cells with UbC and CBA promoters, with GFP^- cells constituting about 6% and 9% of the cells, respectively. Analysis of transcription strength showed that there was a significant difference among these promoters. mCMV and hCMV showed the highest GFP, with MFI=~1500, indicating very high activity of these promoters in hADFs (**Figure 27C**). SFFV had three times, and mEF1α and hEF1α had four times less MFI than CMV promoters. CBA and UbC had the lowest MFI, which was about five times less than CMV. Thus, it was observed that CMV promoter shows robust transcriptional activity in the human adult dermal fibroblasts, whereas other promoters were less efficient.

To evaluate the promoter activity in cells during reprogramming, hADFs transduced with lentiviruses containing different promoters to express GFP were reprogrammed, and GFP expression was analyzed in the different fractions of reprogramming intermediates and reprogrammed cells on days 14 and 18 (**Figure 27D**). The cell fractions analysed included, $CD13^+SSEA4^-$, $CD13^+SSEA4^+$, $CD13^-SSEA\text{-}4^+$ and $SSEA\text{-}4^+TRA\text{-}1\text{-}60^+$. The $CD13^-SSEA\text{-}4^-$ and $SSEA\text{-}4^-TRA\text{-}1\text{-}60^-$ cells dual negative for these markers were omitted from the analysis since feeder cells also belong to this fraction, and they may interfere with data analysis. It was observed that the promoters differed in their transcriptional strength and susceptibility to transcriptional silencing in the reprogramming intermediates. In all the different cell fractions analysed on days 14 and 18, mEF1α and hEF1α promoters were active in 95% of the cells and downregulation of CD13, and induction of SSEA-4 and TRA-1-60 did not induce the silencing of these promoters in these cells. This indicated that mEF1α and hEF1α were the least susceptible for silencing and were consistent with their transcriptional activity throughout the reprogramming process. When $CD13^+SSEA\text{-}4^-$ cell fractions of days 14 and

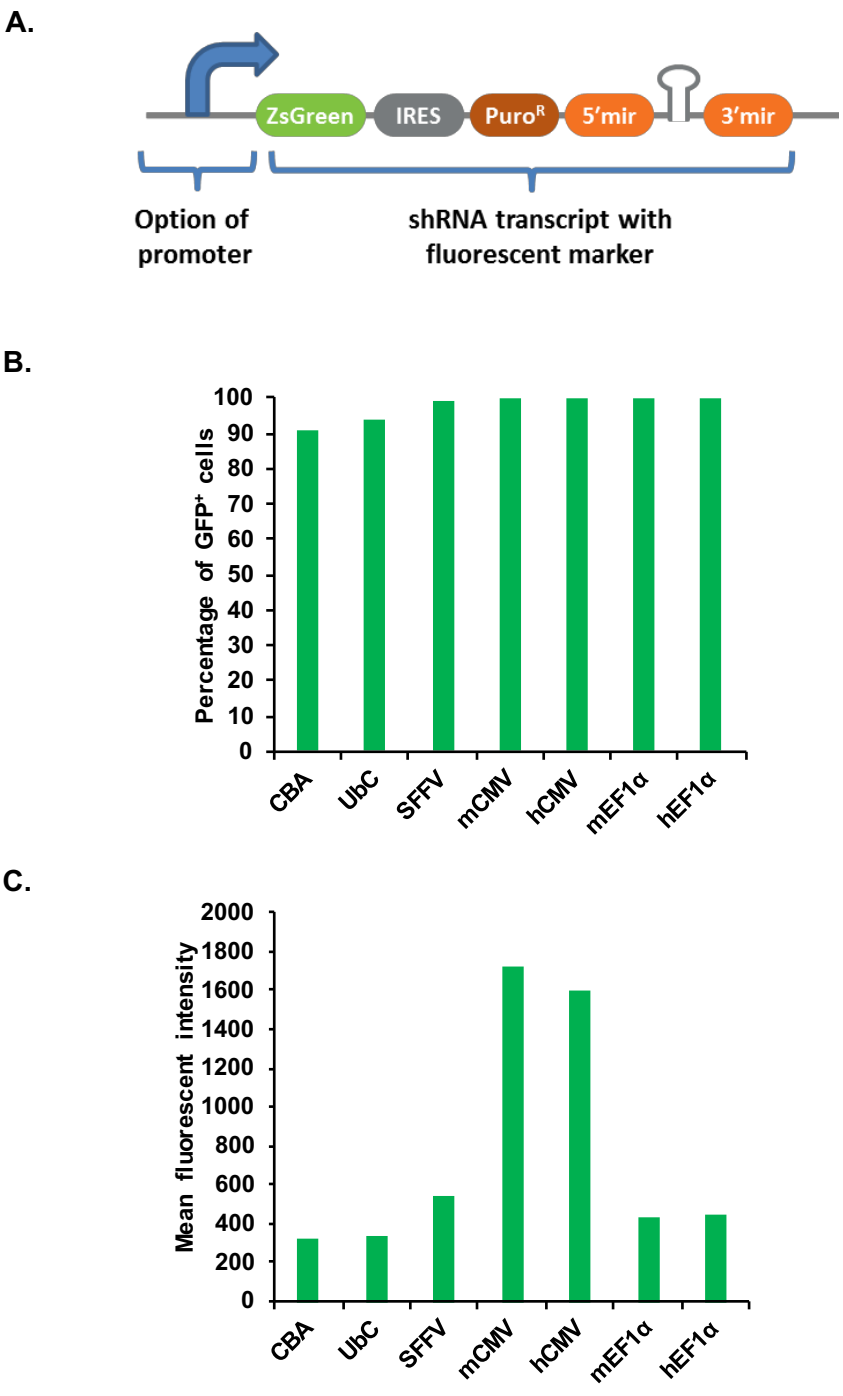
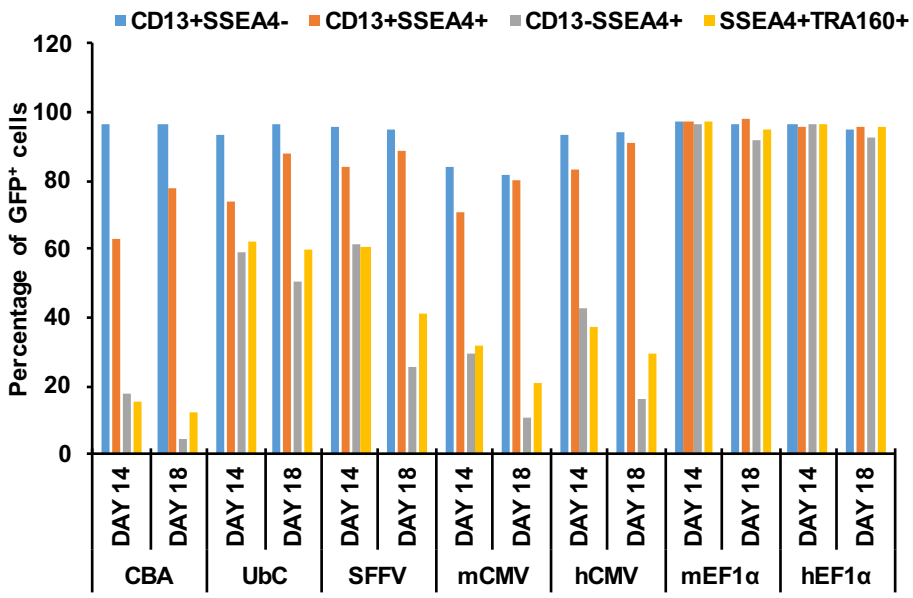


Figure 27: Evaluation of transcriptional activity of the lentiviral promoters in human fibroblasts, reprogramming intermediates and emerging hiPSC colonies

(A) Schematic diagram of pZIP vector that was used in this study. The promoters(CBA, UbC, SFFV, mCMV, hCMV, mEF1α and hEF1α) drive the expression of fluorescent marker, ZsGreen and shRNA as a single transcript. **(B)** Percentage of GFP^+ cells and **(C)** mean fluorescent intensity (MFI) among hADFs transduced with different promoter vectors.

D.



E.

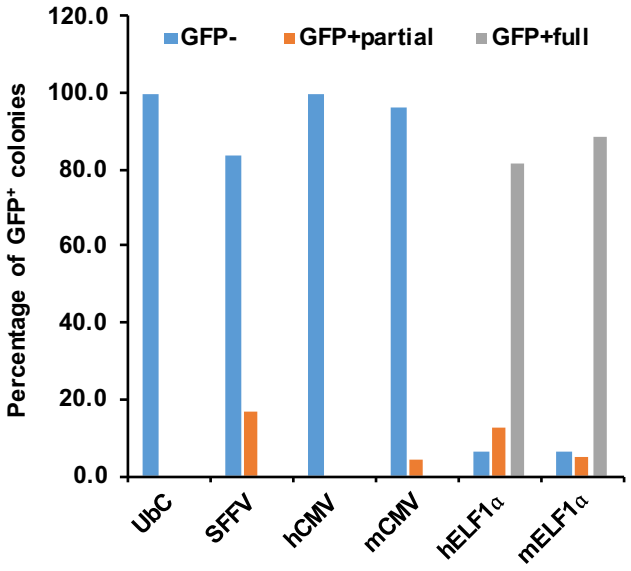


Figure 27 continued: (D) Percentages of GFP+ cells among the reprogramming cells at different stages of reprogramming. **(E)** The percentages of GFP+ colonies among the hiPSC colonies formed on day 18 of reprogramming.

day 18 were analyzed it was observed that 93%-97% of the cells were GFP^+ for all the promoters, except mCMV that showed 16% - 18% of GFP^- cells in this fraction, indicating its silencing in a small fraction of cells.

When the cells transformed from $CD13^+SSEA\text{-}4^-$ stage to $CD13^+SSEA4^+$ and CD13-$SSEA4^+$ stages there was a decrease in the percentages of GFP^+ cells. This decline was very prominent in the CBA promoter where the percentage of the GFP^+ cells decreased from 96% to 63% to 17% on day 14 in the $CD13^+SSEA\text{-}4^-$, $CD13^+SSEA\text{-}4^+$, and $CD13^-SSEA\text{-}4^+$ cells, respectively. The mCMV and hCMV promoters showed relatively less degree of silencing during the transition from $CD13^+SSEA4^-$ stage to $CD13^+SSEA4^+$ stage and then the reduction was drastic from the $CD13^+SSEA4^+$ stage to CD13-$SSEA4^+$ stage. For the mCMV promoter, the percentages of GFP^+ cells declined from 84% to 71 to 29% and for hCMV promoter, it was from 94% to 83% to 43%. UbC and SFFV showed a steady decrease in the percentage of GFP^+ cells, with GFP^+ cells constituting about 60% of $CD13^-SSEA\text{-}4^+$ on day 14. On day 18, the extent of promoter silencing in $CD13^-SSEA\text{-}4^+$ fraction increased further. Among the promoters that showed silencing in the $CD13^-SSEA\text{-}4^+$ fraction of day 18, UbC showed better transcriptional activity with the 50% cells remaining GFP^+, followed by SFFV (26%), hCMV (16%), mCMV (11%) and CBA (5%).

There was no significant difference in the promoter activity between the $SSEA\text{-}4^+TRA\text{-}1\text{-}60^+$ and $CD13^-SSEA\text{-}4^+$ cells on day 14, but significant silencing was observed on day 18 as the reprogramming progressed. The percentage of GFP^+ cells in the $SSEA\text{-}4^+TRA\text{-}1\text{-}60^+$ fraction was the highest in the UbC promoter (60%), followed SFFV (41%), hCMV (30%), mCMV (21%) and CBA (12%).

When the emerging hiPSC colonies were analyzed on day 20 of reprogramming, it was found that more than 80% of the hiPSC colonies were GFP^+ when GFP was driven by hELF1α and mELF1α promoters (**Figure 27E**). At the same time, more than 80% of hiPSC colonies were GFP^- when GFP was driven by UbC, SFFV, hCMV and mCMV promoters. The systematic analysis of the transcriptional activity of promoter has confirmed that each promoter shows a specific kinetics of silencing during the reprogramming process. The mEF1α and hEF1α remained active throughout the reprogramming process and their

promoter activity increased as fibroblast cells reprogrammed to iPSCs. Since hEF1α exhibited consistent activity during reprogramming, the shRNA vector with this promoter was used for the subsequent experiments.

6.3.2. Selection of candidate genes for RNAi studies

In the candidate gene approach for identifying the factors influencing reprogramming, genes were selected based on their likely potential to influence reprogramming. Since MET is one of the initial events in reprogramming, *GRHL1, GRHL2* and *GRHL3* that encode transcription factors regulating epithelial gene expression were selected for the knock-down studies. As the previous studies have shown that the reprogramming process is predominantly associated with epigenetic regulation, the genes associated with histone and DNA modifications were also selected in this study.

To identify the target genes for the RNAi experiments, we analyzed the expression profiles of histone methyltransferases, histone demethylases, DNA methyltransferases and DNA demethylases in the cells from different stages of reprogramming from the published literature through www.stemformatics.com) (**Figure 28A**). The factors that showed a significant difference in the expression levels during reprogramming were selected as the candidate genes for the experiment (**Figures 28B and 28C**). Based on their expression patterns, they were categorized into three groups- genes that showed (a) steady increase, (b) steady decrease and (c) biphasic expression levels (**Figure 28D**). The expression of these selected genes was tested in the reprogramming intermediates (**Figure 28E**). The expression profile of a few genes was found to be similar to that observed in published data sets, whereas a few others were different. This was expected because in published datasets, the bulk population of heterogeneous reprogramming cells was analysed at different time points, whereas in this study, flow-sorted cells expressing specific markers were analyzed.

6.3.3. Construction of lentiviral shRNA vectors and estimation of knockdown efficiency

Following recent advances in the design of synthetic shRNAmir stems, Knott *et al.,* optimized the miR-30 backbone for efficient knockdown of target genes (Knott *et al.,* 2014).

A.

MET related genes		
GRHL1	*GRHL2*	*GRHL3*

Lysine Methyl Transferases and related genes			
ASH1L	*PRDM1*	*SETD1A*	*SETMAR*
DOT1L	*PRDM2*	*SETD1B*	*SUV39H1*
EHMT1	*PRDM3*	*SETD2*	*SUV39H2*
EHMT2	*PRDM4*	*SETD3*	*SUV420H1*
EZH1	*PRDM5*	*SETD4*	*SUV420H2*
EZH2	*PRDM6*	*SETD5*	*SYMD1*
MLL	*PRDM7*	*SETD6*	*SYMD2*
MLL2	*PRDM8*	*SETD7*	*SYMD3*
MLL3	*PRDM9*	*SETD8*	*SYMD4*
MLL4	*PRDM10*	*SETD9*	*SYMD5*
MLL5	*PRDM11*	*SETDB1*	*WHSC1*
NSD1	*PRDM12*	*SETDB2*	*WHSC1L1*
	PRDM13		
	PRDM14		
	PRDM15		

Lysine Demethylases			
KDM1A	*KDM4A*	*KDM5C*	*KDM8*
KDM1B	*KDM4B*	*KDM5D*	*JMJD1C*
KDM2A	*KDM4C*	*KDM6A*	
KDM2B	*KDM4D*	*KDM6B*	
KDM3B	*KDM5B*	*KDM7A*	

DNA methylation related genes		
DNMT3A	*TET1*	*TET2*

OTHERS		
BMI1	*JARID2*	*MECOM*

Figure 28: The genes selected in the candidate gene approach for RNAi
(A) The list of genes involved in the reprogramming process which were analyzed in the data sets through Stemformatics.

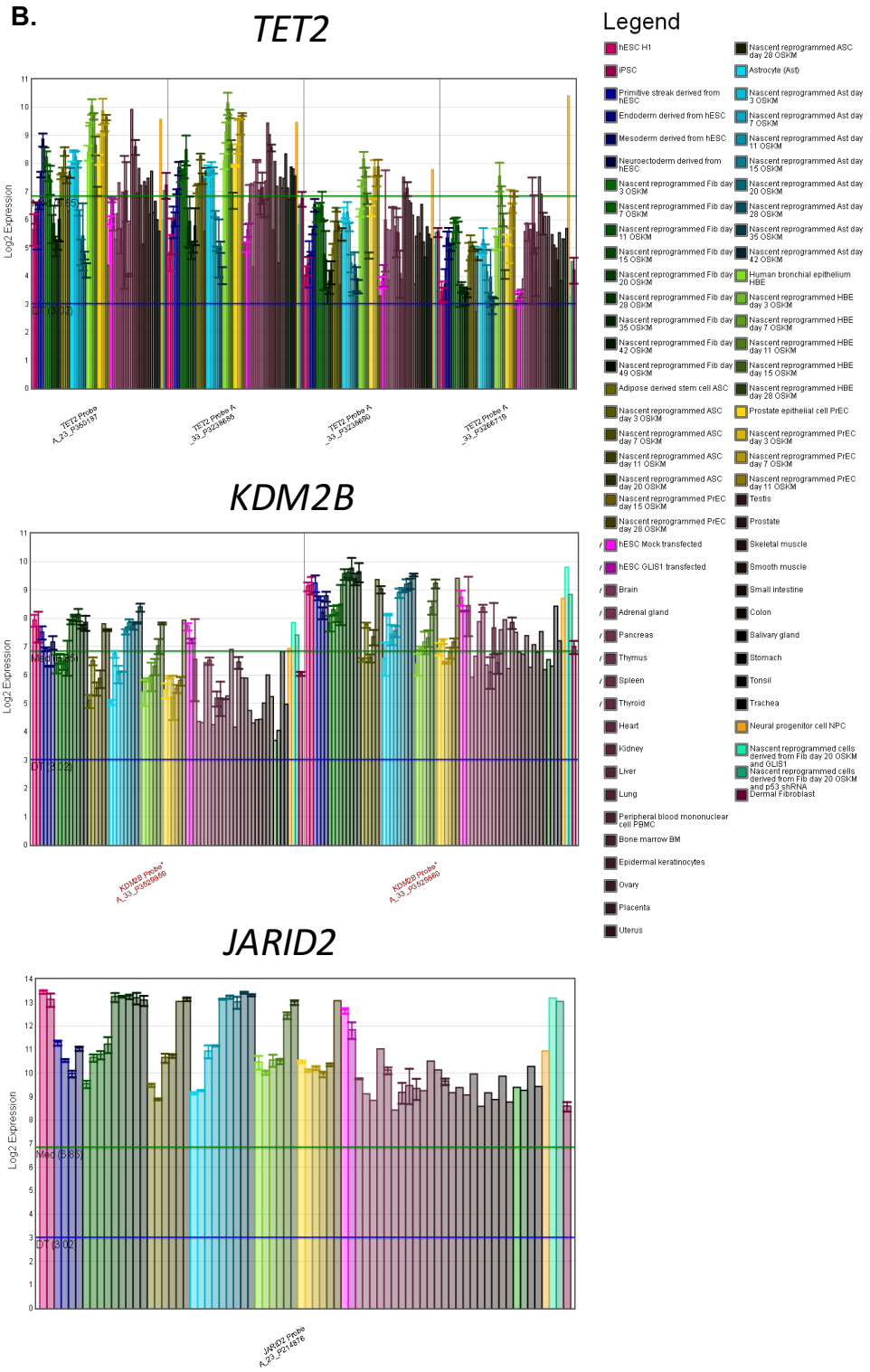


Figure 28 continued: (B) Representative graphs showing the expression profile of genes in the datasets available at Stemformatics.

C.

MET related genes	Lysine methyl transferases and related genes	Lysine Demethylases	DNA methylation genes	Others
GRHL1	*PRDM1*	*KDM1A*	*DNMT3A*	*MECOM*
GRHL2	*PRDM12*	*KDM2A*	*TET1*	*JARID2*
GRHL3	*PRDM14*	*KDM1B*	*TET2*	*BMI1*
	SETD7	*KDM2B*		
	SUV420H2	*KDM4A*		
		KDM6B		
		JMJD1C		

D.

Upregulated	Downregulated	Biphasic
GRHL2	*PRDM12*	*GRHL1*
PRDM14	*SETD7*	*GRHL3*
KDM2B	*KDM1B*	*PRDM1*
KDM6B	*TET2*	*SUV420H2*
DNMT3A	*BMI1*	*KDM1A*
TET1	*MECOM*	*KDM2A*
JMJD1C		*KDM4A*
JARID2		

Figure 28 continued: (C) The list of candidate factors selected for the knockdown experiments. **(D)** Classification of candidate factors based on expression pattern in published datasets.

E.

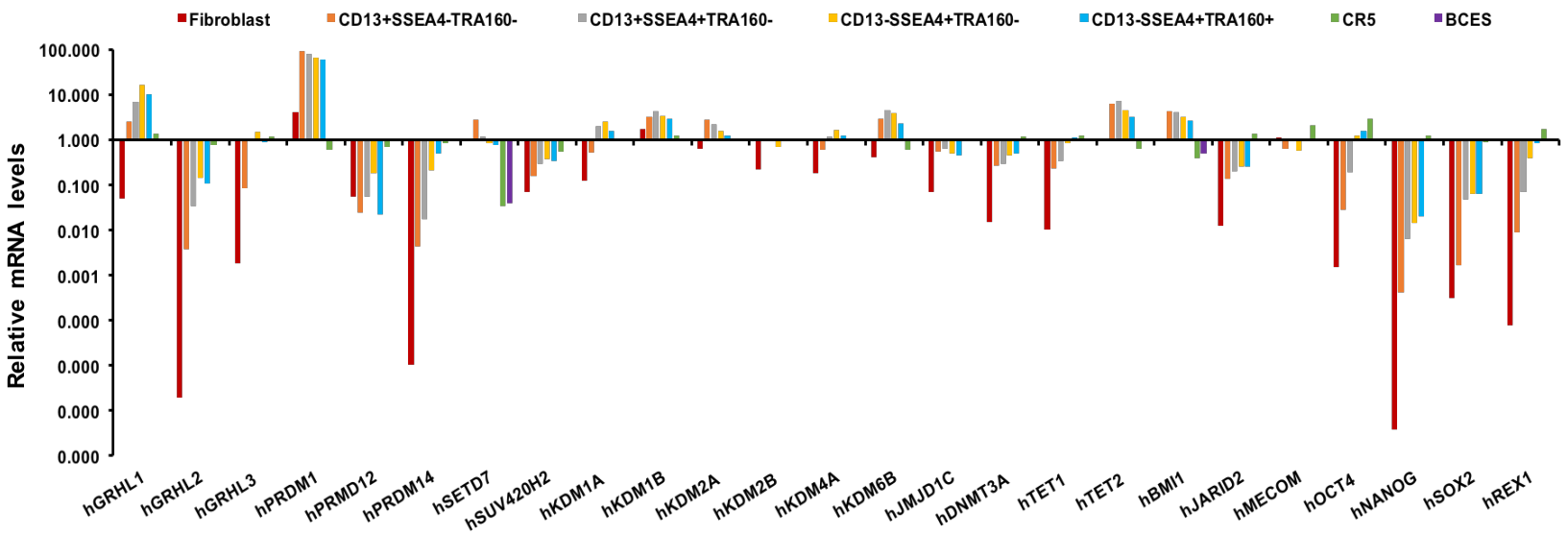


Figure 28 continued: **(E)** Real-time PCR analysis of the genes selected for the knockdown study.

Among several favourable features, they identified a conserved element 3′ of the basal stem as critically required for optimal shRNAmir processing and implement it in an optimized backbone termed “UltramiR,” which strongly increases mature shRNA levels and knockdown efficacy. The shRNA sequences can be easily cloned to these backbones and its combination with sensor based shRNA design rules establishes a validated and accessible platform for generating effective single-copy shRNAs that will facilitate the knockdown studies in the cultured cells.

As the experiments carried out in this thesis identified that hEF1α promoter showed the minimum silencing a lentiviral pZIP-hEF1α plasmid was used for cloning shRNAs. The sensor validated shRNA sequences were obtained from Sherwood (http://sherwood.cshl.edu.). For the estimation of knockdown efficiency, RNA was extracted from transduced HELA or MCF7 cell lines after puromycin selection and real-time PCR was performed to estimate expression levels of target genes. The knockdown efficiencies of the various shRNAs were found to vary from 65%-80% (**Figure 29**).

6.3.4. Identification of the stage-specific role of genes in reprogramming by a candidate gene approach

To identify the role of candidate factors in reprogramming, human adult dermal fibroblasts transduced with lentiviral shRNAs against specific gene were reprogrammed with hSTEMCCA reprogramming vector. The effect of the knockdown of the candidate genes on reprogramming was estimated in terms of their effect on the transition of cells from $CD13^{+}SSEA\text{-}4^{-}TRA\text{-}1\text{-}60^{-}$ fibroblast stage to $CD13^{-}SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{+}$ pluripotent stage, through intermediate $CD13^{+}SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{-}$ and $CD13\text{-}SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{-}$ stages (**Figure 30A**).

The knockdown of GRHL2, SETD7 and KDM2A were found to adversely affect the downregulation of CD13. The percentage of cells that attained CD13 downregulation ($CD13^{-}$) with knockdown of these genes were 16% to 27%, compared to 50% in the control cells on day 14. This adverse effect on the initial stage of reprogramming was reflected in the induction of TRA-1-60 expression. There were only 3% to 6% of the $TRA\text{-}1\text{-}60^{+}$ cells when

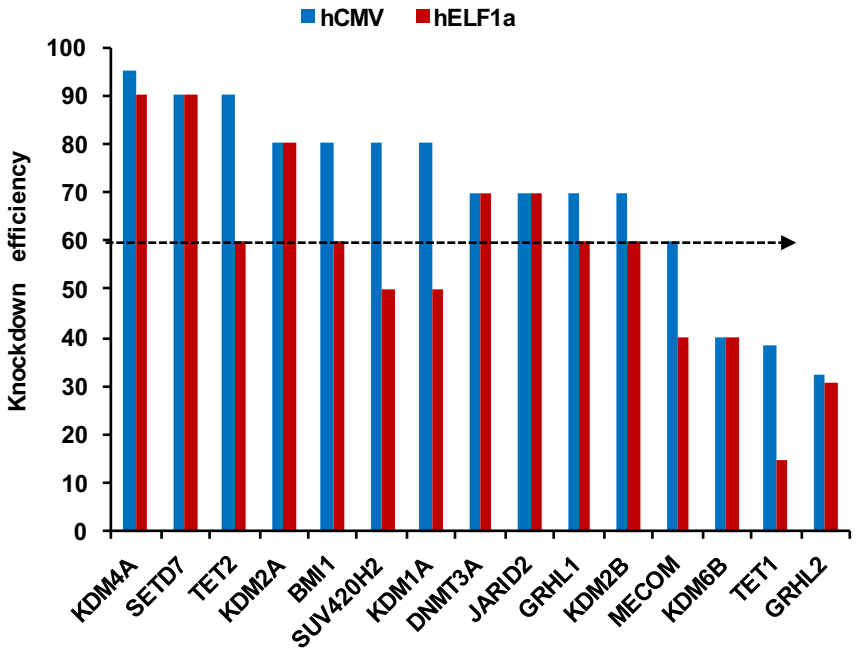


Figure 29: Estimation of knockdown efficiency of pZIP-hEF1α shRNA vector
Percentage of decrease in relative level of target gene mRNA transcripts in HeLA or MCF7 cell lines, transduced with gene specific shRNA, estimated based on real time PCR. HeLA or MCF7 cell lines transduced with pZIP-hEF1α empty vector served as control.

these three factors were knocked down while the control had 11% of TRA-1-60$^+$ cells. This indicated that GRHL2, SETD7 and KDM2A play important roles in the downregulation of somatic cells specific programme (**Figure 30B**).

TET2 knockdown favoured downregulation of CD13 and induction of SSEA-4, with 78% CD13$^-$ cells and 75% SSEA-4$^+$ cells on day 14, compared to 50% and 47% in the controls. However, 44% of the cells remained in the intermediate CD13$^-$SSEA-4$^+$TRA1-60$^-$ state on day 18 without transitioning into the CD13$^-$SSEA-4$^+$TRA1-60$^+$ state. This indicated that TET2 plays a stage-specific role in reprogramming as a negative regulator in the initial stages and a positive regulator in the later stages of reprogramming as it has a critical role in favouring the transition of SSEA-4$^+$ TRA-1-60$^-$ state to SSEA-4$^+$TRA-1-60$^+$ state (**Figure 30C**).

The knockdown of JARID2 and KDM4A enhanced the downregulation of CD13 and induction of SSEA-4$^+$ and TRA-1-60$^+$ cells. On day 14, the percentage of CD13$^-$, SSEA-4$^+$ and TRA160$^+$ cells were 49%, 70% and 18% on JARID2 knockdown and 54%, 57% and 21% on KDM4A knockdown respectively, compared to 23%, 48% and 11% in the controls. Among the SSEA-4$^+$ cells, 71% and 65% were CD13$^-$ with JARID2 and KDM4a knockdown, respectively, compared to 47% in the control on day 14. Thus, knockdown of JARID2 and KDM4a was found to accelerate reprogramming process by favouring downregulation of CD13 and upregulation of SSEA-4 and TRA-1-60 (**Figure 30D**). Moreover, the efficiency of reprogramming was also increased with downregulation of these genes leading to a higher percentage of cells achieving a CD13$^-$SSEA4$^+$TRA160$^+$ state. CD13$^-$SSEA4$^+$TRA160$^+$ constituted about 18% and 21% with JARID knockdown and 21% and 22% with KDM4a knockdown on day 14 and day 18, respectively, compared to 11% and 16% in the control.

Knockdown of DNMT3A, KDM1A and TET1 favoured the induction of SSEA-4 by day 14 and TRA-1-60 by day 18 (**Figure 30E**). The percentage of SSEA-4$^+$ cells was 58% - 60% with the knockdown of DNMT3A, KDM1A and TET1, compared to 50% in the control on day 14. When these genes were knocked down, even though the overall percentage of CD13$^-$ remained the same, the downregulation of CD13 was accelerated within the SSEA4$^+$ cells. Among the SSEA-4$^+$ cells, about 57% - 67% were CD13$^-$ when these genes were

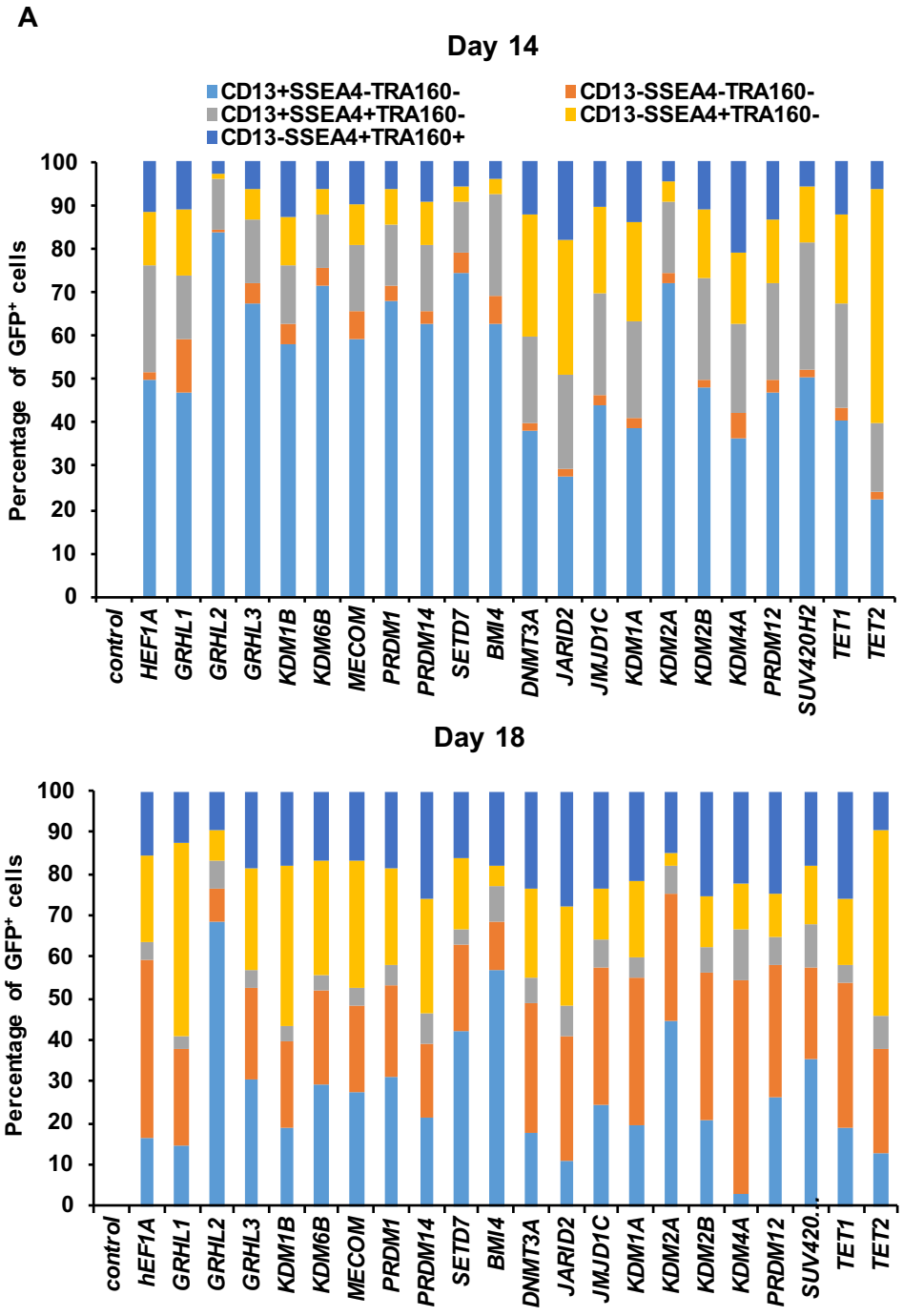
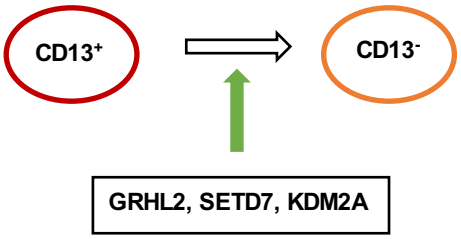


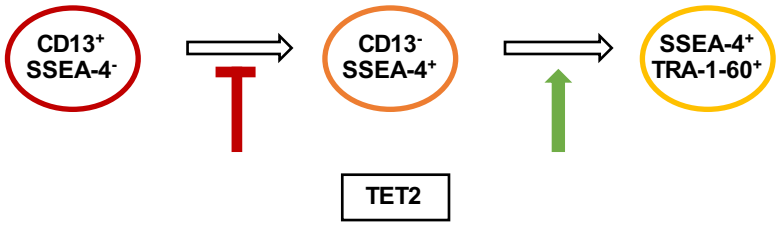
Figure 30: Effect of candidate gene of knock-down on reprogramming efficiency

The percentage of reprogramming cells expressing specific combination of markers, $CD13^{+}SSEA\text{-}4^{-}TRA\text{-}1\text{-}60^{-}$, $CD13^{-}SSEA\text{-}4^{-}TRA\text{-}1\text{-}60^{-}$, $CD13^{+}SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{-}$, $CD13^{-}SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{-}$ and $CD13^{-}SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{+}$ within the GFP^{+} cell fraction on day 14 and day 18 following hSTEMCAA infection of human adult dermal fibroblasts.

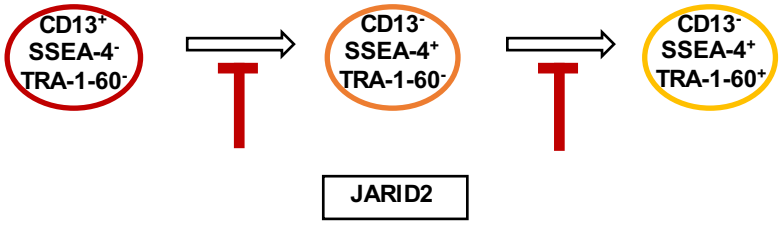
B.



C.



D.



E.

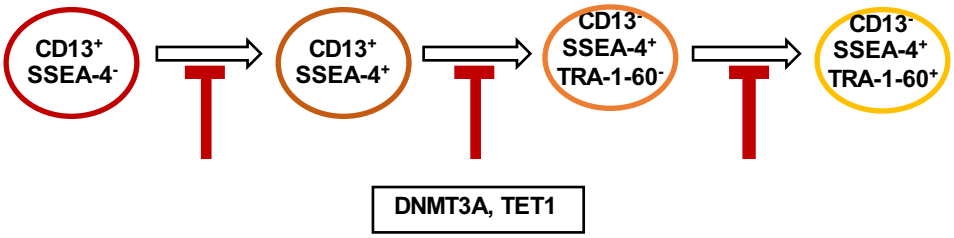


Figure 30 continued: (**B to E**) The schematic of the role of the identified regulators of reprogramming in the transition of cells from one molecular state to another during reprogramming. The expression kinetics of markers, CD13, SSEA-4 and TRA-1-60, were used to identify the transition stages on days 14 and 18 of reprogramming. Red and green lines represent negative and positive regulatory roles of candidate factors, repectively.

knocked down, compared to 47% in controls on day 14. On day 18, the percentage of $CD13^-$ $SSEA4^+TRA160^+$ cells formed from these cells were higher at 23% - 26%, compared to the control (15%). Thus DNMT3A, KDM1A and TET1 were found to play a critical role throughout reprogramming process.

The knockdown of GRHL1 and KDM1B brought about an increase in the percentage of $SSEA4^+$ on day 18, whereas no such change was observed on day 14. The percentage of $SSEA4^+$ cells was found to be increased up to 47% and 39% by GRHL1 and KDM1B shRNAs on day 18, compared to 20% in the control. Despite the increase in the $SSEA4^+$ cells, the percentage of $CD13^-SSEA4^+TRA160^+$ cells remained similar to the control.

The knockdown of PRDM14, KDM2B and PRDM12 did not show any significant effect on reprogramming on day 14. However, on day 18, the percentage of $CD13^-$ $SSEA4^+TRA160^+$ cells were found to be higher at 25-26%. This indicated their role is in the late stage of reprogramming where they promote the formation of $SSEA4^+$TRA-1-60^+ cells. With the knockdown of BMI1 and SUV420H2, even though the relative percentage of $CD13^-$ $SSEA4^+TRA160^+$ formed was similar to that with control, knockdown was found to adversely affect cell proliferation and overall formation of iPSC colonies. Knockdown of MECOM and PRDM1 did not have any significant effect on reprogramming process.

Thus, based on the effect of candidate gene knockdown on the transition of cells from one stage to another, the reprogramming stage in which each factor is involved could be identified.

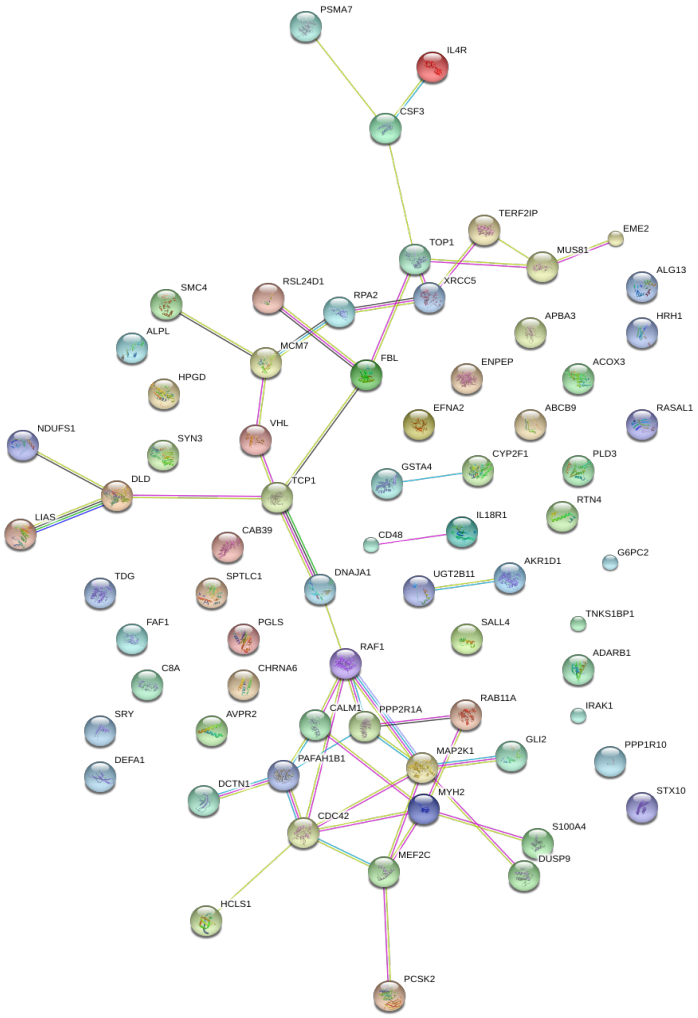
6.3.5. Identification of signaling pathways associated with different stages of reprogramming

To identify the signaling pathways associated with the cells that are unreprogrammed, partially reprogrammed and fully reprogrammed, comprehensive shRNA library screening by RNAi followed by high throughput sequencing was performed. hADFs transduced with shRNA library against 5000 signaling pathway genes were subjected to reprogramming. The cells belonging to three stages, defined by expression profile pluripotency markers SSEA-4 and TRA-1-60, namely SSEA-4^-TRA-1-60^- unreprogrammed stage, SSEA-4^+TRA-1-60^- partially reprogrammed stage and SSEA-4^+TRA-1-60^+ pluripotent

stage were analyzed to identify the integrated shRNAs. The cells were isolated in the third week of reprogramming to ensure that the majority of the cells have reached their final molecular state and to minimize the proportion of cells that are in the process of completing reprogramming. For normalization, hADFs transduced with the shRNA library before reprogramming that represented the entire pool of shRNAs was used. After next generation sequencing, the adapters were trimmed, and the trimmed reads were used for statistical calculations using T test to calculate the *p* values of the shRNA genes that are depleted or enriched in the different fractions of reprogramming cells and the control. The shRNAs genes that showed a *p* value less than 0.01 was chosen as significant and these hits were considered as those that were significantly depleted or enriched in each fraction of the reprogramming cells, compared to the control. The flow sorted fractions SSEA-4$^-$TRA-1-60$^-$ unreprogrammed cells, SSEA-4$^+$TRA-1-60$^-$ cells that represent the intermediate stage and SSEA-4$^+$TRA-1-60$^+$ cells that represent pluripotent stage showed 95, 170 and 164 significant hits, respectively. To identify the pathways to which these genes belong, the genes were analyzed through String-db.org and the top 6 pathways with the least false discovery rate were chosen (**Figures 31A, 31B and 31C**).

In SSEA-4$^-$TRA-1-60$^-$ fibroblast like stage, shRNAs targeting genes involved in Neurotrophin, GnRH, Renal cell carcinoma, Ras signaling and cGMP-PKG pathways were enriched (**Figure 31A**). The shRNAs enriched in SSEA-4$^+$TRA-1-60$^-$ cells were found to downregulate genes associated with pathways related to metabolism, oxytocin signaling, Ras, Oocyte meiosis, PI3K-Akt and purine metabolism (**Figure 31B**). The shRNAs against pathways related to metabolism, steroid hormone biosynthesis, amino acid biosynthesis, oocyte meiosis, Huntington's disease and Ras signaling were found to be enriched in SSEA-4$^+$TRA-1-60+ cells (**Figure 31C**). Thus, using an RNAi screen experiment for signaling pathway genes, specific pathways that are involved in the different stages of reprogramming could be identified.

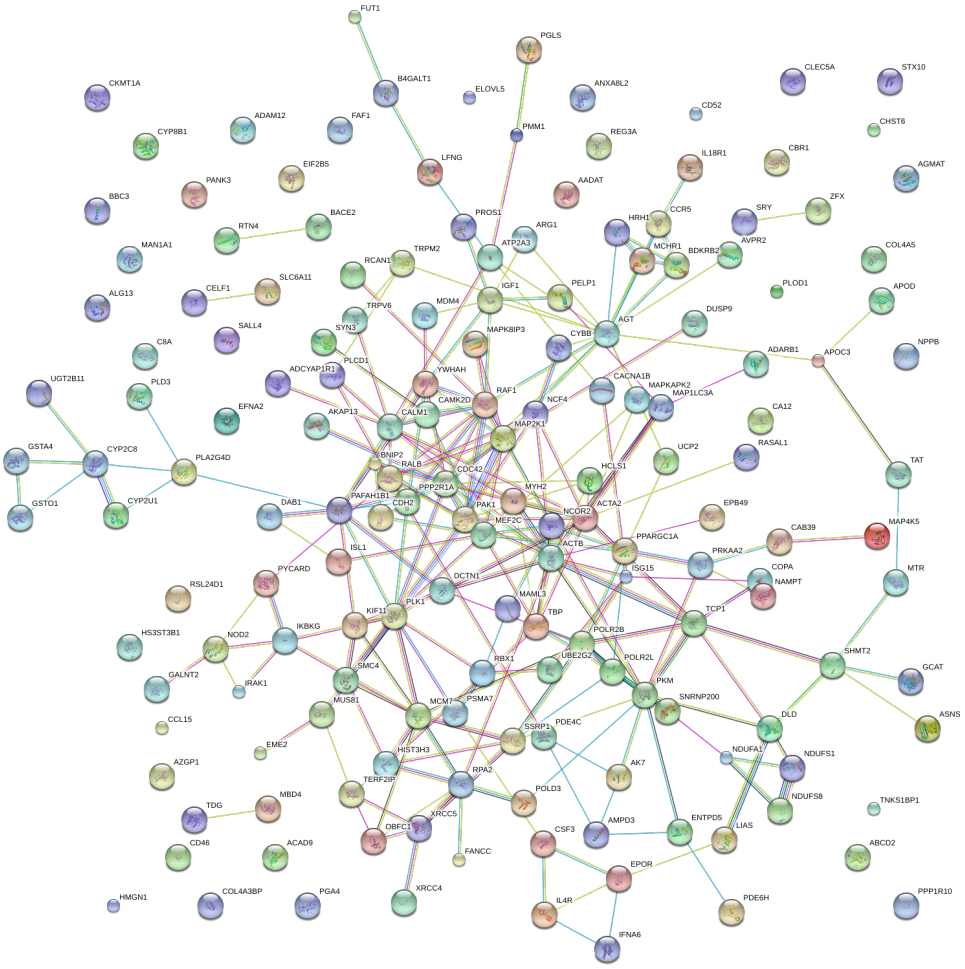
A. **SSEA-4$^-$TRA-1-60$^-$**



Pathway ID	Pathway description	Count in gene set	False discovery rate
4722	Neurotrophin signaling pathway	6	0.0046
4912	GnRH signaling pathway	5	0.00762
4010	MAPK signaling pathway	7	0.0161
5211	Renal cell carcinoma	4	0.0161
4014	Ras signaling pathway	6	0.0317
4022	cGMP-PKG signaling pathway	5	0.034

Figure 31: Signaling pathways associated with different stages of reprogramming
(A) Pathway analysis of significant hits identified in SSEA-4$^-$TRA160$^-$ reprogramming cells on day 15-17 compared to the shRNA library transduced fibroblasts.

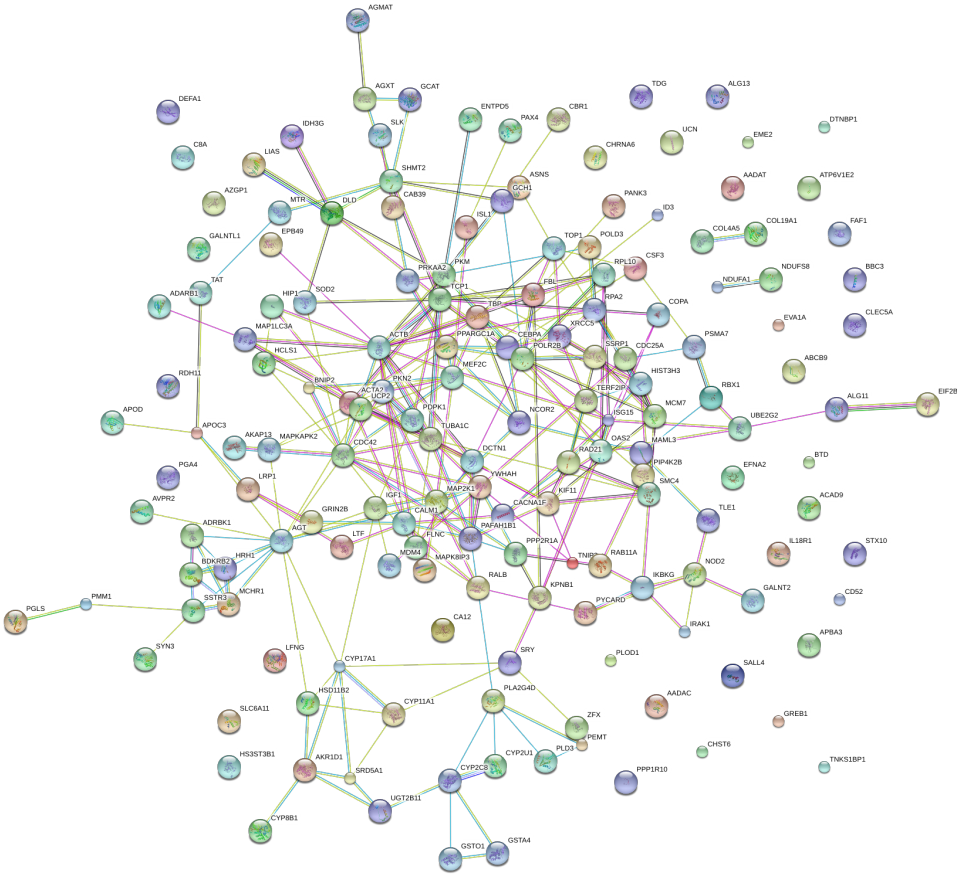
B. **SSEA-4^{+}TRA-1-60^{-}**



Pathway ID	Pathway description	Count in gene set	False discovery rate
1100	Metabolic pathways	34	2.10E-08
4921	Oxytocin signaling pathway	10	7.53E-05
4014	Ras signaling pathway	11	0.000178
4114	Oocyte meiosis	8	0.000178
4151	PI3K-Akt signaling pathway	13	0.000249
230	Purine metabolism	9	0.000341

Figure 31 continued: (B) Pathway analysis of significant hits identified in SSEA-4^{+}TRA160^{-} reprogramming cells on day 15-17 compared to the shRNA library transduced fibroblasts.

C. **SSEA-4$^+$TRA-1-60$^+$**



Pathway ID	Pathway description	Count in gene set	False discovery rate
1100	Metabolic pathways	35	1.56E-09
140	Steroid hormone biosynthesis	6	0.000544
1230	Biosynthesis of amino acids	6	0.00117
4114	Oocyte meiosis	7	0.00117
5016	Huntington s disease	9	0.00117
4014	Ras signaling pathway	9	0.00308

Figure 31 continued: (C) Pathway analysis of significant hits identified in SSEA-4$^+$TRA160$^+$ reprogramming cells on day 15-17 compared to the shRNA library transduced fibroblasts.

6.4. Discussion

Reprogramming of somatic cells to iPSCs is a multi-step process that involves the transition of cells from one state to another to achieve pluripotency. Even with additional manipulations, the efficiency of OSKM mediated reprogramming could be improved only to a very small extent. The complexity of the reprogramming process and the technical limitations of the current mechanistic study methods are the major factors that limit the elucidation of the molecular mechanism of reprogramming.

Since the transcriptional activity of the promoters in the lentiviral vectors depends on cell types, the type of the promoters used for manipulating gene expression can significantly affect the reprogramming outcome. As there are no systematic studies on the activity of the lentiviral vector promoters, in this study, seven different promoters were tested in the reprogramming cells to identify the most efficient promoter that ensures significant knockdown of target genes throughout the reprogramming process. CMV promoter was found to be transcriptionally stronger in fibroblasts, and the cells at the initial stages of reprogramming and therefore this promoter is more suitable for studying the role of target genes in the earlier stages of reprogramming. EF1α promoter was found to remain transcriptionally active without transgene silencing in the cells at all the stages of reprogramming, and it showed very high transcriptional activity and therefore this promoter was chosen for achieving efficient knockdown in this study.Thus, this study is a first systematic evaluation of a promoter that has identified the promoters suited for achieving desired extent and pattern of gene regulation during reprogramming.

The study in this thesis combined strategies from the previous studies to understand the role of factors associated with MET, epigenetic modification, and signaling pathways. In previous studies aiming at elucidating the molecular mechanisms of reprogramming, the role of new factors were identified following their overexpression and knockdown by (i) estimating the number of iPSC colonies formed, (ii) identifying the shRNAs enriched or depleted in the TRA-1-60^{-} and TRA-1-60^{+} cells or (iii) performing a transcriptome analysis of the reprogramming intermediates and reprogrammed cells. In this thesis, the assessment was

based on either calculating the percentage of cells belonging to specific reprogramming stages defined by the expression of fibroblast marker CD13, and early and late pluripotency markers, SSEA-4 and TRA-1-60 or identifying the shRNAs enriched or depleted in these fractions. The assessment of the effect of gene knockdown on the formation of reprogramming intermediates offered the advantage of identifying the stage in which the genes are actively involved.

The role of MET associated genes, GRHL1, GRHL2 and GRHL3, in reprogramming has not been explored so far. When the effect of their knockdown was analysed, it was found that GRHL2 significantly affected reprogramming efficiency by preventing the downregulation of CD13, thereby indicating their role in MET. GRHL1 and GRHL3 did not show significant effect on formation of cells at different stages. However, it was found that knockdown of GRHL1 increases the percentage of SSEA-4^{+} cells.

Knockdown of H3K27 methyltransferases belonging to the PRC1 PcG complex, BMI1, was found to substantially reduce the number of iPSC colonies obtained from human fibroblasts (Onder *et al.,* 2012). This effect was suggested to be due to the role of PcG complex in the transcription repression of the developmental genes regulated by OSN in the pluripotent stem cells. Consistent with this, this thesis found that the knockdown BMI1 reduces the reprogramming efficiency, and this was mainly due to inhibition of cell proliferation. Also, it was found that BMI knockdown did not affect the pace of reprogramming.

It was previously reported that inhibition of H3K36 demethylases, KDM2A and KDM2B decreases reprogramming efficiency. Consistent with this observation, the knockdown of KDM2A brought about a decrease in hiPSC reprogramming efficiency, and this was found to be associated with the inability to downregulate fibroblast programme in the absence of KDM2A. On the contrary, knockdown of KDM2B was not found to have any effect on the pace of hiPSC reprogramming in the initial days but was found to increase the percentage of TRA-1-60+ cells at the late stage of reprogramming. This was unexpected because KDM2B have been found to play a prominent role during the initiation phase of

reprogramming in miPSC reprogramming where it enhances the activation of early responsive genes through binding and demethylating their promoters (Liang *et al.,* 2012).

Downregulation of *DNMT3a* in human and mouse fibroblasts did not have a significant effect on reprogramming (Pawlak *et al.,* 2011). However, in this study, an increase in the pace and efficiency of reprogramming was observed with respect to the downregulation of CD13 and induction of SSEA-4 and TRA-1-60. A similar reprogramming outcome was observed with the TET1 knockdown. This observation is contrary to previous reports on miPSC reprogramming that TET1 can replace OCT4 and overexpression of TET1 can enhance reprogramming efficiency (Chen *et al.,* 2013; Gao *et al.,* 2013). It has been previously reported that knockdown of TET2 adversely affects reprogramming efficiency in MEFs (Hu *et al.,* 2014). Consistent with this, it was found in this thesis study that TET2 knockdown in human fibroblasts considerably decreased the reprogramming efficiency. TET2 knockdown favours CD13 knockdown and SSEA-4 induction, but inhibits TRA-1-60 induction, indicating its role as a positive regulator in the initial stages and as a negative regulator in the late stages. The positive role of TET2 in the late stage of reprogramming is supported by previous studies that have reported a close association between TET2 with pluripotency marker NANOG (Silva *et al.,* 2009; Costa *et al.,* 2013).

Through RNAi screen, the signaling pathways closely linked to partially reprogrammed state, fully reprogrammed state and unreprogrammed states were identified. In the cells at partially and fully reprogrammed state, there is a significant enrichment of shRNAs targeting metabolic genes. This observation highlights the importance of metabolic shift during reprogramming for pluripotency acquisition. Ras signaling found to be associated with all three reprogramming states and pathway was reported earlier to enhance reprogramming efficiency in miPSC reprogramming by improving cell proliferation (Kwon *et al.,* 2015).

Chapter 7

SUMMARY & CONCLUSIONS

- The first goal of the project was to standardize efficient methods for the generation of hiPSCs. OCT4, SOX2, KLF4 AND c-MYC-mediated reprogramming of fibroblasts by different gene delivery methods using retroviruses, lentiviruses, episomal plasmids and Sendai viruses could be achieved. All the methods for the generation, identification, isolation and characterization of hiPSCs could be successfully established. The hiPSC lines were fully characterized by their expression of pluripotency markers, acquisition of hypomethylated *OCT4* and *NANOG* promoters, *in vitro* differentiation potential by EB formation and *in vivo* differentiation potential of teratoma formation. The difference in the reprogramming efficiencies achieved with different methods was estimated to assess their suitability in the subsequent experiments. This was the first study for establishing successful reprogramming methods in India.
- During reprogramming of fibroblasts, it was observed that the initial culture conditions and the culture media of the somatic cells influence the reprogramming efficiency. Fibroblasts maintained in the medium that favoured higher proliferation rate, transduction efficiency, and transgene expression levels exhibited faster kinetics and higher efficiency of reprogramming. Based on these observations, it was confirmed that preconditioning of somatic cells is a major factor for that determines efficient reprogramming.
- Even though hiPSC identification based on morphology, pluripotency marker expression, and retroviral transgene silencing are widely employed a comparison on the suitability of these three markers in identifying true hiPSCs has not been well established. By systematic monitoring of morphology, pluripotency marker expression and retroviral RFP silencing in the colonies generated during reprogramming, a significant correlation was found between the hESC-like morphology of the hiPSC clones and retroviral transgene silencing which was measured by a retrovirally expressed fluorescence protein. These clones expressed all the pluripotency markers. A few colonies that lacked hESC-like morphology showed RFP silencing and/or pluripotency marker expression. Through the routine

observation of reprogramming cells, it was found that emerging hiPSC colonies can be easily identified based on their morphological changes during reprogramming. They emerge dislodging feeder cells around them radially forming flat colonies with defined boundaries and symmetric shape. It was observed that the inclusion of retrovirally expressed RFP as a marker could greatly aid in the identification of such hiPSC colonies as they appeared as patches of RFP^- colonies surrounded by RFP^+ non-iPSCs. Thus by describing the kinetics of acquisition of hESC-like morphology, pluripotency marker expression and retroviral RFP silencing by the reprogramming cells, it was shown that morphology of hiPSCs serves as a more reliable marker in the identification of hiPSC clones. Through this study, it was shown that retroviral transgene silencing is a highly reliable marker for isolation of pluripotent cells on reprogramming dish although morphology is the best criterion for the isolation of pluripotent clones.

- Sorting of reprogramming cells based on a fibroblast marker, CD13, and pluripotency markers, SSEA-4 and TRA-1-60, could facilitate isolation of the cells at different stages of reprogramming. Gene expression analysis and shRNA library screening assays performed on these sorted cells confirmed the suitability of these markers in defining the stages of reprogramming.
- The role of Fanconi Anemia pathway genes in pluripotency and reprogramming was investigated using an inducible gene expression system for complementing defective *FANCA* gene in patient-derived cells. HiPSC lines were derived from FANCA deficient fibroblast complemented with functional *FANCA* gene under an inducible promoter, and the lines were characterized for their pluripotency. The formation of teratoma by the hiPSC line, without FANCA induction during its *in vivo* growth in SCID mice, confirmed their pluripotency. By conditional complementation of FANCA in these hiPSC lines, the functional significance of Fanconi Anemia pathway could be assessed in these cells. In the complete absence of *FANCA* gene and functional Fanconi anemia pathway, hiPSCs failed to survive, thereby mimicking the disease

phenotype. These hiPSC lines can serve as an excellent tool for elucidating the role of FA pathway in pluripotency and disease pathogenesis.

- In the search for promoters of viral vectors that do not get silenced during the entire course of reprogramming, lentiviral vectors with different promoters that drive the expression of GFP were evaluated. It was found that mouse EF1α and human EF1α promoters did not exhibit transgene silencing, and they are the most suitable promoters for consistent knockdown and overexpression experiments in reprogramming.
- The role of GRHL1, GRHL2 and GRHL3 was evaluated based on the effect of their knockdown on the formation of cells at different stages of reprogramming. Analysis of the cells sorted from various stages of reprogramming revealed that GRHL2 shows a constant increase in expression as the cells underwent reprogramming, and this expression pattern was similar to that of pluripotency markers. Knockdown of GRHL2 adversely affected downregulation of fibroblast marker, CD13 and formation of $CD13^{-}SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{+}$ reprogrammed cells, leading to an overall decrease in reprogramming efficiency, indicating their role in MET during reprogramming.
- The role of the epigenetic factors was evaluated based on the effect of their knockdown on the formation of cells at different stages of reprogramming. The epigenetic factors which showed variations in their expression levels during reprogramming were selected after bioinformatics analysis of the gene expression data available from previous publications as well as the gene expression analysis on different cell fractions obtained by flow sorting in this thesis. By knocking down these genes separately with shRNAs, their involvement in reprogramming could be assessed.
- For elucidating the role of various signaling pathways in the reprogramming process, human dermal fibroblasts were infected with lentiviral vectors to express shRNAs against 5000 human signaling pathway genes and then they were reprogrammed with OSKM. The $SSEA\text{-}4^{-}TRA\text{-}1\text{-}60^{-}$, $SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{-}$ and $SSEA\text{-}4^{+}TRA\text{-}1\text{-}60^{+}$

representing different stages of reprogramming were sorted from the reprogramming cells and were analyzed for incorporation of shRNAs by massively parallel sequencing. Based on the specific shRNAs enriched or depleted in the sorted fractions of cells, specific signaling pathways closely linked with each stage of reprogramming could be identified.

- In this study, the stage-specific role of factors in reprogramming could be determined. GRHL2, SETD7 and KDM2A were found to play a critical role in downregulation of fibroblast specific gene expression, whereas PRDM14, PRDM12 and KDM2B were found to act in the late stages inhibiting the induction of TRA-1-60 expression. TET2 was found to be a negative regulator of CD13 downregulation and SSEA-4 expression induction, but a positive regulator for TRA-1-60 expression induction. Thus, the findings from this study could improve our understanding of the molecular mechanism of reprogramming and the knowledge gained will be helpful in developing reprogramming strategies to achieve higher efficiency.

BIBLIOGRAPHY

Aasen T, Raya A, Barrero MJ, Garreta E, Consiglio A, Gonzalez F, Vassena R, Bilić J, Pekarik V, Tiscornia G, Edel M, Boué S, Izpisúa Belmonte JC (2008) Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat Biotechnol 26:1276–84.
Abujarour R, Valamehr B, Robinson M, Rezner B, Vranceanu F, Flynn P (2013) Optimized surface markers for the prospective isolation of high-quality hiPSCs using flow cytometry selection. Sci Rep 3:1179.
Agoulnik AI, Lu B, Zhu Q, Truong C, Ty MT, Arango N, Chada KK, Bishop CE (2002) A novel gene, Pog, is necessary for primordial germ cell proliferation in the mouse and underlies the germ cell deficient mutation, gcd. Hum Mol Genet 11:3047–53.
Alter BP, Caruso JP, Drachtman RA, Uchida T, Velagaleti GVN, Elghetany MT (2000) Fanconi anemia: Myelodysplasia as a predictor of outcome. Cancer Genet Cytogenet 117:125–131.
Ang Y, Tsai S, Lee D, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B, Wang J, Rendl M, Bernstein E, Schaniel C, Lemischka IR (2011) Wdr5 Mediates Self-Renewal and Reprogramming via the Embryonic Stem Cell Core Transcriptional Network. Cell 145:183–197.
Auerbach AD (2009) Fanconi anemia and its diagnosis. Mutat Res - Fundam Mol Mech Mutagen 668:4–10.
Auerbach AD (2015) Diagnosis of Fanconi Anemia by diepoxybutane analysis. Curr Protoc Hum Genet 2015:8.7.1–8.7.17.
Avior Y, Sagi I, Benvenisty N (2016) Pluripotent stem cells in disease modelling and drug discovery. Nat Rev Mol Cell Biol 17:170–182.
Bakker ST, de Winter JP, te Riele H (2013) Learning from a paradox: recent insights into Fanconi anaemia through studying mouse models. Dis Model Mech 6:40–7.
Bakker ST, van de Vrugt HJ, Rooimans MA, Oostra AB, Steltenpool J, Delzenne-Goette E, van der Wal A, van der Valk M, Joenje H, te Riele H, de Winter JP (2009) Fancm-deficient mice reveal unique features of Fanconi anemia complementation group M. Hum Mol Genet 18:3484–3495.
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.
Bandaru V, Sunkara S, Wallace SS, Bond JP (2002) A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to Escherichia coli endonuclease VIII. DNA Repair (Amst) 1:517–529.
Banito A, Rashid ST, Acosta JC, Banito A, Rashid ST, Acosta JC, Li S, Pereira CF, Geti I, Pinho S, Silva JC, Azuara V (2009) Senescence impairs successful reprogramming to pluripotent stem cells service Senescence impairs successful reprogramming to pluripotent stem cells. 2134–2139.
Barker S, Weinfeld M, Murray D (2005) DNA – protein crosslinks : their induction , repair , and biological consequences. 589:111–135.
Barrero MJ, Sese B, Kuebler B, Bilic J, Boue S, Martí M, Izpisua Belmonte JC (2013) Macrohistone Variants Preserve Cell Identity by Preventing the Gain of H3K4me2 during Reprogramming to Pluripotency. Cell Rep 3:1005–1011.
Barroca V, Mouthon MA, Lewandowski D, Brunet de la grange P, Gauthier LR, Pflumio F, Boussin FD, Arwert F, Riou L, Allemand I, Romeo PH, Fouchet P (2012) Impaired functionality and homing of fancg-deficient hematopoietic stem cells. Hum. Mol. Genet. 21:121–135.
Bayart E, Cohen-Haguenauer O (2013) Technological overview of iPS induction from human adult somatic cells. Curr Gene Ther 13:73–92.
Bhagwat N, Olsen AL, Wang AT, Hanada K, Stuckert P, Kanaar R, D'Andrea A, Niedernhofer LJ, McHugh PJ (2009) XPF-ERCC1 participates in the Fanconi anemia pathway of cross-link repair. Mol Cell Biol 29:6427–37.
Bhutani N, Decker MN, Brady JJ, Bussat RT, Burns DM, Corbel SY, Blau HM (2013) A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. FASEB J 27:1107–1113.
Bilic J, Belmonte JCI (2012) Induced Pluripotent Stem Cells Versus Embryonic Stem Cells : Close Enough or Yet Too Far Apart? Stem Cells 30:33–41.

Bogliolo M, Surrallés J (2015) Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. Curr Opin Genet Dev 33:32–40.
Bonnet D, Dick JE (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 3:730–737.
Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, Jaenisch R (2008) Sequential Expression of Pluripotency Markers during Direct Reprogramming of Mouse Somatic Cells. Cell Stem Cell 2:151–159.
Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, Oudenaarden A Van, 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 DA, Jaenisch R (2013) Mechanisms and models of somatic cell reprogramming. Nat Rev Genet 14:427–439.
Buganim Y, Markoulaki S, Van Wietmarschen N, Hoke H, Wu T, Ganz K, Akhtar-Zaidi B, He Y, Abraham BJ, Porubsky D, Kulenkampff E, Faddah DA, Shi L, Gao Q, Sarkar S, Cohen M, Goldmann J, Nery JR, Schultz MD, Ecker JR, Xiao A, Young RA, Lansdorp PM, Jaenisch R (2014) The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. Cell Stem Cell 15:295–309.
Burridge PW, Zambidis ET (2013) A Review of the Methods for Human iPSC Derivation. In: Pluripotent Stem Cells: Methods and Protocols. pp 149–161
Butturini a, Gale RP, Verlander PC, Adler-Brecher B, Gillio a P, Auerbach a D (1994) Hematologic abnormalities in Fanconi anemia: an International Fanconi Anemia Registry study. Blood 84:1650–1655.
Cao L, Tan L, Jiang T, Zhu X, Yu J (2015) Induced Pluripotent Stem Cells for Disease Modeling and Drug Discovery in Neurodegenerative Diseases. Mol Neurobiol 52:244–255.
Carey BW, Markoulaki S, Beard C, Hanna J, Jaenisch R (2010) Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nat Methods 7:56–9.
Carreau M (2004) Not-so-novel phenotypes in the Fanconi anemia group D2 mouse model. Blood 103:2430.
Carreau M, Gan OI, Liu L, Doedens M, McKerlie C, Dick JE, Buchwald M (1998) Bone marrow failure in the Fanconi anemia group C mouse model after DNA damage. Blood 91:2737–2744.
Casado JA, Callen E, Jacome A, Rio P, Castella M, Lobitz S, Ferro T, Munoz A, Sevilla J, Cantalejo A, Cela E, Cervera J, Sanchez-Calero J, Badell I, Estella J, Dasi A, Olive T, Ortega JJ, Rodriguez-Villa A, Tapia M, Molines A, Madero L, Segovia JC, Neveling K, Kalb R, Schindler D, Hanenberg H, Surralles J, Bueren JA (2007) A comprehensive strategy for the subtyping of patients with Fanconi anaemia: conclusions from the Spanish Fanconi Anemia Research Network. J Med Genet 44:241–249.
Ceccaldi R, Parmar K, Mouly E, Delord M, Kim JM, Regairaz M, Pla M, Vasquez N, Zhang QS, Pondarre C, Peffault De Latour R, Gluckman E, Cavazzana-Calvo M, Leblanc T, Larghero J, Grompe M, Socie´ G, D'Andrea AD, Soulier J (2012) Bone marrow failure in fanconi anemia is triggered by an exacerbated p53/p21 DNA damage response that impairs hematopoietic stem and progenitor cells. Cell Stem Cell 11:36–49.
Chan EM, Ratanasirintrawoot S, Park I-H, Manos PD, Loh Y-H, Huo H, Miller JD, Hartung O, Rho J, Ince T a, Daley GQ, Schlaeger TM (2009) Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. Nat Biotechnol 27:1033–1037.
Chandra S, Levran O, Jurickova I, Maas C, Kapur R, Schindler D, Henry R, Milton K, Batish SD, Cancelas JA, Hanenberg H, Auerbach AD, Williams DA (2005) A rapid method for retrovirus-mediated identification of complementation groups in Fanconi anemia patients. Mol Ther 12:976–984.
Chang CW, Lai YS, Pawlik KM, Liu K, Sun CW, Li C, Schoeb TR, Townes TM (2009) Polycistronic lentiviral vector for "hit and run" reprogramming of adult skin fibroblasts to induced pluripotent stem cells. Stem Cells 27:1042–1049.
Chang K, Huang A, Hirata RK, Wang P, Russell DW, Dc W, Wang P, Russell DW (2011) Globin phenotype of erythroid cells derived from human induced pluripotent stem cells. Blood 115:2553–2554.

Chen C mun, Krohn J, Bhattacharya S, Davies B (2011) A comparison of exogenous promoter activity at the ROSA26 locus using a PhiC31 integrase mediated cassette exchange approach in mouse es cells. PLoS One 6:6–13.
Chen J, Guo L, Zhang L, Wu H, Yang J, Liu H, Wang X, Hu X, Gu T, Zhou Z, Liu J, Liu J, Wu H, Mao S-Q, Mo K, Li Y, Lai K, Qi J, Yao H, Pan G, Xu G-L, Pei D (2013) Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet 45:1504–9.
Chen J, Liu H, Liu J, Qi J, Wei B, Yang J, Liang H, Chen Y, Chen J, Wu Y, Guo L, Zhu J, Zhao X, Peng T, Zhang Y, Chen S, Li X, Li D, Wang T, Pei D (2012) H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat Genet 45:34–42.
Chen T, Yuan D, Wei B, Jiang J, Kang J, Ling K, Gu Y, Li J, Xiao L, Pei G (2010) E-cadherin-mediated cell-cell contact is critical for induced pluripotent stem cell generation. Stem Cells 28:1315–1325.
Chen W, Kumar AR, Hudson WA, Li Q, Wu B, Staggs RA, Lund EA, Sam TN, Kersey JH (2008) Malignant Transformation Initiated by Mll-AF9: Gene Dosage and Critical Target Cells. Cancer Cell 13:432–440.
Cheng L, Hansen NF, Zhao L, Du Y, Zou C, Donovan FX, Chou BK, Zhou G, Li S, Dowey SN, Ye Z, Chandrasekharappa SC, Yang H, Mullikin JC, Liu PP (2012) Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. Cell Stem Cell 10:337–344.
Chin MH, Pellegrini M, Plath K, Lowry WE (2010) Molecular analyses of human induced pluripotent stem cells and embryonic stem cells. Cell Stem Cell 7:263–269.
Chou B, Huang X, Ye Z, Dowey SN, Resar LMS (2011) Efficient human iPS cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. Nat Publ Gr 21:518–529.
Cohn MA, Kowal P, Yang K, Haas W, Huang TT, Gygi SP, D'Andrea AD (2007) A UAF1-Containing Multisubunit Protein Complex Regulates the Fanconi Anemia Pathway. Mol Cell 28:786–797.
Cole MF, Johnstone SE, Newman JJ, Kagey MH, Young RA (2008) Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev 22:746–755.
Colman A, Dreesen O (2009) Pluripotent Stem Cells and Disease Modeling. Cell Stem Cell 5:244–247.
Costa Y, Ding J, Theunissen TW, Faiola F, Hore T a, Shliaha P V, Fidalgo M, Saunders A, Lawrence M, Dietmann S, Das S, Levasseur DN, Li Z, Xu M, Reik W, Silva JCR, Wang J (2013) NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature 495:370–4.
Covello KL, Kehler J, Yu H, Gordan JD, Arsham AM, Hu CJ, Labosky PA, Simon MC, Keith B (2006) HIF-2alpha regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. Genes Dev 20:557–570.
Crossan GP, van der Weyden L, Rosado I V, Langevin F, Gaillard P-HL, McIntyre RE, Gallagher F, Kettunen MI, Lewis DY, Brindle K, Arends MJ, Adams DJ, Patel KJ (2011) Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. Nat Genet 43:147–52.
Cusulin C, Chesnelong C, Bose P, Bilenky M, Kopciuk K, Chan JA, Cairncross JG, Jones SJ, Marra MA, Luchman HA, Weiss S (2015) Overcoming Pluripotent Stem Cell Dependence on the Repair of Endogenous DNA Damage. Stem Cell Reports 5:1–9.
David L, Polo JM (2014) Phases of reprogramming. Stem Cell Res 12:754–761.
De Groote FH, Jansen JG, Masuda Y, Shah DM, Kamiya K, De Wind N, Siegal G (2011) The Rev1 translesion synthesis polymerase has multiple distinct DNA binding modes. DNA Repair (Amst) 10:915–925.
Deans AJ, West SC (2011) DNA interstrand crosslink repair and cancer. Nat Rev Cancer 11:467–480.
Dick E, Matsa E, Young LE, Darling D, Denning C (2011) Faster generation of hiPSCs by coupling high-titer lentivirus and column-based positive selection. Nat Protoc 6:701–714.
Ding Q, Lee YK, Schaefer EAK, Peters DT, Veres A, Kim K, Kuperwasser N, Motola DL, Meissner TB, Hendriks WT, Trevisan M, Gupta RM, Moisan A, Banks E, Friesen M, Schinzel RT, Xia F, Tang A, Xia Y, Figueroa E, Wann A, Ahfeldt T, Daheron L, Zhang F, Rubin LL,

Peng LF, Chung RT, Musunuru K, Cowan CA (2013) A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell 12:238–251.
Djuric U, Cheung AYL, Zhang W, Mok RS, Lai W, Piekna A, Hendry JA, Ross PJ, Pasceri P, Kim D-S, Salter MW, Ellis J (2015) MECP2e1 isoform mutation affects the form and function of neurons derived from Rett syndrome patient iPS cells. Neurobiol Dis 76:37–45.
Doege CA, Inoue K, Yamashita T, Rhee DB, Travis S, Fujita R, Guarnieri P, Bhagat G, Vanti WB, Shih A, Levine RL, Nik S, Chen EI, Abeliovich A (2012) Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature 488:652–655.
Dos Santos RL, Tosti L, Radzisheuskaya A, Caballero IM, Kaji K, Hendrich B, Silva JCR (2014) MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell 15:102–110.
Ebert AD, Yu J, Jr FFR, Mattis VB, Lorson CL, Thomson JA, Svendsen CN (2009) Induced pluripotent stem cells from a spinal muscular atrophy patient. Nature 457:277–281.
Eiges R, Urbach A, Malcov M, Frumkin T, Schwartz T, Amit A, Yaron Y, Eden A, Yanuka O, Benvenisty N, Ben-Yosef D (2007) Developmental Study of Fragile X Syndrome Using Human Embryonic Stem Cells Derived from Preimplantation Genetically Diagnosed Embryos. Cell Stem Cell 1:568–577.
Elder DA, D'Alessio DA, Eyal O, Mueller R, Smith FO, Kansra AR, Rose SR (2008) Abnormalities in glucose tolerance are common in children with fanconi anemia and associated with impaired insulin secretion. Pediatr Blood Cancer 51:256–260.
Eminli S, Foudi A, Stadtfeld M, Maherali N, Ahfeldt T, Mostoslavsky G, Hock H, Hochedlinger K (2009) Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. Nat Genet 41:968–76.
Enoiu M, Jiricny J, Schärer OD (2012) Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. Nucleic Acids Res 40:8953–8964.
Esteban MA, 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 (2010) Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. Cell Stem Cell 6:1–9.
Evans E, Fellows J, Coffer A, Wood RD (1997) Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. EMBO J 16:625–638.
Eyal O, Blum S, Mueller R, Smith FO, Rose SR (2008) Improved growth velocity during thyroid hormone therapy in children with Fanconi anemia and borderline thyroid function. Pediatr Blood Cancer 51:652–656.
Fan Q, Zhang F, Barrett B, Ren K, Andreassen PR (2009) A role for monoubiquitinated FANCD2 at telomeres in ALT cells. Nucleic Acids Res 37:1740–1754.
Fekairi S, Scaglione S, Chahwan C, Taylor ER, Tissier A, Coulon S, Dong MQ, Ruse C, Yates JR, Russell P, Fuchs RP, McGowan CH, Gaillard PHL (2009) Human SLX4 Is a Holliday Junction Resolvase Subunit that Binds Multiple DNA Repair/Recombination Endonucleases. Cell 138:78–89.
Feng B, Jiang J, Kraus P, Ng J-H, Heng J-CD, Chan Y-S, Yaw L-P, Zhang W, Loh Y-H, Han J, Vega VB, Cacheux-Rataboul V, Lim B, Lufkin T, Ng H-H (2009a) Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol 11:197–203.
Feng B, Ng J-H, Heng J-CD, Ng H-H (2009b) Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. Cell Stem Cell 4:301–12.
Fernandez TDS, Fernandez CDS, Mencalha AL (2013) Human Induced Pluripotent Stem Cells from Basic Research to Potential Clinical Applications in Cancer. Biomed Res Int 2013:1–11.
Friedel AM, Pike BL, Gasser SM (2009) ATR/Mec1: coordinating fork stability and repair. Curr. Opin. Cell Biol. 21:237–244.
Fusaki N, Ban H, Nishiyama A, Saeki K, Hasegawa M (2009) Efficient 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 Phys Biol Sci 85:348–362.

Gao F, Kwon SW, Zhao Y, Jin Y (2009) PARP1 poly(ADP-ribosyl)ates Sox2 to control Sox2 protein levels and FGF4 expression during embryonic stem cell differentiation. J Biol Chem 284:22263–22273.
Gao Y, Chen J, Li K, Wu T, Huang B, Liu W, Kou X, Zhang Y, Huang H, Jiang Y, Yao C, Liu X, Lu Z, Xu Z, Kang L, Chen J, Wang H, Cai T, Gao S (2013) Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell 12:453–469.
Garaycoechea JI, Patel KJ (2013) Why does the bone marrow fail in Fanconi anemia ? Why does the Bone marrow fail in Fanconi Anemia. Blood. 123:26–34.
Garcia CCM, Angeli JPF, Freitas FP (2011) Acetaldehyde Promotes Unequivocal Formation of 1,N2-Propano-2′-deoxyguanosine in Human Cells.
Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, Heidersbach A, Ramalho-Santos J, McManus MT, Plath K, Meshorer E, Ramalho-Santos M (2009) Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature 460:863–868.
Gaspar-Maia A, Qadeer ZA, Hasson D, Ratnakumar K, Leu NA, Leroy G, Liu S, Costanzi C, Valle-Garcia D, Schaniel C, Lemischka I, Garcia B, Pehrson JR, Bernstein E (2013) MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. Nat Commun 4:1565.
Geng L, Huntoon CJ, Karnitz LM (2010) RAD18-mediated ubiquitination of PCNA activates the Fanconi anemia DNA repair network. J Cell Biol 191:249–257.
Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC (2010) Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. PLoS One. doi: 10.1371/journal.pone.0008975
González F, Boué S, Izpisúa Belmonte JC (2011) Methods for making induced pluripotent stem cells: reprogramming à la carte. Nat Rev Genet 12:231–242.
González F, Georgieva D, Vanoli F, Shi ZD, Stadtfeld M, Ludwig T, Jasin M, Huangfu D (2013) Homologous Recombination DNA Repair Genes Play a Critical Role in Reprogramming to a Pluripotent State. Cell Rep 3:651–660.
González F, Huangfu D (2015) Mechanisms underlying the formation of induced pluripotent stem cells. Wiley Interdiscip Rev Dev Biol 5:39–65.
Gonzalez-Munoz E, Arboleda-Estudillo Y, Out HH, Cibelli JB (2014) Histone chaperone ASF1A is required for maintenance of pluripotency and cellular reprogramming. Science (80-) 345:822–825.
Guervilly JH, Macé-Aimé G, Rosselli F (2008) Loss of CHK1 function impedes DNA damage-induced FANCD2 monoubiquitination but normalizes the abnormal G2 arrest in Fanconi anemia. Hum Mol Genet 17:679–689.
Guo S, Zi X, Schulz VP, Cheng J, Zhong M, Koochaki SHJ, Megyola CM, Pan X, Heydari K, Weissman SM, Gallagher PG, Krause DS, Fan R, Lu J (2014) Nonstochastic reprogramming from a privileged somatic cell state. Cell 156:649–662.
Gustafsson M V., Zheng X, Pereira T, Gradin K, Jin S, Lundkvist J, Ruas JL, Poellinger L, Lendahl U, Bondesson M (2005) Hypoxia requires Notch signaling to maintain the undifferentiated cell state. Dev Cell 9:617–628.
Haase A, Olmer R, Schwanke K, Wunderlich S, Merkert S, Hess C, Zweigerdt R, Gruh I, Meyer J, Wagner S, Maier LS, Han DW, Glage S, Miller K, Fischer P, Sch??ler HR, Martin U (2009) Generation of Induced Pluripotent Stem Cells from Human Cord Blood. Cell Stem Cell 5:434–441.
Habib O, Habib G, Do JT, Moon S-H, Chung H-M (2014) Activation-induced deaminase-coupled DNA demethylation is not crucial for the generation of induced pluripotent stem cells. Stem Cells Dev 23:209–18.
Hadjur S, Ung K, Wadsworth L, Dimmick J, Rajcan-separovic E, Scott RW, Buchwald M, Jirik FR (2012) Defective hematopoiesis and hepatic steatosis in mice with combined deficiencies of the genes encoding Fancc and Cu / Zn superoxide dismutase Defective hematopoiesis and hepatic steatosis in mice with combined deficiencies of the genes encoding Fancc and . 98:1003–1011.
Haneline LS, Broxmeyer HE, Cooper S, Hangoc G, Carreau M, Buchwald M, Clapp DW (1998) Multiple inhibitory cytokines induce deregulated progenitor growth and apoptosis in hematopoietic cells from Fac-/- mice. Blood 91:4092–4098.

Haneline LS, Gobbett TA, Ramani R, Carreau M, Buchwald M, Yoder MC, Clapp DW (1999) Loss of FancC Function Results in Decreased Hematopoietic Stem Cell Repopulating Ability. Blood 94:1–9.
Hanna J, Cheng AW, Saha K, Kim J, Lengner CJ, Soldner F, Cassady JP, Muffat J, Carey BW, Jaenisch R (2010) Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proc Natl Acad Sci U S A 107:9222–9227.
Hanna J, Markoulaki S, Schorderet P, Carey BW, Beard C, Wernig M, Creyghton MP, Steine EJ, Cassady JP, Foreman R, Lengner CJ, Dausman JA, Jaenisch R (2008) Direct Reprogramming of Terminally Differentiated Mature B Lymphocytes to Pluripotency. 250–264.
Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R (2009a) Direct cell reprogramming is a stochastic process amenable to acceleration. Nature 462:595–601.
Hanna J, Saha K, Pando B, Zon J Van, Lengner CJ, Creyghton MP, Oudenaarden A Van, Jaenisch R (2009b) Direct cell reprogramming is a stochastic process amenable to acceleration. Nature. doi: 10.1038/nature08592
Hanna J, Wernig M, Markoulaki S, Sun CW, Meissner A, Cassady JP, Beard C, Brambrink T, Wu L-C, Townes TM, Jaenisch R (2007) Treatment of Sickle Cell Anemia Mouse Model with iPS Cells Generated from Autologous Skin. Science (80-) 318:1920–1923.
Hansson J, Rafiee MR, Reiland S, Polo JM, Gehring J, Okawa S, Huber W, Hochedlinger K, Krijgsveld J (2012) Highly Coordinated Proteome Dynamics during Reprogramming of Somatic Cells to Pluripotency. Cell Rep 2:1579–1592.
Ho GPHGP, Margossian SS, Taniguchi TT, D'Andrea ADAD (2006) Phosphorylation of FANCD2 on two novel sites is required for mitomycin C resistance. Mol Cell Biol 26:7005–7015.
Ho L, Crabtree GR (2010) Chromatin remodelling during development. Nature 463:474–484.
Ho R, Chronis C, Plath K (2011) Mechanistic insights into reprogramming to induced pluripotency. J Cell Physiol 226:868–878.
Ho R, Chronis C, Plath K (2011) Mechanistic insights into reprogramming to induced pluripotency. J Cell Physiol 226:868–878.
Hong H, Takahashi K, Ichisaka T, Aoi T, Kanagawa O, Nakagawa M, Okita K, Yamanaka S (2009) Suppression of induced pluripotent stem cell generation by the p53 – p21 pathway. Nature 460:1132–1135.
Hotta A, Cheung AYL, Farra N, Vijayaragavan K, Séguin C a, Draper JS, Pasceri P, Maksakova I a, Mager DL, Rossant J, Bhatia M, Ellis J (2009) Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. Nat Methods 6:370–376.
Hotta A, Ellis J (2008) Retroviral vector silencing during iPS cell induction: An epigenetic beacon that signals distinct pluripotent states. J Cell Biochem 105:940–948.
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 341:651–4.
Houghtaling S, Timmers C, Noll M, Finegold MJ, Jones SN, Meyn MS, Grompe M (2003) Epithelial cancer in Fanconi anemia complementatioin group D2 (FANCD2) knockout mice. Genes Dev 17:2021–2035.
Howlett NG, Harney JA, Rego MA, Kolling IV FW, Glover TW (2009) Functional interaction between the fanconi anemia D2 protein and proliferating cell nuclear antigen (PCNA) via a conserved putative PCNA interaction motif. J Biol Chem 284:28935–28942.
Hu K, Yu J, Suknuntha K, Tian S, Montgomery K, Choi KD, Stewart R, Thomson JA, Slukvin II (2011) Efficient generation of transgene-free induced pluripotent stem cells from normal and neoplastic bone marrow and cord blood mononuclear cells. Blood. doi: 10.1182/blood-2010-07-298331
Hu X, Zhang L, Mao SQ, Li Z, Chen J, Zhang RR, Wu HP, Gao J, Guo F, Liu W, Xu GF, Dai HQ, Shi YG, Li X, Hu B, Tang F, Pei D, Xu GL (2014) Tet and TDG mediate DNA demethylation essential for mesenchymal-to- epithelial transition in somatic cell reprogramming. Cell Stem Cell 14:512–522.
Huang M, Kim JM, Shiotani B, Yang K, Zou L, D'Andrea AD (2010) The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response. Mol Cell 39:259–268.

Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA (2008b) Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nat Biotechnol 26:1269–1275.
Hussain S, Wilson JB, Medhurst AL, Hejna J, Witt E, Ananth S, Davies A, Masson JY, Moses R, West SC, de Winter JP, Ashworth A, Jones NJ, Mathew CG (2004) Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. Hum Mol Genet 13:1241–1248.
Hussain S, Witt E, Huber PAJ, Medhurst AL, Ashworth A, Mathew CG (2003) Direct interaction of the Fanconi anaemia protein FANCG with BRCA2/FANCD1. Hum Mol Genet 12:2503–2510.
Ichida JK, Blanchard J, Lam K, Son EY, Chung JE, Egli D, Loh KM, Carter AC, Di Giorgio FP, Koszka K, Huangfu D, Akutsu H, Liu DR, Rubin LL, Eggan K (2009) A Small-Molecule Inhibitor of Tgf-Beta Signaling Replaces Sox2 in Reprogramming by Inducing Nanog. Cell Stem Cell 5:491–503.
Ishiai M, Kitao H, Smogorzewska A, Tomida J, Kinomura A, Uchida E, Saberi A, Kinoshita E, Kinoshita-Kikuta E, Koike T, Tashiro S, Elledge SJ, Takata M (2008) FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. Nat Struct Mol Biol 15:1138–1146.
Ishiai M, Tomida J, Itaya A, Hejna J (2016) The Fanconi Anemia Pathway and Interstrand Cross-Link Repair. In: DNA Replication, Recombination, and Repair. pp 175–210
Itzhaki I, Maizels L, Huber I, Zwi-dantsis L, Caspi O, Winterstern A, Feldman O, Gepstein A, Arbel G, Hammerman H, Boulos M, Gepstein L (2011) Modelling the long QT syndrome with induced pluripotent stem cells. Nature 471:225–229.
Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR (2006) Dissecting self-renewal in stem cells with RNA interference. Nature 442:533–538.
Jaenisch R, Young R (2008) Stem Cells, the Molecular Circuitry of Pluripotency and Nuclear Reprogramming. Cell 132:567–582.
Joenje H, Arwert F, Eriksson AW, de Koning H, Oostra AB (1981) Oxygen-dependence of chromosomal aberrations in Fanconi's anaemia. Nature 290:142–3.
Jucker M (2010) The benefits and limitations of animal models for translational research in neurodegenerative diseases. Nat Med 16:1210–1214.
Jung M, Dunbar CE, Winkler T (2015) Modeling Human Bone Marrow Failure Syndromes Using Pluripotent Stem Cells and Genome Engineering. Mol Ther. doi: 10.1038/mt.2015.180
Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA (2010) Mediator and cohesin connect gene expression and chromatin architecture. Nature 467:430–435.
Kahler DJ, Ahmad FS, Ritz A, Hua H, Moroziewicz DN, Sproul AA, Dusenberry CR, Shang L, Paull D, Zimmer M, Weiss KA, Egli D, Noggle SA (2013) Improved methods for reprogramming human dermal fibroblasts using fluorescence activated cell sorting. PLoS One 8:e59867.
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.
Kawamura T, Suzuki J, Wang Y V, Menendez S, Morera LB, Raya A, Wahl GM, Izpisu JC (2009b) Linking the p53 tumour suppressor pathway to somatic cell reprogramming. 460:2–7.
Kee Y, Andrea ADD (2012) Molecular pathogenesis and clinical management of Fanconi anemia. J Clin Invest 122:3799–3806.
Kee Y, D'Andrea AD (2010) Expanded roles of the Fanconi anemia pathway in preserving genomic stability. Genes Dev. 24:1680–1694.
Kim D, Kim C, Moon J, Chung Y, Chang M, Han B, Ko S, Yang E, Cha KY, Lanza R, Kim K (2009a) Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell Stem Cell 4:472–476.
Kim H, D'Andrea AD (2012) Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. Genes Dev. 26:1393–1408.
Kim JB, Sebastiano V, Wu G, Arauzo-Bravo MJ, Sasse P, Gentile L, Ko K, Ruau D, Ehrich M, van den Boom D, Meyer J, Hubner K, Bernemann C, Ortmeier C, Zenke M, Fleischmann BK, Zaehres H, Scholer HR (2009b) Oct4-Induced Pluripotency in Adult Neural Stem Cells. Cell 136:411–419.

Kim JB, Zaehres H, Wu G, Gentile L, Ko K, Sebastiano V, Ruau D, Han DW, Zenke M, Scho HR, Arau MJ (2008) Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. Nature. 454:646-650.
Kim JM, Parmar K, Huang M, Weinstock DM, Ruit CA, Kutok JL, D'Andrea AD (2009c) Inactivation of Murine Usp1 Results in Genomic Instability and a Fanconi Anemia Phenotype. Dev Cell 16:314–320.
Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LI, Yabuuchi A, Takeuchi A, Cunniff KC, Hongguang H, McKinney-Freeman S, Naveiras O, Yoon TJ, Irizarry RA, Jung N, Seita J, Hanna J, Murakami P, Jaenisch R, Weissleder R, Orkin SH, Weissman IL, Feinberg AP, Daley GQ (2010) Epigenetic memory in induced pluripotent stem cells. Nature 467:285–290.
Kim K, Zhao R, Doi A, Ng K, Unternaehrer J, Cahan P, Hongguang H, Loh Y, Aryee MJ, Lensch MW, Li H, Collins JJ, Feinberg AP, Daley GQ (2012) Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. 29:1117–1120.
Kim S, Kim GJ, Miyoshi H, Moon S-H, Ahn SE, Lee JH, Lee HJ, Cha K-Y, Chung HM (2007) Efficiency of the elongation factor-1alpha promoter in mammalian embryonic stem cells using lentiviral gene delivery systems. Stem Cells Dev 16:537–545.
Knipscheer P, Räschle M, Smogorzewska A, Enoiu M, Ho TV, Schärer OD, Elledge SJ, Walter JC (2009) The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. Science 326:1698–1701.
Knott SR V, Maceli AR, Erard N, Chang K, Marran K, Zhou X, Gordon A, ElDemerdash O, Wagenblast E, Kim S, Fellmann C, Hannon GJ (2014) A Computational Algorithm to Predict shRNA Potency. Mol Cell 56:796–807.
Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, Bernstein BE, Meissner A (2010) Reprogramming Factor Expression Initiates Widespread Targeted Chromatin Remodeling. Stem Cell 8:96–105.
Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502:472–9.
Kottemann MC, Smogorzewska A (2013) Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. 0–7.
Kumar R, DiMenna L, Schrode N, Liu T-C, Franck P, Muñoz-Descalzo S, Hadjantonakis A-K, Zarrin AA, Chaudhuri J, Elemento O, Evans T (2013) AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. Nature 500:89–92.
Kuraoka I, Kobertz WR, Ariza RR, Biggerstaff M, Essigmann JM, Wood RD (2000) Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. J Biol Chem 275:26632–26636.
Kutler DI, Singh B, Satagopan J, Batish SD, Berwick M, Giampietro PF, Hanenberg H, Auerbach AD (2003) A 20-year perspective on the International Fanconi Anemia Registry (IFAR). Blood 101:1249–1256.
Kwon YW, Jang S, Paek JS, Lee JW, Cho HJ, Yang HM, Kim HS (2015) E-Ras improves the efficiency of reprogramming by facilitating cell cycle progression through JNK-Sp1 pathway. Stem Cell Res 15:481–494.
Langevin F, Crossan GP, Rosado I V, Arends MJ, Patel KJ (2011) Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. Nature 475:53–58.
Li H, Collado M, Villasante A, Strati K, Ortega S, Can M, Blasco MA, Serrano M (2009) The Ink4 / Arf locus is a barrier for iPS cell reprogramming. 460:2–7.
Li R, Liang J, Ni S, Zhou T, Qing X, Li H, He W, Chen J, Li F, Zhuang Q, Qin B, Xu J, Li W, Yang J, Gan Y, Qin D, Feng S, Song H, Yang D, Zhang B, Zeng L, Lai L, Esteban MA, Pei D (2010) A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. Cell Stem Cell 7:51–63.
Li X, Yang Y, Yuan J, Hong P, Freie B, Orazi A, Haneline LS, Clapp DW (2004) Continuous in vivo infusion of interferon-gamma (IFN-y) preferentially reduces myeloid progenitor numbers and enhances engraftment of syngeneic wild-type cells in Fancc-/- mice. Blood 104:1204–1209.
Li Z, Rana TM (2012) A kinase inhibitor screen identifies small-molecule enhancers of reprogramming and iPS cell generation. Nat Commun 3:1011–1085.

Lian I, Kim J, Okazawa H, Zhao J, Zhao B, Yu J, Chinnaiyan A, Israel MA, Goldstein LSB, Abujarour R, Ding S, Guan KL (2010) The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. Genes Dev 24:1106–1118.
Liang G, He J, Zhang Y (2012) Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. Nat Cel Bio. 14:457–466.
Liang G, Taranova O, Xia K, Zhang Y (2010) Butyrate promotes induced pluripotent stem cell generation. J Biol Chem 285:25516–21.
Liang G, Zhang Y (2012) Embryonic stem cell and induced pluripotent stem cell : an epigenetic perspective. Cell Res 23:49–69.
Liang G, Zhang Y (2013) Genetic and epigenetic variations in iPSCs: Potential causes and implications for application. Cell Stem Cell 13:149–159.
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 MA, Pei D (2011) MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. J Biol Chem 286:17359–17364.
Liao J, Marumoto T, Yamaguchi S, Okano S, Takeda N, Sakamoto C, Kawano H, Nii T, Miyamato S, Nagai Y, Okada M, Inoue H, Kawahara K, Suzuki A, Miura Y, Tani K (2013) Inhibition of PTEN Tumor Suppressor Promotes the Generation of Induced Pluripotent Stem Cells. Mol Ther 21:1242–1250.
Lim ET, Wurtz P, Havulinna AS, Palta P, Tukiainen T, Rehnstrom K, Esko T, Magi R, Inouye M, Lappalainen T, Chan Y, Salem RM, Lek M, Flannick J, Sim X, Manning A, Ladenvall C, Bumpstead S, Hamalainen E, Aalto K, Maksimow M, Salmi M, Blankenberg S, Ardissino D, Shah S, Horne B, McPherson R, Hovingh GK, Reilly MP, Watkins H, Goel A, Farrall M, Girelli D, Reiner AP, Stitziel NO, Kathiresan S, Gabriel S, Barrett JC, Lehtimaki T, Laakso M, Groop L, Kaprio J, Perola M, McCarthy MI, Boehnke M, Altshuler DM, Lindgren CM, Hirschhorn JN, Metspalu A, Freimer NB, Zeller T, Jalkanen S, Koskinen S, Raitakari O, Durbin R, MacArthur DG, Salomaa V, Ripatti S, Daly MJ, Palotie A (2014) Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population. PLoS Genet. 7: e1004494-e1004494.
Lin C, Garruss AS, Luo Z, Guo F, Shilatifard A (2013) The RNA Pol II elongation factor Ell3 marks enhancers in ES cells and primes future gene activation. Cell 152:144–156.
Lin CY, Lovén J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA (2012) Transcriptional amplification in tumor cells with elevated c-Myc. Cell 151:56–67.
Liu G, Suzuki K, Li M, Qu J, Montserrat N, Tarantino C, Gu Y, Yi F, Xu X, Zhang W, Ruiz S, Plongthongkum N, Zhang K, Masuda S, Nivet E, Tsunekawa Y, Soligalla RD, Goebl A, Aizawa E, Kim NY, Kim J, Dubova I, Li Y, Ren R, Benner C, Del Sol A, Bueren J, Trujillo JP, Surralles J, Cappelli E, Dufour C, Esteban CR, Izpisua Belmonte JC (2014) Modelling Fanconi anemia pathogenesis and therapeutics using integration-free patient-derived iPSCs. Nat Commun 5:4330.
Liu G, Suzuki K, Li M, Qu J, Montserrat N, Tarantino C, Gu Y, Yi F, Xu X, Zhang W, Ruiz S, Plongthongkum N, Zhang K, Masuda S, Nivet E, Tsunekawa Y, Soligalla RD, Goebl A, Aizawa E, Kim NY, Kim J, Dubova I, Li Y, Ren R, Benner C, Del Sol A, Bueren J, Trujillo JP, Surralles J, Cappelli E, Dufour C, Esteban CR, Izpisua Belmonte JC (2014) Modelling Fanconi anemia pathogenesis and therapeutics using integration-free patient-derived iPSCs. Nat Commun 5:4330.
Liu GH, Ding Z, Izpisua Belmonte JC (2012) IPSC technology to study human aging and aging-related disorders. Curr Opin Cell Biol 24:765–774.
Liu T, Ghosal G, Yuan J, Chen J, Huang J (2010) FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. Science 329:693–696.
Loh KM, Lim B (2011) A precarious balance: Pluripotency factors as lineage specifiers. Cell Stem Cell 8:363–369.
Loh Y-H, Agarwal S, Park I-H, Urbach A, Huo H, Heffner GC, Kim K, Miller JD, Ng K, Daley GQ (2009) Generation of induced pluripotent stem cells from human blood. Blood 113:5476–5479.
Loh Y-H, Hartung O, Li H, Guo C, Sahalie JM, Manos PD, Urbach A, Heffner GC, Grskovic M, Vigneault F, Lensch MW, Park I-H, Agarwal S, Church GM, Collins JJ, Irion S, Daley GQ (2010) Reprogramming of T cells from human peripheral blood. Cell Stem Cell 7:15–19.

Loh Y-H, Wu Q, Chew J-L, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong K-Y, Sung KW, Lee CWH, Zhao X-D, Chiu K-P, Lipovich L, Kuznetsov V a, Robson P, Stanton LW, Wei C-L, Ruan Y, Lim B, Ng H-H (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38:431–440.
Long DT, Räschle M, Joukov V, Walter JC (2011) Mechanism of RAD51-dependent DNA interstrand cross-link repair. Science 333:84–87.
Lowry WE, Richter L, Yachechko R, Pyle a D, Tchieu J, Sridharan R, Clark a T, Plath K (2008) Generation of human induced pluripotent stem cells from dermal fibroblasts. Proc Natl Acad Sci U S A 105:2883–2888.
Luo M, Ling T, Xie W, Sun H, Zhou Y, Zhu Q, Shen M, Zong L, Lyu G, Zhao Y, Ye T, Gu J, Tao W, Lu Z, Grummt I (2013) NuRD blocks reprogramming of mouse somatic cells into Pluripotent stem cells. Stem Cells 31:1278–1286.
Ma T, Xie M, Laurent T, Ding S (2013) Progress in the reprogramming of somatic cells. Circ Res 112:562–574.
Machida YJ, Machida Y, Chen Y, Gurtan AM, Kupfer GM, D'Andrea AD, Dutta A (2006) UBE2T Is the E2 in the Fanconi Anemia Pathway and Undergoes Negative Autoregulation. Mol Cell 23:589–596.
MacKay C, Declais AC, Lundin C, Agostinho A, Deans AJ, MacArtney TJ, Hofmann K, Gartner A, West SC, Helleday T, Lilley DMJ, Rouse J (2010) Identification of KIAA1018/FAN1, a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2. Cell 142:65–76.
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 474:225–229.
Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R, Plath K (2007) Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. Cell Stem Cell 1:55–70.
Mali P, Chou BK, Yen J, Ye Z, Zou J, Dowey S, Brodsky RA, Ohm JE, Yu W, Baylin SB, Yusa K, Bradley A, Meyers DJ, Mukherjee C, Cole PA, Cheng L (2010) Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. Stem Cells 28:713–720.
Mancarci OB, Unternaehrer J, Gupta PB, Lander ES, Armstrong SA (2012) Chromatin-modifying enzymes asmodulators of reprogramming. Nature. doi: 10.1038/nature10953
Mansour AA, Gafni O, Weinberger L, Zviran A, Ayyash M, Rais Y, Krupalnik V, Zerbib M, Amann-zalcenstein D, Maza I, Geula S, Viukov S, Holtzman L, Pribluda A, Canaani E, Horn-saban S, Amit I, Novershtern N, Hanna JH (2012) The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. Nature 488:409–413.
Marchitti SA, Brocker C, Stagos D, Vasiliou V (2008) Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. Expert Opin Drug Metab Toxicol 4:697–720.
Marson A, Foreman R, Chevalier B, Bilodeau S, Kahn M, Young RA, Jaenisch R (2008) Wnt Signaling Promotes Reprogramming of Somatic Cells to Pluripotency. Cell Stem Cell 3:132–135.
Marucci L, Pedone E, Di Vicino U, Sanuy-Escribano B, Isalan M, Cosma MP (2014) β-Catenin Fluctuates in mouse ESCs and is essential for Nanog-mediated reprogramming of somatic cells to pluripotency. Cell Rep 8:1686–1696.
Mathieu J, Zhou W, Xing Y, Sperber H, Ferreccio A, Agoston Z, Kuppusamy KT, Moon RT, Ruohola-Baker H (2014) Hypoxia-inducible factors have distinct and stage-specific roles during reprogramming of human cells to pluripotency. Cell Stem Cell 14:592–605.
Mayshar Y, Ben-david U, Lavon N, Biancotti J, Yakir B, Clark AT, Plath K, Lowry WE, Benvenisty N (2010) Identification and Classification of Chromosomal Aberrations in Human Induced Pluripotent Stem Cells. Stem Cell 7:521–531.
Mechilli M, Schinoppi A, Kobos K, Natarajan AT, Palitti F (2008) DNA repair deficiency and acetaldehyde-induced chromosomal alterations in CHO cells. Mutagenesis 23:51–56.
Medhurst AL, Laghmani EH, Steltenpool J, Ferrer M, Fontaine C, De Groot J, Rooimans MA, Scheper RJ, Meetei AR, Wang W, Joenje H, De Winter JP (2006) Evidence for subcomplexes in the Fanconi anemia pathway. Blood 108:2072–2080.

Meetei AR, de Winter JP, Medhurst AL, Wallisch M, Waisfisz Q, van de Vrugt HJ, Oostra AB, Yan Z, Ling C, Bishop CE, Hoatlin ME, Joenje H, Wang W (2003) A novel ubiquitin ligase is deficient in Fanconi anemia. Nat Genet 35:165–170.
Meetei AR, Medhurst AL, Ling C, Xue Y, Singh TR, Bier P, Steltenpool J, Stone S, Dokal I, Mathew CG, Hoatlin M, Joenje H, de Winter JP, Wang W (2005) A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. Nat Genet 37:958–63.
Mekhoubad S, Bock C, Boer AS De, Kiskinis E, Meissner A (2012) Erosion of Dosage Compensation Impacts Human iPSC Disease Modeling. Cell Stem Cell 10:595–609.
Meng X, Shen J, Kawagoe S, Ohashi T, Brady RO, Eto Y (2010) Induced pluripotent stem cells derived from mouse models of lysosomal storage disorders. 2–7.
Meyer N, Penn LZ (2008) Reflecting on 25 years with MYC. Nat Rev Cancer 8:976–990.
Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A (2008a) Dissecting direct reprogramming through integrative genomic analysis. 454:49–56.
Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A (2008b) Dissecting direct reprogramming through integrative genomic analysis. Nature 1–8.
Mills JA, Paluru P, Weiss MJ, Gadue P, French DL (2014) Hematopoietic Differentiation of Pluripotent Stem Cells in Culture. In: Hematopoietic Stem Cell Protocols. pp 181–194
Milyavsky M, Gan OI, Trottier M, Komosa M, Tabach O, Notta F, Lechman E, Hermans KG, Eppert K, Konovalova Z, Ornatsky O, Domany E, Meyn MS, Dick JE (2010) A distinctive DNA damage response in human hematopoietic stem cells reveals an apoptosis-independent role for p53 in self-renewal. Cell Stem Cell 7:186–197.
Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S (2003) The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113:631–642.
Mohyeldin A, Garzon-Muvdi T, Quinones-Hinojosa A (2010) Oxygen in stem cell biology: A critical component of the stem cell niche. Cell Stem Cell 7:150–161.
Molina-Estevez FJ, Lozano ML, Navarro S, Torres Y, Grabundzija I, Ivics Z, Samper E, Bueren JA, Guenechea G (2013) Brief report: Impaired cell reprogramming in nonhomologous end joining deficient cells. Stem Cells 31:1726–1730.
Müller LUW, Milsom MD, Harris CE, Vyas R, Brumme KM, Parmar K, Moreau LA, Schambach A, Park I, London WB, Strait K, Schlaeger T, Devine AL, Grassman E, Andrea AD, Daley GQ, Williams DA (2012a) Overcoming Reprogramming Resistance of Fanconi Anemia Cells. Blood 23:5449–5457.
Müller LUW, Schlaeger TM, Devine AL, David A (2012b) Induced pluripotent stem cells as a tool for gaining new insights into Fanconi anemia. Cell Cycle 11:16:2985–2990.
Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol 26:101–106.
Nakayama M (2010) Homologous recombination in human iPS and ES cells for use in gene correction therapy. Drug Discov Today 00:1–5.
Narsinh KH, Jia F, Robbins RC, Kay MA, Longaker MT, Wu JC (2011) Generation of adult human induced pluripotent stem cells using nonviral minicircle DNA vectors. Nat Protoc 6:78–88.
Navarro S, Meza NW, Quintana-Bustamante O, Casado JA, Jacome A, McAllister K, Puerto S, Surralles J, Segovia JC, Bueren JA (2006) Hematopoietic Dysfunction in a Mouse Model for Fanconi Anemia Group D1. Mol Ther 14:525–535.
Navarro S, Moleiro V, Molina-Estevez FJ, Lozano ML, Chinchon R, Almarza E, Quintana-Bustamante O, Mostoslavsky G, Maetzig T, Galla M, Heinz N, Schiedlmeier B, Torres Y, Modlich U, Samper E, Rio P, Segovia JC, Raya A, Guenechea G, Izpisua-Belmonte JC, Bueren JA (2014) Generation of iPSCs from genetically corrected Brca2 hypomorphic cells: Implications in cell reprogramming and stem cell therapy. Stem Cells 32:436–446.
Neveling K, Endt D, Hoehn H, Schindler D (2009) Genotype – phenotype correlations in Fanconi anemia. 668:73–91.

Nie Z, Hu G, Wei G, Cui K, Yamane A, Resch W, Wang R, Green DR, Tessarollo L, Casellas R, Zhao K, Levens D (2012) c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. Cell 151:68–79.
Niedernhofer LJ (2008) DNA repair is crucial for maintaining hematopoietic stem cell function. DNA Repair (Amst) 7:523–529.
Nijman SMB, Huang TT, Dirac AMG, Brummelkamp TR, Kerkhoven RM, D'Andrea AD, Bernards R (2005) The deubiquitinating enzyme USP1 regulates the fanconi anemia pathway. Mol Cell 17:331–339.
Nilsson L, Astrand-Grundström I, Arvidsson I, Jacobsson B, Hellström-Lindberg E, Hast R, Jacobsen SE (2000) Isolation and characterization of hematopoietic progenitor/stem cells in 5q-deleted myelodysplastic syndromes: evidence for involvement at the hematopoietic stem cell level. Blood 96:2012–2021.
Nilsson L, Edén P, Olsson E, Månsson R, Åstrand-Grundström I, Strömbeck B, Theilgaard-Mönch K, Anderson K, Hast R, Hellström-Lindberg E, Samuelsson J, Bergh G, Nerlov C, Johansson B, Sigvardsson M, Borg Å, Jacobsen SEW (2007) The molecular signature of MDS stem cells supports a stem-cell origin of 5q-myelodysplastic syndromes. Blood 110:3005–3014.
Noll M, Battaile KP, Bateman R, Lax TP, Rathbun K, Reifsteck C, Bagby G, Finegold M, Olson S, Grompe M (2002) Fanconi anemia group A and C double-mutant mice: Functional evidence for a multi-protein Fanconi anemia complex. Exp Hematol 30:679–688.
O'Malley J, Skylaki S, Iwabuchi KA, Chantzoura E, Ruetz T, Johnsson A, Tomlinson SR, Linnarsson S, Kaji K (2013) High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature 499:88–91.
Oestergaard VH, Langevin F, Kuiken HJ, Pace P, Niedzwiedz W, Simpson LJ, Ohzeki M, Takata M, Sale JE, Patel KJ (2007) Deubiquitination of FANCD2 Is Required for DNA Crosslink Repair. Mol Cell 28:798–809.
Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ, Yu J, Hebrok M (2011) Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. doi: 10.1038/ncb2239
Ohnuki M, Takahashi K, Yamanaka S (2009) Generation and characterization of human induced pluripotent stem cells. Curr Protoc Stem Cell Biol Chapter 4:Unit 4A.2.
Okabe M, Otsu M, Ahn DH, Kobayashi T, Morita Y, Wakiyama Y, Onodera M, Eto K, Ema H, Nakauchi H (2009) Definitive proof for direct reprogramming of hematopoietic cells to pluripotency. Blood 114:1764–1767.
Okada M, Yoneda Y (2011) The timing of retroviral silencing correlates with the quality of induced pluripotent stem cell lines. Biochim Biophys Acta 1810:226–35.
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, Shibata T, Kunisada T, Takahashi M, Takahashi J, Saji H, Yamanaka S (2011) A more efficient method to generate integration-free human iPS cells. Nat Methods 8:409–412.
Okita K, Nakagawa M, Hong H, Ichisaka T, Yamanaka S (2009) Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors. Science (80-) 949:949–953.
Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, Gupta PB, Lander ES, Armstrong S a., Daley GQ (2012) Chromatin-modifying enzymes as modulators of reprogramming. Nature 483:598–602.
Ozolek JA, Castro CA (2011) Teratomas Derived from Embryonic Stem Cells as Models for Embryonic Development , Disease , and Tumorigenesis. In: Kallos M (ed). InTech,
Pan G, Thomson J a (2007) Nanog and transcriptional networks in embryonic stem cell pluripotency. Cell Res 17:42–49.
Panopoulos AD, Yanes O, Ruiz S, Kida YS, Diep D, Tautenhahn R, Herrerias A, Batchelder EM, Plongthongkum N, Lutz M, Berggren WT, Zhang K, Evans RM, Siuzdak G, Izpisua Belmonte JC (2012) The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. Cell Res 22:168–177.

Park IH, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, Lensch MW, Cowan C, Hochedlinger K, Daley GQ (2008b) Disease-specific induced pluripotent stem cells. Cell 134:877–886.
Park I-H, Zhao R, West J a, Yabuuchi A, Huo H, Ince T a, Lerou PH, Lensch MW, Daley GQ (2008a) Reprogramming of human somatic cells to pluripotency with defined factors. Nature 451:141–146.
Parmar K, Andrea AD, Niedernhofer LJ (2009) Mouse models of Fanconi anemia. Mutat Res 668:133–140.
Pasque V, Radzisheuskaya A, Gillich A, Halley-stott RP (2012) Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. 2:6094–6104.
Pawlak M, Jaenisch R, Pawlak M, Jaenisch R (2011) De novo DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state De novo DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state. 1035–1040.
Pinto FO, Leblanc T, Chamousset D, Roux G Le, Brethon B, Cassinat B, Larghero J, Villartay JP De, Stoppa-Lyonnet D, Baruchel A, Socié G, Gluckman E, Soulier J (2009) Diagnosis of Fanconi anemia in Patients with bone marrow failure. Haematologica 94:487–495.
Polak U, Hirsch C, Ku S, Gottesfeld J, Dent SYR, Napierala M (2012) Selecting and isolating colonies of human induced pluripotent stem cells reprogrammed from adult fibroblasts. J Vis Exp 3–7.
Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, Bar-nur O, Cheloufi S, Stadtfeld M, Figueroa ME, Robinton D, Natesan S, Melnick A, Zhu J (2012) A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. Cell 151:1617–1632.
Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natesan S, Wagers AJ, Melnick A, Evans T, Hochedlinger K (2010) Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Publ Gr 1–10.
Prokhorova T a, Harkness LM, Frandsen U, Ditzel N, Schrøder HD, Burns JS, Kassem M (2009) Teratoma formation by human embryonic stem cells is site dependent and enhanced by the presence of Matrigel. Stem Cells Dev 18:47–54.
Pulliam-Leath AC, Ciccone SL, Nalepa G, Li X, Si Y, Miravalle L, Smith D, Yuan J, Li J, Anur P, Orazi A, Vance GH, Yang FC, Hanenberg H, Bagby GC, Clapp DW (2010) Genetic disruption of both Fancc and Fancg in mice recapitulates the hematopoietic manifestations of Fanconi anemia. Blood 116:2915–2920.
Qin H, Diaz A, Blouin L, Lebbink RJ, Patena W, Tanbun P (2014) Systematic Identification of Barriers to Human iPSC Generation. Cell 158:449–461.
Qin JY, Zhang L, Clift KL, Hulur I, Xiang AP, Ren BZ, Lahn BT (2010) Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. PLoS One 5:3–6.
Quintanilla RH, Asprer JST, Vaz C, Tanavde V, Lakshmipathy U (2014) CD44 is a negative cell surface marker for pluripotent stem cell identification during human fibroblast reprogramming. PLoS One 9:e85419.
Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, Maza I, Mor N, Baran D, Weinberger L, Jaitin D a., Lara-Astiaso D, Blecher-Gonen R, Shipony Z, Mukamel Z, Hagai T, Gilad S, Amann-Zalcenstein D, Tanay A, Amit I, Novershtern N, Hanna JH (2013) Deterministic direct reprogramming of somatic cells to pluripotency. Nature 502:65–70.
Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods 5:877–9.
Ramezani A, Hawley TS, Robert G. Hawley (2000) Lentiviral Vectors for Enhanced Gene Expression in Human Hematopoietic Cells. Mol Ther 2:458–469.
Rao RA, Dhele N, Cheemadan S, Ketkar A, Jayandharan GR, Palakodeti D, Rampalli S (2015) Ezh2 mediated H3K27me3 activity facilitates somatic transition during human pluripotent reprogramming. Sci Rep 5:8229.

Raschle M, Knipsheer P, Enoiu M, Angelov T, Sun J, Griffith JD, Ellenberger TE, Scharer OD, Walter JC (2008) Mechanism of Replication-Coupled DNA Interstrand Crosslink Repair. Cell 134:969–980.
Rathbun RK, Faulkner GR, Ostroski MH, Christianson T a, Hughes G, Jones G, Cahn R, Maziarz R, Royle G, Keeble W, Heinrich MC, Grompe M, Tower P a, Bagby GC (1997) Inactivation of the Fanconi anemia group C gene augments interferon-gamma-induced apoptotic responses in hematopoietic cells. Blood 90:974–985.
Raya A, Rodríguez-Pizà I, Guenechea G, Vassena R, Navarro S, Barrero MJ, Consiglio A, Castellà M, Río P, Sleep E, González F, Tiscornia G, Garreta E, Aasen T, Veiga A, Verma IM, Surrallés J, Bueren J, Izpisúa Belmonte JC (2009) Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. Nature 460:53–59.
Redmer T, Diecke S, Grigoryan T, Quiroga-Negreira A, Birchmeier W, Besser D (2011) E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. EMBO Rep 12:720–6.
Rhee DB, Wang Y, Mizesko M, Zhou F, Haneline L, Liu Y (2010) FANCC suppresses short telomere-initiated telomere sister chromatid exchange. Hum Mol Genet 19:879–887.
Ridpath JR, Nakamura A, Tano K, Luke AM, Sonoda E, Arakawa H, Buerstedde JM, Gillespie DAF, Sale JE, Yamazoe M, Bishop DK, Takata M, Takeda S, Watanabe M, Swenberg JA, Nakamura J (2007) Cells deficient in the FANC/BRCA pathway are hypersensitive to plasma levels of formaldehyde. Cancer Res 67:11117–11122.
Rio P, Baños R, Lombardo A, Quintana-Bustamante O, Alvarez L, Garate Z, Genovese P, Almarza E, Valeri A, Díez B, Navarro S, Torres Y, Trujillo JP, Murillas R, Segovia JC, Samper E, Surralles J, Gregory PD, Holmes MC, Naldini L, Bueren J a (2014) Targeted gene therapy and cell reprogramming in Fanconi anemia. EMBO Mol Med 6:835–48.
Robinton D a., Daley GQ (2012) The promise of induced pluripotent stem cells in research and therapy. Nature 481:295–305.
Rodrı I, Guenechea G, Vassena R, Navarro S, Consiglio A, Castella M, Sleep E, Gonza F, Bueren J, Izpisu JC (2009) Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. doi: 10.1038/nature08129
Rosado I V, Langevin F, Crossan GP, Takata M, Patel KJ (2011) Formaldehyde catabolism is essential in cells deficient for the Fanconi anemia DNA-repair pathway. Nat Struct Mol Biol 18:1432–4.
Rosenberg PS, Greene MH, Alter BP (2003) Cancer incidence in persons with Fanconi anemia. Blood 101:822–826.
Rosselli F, Sanceau J, Wietzerbin J, Moustacchi E (1992) Abnormal lymphokine production: a novel feature of the genetic disease Fanconi anemia - I. Involvement of interleukin-6. Hum Genet 89:42–48.
Ruiz S, Panopoulos AD, Herrerías A, Bissig K-D, Lutz M, Berggren WT, Verma IM, Izpisua Belmonte JC (2011) A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. Curr Biol 21:45–52.
Ruzankina Y, Pinzon-Guzman C, Asare A, Ong T, Pontano L, Cotsarelis G, Zediak VP, Velez M, Bhandoola A, Brown EJ (2007) Deletion of the Developmentally Essential Gene ATR in Adult Mice Leads to Age-Related Phenotypes and Stem Cell Loss. Cell Stem Cell 1:113–126.
Sakurai K, Talukdar I, Patil VS, Dang J, Li Z, Chang KY, Lu CC, Delorme-Walker V, Dermardirossian C, Anderson K, Hanein D, Yang CS, Wu D, Liu Y, Rana TM (2014) Kinome-wide functional analysis highlights the role of cytoskeletal remodeling in somatic cell reprogramming. Cell Stem Cell 14:523–534.
Samavarchi-tehrani P, Golipour A, David L, Sung H, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL (2010) Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming. Stem Cell 7:64–77.
Sanges D, Cosma MP (2011) Reprogramming cell fate to pluripotency: The decision-making signalling pathways. Int J Dev Biol 54:1575–1587.
Sarkies P, Murat P, Phillips LG, Patel KJ, Balasubramanian S, Sale JE (2012) FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. Nucleic Acids Res 40:1485–1498.
Sasaki MS, Tonomura A (1973) A high susceptibility of Fanconi's anemia to chromosome breakage by DNA cross linking agents. Cancer Res 33:1829–1836.

Sawyer SL, Tian L, Kähkönen M, Schwartzentruber J, Kircher M, Majewski J, Dyment DA, Innes AM, Boycott KM, Moreau LA, Moilanen JS, Greenberg RA, Boycott K, Friedman J, Michaud J, Bernier F, Brudno M, Fernandez B, Knoppers B, Samuels M, Scherer S (2015) Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. Cancer Discov 5:135–142.
Schindler D, Friedl R, Gavvovidis I, Kalb R, Neveling K, Linka Y, Hanenberg H, Kubbies M, Hoehn H (2007) Applications of Cell Cycle Testing in Fanconi Anemia. In: Schindler D, Hoehn H (eds) Fanconi Anemia. A Paradigmatic Disease for the Understanding of Cance and Aging. Monogr Hum Genet. Basel, Karger, pp 110–130
Schindler D, Hoehn H (1988) Fanconi anemia mutation causes cellular susceptibility to ambient oxygen. Am J Hum Genet 43:429–35.
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) Generation of induced pluripotent stem cells from human terminally differentiated circulating t cells. Cell Stem Cell 7:11–13.
Seyschab H, Friedl R, Sun Y, Schindler D, Hoehn H, Hentze S, Schroeder-Kurth T (1995) Comparative evaluation of diepoxybutane sensitivity and cell cycle blockage in the diagnosis of Fanconi anemia. Blood 85:2233–2237.
Shen X, Do H, Li Y, Chung WH, Tomasz M, de Winter JP, Xia B, Elledge SJ, Wang W, Li L (2009) Recruitment of Fanconi Anemia and Breast Cancer Proteins to DNA Damage Sites Is Differentially Governed by Replication. Mol Cell 35:716–723.
Sherafat-Kazemzadeh R, Mehta SN, Care MM, Kim MO, Williams DA, Rose SR, Breech L, Callen M, Davies S, Gluckman J, Groh E, Harris R, Heist S, Hillard P, Hopkin R, Kelly P, Kiefhaber T, Leslie N, Mehlman C, Meyer R, Minevich E, Mueller R, Rutter M, Schmid C, Smith F, Strife F, Sutton M, Walker M, Williams D, Wood R, Yazigi N (2007) Small pituitary size in children with fanconi anemia. Pediatr Blood Cancer 49:166–170.
Shi Y, Do JT, Desponts C, Hahm HS, Scholer HR, Ding S (2008) A Combined Chemical and Genetic Approach for the Generation of Induced Pluripotent Stem Cells. Cell Stem Cell 2:525–528.
Shimamoto R, Amano N, Ichisaka T, Watanabe A, Yamanaka S, Okita K (2014) Generation and characterization of induced pluripotent stem cells from aid-deficient mice. PLoS One. doi: 10.1371/journal.pone.0094735
Shinagawa T, Takagi T, Tsukamoto D, Tomaru C, Huynh LM, Sivaraman P, Kumarevel T, Inoue K, Nakato R, Katou Y, Sado T, Takahashi S, Ogura A, Shirahige K, Ishii S (2014) Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells. Cell Stem Cell 14:217–227.
Shu J, Wu C, Wu Y, Li Z, Shao S, Zhao W, Tang X, Yang H, Shen L, Zuo X, Yang W, Shi Y, Chi X, Zhang H, Gao G, Shu Y, Yuan K, He W, Tang C, Zhao Y, Deng H (2013) Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell. doi: 10.1016/j.cell.2013.05.001
Shyh-Chang N, Daley GQ (2013) Lin28: Primal regulator of growth and metabolism in stem cells. Cell Stem Cell 12:395–406.
Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A (2008a) Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol 6:2237–2247.
Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, Wray J, Yamanaka S, Chambers I, Smith A (2009) Nanog Is the Gateway to the Pluripotent Ground State. Cell 138:722–737.
Silva SS, Rowntree RK, Mekhoubad S, Lee JT (2008b) X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. Proc Natl Acad Sci U S A 105:4820–5.
Singhal N, Graumann J, Wu G, Arauzo-Bravo MJ, Han DW, Greber B, Gentile L, Mann M, Scholer HR (2010) Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. Cell 141:943–955.
Smith ZD, Nachman I, Regev A, Meissner A (2010) Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. Nat Biotechnol 28:521–526.
Smith ZD, Sindhu C, Meissner A (2016) Molecular features of cellular reprogramming and development. Nat Rev Mol Cell Biol 17:139–154.

Smogorzewska A, Matsuoka S, Vinciguerra P, McDonald ER, Hurov KE, Luo J, Ballif BA, Gygi SP, Hofmann K, D'Andrea AD, Elledge SJ (2007) Identification of the FANCI Protein, a Monoubiquitinated FANCD2 Paralog Required for DNA Repair. Cell 129:289–301.
Soldner F, Hockemeyer D, Beard C, Gao Q, Bell GW, Cook EG, Hargus G, Blak A, Cooper O, Mitalipova M, Isacson O, Jaenisch R (2009) Parkinson's Disease Patient-Derived Induced Pluripotent Stem Cells Free of Viral Reprogramming Factors. Cell 136:964–977.
Somers A, Jean J, Sommer CA, Omari A, Ford CC, Jason A, Ying L, Sommer AG, Jean JM, Smith BW, Robert A, Demierre M, Weiss DJ, French DL, Gadue P, George J, Mostoslavsky G, Kotton DN (2010) Generation of Transgene-Free Lung Disease-Specific Human iPS Cells. 617–638.
Somers A, Jean JC, Sommer CA, Omari A, Ford CC, Mills JA, Ying L, Sommer AG, Jean JM, Smith BW, Lafyatis R, Demierre MF, Weiss DJ, French DL, Gadue P, Murphy GJ, Mostoslavsky G, Kotton DN (2010) Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. Stem Cells 28:1728–1740.
Sommer C a, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G (2009) Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells 27:543–9.
Sommer C a., Sommer AG, Longmire T a., Christodoulou C, Thomas DD, Gostissa M, Alt FW, Murphy GJ, Kotton DN, Mostoslavsky G (2010) Excision of reprogramming transgenes improves the differentiation potential of iPS cells generated with a single excisable vector. Stem Cells 28:64–74.
Soufi A, Donahue G, Zaret KS (2012) Facilitators and Impediments of the Pluripotency Reprogramming Factors ' Initial Engagement with the Genome. Cell 151:994–1004.
Soyombo AA, Wu Y, Kolski L, Rios JJ, Rakheja D, Chen A, Kehler J, Hampel H, Coughran A, Ross TS (2013) Analysis of induced pluripotent stem cells from a BRCA1 mutant family. Stem Cell Reports 1:336–349.
Spardy N, Duensing A, Hoskins EE, Wells SI, Duensing S (2008) HPV-16 E7 reveals a link between DNA replication stress, Fanconi anemia D2 protein, and alternative lengthening of telomere-associated promyelocytic leukemia bodies. Cancer Res 68:9954–9963.
Sridharan R, Gonzales-Cope M, Chronis C, Bonora G, McKee R, Huang C, Patel S, Lopez D, Mishra N, Pellegrini M, Carey M, Garcia B a, Plath K (2013) Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat Cell Biol 15:872–82.
Sridharan R, Tchieu J, Mason MJ, 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, Apostolou E, Ferrari F, Choi J, Walsh RM, Chen T, Ooi SSK, Kim SY, Bestor TH, Shioda T, Park PJ, Hochedlinger K (2012) Ascorbic acid prevents loss of Dlk1 - Dio3 imprinting and facilitates generation of all – iPS cell mice from terminally differentiated B cells. doi: 10.1038/ng.1110
Stadtfeld M, Hochedlinger K, Stadtfeld M, Hochedlinger K (2010a) Induced pluripotency : history , mechanisms , and applications Induced pluripotency : history , mechanisms , and applications. 2239–2263.
Stadtfeld M, Maherali N, Borkent M, Hochedlinger K (2010b) A reprogrammable mouse strain from gene-targeted embryonic stem cells. Nat Methods 7:53–5.
Stadtfeld M, Maherali N, Breault DT, Hochedlinger K (2008a) Defining Molecular Cornerstones during Fibroblast to iPS Cell Reprogramming in Mouse. Cell Stem Cell 2:230–240.
Stadtfeld M, Nagaya M, Utikal J, Weir G, Hochedlinger K (2008b) Induced Pluripotent Stem Cells Generated Without Viral Integration. Science (80-) 322:945–949.
Staerk J, Dawlaty MM, Gao Q, Maetzel D, Hanna J, Sommer CA, Mostoslavsky G, Jaenisch R (2010) Brief Report Reprogramming of Human Peripheral Blood Cells to Induced Pluripotent Stem Cells. Stem Cell 7:20–24.
Sterneckert JL, Reinhardt P, Schöler HR (2014) Investigating human disease using stem cell models. doi: 10.1038/nrg3764

Strati K, Li H, Murga M, Blanco R, Ortega S, Mario RM, Fernandez-capetillo O, Serrano M, Blasco MA (2009) A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. 460:1149–1153.
Su X, Huang J (2011) The Fanconi anemia pathway and DNA interstrand cross-link repair. 2:704–711.
Subramanyam D, Blelloch R (2009) Watching reprogramming in real time. Nat Biotechnol 27:997–998.
Sun N, Panetta NJ, Gupta DM, Wilson KD, Lee A, Jia F, Hu S, Cherry AM, Robbins RC, Longaker MT, Wu JC (2009) Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells. Proc Natl Acad Sci U S A 106:15720–5.
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 131:861–72.
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 5:1–9.
Takahashi K, Yamanaka S (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126:663–676.
Talmoudi F, Kilani O, Ayed W, Ben Halim N, Mellouli F, Torjmane L, Aissaoui L, Ben Youssef Y, Kammoun L, Ben Othmane T, Bejaoui M, Ben Romdhane N, Elloumi M, Hadiji S, Hentati S, Chemkhi I, Abidli N, Guermani H, Abdelhak S, Amouri A (2013) Differentiation of Fanconi anemia and aplastic anemia using mitomycin C test in Tunisia. Comptes Rendus - Biol 336:29–33.
Tanabe K, Nakamura M, Narita M, Takahashi K, Yamanaka S (2013) Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. Proc Natl Acad Sci 110:12172–12179.
Tanaka Y, Hysolli E, Su J, Xiang Y, Kim K, Zhong M, Li Y, Heydari K, Euskirchen G, Snyder MP, Pan X, Weissman SM, Park I (2015a) Transcriptome Signature and Regulation in Human Somatic Cell Reprogramming. Stem Cell Reports 4:1125–1139.
Tanaka Y, Hysolli E, Su J, Xiang Y, Kim K-Y, Zhong M, Li Y, Heydari K, Euskirchen G, Snyder MP, Pan X, Weissman SM, Park I-H (2015b) Transcriptome Signature and Regulation in Human Somatic Cell Reprogramming. Stem cell reports 4:1125–1139.
Tchieu J, Kuoy E, Chin MH, Trinh H, Patterson M, Sherman SP, Aimiuwu O, Lindgren A, Hakimian S, Zack JA, Clark AT, Pyle AD, Lowry WE, Plath K (2010) Female human iPSCs retain an inactive X chromosome. Cell Stem Cell 7:329–342.
Theunissen TW, Powell BE, Wang H, Mitalipova M, Faddah DA, Reddy J, Fan ZP, Maetzel D, Ganz K, Shi L, Lungjangwa T, Imsoonthornruksa S, Stelzer Y, Rangarajan S, D'Alessio A, Zhang J, Gao Q, Dawlaty MM, Young RA, Gray NS, Jaenisch R (2014) Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell 15:471–487.
Thompson LH, Hinz JM, Yamada NA, Jones NJ (2005) How Fanconi anemia proteins promote the four Rs: Replication, recombination, repair, and recovery. Environ. Mol. Mutagen. 45:128–142.
Tischkowitz M, Winqvist R (2011) Using mouse models to investigate the biological and physiological consequences of defects in the Fanconi anaemia/breast cancer DNA repair signalling pathway. J Pathol 224:301–305.
Tiscornia G, Vivas EL, Belmonte JCI (2011) Diseases in a dish: modeling human genetic disorders using induced pluripotent cells. Nat Med 17:1570–1576.
Tönnies H, Huber S, Kühl JS, Gerlach A, Ebell W, Neitzel H (2003) Clonal chromosomal aberrations in bone marrow cells of Fanconi anemia patients: Gains of the chromosomal segment 3q26q29 as an adverse risk factor. Blood 101:3872–3874.
Trounson A, Dewitt ND (2016) Pluripotent stem cells progressing to the clinic. Nat Rev Mol Cell Biol 17:194–200.
Tsai S-Y, Clavel C, Kim S, Ang Y-S, Grisanti L, Lee D-F, Kelley K, Rendl M (2010) Oct4 and klf4 reprogram dermal papilla cells into induced pluripotent stem cells. Stem Cells 28:221–8.

Tsubooka N, Ichisaka T, Okita K, Takahashi K, Nakagawa M, Yamanaka S (2009) Roles of Sall4 in the generation of pluripotent stem cells from blastocysts and fibroblasts. Genes to Cells 14:683–694.
Tulpule A, Kelley JM, Lensch MW, McPherson J, Park IH, Hartung O, Nakamura T, Schlaeger TM, Shimamura A, Daley GQ (2013) Pluripotent stem cell models of shwachman-diamond syndrome reveal a common mechanism for pancreatic and hematopoietic dysfunction. Cell Stem Cell 12:727–736.
Tulpule A, William Lensch M, Miller JD, Austin K, D'Andrea A, Schlaeger TM, Shimamura A, Daley GQ (2010) Knockdown of Fanconi anemia genes in human embryonic stem cells reveals early developmental defects in the hematopoietic lineage. Blood 115:3453–3462.
Urbach A, Schuldiner M, Benvenisty N (2004) Modeling for Lesch-Nyhan disease by gene targeting in human embryonic stem cells. Stem Cells 22:635–41.
Utikal J, Polo JM, Stadtfeld M, Maherali N, Kulalert W, Walsh RM, Khalil A, Rheinwald JG, Hochedlinger K (2009) Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature 1–5.
Valamehr B, Abujarour R, Robinson M, Le T, Robbins D, Shoemaker D, Flynn P (2012) A novel platform to enable the high-throughput derivation and characterization of feeder-free human iPSCs. Sci Rep 2:1–11.
Valamehr B, Abujarour R, Robinson M, Le T, Robbins D, Shoemaker D, Flynn P (2012) A novel platform to enable the high-throughput derivation and characterization of feeder-free human iPSCs. Sci Rep 2:1–11.
Valamehr B, Abujarour R, Robinson M, Le T, Robbins D, Shoemaker D, Flynn P (2012) A novel platform to enable the high-throughput derivation and characterization of feeder-free human iPSCs. Sci Rep 2:1–11.
Van de Vrugt HJ, Koomen M, Bakker S, Berns MAD, Cheng NC, van der Valk MA, de Vries Y, Rooimans MA, Oostra AB, Hoatlin ME, te Riele H, Joenje H, Arwert F (2011) Evidence for complete epistasis of null mutations in murine Fanconi anemia genes Fanca and Fancg. DNA Repair (Amst) 10:1252–1261.
Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, Erven V, Neveling K, Endt D, Kesterton I, Autore F, Fraternali F, Freund M, Hartmann L, Grimwade D, Roberts RG, Schaal H, Mohammed S, Rahman N, Schindler D, Mathew CG (2010) Mutation of the RAD51C gene in a Fanconi anemia-like disorder. Nat Genet 42:406–409.
Venkitaraman AR (2004) Tracing the network connecting BRCA and Fanconi anaemia proteins. Nat Rev Cancer 4:266–276.
Viswanathan SR, Daley GQ (2010) Lin28: A MicroRNA Regulator with a Macro Role. Cell 140:445–449.
Walden H, Deans a J (2014) The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. Annu Rev Biophys 43:257–278.
Wang AT, Smogorzewska A (2015) SnapShot : Fanconi Anemia and Associated Proteins. Cell 160:354–354.e1.
Wang AT, Taeho K, Wagner JE, Conti BA, Lach FP, Huang AL, Molina H, Sanborn EM, Zierhut H, Cornes BK, Abhyankar A, Sougnez C, Gabriel SB, Auerbach AD, Kowalczykowski SC, Smogorzewska A (2015) A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. Mol Cell 59:478–490.
Wang J, Otsuki T, Youssoufian H, Lo Ten Foe J, Kim S, Devetten M, Yu J, Li Y, Dunn D, Liu JM (1998) Overexpression of the Fanconi anemia group C gene (FAC) protects hematopoietic progenitors from death induced by Fas-mediated apoptosis. Cancer Res 58:3538–3541.
Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW, Orkin SH (2006) A protein interaction network for pluripotency of embryonic stem cells. Nature 444:364–368.
Wang L, Du Y, Ward JM, Shimbo T, Lackford B, Zheng X, Miao YL, Zhou B, Han L, Fargo DC, Jothi R, Williams CJ, Wade PA, Hu G (2014) INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. Cell Stem Cell 14:575–591.
Wang L, Wang L, Huang W, Su H, Xue Y, Su Z, Liao B, Wang H, Bao X, Qin D, He J, Wu W, So KF, Pan G, Pei D (2013) Generation of integration-free neural progenitor cells from cells in human urine. Nat Methods 10:84–9.

Wang M, McIntee EJ, Cheng G, Shi Y, Villalta PW, Hecht SS (2000) Identification of DNA adducts of acetaldehyde. Chem Res Toxicol 13:1149–1157.
Wang R, Liang J, Jiang H, Qin L-J, Yang H-T (2008) Promoter-dependent EGFP expression during embryonic stem cell propagation and differentiation. Stem Cells Dev 17:279–289.
Wang T, Chen K, Zeng X, Yang J, Wu Y, Shi X, Qin B, Zeng L, Esteban MA, Pan G, Pei D (2011) The Histone Demethylases Jhdm1a / 1b Enhance Somatic Cell Reprogramming in a Vitamin-C-Dependent Manner. Stem Cell 9:1–13.
Wang W (2007) Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. Nat Rev Genet 8:735–748.
Wang X, Andreassen PR, Andrea ADD (2004) Functional Interaction of Monoubiquitinated FANCD2 and BRCA2 / FANCD1 in Chromatin. Mol Cell Biol 24:5850–62.
Wang X, Peterson C a, Zheng H, Nairn RS, Legerski RJ, Li LEI, Wang XIN (2001) Involvement of Nucleotide Excision Repair in a Recombination-Independent and Error-Prone Pathway of DNA Interstrand Cross-Link Repair. Mol Cell Biol 21:713–720.
Wang Y, Adjaye J (2011) A Cyclic AMP Analog, 8-Br-cAMP, Enhances the Induction of Pluripotency in Human Fibroblast Cells. Stem Cell Rev Reports 7:331–341.
Ware CB, Nelson AM, Mecham B, Hesson J, Zhou W, Jonlin EC, Jimenez-Caliani AJ, Deng X, Cavanaugh C, Cook S, Tesar PJ, Okada J, Margaretha L, Sperber H, Choi M, Blau CA, Treuting PM, Hawkins RD, Cirulli V, Ruohola-Baker H (2014) Derivation of naive human embryonic stem cells. Proc Natl Acad Sci U S A 111:4484–9.
Warren L, Manos PD, Ahfeldt T, Loh Y, Li H, Lau F, Ebina W (2010) Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. 618–630.
Waters LS, Minesinger BK, Wiltrout ME, D'Souza S, Woodruff R V, Walker GC (2009) Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. Microbiol Mol Biol Rev 73:134–54.
Weber FA, Bartolomei G, Hottiger MO, Cinelli P (2013) Artd1/Parp1 regulates reprogramming by transcriptional regulation of Fgf4 Via Sox2 ADP-Ribosylation. Stem Cells 31:2364–2373.
Weinberger L, Ayyash M, Novershtern N, Hanna JH (2016) Dynamic stem cell states : naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Biol 17:155–169.
Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J, Kandoth C, Fulton RS, McLellan MD, Dooling DJ, Wallis JW, Chen K, Harris CC, Schmidt HK, Kalicki-Veizer JM, Lu C, Zhang Q, Lin L, O'Laughlin MD, McMichael JF, Delehaunty KD, Fulton LA, Magrini VJ, McGrath SD, Demeter RT, Vickery TL, Hundal J, Cook LL, Swift GW, Reed JP, Alldredge PA, Wylie TN, Walker JR, Watson MA, Heath SE, Shannon WD, Varghese N, Nagarajan R, Payton JE, Baty JD, Kulkarni S, Klco JM, Tomasson MH, Westervelt P, Walter MJ, Graubert TA, Dipersio JF, Ding L, Mardis ER, Wilson RK (2012) The origin and evolution of mutations in acute myeloid leukemia. Cell 150:264–278.
Wernig M, Lengner CJ, Hanna J, Lodato M a, Steine E, Foreman R, Staerk J, Markoulaki S, Jaenisch R (2008) A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. Nat Biotechnol 26:916–24.
Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature 448:318–324.
Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Ha R, Cowling R, Wang W, Liu P, Gertsenstein M, Kaji K, Sung H, Nagy A (2009) piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. doi: 10.1038/nature07863
Wong JCY, Alon N, Mckerlie C, Huang JR, Meyn MS, Buchwald M (2003) Targeted disruption of exons 1 to 6 of the Fanconi Anemia group A gene leads to growth retardation, strain-specific microphthalmia, meiotic defects and primordial germ cell hypoplasia. Hum Mol Genet 12:2063–2076.
Wu T, Liu Y, Wen D, Tseng Z, Tahmasian M, Zhong M, Rafii S, Stadtfeld M, Hochedlinger K, Xiao A (2014) Histone variant H2A.X deposition pattern serves as a functional epigenetic mark for distinguishing the developmental potentials of iPSCs. Cell Stem Cell 15:281–294.
Yamanaka S (2009) Elite and stochastic models for induced pluripotent stem cell generation. Nature 460:49–52.

Yamanaka S, Blau HM (2010) Nuclear reprogramming to a pluripotent state by three approaches. doi: 10.1038/nature09229
Yang C, Chang K, Rana TM (2014) Genome-wide Functional Analysis Reveals Factors Needed at the Transition Steps of Induced Reprogramming. 327–337.
Yang J, Van Oosten AL, Theunissen TW, Guo G, Silva JCR, Smith A (2010) Stat3 activation is limiting for reprogramming to ground state pluripotency. Cell Stem Cell 7:319–328.
Yoshida Y, Takahashi K, Okita K, Ichisaka T, Yamanaka S (2009) Hypoxia Enhances the Generation of Induced Pluripotent Stem Cells. Cell Stem Cell 5:237–241.
Young RA (2011) Control of the embryonic stem cell state. Cell 144:940–954.
Yu J, Hu K, Smuga-otto K, Tian S, Stewart R, Slukvin II, James 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 JL, Tian S, Nie J, Jonsdottir G a, Ruotti V, Stewart R, Slukvin II, Thomson J a (2007) Induced pluripotent stem cell lines derived from human somatic cells. Science 318:1917–20.
Yu J, Vodyanik MA, Smuga-otto K, Antosiewicz-bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA (2011) Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science (80-) 1917:1917–1920.
Yung SK, Tilgner K, Ledran MH, Habibollah S, Neganova I, Singhapol C, Saretzki G, Stojkovic M, Armstrong L, Przyborski S, Lako M (2013) Human Pluripotent Stem Cell Models of Fanconi Anemia Deficiency Reveal an Important Role for Fanconi Anemia Proteins in Cellular Reprogramming and Survival of Hematopoietic Progenitors. Stem Cells 31:1022–1029.
Zhang F, Fan Q, Ren K, Andreassen PR (2009) PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. Mol Cancer Res 7:1110–8.
Zhang QS, Eaton L, Snyder ER, Houghtaling S, Mitchell JB, Finegold M, Van Waes C, Grompe M (2008) Tempol protects against oxidative damage and delays epithelial tumor onset in Fanconi anemia mice. Cancer Res 68:1601–1608.
Zhang QS, Marquez-Loza L, Eaton L, Duncan AW, Goldman DC, Anur P, Watanabe-Smith K, Rathbun RK, Fleming WH, Bagby GC, Grompe M (2010) Fancd2-/- mice have hematopoietic defects that can be partially corrected by resveratrol. Blood 116:5140–5148.
Zhang S, Yajima H, Huynh H, Zheng J, Callen E, Chen HT, Wong N, Bunting S, Lin YF, Li M, Lee KJ, Story M, Gapud E, Sleckman BP, Nussenzweig A, Zhang CC, Chen DJ, Chen BPC (2011) Congenital bone marrow failure in DNA-PKcs mutant mice associated with deficiencies in DNA repair. J Cell Biol 193:295–305.
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 125:5609–20.
Zhang YZXRX (2014) Neurodegenerative diseases in a dish : the promise of iPSC technology in disease modeling and therapeutic discovery. doi: 10.1007/s10072-014-1989-9
Zhou H, Wu S, Joo JY, Zhu S, Han DW, Lin T, Trauger S, Duan L, Ding S, Bien G, Yao S, Zhu Y, Siuzdak G, Scho HR, Duan L, Ding S (2009) Generation of Induced Pluripotent Stem Cells Using Recombinant Proteins. Cell Stem Cell 4:1–4.
Zhou T, Benda C, Dunzinger S, Huang Y, Ho JC, Yang J, Wang Y, Zhang Y, Zhuang Q, Li Y, Bao X, Tse H-F, Grillari J, Grillari-Voglauer R, Pei D, Esteban M a (2012) Generation of human induced pluripotent stem cells from urine samples. Nat Protoc 7:2080–9.
Zhou W, Freed CR (2009) Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells. Stem Cells 27:2667–2674.
Zhu S, Li W, Zhou H, Wei W, Ambasudhan R, Lin T, Kim J, Zhang K, Ding S (2010) Reprogramming of human primary somatic cells by OCT4 and chemical compounds. Cell Stem Cell 7:651–655.

LIST OF PUBLICATIONS

- **Bharathan SP**, Nandy K, Palani D, Janet NB, Natarajan K, George B, Srivastava A, Velayudhan SR. (2017) Generation of an induced pluripotent stem cell line that mimics the disease phenotypes from a patient with Fanconi Anaemia by conditional complementation**.** Stem Cell Research. 20: 54-57; doi: 10.1016/j.scr.2017.02.006

- **Bharathan SP**, Manian KV, Aalam SM, Palani D, Deshpande PA, Pratheesh MD, Srivastava A and Velayudhan SR. (2017) Systematic evaluation of markers used for the identification of human induced pluripotent stem cells. Biology Open. 6: 100-108; doi:10.1242/bio.022111

- Manian KV, Aalam SM, **Bharathan SP**, Srivastava A, and Velayudhan SR. (2015). Understanding the Molecular Basis of Heterogeneity in Induced Pluripotent Stem Cells. Cellular Reprogramming. 17(6):427-440. doi:10.1089/cell.2015.0013

- Aalam SM, Manian KV, **Bharathan SP,** Mayuranathan T and Velayudhan, SR. (2015). Identification of Stable OCT4+NANOG- State in Somatic Cell Reprogramming. Cellular Reprogramming. 18(6):367-368. doi:10.1089/cell.2016.0018