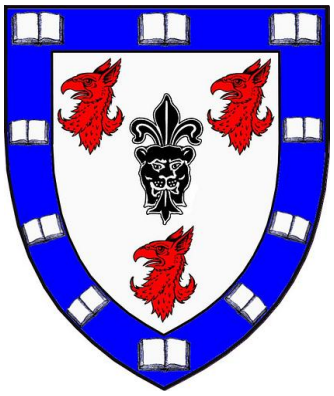
# The Role of Retinoic Acid Receptors in the Replacement of Oct4 during the Generation of Induced Pluripotent Stem Cells

**Ooi Yu Zhu, Jolene**



**This dissertation is submitted for the degree of Doctor of Philosophy**

**Homerton College, University of Cambridge**

**The Wellcome Trust Sanger Institute**

UNIVERSITY OF CAMBRIDGE

## DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text and acknowledgements.

This thesis does not exceed the word limit of 60,000 as set by the Degree Committee for the Faculty of Biology.

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my heartfelt thanks to my supervisors, Drs Neal Copeland, Nancy Jenkins and Pentao Liu. They have been extremely supportive and have offered voluminous amounts of advice during the course of my study. I am very thankful that I managed to secure a wonderful project in an institute that participates in the cutting edge of Science, with guidance from brilliant minds who possess a wealth of experience.

I would also like to show my gratitude to the Wellcome Trust Sanger Institute and Agency for Science Technology and Research for believing in my ability and funding my Ph.D. It has been a steep learning curve and I am appreciative of this wonderful opportunity. In particular, I would like to thank Drs Alex Bateman, Christina Hedberg-Delouka and Annabel Smith for their continuous support and kind understanding.

In addition, I am grateful to Drs Allan Bradley, Andy Futreal and John Stingl who have continually shown concern about my project and wellbeing as a student. I am also very thankful for the time that they took out of their busy schedules to learn about my work and for all the advice that they have offered.

This project could not have been completed without the help of several individuals. These include Wei Wang and Hui Liu who established the project and generated the screening construct before I joined the laboratory. Together with Jian Yang, they have provided constructive suggestions and advice as the project unfolded. In addition, I would like to express thanks to Stephen Rice who kindly assisted with the bioinformatics analysis of transposon integration sites. Yvette Hooks and Lia Campos were extremely helpful in the preparation and analysis of histological sections. Beiyuan Fu and Fengtang Yang provided immense help in the karyotype analysis of iPSCs and splenocytes. As a significant amount of work involved the generation and maintenance of mouse colonies, acknowledgements would

not be complete without thanks to members of the Research Support Facility (Mike, Marie, Rob, Andrea, Nick) who were extremely helpful and dedicated.

Last but not least, I would like to express my appreciation for my family, friends and partner for all the encouragement and concern that I have been showered with. Friends made at Homerton College, WTSI and CUMSA have played important roles during the four years of study, bearing with my complaints and amusing me with interesting anecdotes. My family, though 6000 miles away, has been constantly in touch to ensure that I was in good shape. Johan has been by my side, providing me with endless laughs and love.

The abovementioned people made a huge impact during my life as a Ph.D. student and I am grateful that I had the opportunity to be influenced by each and every one.

## ABSTRACT

Reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) has been accomplished through the delivery of four transcription factors, Oct4, c-Myc, Klf4 and Sox2. Although the reprogramming capacity of the four genes has been recapitulated across cell types and species, the molecular intricacies involved in the reprogramming process are not fully delineated. To address this, genetic analyses have been conducted to identify factors that can replace or enhance the defined set of four genes. Of the reprogramming cocktail, Oct4 has been most recalcitrant to replacement, as suggested by the inability of its family members, Oct1 and Oct6, to act as substitutes. The importance of Oct4 is highlighted in its obligatory need during development and embryonic lethality by ablation of Oct4. To decipher the role of Oct4 during the establishment of pluripotency, this project aims to identify genetic replacements of exogenous Oct4 during the generation of iPSCs using a genome-wide piggyBac transposon-assisted mutagenesis approach.

Transposons are mobile genetic elements that integrate into the host genome in a random fashion. The consequent disruption of genes can facilitate the identification of critical genes in a process-of-interest, and acts as the basis to my project. Mouse fibroblasts were mutagenised and assessed for their ability to generate iPSCs in the presence of ectopic c-Myc, Klf4 and Sox2. Basonuclin-2 (BNC2) and Retinoic Acid Receptor beta (RARβ) were identified to be novel candidate genes involved in reprogramming. Both genes were validated and able to generate iPSCs from mouse fibroblasts in cooperation with c-Myc, Klf4 and Sox2, hence abolishing the need for the ectopic expression of Oct4. To demonstrate the pluripotent potential of these iPSCs, I showed that these iPSCs resembled ESCs, and displayed the capacity of contributing to both somatic lineages and the germline, when introduced into mouse blastocysts.

As the retinoic acid pathway was recently implicated in the reprogramming process, the molecular mechanism behind the action of RARβ was pursued. Bioinformatics and experimental data have described several Retinoic Acid Response Elements (RAREs) in the

enhancer and promoter regions of Oct4. Given that RAREs are motifs recognisable by Retinoic Acid Receptors, I demonstrated that RARβ binds to the RARE within the distal enhancer of Oct4 and activates Oct4 expression. I also examined RARγ, another RAR family member, which has been described to promote reprogramming speed and efficiencies in synergy with LRH1. Similar to RARβ, RARγ also had the capacity to positively regulate Oct4 levels through the Oct4 distal enhancer. Moreover, a combination of RARγ, LRH1, Oct4, c-Myc, Klf4 and Sox2 (6F) represents the most efficient was to activate Oct4 expression in the luciferase assay.

As the 6F reprogramming cocktail (RARγ, LRH1, Oct4, c-Myc, Klf4, Sox2) improves the speed, quality, and efficiency of reprogramming, I next investigated if these properties were partly attributed to rapid epigenetic changes at the Oct4 distal enhancer. By examining this enhancer locus over the first three days of reprogramming, I demonstrated that the region was associated to activating histone marks within 24 hours of the ectopic expression of 6F, drawing a parallel to rapid reprogramming. In contrast, the Oct4 distal enhancer remained silent for the first 3 days when only four reprogramming factors (Oct4, c-Myc, Klf4, Sox2) were used. These findings represent a unique approach to dissect the kinetics of reprogramming through the observation of epigenomic changes at the Oct4 locus, culminating in the better understanding of molecular events during nuclear reprogramming.

**CONTENTS**

## FIGURES

### CHAPTER 1: INTRODUCTION



### CHAPTER 3: PIGGYBAC-ASSISTED SCREEN TO IDENTIFY GENETIC REPLACEMENTS OF EXOGENOUS OCT4 DURING REPROGRAMMING

**CHAPTER 4: VALIDATION OF CANDIDATE GENES**

**CHAPTER 5: RETINOIC ACID RECEPTORS and the REGULATION OF OCT4 EXPRESSION**

## TABLES

## ABBREVIATIONS

| | |
|---|---|
| 2i | 2 inhibitors (CHIR99021 and PD0325901) |
| 4F | Oct4, c-Myc, Klf4, Sox3 |
| 6F | Oct4, c-Myc, Klf4, Sox2, RARγ, LRH1 |
| Ac | Activator |
| ADH | Alcohol Dehydrogenase |
| AID | Activation-Induced cytidine Deaminase |
| AP | Alkaline Phosphatase |
| ATRA | All-Trans Retinoic Acid |
| AZA | 5-aza-cytidine |
| BAF | Brahma-related gene 1-Associated Factors |
| BMP | Bone Morphogenetic Protein |
| BNC1 | Basonuclin 1 |
| BNC2 | Basonuclin 2 |
| bp | base pairs |
| CAG | Composite CMV early enhancer/chicken beta actin promoter |
| Cald1 | Caldesmon 1 |
| CAT | Chloramphenicol AcetylTransferase |
| cDNA | complementary Deoxyribonucleic Acid |
| ChIP | Chromatin Immunoprecipitation |
| CKS | c-Myc, Klf4, Sox2 |

| | |
|---|---|
| CKS-B | c-Myc, Klf4, Sox2, BNC2 |
| CKS-Rβ | c-Myc, Klf4, Sox2, RARβ |
| cM | centimorgan |
| CMV | Cytomegalovirus |
| CR | Conserved Region |
| CRABP | Cellular Retinoic Acid Binding Protein |
| CRBP | Cellular Retinol Binding Protein |
| DBD | DNA Binding Domain |
| DCAF5 | DDB1-CUL4A-associated factor 5 |
| DE | Distal Enhancer |
| DNA | Deoxyribonucleic Acid |
| DOXYCYCLINE | Doxycycline |
| dpc | days post coitum |
| DR | Direct Repetition |
| Ds | Dissociator |
| EC cells | Embryonic Carcinoma cells |
| EG cells | Embryonic Germ cells |
| EGFP | Enhanced Green Fluorescence |
| ENU | N-ethyl-N-nitrosourea |
| EpiSCs | Epiblast Stem cells |
| ER | Estrogen Receptor |

| | |
|---|---|
| ESCs | Embryonic Stem cells |
| ESRRβ | Estrogen Receptor Related Receptor-beta |
| FA | Fanconi's Anemia |
| FGF | Fibroblast Growth Factor |
| FX | Fragile X |
| Gb | Gigabases |
| H3K4me1 | monomethylated lysine 4 of histone 3 |
| H3K4me2 | dimethylated lysine 4 of histone 3 |
| H3K4me3 | trimethylated lysine 4 of histone 3 |
| HBV | Hepatitis B Virus |
| HSV | Herpes Simplex Virus |
| Id | Inhibitor of differentiation |
| iPSCs | Induced Pluripotent Stem cells |
| IRES | Internal Ribosomal Entry Site |
| kb | kilobases |
| Klf4 | Kruppel-Like Factor 4 |
| LBD | Ligand Binding Domain |
| LIF | Leukemia Inhibitory Factor |
| lincRNA | large intergenic non-coding Ribonucleic Acid |
| LRH1 | Liver Receptor Homologue |
| MAPK | Mitogen Associated Protein Kinase |

| | |
|---|---|
| Mb | Megabases |
| MEF | Mouse Embryonic Fibroblasts |
| MET | Mesenchymal-to-Epithelial Transition |
| mFx | murine factors |
| MHC | Major Histocompatibility Complex |
| minP | minimal promoter |
| mRNA | messenger Ribonucleic Acid |
| MSCV | Murine Stem Cell Virus |
| OCKS | Oct4, c-Myc, Klf4, Sox2 |
| Oct4 | Octamer-binding transcription factor 4 |
| PB | piggyBac |
| PBase | piggyBac transposase |
| PBS | Phosphate Buffered Saline |
| PCR | polymerase chain reaction |
| PE | Proximal Enhancer |
| PGCs | Primordial Germ Cells |
| PGK | Phosphoglycerate Kinase |
| PI3K | Phosphoinositide 3-Kinase |
| POU | Pit-Oct-Unc |
| PP | Proximal Promoter |
| RA | Retinoic Acid |

| | |
|---|---|
| RALDH | Retinaldehyde Dehydrogenase |
| RARα | Retinoic Acid Receptor alpha |
| RARβ | Retinoic Acid Receptor beta |
| RARE | Retinoic Acid Response Element |
| RARγ | Retinoic Acid Receptor gamma |
| RBP4 | Retinol Binding Protein 4 |
| RDH | Retinol Dehydrogenase |
| RNA | Ribonucleic Acid |
| RORα | Retinoic Acid Receptor-related Orphan Receptor alpha |
| RT | Reverse Transcriptase |
| RT-PCR | reverse transcription-polymerase chain reaction |
| rtTA | Reverse Tetracycyline Transactivator |
| RXR | Retinoic X Receptor |
| SA | Splice Acceptor |
| SCNT | Somatic Cell Nuclear Transfer |
| SD | Splice Donor |
| SF | Steel Factor |
| SF1 | Steroidogenic Factor 1 |
| Sox2 | SRY (sex determining region Y)-box 2 |
| TAF | TBP Associated Factors |
| TALEs | transcription activator-like effectors |

| | |
|---|---|
| TBP | TATA Binding Proteins |
| Tc | Tetracycline |
| tetR | tet Repressor |
| TGFβR | Transforming Growth Factor beta Receptor |
| TRE | Tetracycline Response Element |
| TSS | Transcriptional Start Site |
| tTA | Tetracycline Transactivator |
| UTR | Untranslated Regions |
| VPA | Valproic Acid |

# CHAPTER 1 : INTRODUCTION

## 1.1 Pluripotency

### 1.1.1 History behind the study of Vertebrate Development

The development of vertebrates involves the orchestration of a series of steps in a tightly regulated process that determines cell lineage specification into endodermal, ectodermal and mesodermal derivatives. Imprinted into the operational dogma of modern developmental biology, conception of these notions has been accompanied by a history of key observations and controversies (**Figure 1.1**).

Originating from examinations of the chick embryo, Aristotle witnessed the development of a palpitating heart, head and eyes, laying ink on a clean palette of embryology (Aristotle, 1979). With the proposition of epigenesis, he described development as a sequential process involving the formation of organs to construct a complete organism. Almost two thousand years after these initial recordings, the field was reawakened and the mechanisms behind these phenomena were questioned. To examine the root of development, Girolamo Fibrici performed dissections on cadavers of pregnant mammals, providing comparisons between anatomical structures of uteri (Fibrici, 1621). This work was advanced by his student William Harvey who hypothesised the presence of female germ cells within uteri that hold the capacity to constitute a new organism (Harvey, 1651). Furthermore, identification of budding and subdivision during primary stages of embryonic development of the chick led him to be a strong advocate of epigenesis. These findings revived Aristotle's theory and provoked collision against preformation views. Preformationism was held as the dominant perception of development, and describes the existence of a miniature organism that expands without increasing complexity within the germ cell. Although epigenesis perceptions were resurrected, it was not received warmly. Transformation of the field of development biology was invoked by subsequent experiments led by Caspar Frederich Wolff and Karl Ernst von Baer. Using plants as a surrogate organism for study, Wolff explained the ability of differentiated plant root to regenerate a new organism. This study was traversed to chick embryos where Wolff

studied the formation of embryonic kidneys (Wolff, 1774). Building on the scaffold of information uncovered by his predecessors, von Baer discovered the presence of primitive germ cells and ultimately eclipsed any influence of preformationism (von Baer, 1827).

Simultaneous work by Rathke and Pander (von Baer, 1828; Pander, 1817) reinforced the epigenesis model. Through the examination of chick embryos, Pander noted that specific regions within the embryo gave rise to specialised organs. This led to the notion of three germ layers and is routinely referenced in modern developmental biology. Inspecting a broader spectrum of vertebrates, Rathke observed similarities in the development of frogs, salamanders, fish, birds, and mammals, and described the formation of pharyngeal arches, skull and reproductive, excretory, and respiratory systems, completely refuting notions of preformation and providing the basis of developmental biology as we know it today.

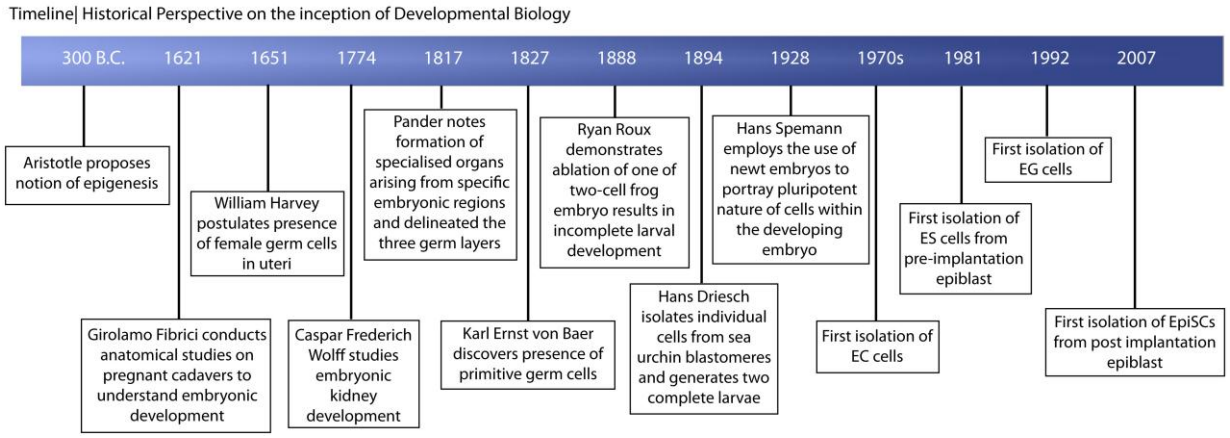



**Figure 1.1 Timeline displaying milestones that led to the inception of developmental biology**

### 1.1.2 Pluripotent stem cells

Vertebrate development as delineated by epigenesis proposes the presence of pluripotent cells which participate in the contribution of the cellular entirety of an adult organism. Analagous to the establishment of epigenesis, the proposition of pluripotent stem cells was also met with opposing opinions. Coining the germ plasm theory, August Weismann believed that development occurred in a unilateral fashion where germ cells were the only cells capable of hereditary potential and somatic cells received permanent modifications which rendered them committed to perform limited functions. This was further supported by Wilhelm Roux who showed that ablation of one cell of the two-cell frog embryo with a hot needle resulted in incomplete larval development (Roux, 1888). These beliefs were quickly confounded by Hans Driesch who turned to an invertebrate system and conducted a similar experiment using sea urchin blastomeres. Isolation of individual cells at its 2-cell stage culminated in the generation of two but small complete larvae (Driesch, 1894). In parallel, Hans Spemann duplicated the findings using newt embryos (Spemann, 1928), where separation of the embryo at its 2-cell stage gave rise to two intact organisms. Poised with micro-surgical skills, Spemann extended his findings in embryology through the constriction of developing embryos using baby hair. Using this technique, he was able to restrict the position of the nucleus to one side of the cytoplasm, leaving solely cytoplasmic material in the other. As the embryo divides successively into the 16-cell stage, one cell would escape to the other side and the hair knot would be tightened to prevent any further transfer to occur. As a result, twin larvae would form, suggesting the pluripotent capacity of cells within the developing embryo. These pieces of evidence refuted the germ plasm theory, illuminating a new era of embryology study.

#### 1.1.2.1 Embryonic Carcinoma Cells

To reinforce the notion of pluripotency, pluripotent cells have been successfully established on the Petri dish (**Figure 1.2**). Teratocarcinomas are tumours discovered in humans and mice (Stevens and Little, 1954; Giovanella et al., 1974). Inspection of these cellular masses reveals the presence of a plethora of organised structures, including teeth, fingers and hair, suggesting the presence of pluripotent cells within the tumour. This was

corroborated through the determination that intraperitoneal injection of a single cell could generate teratocarcinoma consisting of an array of differentiated tissues (Kleinsmith and Pierce, 1964). Likewise, grafting of mouse embryos into adult mice also leads to the formation of teratomas, reinforcing the existence of all-encompassing cells (Stevens et al, 1970; Solter et al., 1970). Teratomas in both contexts have been successfully maintained in culture (Kahan and Ephrussi, 1970; Rosenthal et al., 1970; Evans, 1972; Martin and Evans, 1974). Designated as embryonic carcinoma (EC) cells, these cells exhibit pluripotent properties including the ability to form teratomas in immune-compromised mice and serve as the first platform to study embryonic development of mice *in vitro*.

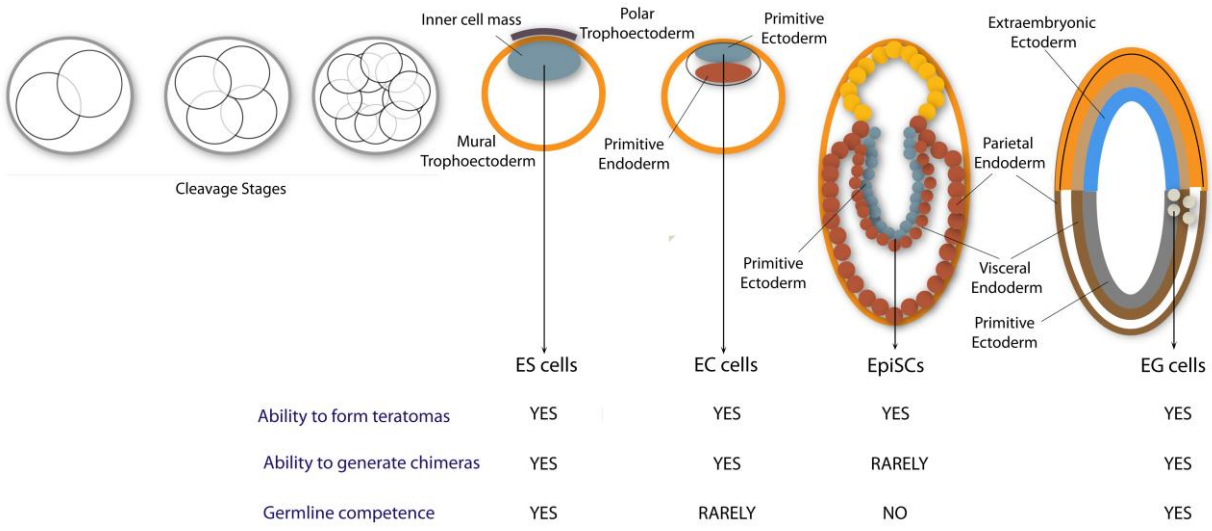



**Figure 1.2 Comparison of pluripotent cells established from various stages of embryonic development.** Once fertilised, the zygote undergoes several rounds of cell division and forms an inner cell mass, as indicated in grey. Cultivation of the inner cell mass results in the establishment of ESCs. The inner cell mass generates two cell types, (i) the primitive endoderm (brown) which forms the parietal and visceral endoderm, and (ii) the primitive ectoderm (grey) which gives rise to the fetus. EpiSCs and EC cells can be achieved from various stages of the primitive ectoderm. As the mesoderm starts to form at the boundary between the endoderm and ectoderm, primordial germ cells are generated and EG cells can subsequently be isolated.

### 1.1.2.1 Embryonic Stem Cells

Soon after, the inner cell mass of mouse blastocysts was demonstrated to be sustained on a petri dish, recapitulating an early developmental event *in vitro* (Evans and Kaufman, 1981; Martin, 1981). Labelled as embryonic stem (ES) cells, these cells were competent at contributing to the three germ layers in teratomas, when injected into immune-compromised or syngenic mice. Furthermore, re-introduction of these cells into the mouse blastocyst led to the formation of high percentage chimeras, indicating their ability to participate in normal murine development, a property not frequently shared with EC cells. To satisfy stringent pluripotent stipulations, ESCs were also studied for their ability to contribute to the germline and an intact embryo. The former was approached through the cross of chimeras to phenotypically distinct wildtype mice (Bradley et al., 1984), whereas the latter was addressed through tetraploid complementation assays (Nagy et al., 1990; Nagy et al., 1993).

Amenable to modifications, ESCs embody a useful tool for genetic alterations (reviewed in Evans, 2011). With the elucidation of the genetic composition of the mouse in 2002 (Waterston et al., 2002), the genomic content of ESCs has been frequently disrupted in a precise fashion to study gene function. The capacity for germline transmission results in the establishment of intact mice harbouring any desired genetic mutation in the germline (Smithies et al., 1985; Kuehn et al., 1987; Thomas and Capecchi, 1987; Skarnes et al., 2011).

### 1.1.2.3 Epiblast Stem Cells

ESCs represent a subset of cells isolated from the epiblast in pre-implantation blastocysts and depict a primitive developmental stage of the developing embryo. To recapitulate late phases, two independent groups have segregated the columnar epithelial epiblast of the early post-implantation embryo and cultivated it on a petri dish (Tesar et al., 2007; Brons et al., 2007). Termed as Epiblast stem cells (EpiSCs), they behave distinctly from ESCs and are rarely able to generate chimeras. However, both ESCs and EpiSCs are competent

in multi-lineage differentiation, where injection of these cells into immune-compromised mice results in the development of teratomas comprised of tissue types characteristic of the three germ layers (Tesar et al., 2007; Brons et al., 2007).

Examination of the molecular circuitry within these cells revealed some similarities to ESCs, where the core transcriptional machinery consisting of Oct4, Sox2 and Nanog was expressed (Boyer et al., 2005; Loh et al., 2006; Wang et al., 2006). However, ESCs and EpiSCs exhibit disparities in transcript and epigenetic levels of markers associated to the inner cell mass and early germ layers (Tesar et al., 2007), highlighting distinctions in their original developmental stages.

Differences between ESCs and EpiSCs also lie in epigenomic modifications of pluripotency markers. Although Oct4 is detected at similar levels in ESCs and EpiSCs, Oct4 is differentially regulated, as disclosed by chromatin immunoprecipitation experiments which illustrate that the proximal and distal enhancers of Oct4 control its expression in EpiSCs and ESCs respectively (Tesar et al., 2007). These pieces of evidence uncover variations in the maintenance of different stages of pluripotency, placing emphasis on an assortment of developmental cues spanning embryogenesis.

#### 1.1.2.4 Embryonic Germ Cells

Pluripotent stem cells divergent from the mouse blastocyst were first derived from primordial germ cells (PGCs) (Matsui et al., 1992; Resnick et al., 1992). Emergent at 7 days post coitum (dpc), these cells are represented by a small population of alkaline-phosphatase positive cells (Ginsburg et al., 1990). In a span of six days, these cells undergo extensive proliferation every 16 hours to comprise of 25,000 PGCs (Tam and Snow, 1981), and eventually reside in either the testis or ovary of the mouse.

Extraction of PGCs at 8.5-12.5 dpc from the posterior fragment of the embryo and cultivation in the presence of soluble factors such as leukemia inhibitory factor (LIF), steel factor (SF) and fibroblast growth factor (FGF), results in a population of cells that exhibits self-renewal and limitless proliferation (Godlin et al., 1991; Matsui et al., 1991; Matsui et al., 1992; Resnick et al., 1992). Coined as embryonic germ (EG) cells, these cells resemble ESCs and are capable of generating chimeras and contributing to the mouse germline (Matsui et al., 1992; Stewart et al., 1994; Labosky et al., 1994).

### 1.1.3 Chemicals and Pathways associated with Murine Pluripotent Stem Cells

Pluripotent cells exist in a fleeting manner within the mouse embryo, placing emphasis on the remarkable extension of their life in culture. Supporting chemicals or matrix are necessary for the maintenance of pluripotency in culture, as exemplified in the reliance on fibroblasts and serum in primary studies describing ESCs, EC cells and EG cells, suggesting a non-cell autonomous mechanism in self-renewal. Extrication of components that support pluripotency, in concert with our current understanding of developmental pathways, can lead to the improvement of growth parameters of pluripotent cells, and augment our knowledge on embryonic development.

Stemming from the discovery that medium conditioned by Buffalo rat liver cells was sufficient to retain pluripotency (Smith and Hooper, 1987), the active component necessary for this phenomenon was narrowed down to leukemia inhibitory factor (LIF) (Smith et al., 1988; Williams et al., 1988). The importance of LIF has been portrayed by its obligatory need in the culture medium (Stewart et al., 1992), acting through gp130 and the recruitment of JAK kinase and STAT3 (Yoshida et al., 1994; Niwa et al., 1998). The ability of LIF to maintain pluripotency in culture is mirrored in a physiological context, where LIF and gp130 are expressed in early embryos and during diapause (Nichols et al., 2001). This is reinforced by the observation that female mice expressing truncated LIF mutants are fertile but the inability of blastocysts to implant into the uterus leads to developmental arrest. On the contrary, transient expression of LIF exhibits a rescued

phenotype where implantation and development of the blastocysts are permitted (Stewart et al., 1992).

Although LIF/gp130 and their related pathways are pivotal in the maintenance of pluripotency, the use of chemically defined basal media supplemented with N2, B27 and LIF is unable to impede differentiation of ESCs into neuronal derivatives (Ying et al., 2003). This propensity to differentiate can be restrained by the addition of bone morphogenetic protein (BMP), an anti-neural factor in vertebrate development (Wilson and Hemmati-Brivanlou, 1995). Functioning through Inhibitor of differentiation (Id), BMP together with LIF are sufficient to drive ESCs into self-renewal without differentiation (Ying et al., 2003).

Interaction between LIF and gp130 triggers a conflicting response, where both the JAK/STAT and ERK1/2 pathways are activated (Burdon et al., 1999; Kunath et al., 2007). As the latter stimulates differentiation, ERK or FGF inhibitors were demonstrated to circumvent this impediment and support the maintenance of ESCs (Ying et al., 2008). Activation of the Wnt pathway through the inhibition of GSK3β also assists in the sustenance of an undifferentiated state. The effects of ERK and GSK3β inhibition (2i) are compounded when used in combination, and results in a homogenous population of primitive cells designated as ground state pluripotency (Silva et al., 2008). Mirroring this *in vivo*, addition of 2i to early mouse embryos in culture causes an expansion of the Nanog-expressing epiblast at the expense of the hypoblast and trophoectoderm compartments (Nichols et al., 2009b). Growth media containing 2i and LIF has also made it possible to derive ESCs from mouse strains, such as CBA and NOD, and rats, which have been recalcitrant to previous methods (Ying et al., 2008; Nichols et al., 2009a; Buehr et al., 2008; Li et al., 2008).

The culture conditions of EpiSCs are distinct from ESCs. FGF and Activin are necessary to preserve EpiSCs (Brons et al., 2007; Tesar et al., 2007), whereas the addition of 2i and

LIF results in cell death (Guo et al., 2009). In contrast, addition of an Activin inhibitor leads to widespread differentiation, suggesting reliance on Nodal/Activin signalling (Tesar et al., 2007).

The disparities reflected by dissimilar developmental potential and growth conditions of mouse ESCs and EpiSCs has led to the notion of naïve and primed pluripotency (Nichols and Smith, 2009c). Originating from the pre-implantation epiblast, ESCs display complete pluripotent potential and are capable of germline contribution. In contrast, EpiSCs derived from the post-implantation epiblast are incapable of germline contribution, exhibiting limited pluripotent potential.

### 1.1.4 Establishment of ESCs from other species

After successful isolation and culture of mouse ESCs, there have been several attempts at engineering an equivalent in various species, including rodents such as hamsters and rats (Doetschman et al., 1988; Buehr et al., 2008; Li et al., 2008), non-rodents such as rabbits, minks, chickens, pigs and cows (Graves and Moreadith, 1993; Sukoyan et al., 1993; Pain et al., 1996; Notarianni et al., 1990; Notarianni et al., 1991; Saito et al., 1992; Stice et al., 1996), and primates such as rhesus monkeys and the common marmosets (Thomson et al., 1995; Thomson et al., 1996).

Heightened interest in the generation of an array of pluripotent stem cells can be attributed to its potential to differentiate into an array of cell types, representative of the three germ layers. Application of this technology to humans illuminates the possibility of regenerative medicine. To address this, human ESCs were derived from cleavage stage human embryos that were acquired from *in vitro* fertilization donors (Thomson et al., 1998). However, human ESCs are distinct from mouse ESCs, where LIF and BMP signalling pathways are incapable of sustaining undifferentiated human ESCs (Daheron et al., 2004; Humphrey et al., 2004; Xu et al., 2002; Gerami-Naini et al., 2004). Instead, growth conditions and gene expression patterns of human ESCs are reminiscent of EpiSCs (Beattie et al., 2005; Vallier et al., 2005; Wang et al., 2005; Xu et al., 2005; Adewumi et al., 2007). In a bid to confront

the possibility of an unexplored naïve human pluripotent state which resembles mouse ESCs, recent studies have described the presence of a Nanog-expressing epiblast compartment in human embryos that is recalcitrant to inhibitors against FGF, ERK and GSK3. This observation highlights a stark contrast to conventional human pluripotent cells which readily differentiate in similar conditions (Roode et al., 2012; Kuijk et al., 2012). An additional study proposes that establishment of human ES cell lines implicates the transition into a post-ICM intermediate displaying X-reactivation, reminiscent of mouse ESCs but not EpiSCs (O'Leary et al., 2012). Overall, these findings broaden our understanding of human embryonic development and surmise the existence of a population of naïve pluripotent stem cells which have yet to be captured from the developing embryo.

## 1.2 Cellular Plasticity

### 1.2.1 Cellular Regeneration

To pursue survival, organisms have evolved strategies to maintain their functional and physical integrities when damage is inflicted or as aging takes its toll. This is exemplified in mammals by the continuous rejuvenation of the circulatory system, gut and skin epithelium. Time-lapse imaging and visualisation of cultured endothelial sheet colonies reveals the disengagement of tight junctions and the display of erythroid and monocytic haematopoietic antigens (Eilken et al., 2009). In parallel, conversion of glucagon-producing α-cells into pancreatic insulin-producing β-cells upon a diphtheria toxin insult has also been described (Thorel et al., 2010). These examples reflect a dynamic environment to maintain homeostasis.

However the inherent ability of mammals to regenerate functional organs is limited, and can be better delineated using invertebrates and amphibians. The study of cellular regeneration can be traced to the 18$^{th}$ century when Abraham Trembley pioneered studies on Hydra, a small freshwater animal. Through the dissection of its polyps, Trembley noticed that both halves developed into complete organisms, documenting the first observations of cellular regeneration (Lenhoff and Lenhoff, 1986). This phenomenon was mirrored in a spectrum of invertebrates such as planarians and annelid worms. Vertebrates

such as amphibia, have also demonstrated their ability to replace lost limbs, tails, lens, retina, and several internal organs and repair skeletal muscle, peripheral nervous system and CNS. The ability of vertebrate regeneration is illustrated in newts by the regeneration of its eye lens upon lentectomy (Tsonis et al., 2004). Pigmented epithelial cells contribute to the formation of the replacement lens by losing their pigment and exiting from the quiescent state. Thereafter, the cells proliferate and redifferentiate into lens epithelial cells and subsequently into the mature lens fibers that comprise the regenerated lens.

The plasticity of cellular identity was once perceived to be impossible amidst the proposition of Weismann's germ plasm theory. This ability to transcend lineage restriction barriers and generate distinct cell types in a natural setting has ignited an undulating interest in the field of reprogramming. Emanating from naturally occurring examples of cellular regeneration, attempts at replicating these observations through artificial means have led to a series of landmark studies.

### 1.2.2 Nuclear Reprogramming

Nuclear reprogramming is depicted by the transition between distinct cell types introduced by switches in gene expression patterns (**Figure 1.3**). Contesting against the germ plasm theory, preliminary nuclear reprogramming experiments sought to contradict the belief that cells accumulate permanent modifications during development and prevent reversion to a primitive state. The successful transfer of nuclei from cells of a developing embryo into enucleated oocytes provided confounding evidence and the notion behind the germ plasm theory was effectively eradicated. This procedure was coined as somatic cell nuclear transfer (SCNT) and paved the way for cell fusion and lineage switching experiments, culminating in the recent discovery that introduction of transcription factors can drive somatic cells into pluripotency.

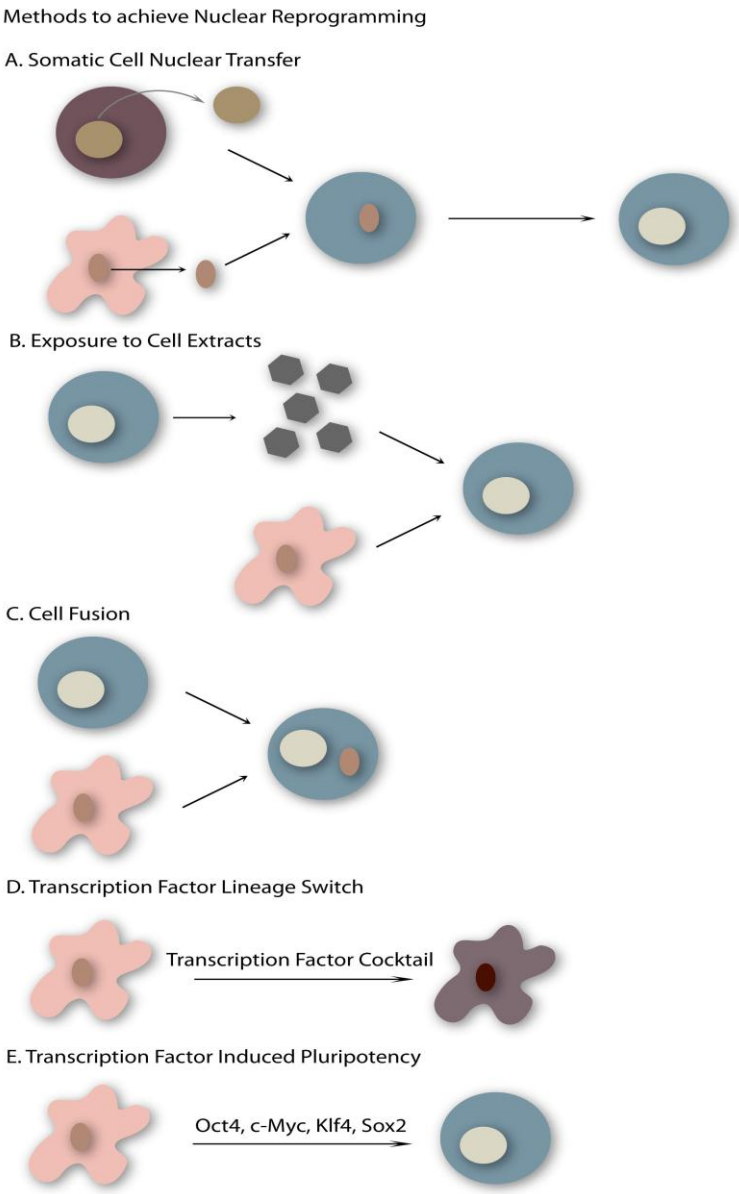



**Figure 1.3 Methods to achieve nuclear reprogramming** Nuclear reprogramming is defined as the transition between distinct cell types introduced by switches in gene patterns. This can be accomplished through various methods. (A) Somatic cell nuclear transfer entails the transfer of a somatic cell nucleus into an enucleated oocyte. (B) Exposure of cellular components within pluripotent cells to somatic cells triggers the loss of somatic identity. (C) The fusion between somatic and pluripotent cells results in the generation of heterokaryons that exhibit pluripotent properties. (D) Expression of transcription factors that are lineage specific results in the conversion of cell types across lineage barriers (E) Introduction of Oct4, c-Myc, Klf4 and Sox2 into somatic cells results in the reprogramming into an ES-like state.

### 1.2.2.1 Somatic Cell Nuclear Transfer

The first observations of nuclear reprogramming *in vitro* were witnessed in 1952 when Briggs and King successfully transplanted a nucleus from a *Rana pipiens* embryo into an enucleated oocyte to produce adult organisms (Briggs and King, 1952). A later study by the pair revealed that endoderm nuclei from an early tail-bud stage resulted in arrest at blastula gastrula and early-neurula stages (King and Briggs, 1955). Surprisingly, when similar experiments were carried out in *Xenopus laevis*, nuclei from terminally differentiated intestinal epithelial cells were successfully transplanted into irradiated oocytes and developed into fertile male and female frogs (Gurdon, 1958; Gurdon, 1962). Intestinal epithelium cells of feeding tadpoles which correspond to terminally differentiated endoderm cells gave rise to normal feeding tadpoles at a frequency of 1.5% when transplanted into unfertilised oocytes. This was noticeably lower than nuclear transfers from the blastula and gastrula endoderm which led to a proportion of 36% feeding tadpoles. Interestingly, serial transplantations of intestine nuclei which promoted abnormal development after the first transfer could support the formation of normal feeding tadpoles, suggesting that disparities in efficiencies between differentiated and embryonic nuclei were partially explained by non-genetic causes such as technical limitations and incompatible mitotic properties. Collectively, first and serial transplantations, in concert with grafts, generated functional tadpoles at a frequency of 30% (Gurdon, 1962; Byrne et al., 2003).

Shortly after came reports of successful transplantations of nuclei from a spectrum of cell origins, such as kidney, lung and skin (Gurdon et al., 1975; Laskey and Gurdon, 1970). Non-dividing erythrocytes from *Rana pipiens* were demonstrated to produce normal larvae upon serial transplantations (DiBerardino and Hoffner, 1983). Nuclear transfer from myotomes could also effectively produce swimming tadpoles at a frequency of 2% (Gurdon et al., 1984). Transcription of muscle differentiation markers ceased upon entry into the oocyte and re-activated once gastrulation occurred. Taken together, these studies suggest that the chemical milieu in the oocyte is able to perturb the transcriptional machinery of the donor cell and reverse its differentiation status.

Alternatively, SCNT can be performed using oocytes in their first meiotic prophase (Byrne, 2003). This manipulation results in the swelling of the cell due to a stark increase in mRNA synthesis. DNA replication is reversely synchronised and pauses, contrary to the use of oocytes in the second meiotic metaphase which leads to multiple rounds of cell division and little transcription. Although mechanistically distinct, this reaffirms the intrinsic ability of the oocyte to influence the donor cell nucleus.

Concurrent to the described experiments, other groups ventured away from amphibians. Using one-cell mouse embryos as donor and host, adult mice were generated at a success ratio of 13% (McGrath and Solter, 1983). In contrast, transplantation of nuclei from tail-tip fibroblasts plunged the reprogramming efficiency to 0.5%. Mirroring this phenomenon in sheep, Willadsen demonstrated the ability to produce cloned lambs by transplanting a nucleus of an 8-cell embryo to enucleated oocyte (Willadsen, 1986). Mammary gland cell nuclei were also competent at producing an adult fertile female sheep, albeit at a low efficiency of 0.2% (Wilmut et al., 1997).

Several attempts have been made with the use of human oocytes. Injection of donor nuclei into enucleated oocytes gave rise to blastocysts at low frequencies (Stojkovic et al., 2005; Hall et al., 2007; French et al., 2008). Low reprogramming efficiency was alleviated by leaving the nuclei of the oocyte intact (Noggle et al., 2011), leading to the successful development of blastocysts. The inner cell mass was stable in culture and triploid in karyotype, with one set of diploid chromosomes from the somatic nucleus and one set of haploid chromosomes from the oocyte. Competent in producing teratomas consisting of cell types representative of all three germ layers, these cells display similar transcriptional patterns to pluripotent stem cells derived from *in vitro* fertilised blastocysts.

These findings present evidence that somatic cells encompass the capacity to return to pluripotency and can be evaluated in two ways. First, cells which have veered away from pluripotency require strict conditions for reprogramming. Second, the egg naturally

contains factors which dominate the performance of the somatic nucleus. Stripping the composition of the egg would allow us to comprehend the intricacies that trigger rapid dedifferentiation. To this end, histones 1 and B4 and the Tet proteins have been illustrated to play roles both in the oocyte and male pronucleus, highlighting the importance of epigenetic remodelling interactions during reprogramming (Teranishi et al. 2004; Gao et al., 2004; Becker et al., 2005; Jullien et al., 2010; Maki et al., 2010; Inoue and Zhang, 2011; Wu and Zhang, 2011).

#### 1.2.2.2 Exposure to Cell Extracts and Cell Fusion

Re-illustrating the innate ability of the chemical milieu within an oocyte to instigate transcriptional alteration, exposure of cytoplasmic contents from *Xenopus laevis* oocytes and early embryos to human somatic cells elicited the expression of pluripotency markers (Hansis et al., 2004). This capacity extended till late blastula-stages where extracts were inhibitory to reprogramming. Human embryonic kidney cells incubated with cell extracts from embryonic carcinoma cells for a mere hour also resulted in transcriptional activation and partial demethylation of pluripotency genes within a week (Flasza et al., 2003; Taranger et al., 2005; Freberg et al., 2007).

Corroborating these observations, cell fusions between human lymphocytes and mouse ESCs have been competent at triggering the reactivation of Oct4. However, the level of transcriptional induction was <1% compared to human ESCs (Pereira et al., 2008). The capacity of cell fusion induced reprogramming has been further exemplified through the combination of male ESCs and female thymocytes. Resulting heterokaryons expressed pluripotency genes and reactivated the silent X chromosome (Tada et al., 2001). In addition, spontaneous fusion between mouse neural progenitors and mouse ESCs led to the silencing of neural markers (Ying et al., 2002). Replicated in human cells, human ESCs fused to human foreskin fibroblasts generated hybrid pluripotent cells that exhibited the capacity to develop into three germ layers (Cowan et al., 2005). In general, cell fusion experiments displayed trends analagous to SCNT. First, terminally differentiated cells were more demanding to transform. Second, cell fusion leads to rapid reprogramming, where genes are expressed within two days and independent of DNA replication (Pereira et

al., 2008; Do and Schöler, 2010). Despite success in generating pluripotent stem cells, reprogramming through cell fusion is inefficient (~1%), and creates a bottleneck during the purification of stable hybrids for analysis. Introduction of selection markers in the two starting populations can overcome this problem, but does not alleviate the eventual low population of reprogrammed cells.

Apart from establishing pluripotency, other cell types have also been successfully reprogrammed through cell fusion. Stemming from early experiments, hetreokaryons between erythrocytes from a mature hen and HeLa cells have been described to swell in size and prompt transcriptional activity in terminally differentiated erythrocytes, reminiscent of the observations in SCNT (Harris, 1965; Harris et al., 1966). Surprisingly, the combination of human amniotic cells and mouse muscle cells triggered the expression of genes specific to muscle cells, albeit the naïve nature of amniotic cells (Blau et al., 1983). This draws attention to inherent factors in somatic cells that safeguard the integrity and lineage of the cell. Besides sharing similarities to SCNT, it is useful to note that cell fusion encompasses unique qualities such as the ability to generate heterokaryons using multiple cell-nuclei. As a result, gene dosage influences the reprogramming efficiencies and the presence of the dominant nuclei in more than two fold will delay transcriptional activation (Pavlath and Blau, 1986).

#### 1.2.2.3 Transcription Factor Induced Lineage-Switch

It is evident from cell fusion experiments that there are specific components within each branch of cell lineage, shielding it from external influences and guarding its designed function. These intrinsic cell specific properties can be harnessed to redirect differentiation signals. Using observations that DNA demethylating agent, 5-azacytidine, enables efficient reprogramming into muscle lineage (Chiu and Blau, 1985), Weintraub and colleagues identified MyoD as a master regulator of myogenesis, and its ectopic expression forced recipient non-muscle cells to convert into myocytes (Weintraub et al., 1989). Although cells from three germ layers were successfully reprogrammed, starting material originating from the mesoderm was effectively converted, whereas cells from endoderm and ectoderm

lineages retained their original gene signatures. This suggests the necessity of additional factors to induce a complete reform of the pre-existing gene expression profile.

Intuitively, forced expression of transcription factors that naturally occur during developmental programs can incite lineage-switches. As described above, expression of MyoD can redirect non-muscle cells into the muscle lineage (Weintraub et al., 1989). In addition, overexpression of adipocyte-specific nuclear hormone receptor, PPARγ2, triggers adipogenesis in fibroblasts (Tontonoz et al., 1994). Although single genetic factors have been described to control cell fate in adipocytes and myocytes, dermal fibroblasts require a combination of Gata4, Mef2c and Tbx5 to be converted into cardiomyocytes (Ieda et al., 2010). Likewise, ectopic expression of separate sets of transcription factors have successfully triggered transdifferentiation of dermal fibroblasts into neuron-like, blood progenitors, insulin secreting β-cells, brown adipose cells and cardiac-like myocytes (Vierbuchen et al., 2010; Szabo et al., 2010; Zhou et al., 2008; Kajimura et al., 2009; Song et al., 2012). Further exploration in the realm of lineage switching has placed emphasis on the importance of spatial, temporal and quantitative control of transcription factor expression in the generation of an array of cell types (Kulessa et al., 1995; Xie et al., 2004; Laiosa et al., 2006; Iwasaki et al., 2006).

Lineage switching mimics the acquisition of pluripotency via SCNT and cell fusion closely. First, cells which have climbed the differentiation ladder face more roadblocks during reprogramming (Chickarmane et al., 2009). Second, lineage-switching occurs without the need for cell division. Markers of progenitor cells are often not detected, suggesting that reprogramming may occur in a direct fashion without the presence of an undifferentiated intermediate (Thorel et al., 2010). However, this does not exclude the possibility of a transient intermediate state.

#### 1.2.2.4 Transcription Factor Induced Pluripotency

The ability of transcription factors to impose a lineage switch suggests that the right cocktail of factors would be able to erase all differentiation marks in somatic cells, driving the cell to pluripotency. In 2006, Takahashi and Yamanaka conducted a systematic screen

of 24 transcription factors associated with pluripotency to identify a distinct set of genes able to activate the endogenous Fbx15 gene locus of mouse embryonic fibroblasts (MEFs) (Takahashi and Yamanaka, 2006). This was narrowed to 4 genes, Oct4, c-Myc, Klf4 and Sox2, and produced cells that resembled ESCs. These cells were able to differentiate into three germ layers when injected into immune-compromised mice, portraying its pluripotent potential.

Shortly after, the phenomenon was replicated in human fibroblasts (Takahashi et al., 2007). Concurrently, an independent laboratory reported the generation of human ES-like cells through a discrete set of genes, Oct4, Sox2, Nanog, and Lin28 (Yu et al., 2007), reinforcing the notion that somatic cells can be reprogrammed to an undifferentiated state. These groundbreaking discoveries indicated that mere four genes can elicit pluripotency, igniting a surge of interest in this sphere of reprogramming. For the ease of nomenclature, these cells have since been termed as induced pluripotent stem (iPS) cells.

Several advancements have been made to the generation of iPSCs (**Figure 1.4**). First, alternative delivery methods were discovered, circumventing the initial use of retroviruses that leaves a permanent footprint in the recipient genome. As the reprogramming cocktails are made up of genes implicated in diseases such as cancer, the use of iPSCs may lead to oncogenic transformation (Okita et al., 2007). Surrogate gene delivery tools include polycistronic lentiviruses, piggyBac mediated transposition and non integrating approaches such as adenoviruses, Sendai viruses, episomal DNA, repeated transfections of mRNA or plasmid DNA and purified proteins (Sommer et al., 2009; Carey et al., 2009; Fusaki et al., 2009; Stadtfeld et al., 2008; Yu et al., 2009; Warren et al., 2010; Okita et al., 2008; Kim et al., 2009a; Zhou et al., 2009; Woltjen et al., 2009; Kaji et al., 2009; Yusa et al., 2009). Albeit the low efficiency of producing iPSCs, the employment of non integrating methods allows the generation of colonies which are completely free of exogenous material.

Second, the selection procedure for reprogrammed cells has been enhanced. Population of cells selected based on re-activation of Fbx15 displayed an incomplete erasure of somatic methylation marks and could not contribute to the mouse germline (Takahashi and Yamanaka, 2006). Subsequent studies used the activation of endogenous Nanog, Oct4 or

Rex1 as an indication of pluripotency (Okita et al., 2007; Meissner et al., 2007; Wang et al., 2011c).

To illustrate the robustness of the protocol, somatic cells originating from the three germ layers have been described to be pliable to reprogramming (Aasen et al., 2008; Shi et al., 2008; Hanna et al., 2008; Aoi et al., 2008). In addition, iPSCs from an array of organisms, including pigs, monkeys and rats, have also been achieved (Esteban et al., 2009; Liu et al., 2008; Liao et al., 2009). Akin to SCNT and cell fusion experiments, terminally differentiated cells have been proven difficult to reprogram (Hanna et al., 2008). Cell types which possess high intrinsic expression of certain factors negotiate the need for the complete set of reprogramming cocktail (Tsai et al., 2010; Kim et al., 2009b; Kim et al., 2009c; Giorgetti et al., 2009). Overall, these studies display the inherent ability of a spectrum of somatic cell types to revert to its pluripotent state.

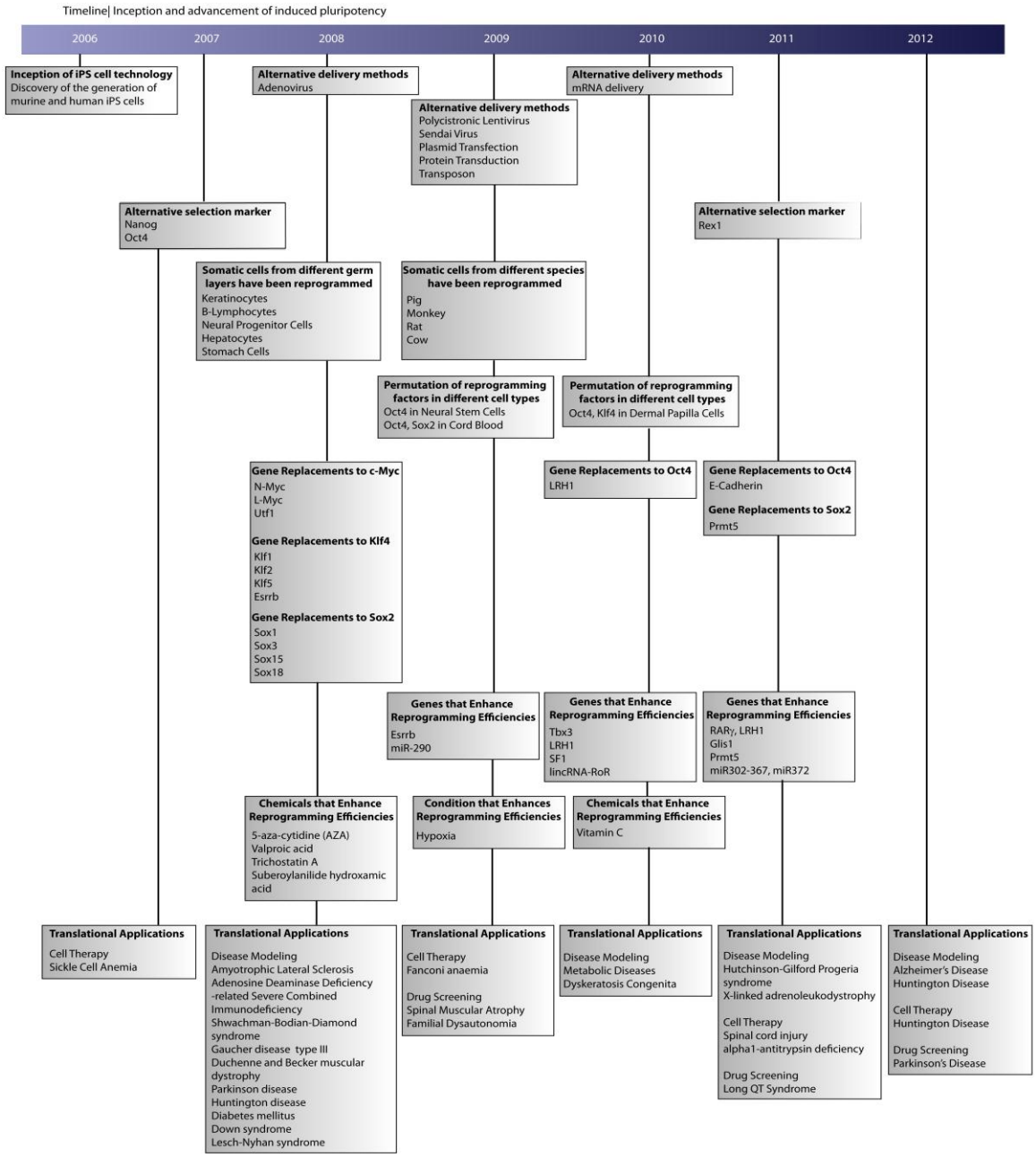



**Figure 1.4 Timeline highlighting the milestones achieved since the inception of transcription factor induced pluripotency** Transcription factor induced pluripotency was conceived in 2006 through the introduction of four transcription factors, Oct4, c-Myc, Klf4 and Sox2 into mouse fibroblasts. Since its conception, a variety of somatic cell types have been successfully reprogrammed. Various refinements have been demonstrated to enhance the initial protocol. This includes the use of alternative delivery methods and selection markers. In addition, supplementary factors have been identified to improve reprogramming efficiencies or replace the ectopic necessity of the initial set of transcription factors. Ultimately, application of this technology has been demonstrated in disease modelling, drug screening and cell therapy.

### 1.3 Induced Pluripotent Stem Cells

The advent of iPSC technology holds many promises in the clinic. A plethora of cell types, ranging from dermal fibroblasts, primary hepatocytes, blood and keratinocytes exhibits amenability to reprogramming (Dimos et al., 2008; Liu et al., 2010; Loh et al., 2009; Giorgetti et al., 2009; Aasen et al., 2008). In addition, manipulation of fibroblasts derived from patients inflicted with disease allows faithful recapitulation of the disease phenotype *in vitro* (Park et al., 2008; Dimos et al., 2008; Rashid et al., 2010; Jang et al., 2011; Agarwal et al., 2010; Zhang et al., 2011; Shi et al., 2012; HD iPSC consortium, 2012). The potential of patient-derived iPSCs can be illuminated two-fold, as reflected in the ability to act as a platform for drug screening (Itzhaki et al., 2011; Moretti et al., 2011; Ebert et al., 2009; Lee et al., 2009; Cooper et al., 2012), and cell therapy (Hanna et al., 2007; Raya et al., 2009; Nori et al., 2011; Yusa et al., 2011; An et al., 2012).

Although the use of iPSCs in the clinic is regarded attractively, there are several anomalies which have to be addressed. Cells derived from patients suffering from Fanconi's Anemia (FA) and Fragile X (FX) syndrome are not easily amenable (Raya et al., 2009; Urbach et al., 2010). These bottlenecks impede disease modelling and amendments to the existing protocol have to be taken into consideration. In addition, although comprehensive studies of iPSCs and ESCs reveal minute differences in phosphoproteomic and transcriptomic components that were statistically disregarded (Phanstiel et al., 2011), epigenomic analysis at high resolution expose subtle differences between iPSCs and ESCs (Lister et al., 2011; Stadtfeld et al., 2010) and continuous *in vitro* culture could incur genomic aberrations (Gore et al., 2011; Hussein et al., 2011; Lister et al., 2011). These discrepancies could account for functional disparities such as epigenetic memory (Kim et al., 2010; Ohi et al., 2011; Polo et al., 2010) and immunogenicity (Zhao et al., 2011), imposing barriers in clinical applications.

From disease modelling to drug screening and cell therapy, the ability to generate pluripotent stem cells from human somatic cells displays volumes of potential. However, conceived in 2006, this platform is still in its infancy and harbours many unknowns. To address our lack of knowledge on the reprogramming process, it is crucial to elucidate

factors and chemicals which play a complementary role to the defined cocktail of reprogramming factors.

### 1.3.1 Molecular mechanisms behind the initial set of four reprogramming factors

Before we endeavour to expand our knowledge of the reprogramming process, it is crucial to understand the roles undertaken by each of the four transcription factors. Naturally present in ESCs, the regulatory roles of these transcription factors create a ripple effect, instigating a cascade of molecular events that preserves the functional integrity of the ESCs (Loh et al., 2006; Nichols et al., 1998; Avilion et al., 2003).

Oct4 and Sox2 act in concert with Nanog to activate several pluripotency-associated elements (Boyer et al., 2005; Kim et al., 2008; Masui et al., 2007). The expression levels of both genes are tightly regulated, where the absence of Oct4 and Sox2 in ESCs results in the loss of pluripotency and differentiation into trophoblast cells (Chew et al., 2005; Nichols et al., 1998; Niwa et al., 2000; Tomioka et al., 2002). Conversely, surplus amounts of Oct4 results in differentiation into the primitive endoderm (Niwa et al., 2000; Kopp et al., 2008), while elevated Sox2 levels induce neuroectoderm, trophectoderm and mesoderm lineage specification.

Klf4 belongs to the family of Kruppel like factors which display functional redundancy to maintain pluripotency (Jiang et al., 2008). Occupying genomic sites similar to the Oct4-Sox2-Nanog cluster, it acts as part of the foundation preserving the undifferentiated environment within ESCs (Jiang et al., 2008). The Klf family has been described to act downstream of Oct4, p53 and leukemia inhibitory factor (LIF)/Stat pathways (Hall et al., 2009; Rowland et al., 2005). In addition, overexpression of Klf4 leads to sustained levels of Oct4 and inhibits differentiation of ESCs.

c-Myc is a helix-loop-helix/leucine zipper transcription factor that mediates pleiotropic cellular functions including metabolism, cell cycle, splicing, translation and oncogenic transformation (Kim et al., 2008; Sridharan et al., 2009; Klein, 1983). It has been implicated in the transcriptional elongation in ESCs (Rahl et al., 2010) and shares similar

genetic targets with other pluripotency associated transcription factors (Chen et al., 2008; Kim et al., 2008; Dejosez et al,, 2008). In ESCs, c-Myc resembles Klf4, acting in the LIF/Stat pathway (Cartwright et al., 2005) and opposing anti-proliferative properties of p21Cip1 (Claassen and Haan, 2000). c-Myc also recruits histone acetylase complexes (Bouchard et al., 2001; Frank et al., 2001), unfolds the chromatin structure in somatic cells and enables under-privileged sites to undergo remodelling.

The four transcription factors participate in the maintenance of the pluripotent state, where Oct4, Sox2 and Klf4 constitute the core apparatus and c-Myc manipulates a wider scope of cellular functions. Although the genomic targets of each transcription factor act as jigsaw pieces to the pluripotency network puzzle, they do not necessarily reflect the molecular changes coupled to reprogramming in vivo. To acquire a mechanistic view of reprogramming, additional components that participate in the process have been unravelled.

### 1.3.2 Chemicals that enhance reprogramming

The route to pluripotency is multi-faceted. In addition to the identification of the four transcription factors that encompass the reprogramming mix, elucidation of chemicals that affect the reprogramming process may illuminate potential approaches to improve the reprogramming protocol.

Transcription factor induced pluripotency occurs at sub-optimal frequencies of 0.01-0.2% (reviewed by Hochedlinger and Plath, 2009). Forced re-expression of reprogramming factors in secondary fibroblasts resulted in 20% of cells displaying early pluripotency markers but a mere 1.2% of these cells re-activating their endogenous Nanog locus after 16 days (Mikkelsen et al., 2008). Amalgamating this to the knowledge that primitive cells which encompass fewer somatic epigenetic marks are more pliable to dedifferentiation and iPSCs which harbour incomplete erasure of silencing modifications were unable to contribute to the mouse germline (Takahashi and Yamanaka, 2006), it is tempting to speculate that overcoming the epigenetic barrier would enhance reprogramming.

To address this, a DNA methyltransferase inhibitor, 5-aza-cytidine (AZA), was described to propel partially reprogrammed cells into pluripotency and improved the efficiency of producing iPSC colonies by 4-fold (Mikkelsen et al., 2008; Huangfu et al., 2008a). Given that c-Myc regulates histone deacetylation and greatly improves reprogramming efficiencies, histone deacetylase inhibitors, such as valproic acid (VPA), trichostatin A, and suberoylanilide hydroxamic acid, have been proven to be efficacious in increasing the kinetics and efficiencies of generating iPSCs (Huangfu et al., 2008a). Moreover, VPA ablates the need for c-Myc and Klf4, producing ES-like cells solely in the presence of Oct4 and Sox2.

As reprogramming leads to the accumulation of reactive oxygen species, antioxidant vitamin C was described to increase the efficiency of generating iPSCs (Esteban et al., 2010). This is triggered by vitamin C-dependent H3K36me2/3 demethylation and the repression of p53/p21 and Ink4/Arf loci (Wang et al., 2011a).

Besides the addition of chemicals, altering environmental conditions during reprogramming can also benefit the generation of iPSCs. In a natural setting, development of embryos occurs in physiologically hypoxic conditions. Recapitulating low oxygen concentrations during reprogramming experiments increases the efficiency of obtaining iPSCs (Yoshida et al., 2009). Furthermore, resultant female human iPSCs exhibit two activated copies of X chromosomes, indicating the attainment of ground state pluripotency (Lengner et al., 2010). Possible explanations behind this phenomenon include decreased accumulation of chromosomal abnormalities (Forsyth et al., 2006) and protection against spontaneous differentiation (Ezashi et al., 2005; Prasad et al., 2009).

### 1.3.3 Cellular processes associated to reprogramming

Although the search for chemicals that enhance the creation of iPSCs has yielded a considerable amount of information, other avenues of studying the dedifferentiation process have also provided us with much insight. Two complementary approaches were used to determine the importance of mesenchymal-to-epithelial transition (MET) during the reprogramming process (Samavarchi-Tehrani et al., 2010; Li et al., 2010). Molecular

dissection of the process identified BMP as a key mediator, leading to a later discovery that BMP can replace Klf4 in the reprogramming process, and its expression with Oct4 alone is sufficient to generate iPSCs (Chen et al., 2011). These findings were corroborated by two independent studies describing the essential roles of E-cadherin (Redmer et al., 2011) and the TGFβ pathway (Ichida et al., 2009) in driving partially reprogrammed cells to pluripotency.

Secondary cells isolated from chimeras act as a homogenous platform to study the mechanics behind induced pluripotency. Silencing p53 or p21, or over-expressing Lin28 in secondary B-cells improved reprogramming efficiencies in a cell division dependent manner (Hanna et al., 2009). This validated an initial finding that silencing p53 in primary fibroblasts improved reprogramming conditions (Zhao et al., 2008). An addition five studies also illustrated the role of DNA damage response and immortalization in reprogramming (Utikal et al., 2009; Marión et al., 2009; Li et al., 2009; Kawamura et al., 2009; Hong et al., 2009).

The importance of cell division in generating iPSCs is disparate to the mechanics behind SCNT and cell fusion, insinuating that the current recipe of transcription factors is unable to replicate natural reprogramming events. Interestingly, delivery of Nanog into secondary B-cells increased the number of reprogrammed colonies, independent of cell division (Hanna et al., 2009). Nanog is known to play essential roles in reprogramming and cell fusion (Yu et al., 2007; Chambers et al., 2007; Silva et al., 2006; Silva et al., 2009) despite a deficiency in its expression in mouse oocytes (Chambers et al., 2003; Yamaguchi et al., 2005), postulating that it functions in a discrete role, providing a new dimension to the process.

### 1.3.4 Genetic factors that improve the reprogramming process

As a surrogate method to peel away additional layers of complexity behind induced pluripotency, attempts to explore beyond the initial set of 24 ES cell specific transcription factors have been made to identify genetic components that contribute to the reprogramming process.

To delve into an expanded list of genes, several screens of varying scopes have been conducted. From a narrow screen examining three genes known to participate in primodial germ cell development and EG cell derivation, Prmt5 was discovered to improve reprogramming efficiencies and ablation of its expression led to decreased number of reprogrammed colonies (Nagamatsu et al., 2011). In contrast, the study of a wide library consisting of 1,437 transcription factors revealed that Glis1 can increase the generation of iPSC colonies (Maekawa et al., 2011). Additional screens have delineated the roles played by Tbx3, Esrrb and Nr5a2 in the acquisition of pluripotency (Han et al., 2010b; Heng et al., 2010; Feng et al., 2009).

Using an unbiased piggyBac assisted approach, co-operation between RARγ and Nr5a2 was demonstrated to augment the transcriptional activity of Oct4 and trigger rapid dedifferentiation (Wang et al., 2011c and unpublished). Although implication of the retinoic acid pathway in the reprogramming process is surprising, it is not completely unexpected (Ben-Shushan et al., 1995; Barnea et al., 2000). As the dedifferentiation process is distinct to the maintenance of ES cell pluripotency (Mansour et al., 2012), this study may shed light from a new perspective on the molecular mechanisms behind reprogramming.

Aside from transcription factors, microRNAs have also been addressed. The miR-290 cluster constitutes more than 70% of the entire miRNA population in mouse ESCs and its overexpression amplifies the reprogramming efficiencies due to Oct4, Sox2 and Klf4, but not c-Myc, potentially due to its regulatory role (Judson et al., 2009). miR302–367 and miR-372 also display abilities to enhance the dedifferentiation process (Subramanyam et al., 2011; Liao et al., 2011). Two independent studies have reported that microRNAs alone could reprogram both mouse and human somatic cells (Miyoshi et al., 2011; Anokye-Danso et al., 2011). Extending this discovery, a new class of large intergenic non coding RNAs (lincRNAs) associated with epigenetic regulators were portrayed to be involved in pluripotency. Of 10 lincRNAs identified to be differentially expressed in iPS and ESCs, deregulation of lincRNA-RoR in human fibroblasts has been described to affect iPSC formation (Loewer et al., 2010).

### 1.3.5 Genetic factors that replace the exogenous requirement of reprogramming factors

The roles of each of the four reprogramming factors have been well described in literature and their genomic targets in ESCs are clearly depicted. However, the reprogramming process begins with somatic cells and adds a layer of uncertainty to our understanding of their functions while triggering dedifferentiation (**Figure 1.5**). To address this, genes which act as proficient substitutes to the conventional reprogramming components may provide insight to the roles performed by the four transcription factors during the acquisition of pluripotency.

Inspection of functional redundancy among family members reveals that relatives of Klf4, Sox2 and c-Myc could successfully operate as substitutes, whereas Oct1 and Oct6 were unable to replace Oct4 (Nakagawa et al., 2008).

In addition to family members, directed screens have been employed to uncover genes competent at substituting components of the reprogramming cocktail. Through the study of 18 ES cell-associated transcription factors that were absent in the original screen (Takahashi and Yamanaka, 2006), Esrrb was demonstrated to possess the ability of replacing the exogenous requirement of Klf4 during reprogramming, albeit at a lower reprogramming efficiency of 50% (Feng et al., 2008). In addition, ES cell differentiation induced by triple knockdown of Klf2, Klf4 and Klf5 could be rescued through the co-expression of Esrrb, indicating that Esrrb regulates cellular pathways independent of Klf family members. Both a genomic target (Jiang et al., 2008) and a regulatory component of Klf4, the study suggests a reciprocal relationship between the two genes, hence providing some mechanistic insight to the role of Klf4 during reprogramming.

In a separate directed screen, 17 transcription factors were utilised to determine their effects on the reprogramming efficiencies of adult human foreskin fibroblasts (Zhao et al., 2008). Two genes, Utf1 and p53 siRNA, were isolated and their co-operation with Oct4, c-Myc, Klf4 and Sox2 could escalate reprogramming efficiencies by 100-fold. It was

subsequently discovered that both genes could effectively reprogram human fibroblasts in the absence of c-Myc.

Two independent studies have identified factors capable of substituting the exogenous requirement of Sox2. As described above, three genes associated with primodial germ cell development and EG cells derivation were tasked to trigger pluripotency in MEFs (Nagamatsu et al., 2011). The co-introduction of the three genes with Oct4, c-Myc, Klf4 and Sox2 led to the formation of iPSC colonies but only a small population expressed Nanog. Circumventing this, Prmt5, Oct4 and Klf4 alone could generate a higher proportion of iPSCs which expressed Nanog. In addition, silencing of Prmt5 led to a reduction of Nanog-positive iPSCs, reinforcing its importance in reprogramming. The second study explored a chemical library to search for compounds that can elicit dedifferentiation and activate the expression of endogenous Oct4 in the presence of VPA (Ichida et al., 2009). Three compounds were isolated from the screen, where two were not able to induce reprogramming in the absence of VPA. The remaining candidate was an inhibitor of TGFβR (RepSox) that could trigger the acquisition of pluripotency in the absence of Sox2. Further investigation of RepSox revealed that it could promote reprogramming at intermediate stages through the induction of Nanog. The genetic intersection between Nanog and Sox2 has been reaffirmed independently (Rodda et al., 2005).

In search of replacements of exogenous Oct4, a directed screen using 19 transcription factors was conducted (Heng et al., 2010). Initially, the set of candidate genes was examined for its ability to enhance reprogramming efficiencies piloted by the conventional reprogramming cocktail. Nr5a2 was identified and subsequently verified to exhibit the ability of generating iPSCs in the presence of Klf4 and Sox2 alone. In addition, Nr5a1, a close relative of Nr5a2, was also able to substitute for exogenous Oct4. Using chromatin immunoprecipitation, the authors postulate that Nr5a2 regulates and acts in concert with Nanog to drive pluripotency. In addition, Nr5a2 has previously been demonstrated to play a positive regulatory role on the expression of Oct4 (Gu et al., 2005), suggesting a spectrum of effects triggered by the introduction of Nr5a2. Extending these findings, E-cadherin was also identified to elicit reprogramming events in the absence of exogenous Oct4 and function as an indicator for fully reprogrammed cells (Redmer et al., 2011).

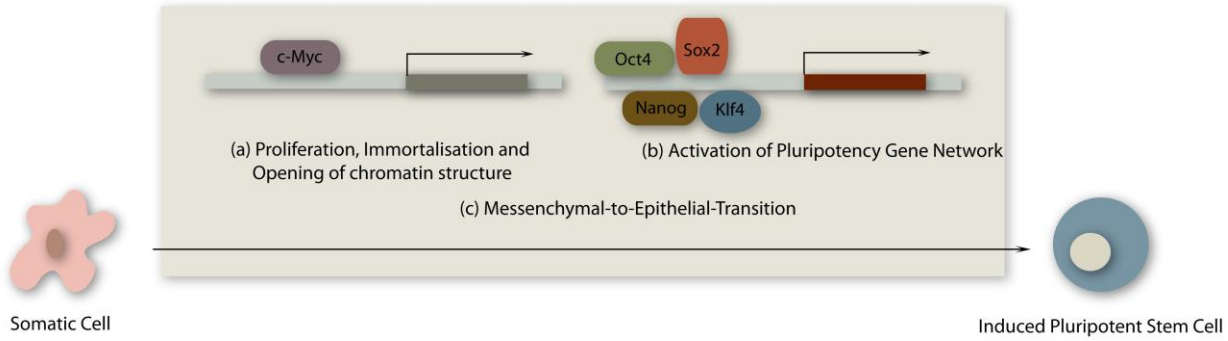



**Figure 1.5 Schematic diagram portraying the current knowledge on the molecular mechanisms behind induced pluripotency** The conversion of somatic cells to iPSCs involves a radical change in the epigenomic landscape and cellular components. To accomplish this, the cell has to overcome cellular processes such as proliferation, immortalisation, unraveling of chromatin, activation of pluripotency transcriptional networks and the initiation of mesenchymal-to-epithelial-transition. Although the changes in cellular processes have been delineated, there exists a grey area over the molecular intricacies behind induced pluripotency.

## 1.4 Oct4 and Reprogramming

### 1.4.1 Discovery of Oct4

Oct4 (Octamer-binding transcription factor-4) belongs to class V of the POU transcription factor family (Okamoto et al., 1990; Schöler et al., 1990a; Schöler et al., 1990b; Rosner et al., 1990). POU transcription factors are DNA binding proteins that recognise a consensus octamer motif ATGCAAAT (Klemm et al., 1994) within cis-acting elements and elicit transcriptional activation. The identification of the binding motif is facilitated by the POU domain. As a bipartite domain, the POU domain consists of a POU-specific and a POU homeo-domain connected by a flexible linker. The presence of the linker allows both domains to bind genomic regions irrespective of each other, amplifying the possible combinations of recognition sites (Kemler et al., 1989; Bendall et al., 1993) and ultimately leads to a plethora of interactions with various co-activators and repressors.

A subset of POU transcription factors was identified to tether to an octamer motif, ATTTGCAT, which positively regulates the activation of immunoglobulin gene promoters in lymphoid cells (Singh et al., 1986; Staudt et al., 1986). These transcription factors were subsequently defined as octamer binding transcription factors. Oct1 and Oct2 were the first family members identified (Singh et al., 1986; Staudt et al., 1986), where Oct1 is ubiquitously expressed and Oct2 is confined to B-cells (Schöler et al., 1989). Through the application of electromobility shift assays accompanied by the presence radioactively labelled immunoglobulin enhancer consisting of the octamer motif, a spectrum of 10 members were identified across tissues and developmental stages (Schöler et al., 1989). Three isoforms, Oct4, Oct5 and Oct6 were specifically expressed in F9 EC cells.

Oct4 was concurrently isolated in three independent laboratories using probes specific to the POU homeo-domain of Oct2 (Okamoto et al., 1990; Schöler et al., 1990a; Schöler et al., 1990b; Rosner et al., 1990). Comparison of the structure and sequence of Oct4 to its family members reveals high resemblance within the POU homeo-domain, but distinct differences within the POU-specific domain and its adjacent linker (**Figure 1.6**). The remainder of the protein is also varied, displaying a high proportion of proline and glycine residues in the N-terminus and proline, glycine and serine/threonine-rich regions in the C-

terminus (Rosner et al., 1990), surmising its function as a transcriptional activator (Mermod et al., 1989). Oct4 was first examined for its ability to associate with the octamer containing DNA probes (Rosner et al., 1990). Oct4 demonstrated specificity to these octamer motifs, but not mutant probes. To further delineate its transcriptional ability, the fusion of the proline rich N-terminus to the DNA binding domain of c-Jun resulted in the initiation of transcription (Okamoto et al., 1990). Moreover, ectopic expression of Oct4 into HeLa cells also elicited transcriptional activation, as detected using Chloramphenicol AcetylTransferase (CAT) assays (Schöler et al., 1990a).

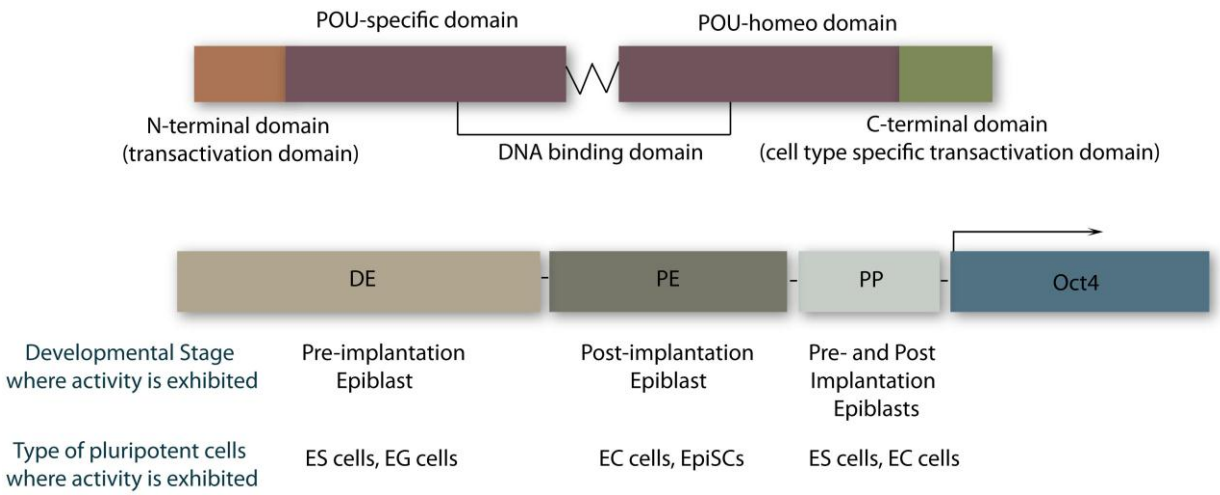



**Figure 1.6 Schematic diagram of the structure of Oct4 and its regulatory elements** (Top) Oct4 is comprised of N- and C-terminal domains flanking both ends. These regions display high proportions of proline and glycine, suggesting their function as transactivators. Oct4 also bears bipartite POU domains connected by a linker. The presence of the linker allows flexibility when tethering to its recognition motifs. (Bottom) Schematic diagram reflecting the regulatory elements upstream of the Oct4 coding sequence. These are namely the distal enhancer (DE), proximal enhancer (PE) and proximal promoter (PP), approximately 2kb, 1kb and 100bp upstream of the Oct4 transcriptional start site respectively. Each element controls Oct4 expression in distinct cell types, as described in the table.

### 1.4.2 Oct4 during embryogenesis

Discovered in EC cells, Oct4 was the first transcription factor identified to play a role in pluripotency. Maternal transcripts of Oct4 are present within oocytes and the expression of Oct4 persists until fertilization (Rosner et al., 1990; Pesce et al., 1998). During murine embryogenesis, Oct4 is exhibited at low levels during 2-cell and 4-cell stages, but its expression progressively increases before the 8-cell stage. Subsequently, Oct4 expression is confined within the inner cell mass of the blastocyst, and not the trophectoderm and primitive endoderm. Levels of Oct4 persevere after implantation and remains in the epiblast (Rosner et al., 1990; Schöler et al., 1990a; Yeom et al., 1991). Once gastrulation is initiated, Oct4 levels reduce in an anterior to posterior approach, such that primodial germ cells are the only cell population which expresses Oct4 beyond 8 days post coitum (dpc), (Pesce et al., 1998). Human blastocysts display analogous patterns where Oct4 is expressed 30 fold higher in the inner cell mass, as opposed to the trophoectoderm compartment (Hansis et al., 2000; Hansis et al., 2001).

Abrogation of Oct4 using murine models by generating a mutant allele where exons 2-5 are replaced by an internal ribosome entry site (IRES)-βgeo cassette, results in improper embryonic development and offspring homozygous for the mutation was unattainable from crosses between heterozygous parents (Nichols et al., 1998). Analysis of 61 implantation sites at 5.5dpc revealed only 43 prestreak embryos consisting of 11 wild-type and 32 heterozygous for the mutant allele. In contrast, at 3.5dpc, preimplantation embryos homozygous for the mutant allele were observed at expected Mendelian ratios. Cultivation of homozygous mutant embryos generated giant trophoblast cells, whereas wildtype and heterozygous embryos displayed expanded inner cell mass compartments and differentiated parietal endoderm cells. Likewise, extraction of the internal population of cells within 3.5 dpc embryos led to a similar phenomenon, where cells homozygous for the mutation developed into giant trophoblast cells, and one copy of the wild type allele is sufficient to generate parietal endoderm and visceral endoderm cells.

The importance of Oct4 in cell fate decision events during embryogenesis is highlighted through the temporal control of its expression in ESCs (Niwa et al., 2000). ESCs which

lack one or both allelic copies of endogenous Oct4, and comprise of an additional tetracycline (Tc) controlled cassette driving exogenous Oct4 expression, were generated and labelled as ZHTc6 and ZHBTc4 ESCs respectively. Transgenic expression of Oct4 can be efficiently initiated or repressed in the absence or presence of Tc. In ZHTc6 ESCs, triggering ectopic expression of Oct4 resulted in less than a fold increment of protein level, but led to differentiation into extra-embryonic endoderm and mesoderm lineages. On the contrary, the addition of Tc and repression of exogenous Oct4 expression in ZHBTc4 ESCs instigated differentiation into the trophoectoderm lineage. These pieces of evidence reflect the importance of Oct4 dosage in lineage commitment decisions during embryogenesis.

### 1.4.3 Regulation of Oct4

To elucidate the mechanisms behind Oct4 regulation, cis-acting elements located upstream of Oct4 were fused to LacZ reporter constructs and introduced into the pronuclei of oocytes by microinjection (Yeom et al., 1996). Fragments of 32, 18 and 12 kilobases (kb) in length were employed to investigate their effects on β-galactosidase activity in the developing embryo. All three constructs faithfully reproduced endogenous Oct4 expression within the embryo. A series of deletion constructs were generated to identify regions within the genomic region that had influence on the expression patterns. Analysis of changes in transcriptional activity imposed by these modifications revealed that deletion of 3 kb starting from 1.2 kb away from the transcriptional start site (TSS) reduced Oct4 expression levels in ESCs and EG cells. Due to its location and ability to enhance transcription in both orientations, this region was subsequently designated as the distal enhancer (**Figure 1.6**). Interestingly, the deletion of this region did not affect Oct4 expression in EC cells. Instead, a separate 1 kb element, 300 base pairs (bp) upstream of the TSS was crucial for the expression of Oct4 in EC cells, post implantation embryos and EpiSCs (Yeom et al., 1996; Guo et al., 2010; Tesar et al., 2007), and was termed as the proximal enhancer. In addition, a 230 bp fragment adjacent to the TSS, designated as the proximal promoter, was described to govern the expression of Oct4 in both ESCs and EC cells. These findings suggest an enhancer switch during the progression of embryonic development but the mechanism behind the phenomenon has yet to be identified.

Extending these observations, genome wide pursuits of transcription factor binding sites have established that the distal enhancer region of Oct4 is the densest binding locus for the key pluripotency-specific transcripts in ESCs (Chen et al., 2008), reinforcing the importance of this site. Likewise, the proximal promoter of Oct4 has been studied in several instances and is postulated to exhibit susceptibility to high concentrations of retinoic acid (Ben-Shushan et al., 1995; Barnea and Bergman, 2000; Pikarsky et al., 1994; Schoorlemmer et al., 1994). During retinoic acid induced differentiation, the proximal enhancer has been described to be essential for the repression of Oct4 (Pikarsky et al., 1994; Schoorlemmer et al., 1994). The Oct4 proximal promoter lacks a canonical TATA box but comprises of a GC-rich box, suggesting the binding of hormone receptors. Not surprisingly, the presence of Retinoic Acid Receptors (RARs) and Retinoic X Receptors (RXRs) activates transcription by binding to the proximal promoter (Ben-shushan et al., 1995), and introduction of Steroidogenic Factor 1 (SF1) promotes this activation (Barnea and Bergman, 2000). These binding events serve as competition against COUP transcription factors which act as repressors to Oct4 expression.

### 1.4.4 Resolving the function of Oct4

To study the function of Oct4, various approaches to identify its genetic targets and interacting partners have been employed. The former can be addressed through the execution of chromatin-immunoprecipitation (ChIP) coupled with sequencing (ChIP-seq). There have been several large scale endeavours to extricate pathways involved in pluripotency (Boyer et al., 2005; Loh et al., 2006; Chen et al., 2008; Kim et al., 2008). Close inspection of binding sites tethered to Oct4 highlights close proximity to regions occupied by Sox2 and Nanog, implicating their co-operation as a complex. The obligatory requirement of the three transcription factors suggest the importance of their involvement in the maintenance of pluripotency, leading to the inference that they comprise of the core transcriptional machinery and control the cellular integrity of pluripotent stem cells.

From a different perspective, the importance and role of Oct4 occupied genomic sites can be portrayed through knockdown studies (Loh et al., 2006). Esrrb and Rif1 are examples of

genes positively regulated by Oct4, and enforced silencing of their expression elicits differentiation into trophectodermal lineages. Extraction of these pieces of information enables us to illuminate the significance of Oct4.

Apart from exploring genetic networks, proteomic studies illuminate a separate dimension to the elucidation of Oct4 function. To expand our knowledge on the proteomic landscape in ESCs, proteins synthesised from genes associated with pluripotency, Oct4, Nanog, Dax1, Nac1, Zfp281 and Rex1, were coupled to biotin and precipitated (Wang et al., 2006). Assessment of their interacting partners using mass spectrometry techniques verify the existence of transcriptional complexes, reinforcing findings gathered from genomic studies. Recently, two independent groups have executed mass spectrometry in mouse embryonic stem cells to unravel a magnified list of Oct4 interacting partners (Pardo et al., 2010; van den Berg et al., 2010). Employing an Oct4-centric approach, both groups found a total of 128 binding partners of Oct4, where 20 were identified in both studies. Well-established co-operations with Sox2, Dax1, Sall4 and chromatin-modifying complexes such as NuRD and SWI/SNF were illustrated, although other primary candidates such as Nanog and Esrrb were only uncovered in one of the two studies. These studies reflect an average snapshot of the protein complexes present within the cell at a specific time in a defined environmental milieu. However, overall, these studies provide clues to role of Oct4 in the maintenance of pluripotency.

### 1.4.5 Importance of Oct4 in reprogramming

The significance of Oct4 in the maintenance of pluripotency has been fortified in its participation in the establishment of pluripotency (Takahashi and Yamanaka, 2006; Yu et al., 2007; Takahashi et al., 2007). Oct4 acts in concert with c-Myc, Klf4 and Sox2 to induce the dedifferentiation of a spectrum of somatic cell types into ES-like cells. Among the four transcription factors, Oct4 bears a unique quality and is resistant to substitution by its family members, Oct1 and Oct6 (Nakagawa et al., 2008). Omission of components within the classical reprogramming cocktail has been demonstrated to lead to the acquisition of pluripotency in certain cell types, but exclusion of Oct4 has not been exemplified (Tsai et al., 2010; Kim et al., 2009b; Kim et al., 2009c; Giorgetti et al., 2009).

In neural stem cells, sole expression of Oct4 is sufficient to trigger the reprogramming cascade of events (Kim et al., 2009b; Kim et al., 2009c).

As previously described, genes which are capable of replacing the exogenous requirement of Oct4 have been uncovered (Heng et al., 2010; Redmer et al., 2011), and the mechanisms behind their substitution potentially involve the activation of endogenous Oct4 loci. A recent study attributes the speed of reprogramming to the initiation of endogenous Oct4 expression elicited by RARγ and Nr5a2 (Wang et al., 2011c). These findings provide insight to the roles undertaken by Oct4 in the reprogramming process.

Although a wealth of genetic and proteomic information pertaining to Oct4 has been amassed, these findings were obtained from the study of ESCs and do not necessarily depict the molecular changes coupled to reprogramming. With this in mind, it is necessary to expand our knowledge on the molecular changes associated with reprogramming. One approach involves the identification of factors that can act as replacements to Oct4 during reprogramming. Although there have been various attempts to scout for substitutes, the strategies employed are limited and narrow. In contrast, a genome wide approach will enable the surveyance of genes which would have otherwise been ignored in a candidate gene approach.

### 1.5 Retinoic Acid signalling

Amid the plethora of developmental pathways involved in embryogenesis, the obligatory role executed by vitamin A emerged more than half a century ago (Hale, 1933; Warkany and Schraffenberger, 1946; Wilson and Warkany, 1948; Wilson et al., 1953). Exposure of pregnant rats and pigs to vitamin A deficient diets led to litters which exhibited malformations of the eye and genito-urinary tract. These anomalies were subsequently attributed to the absence of retinoic acid (RA) (Thaller and Eichele, 1987; Dickman et al., 1997; White et al., 1998). RA is an active acidic metabolite of vitamin A. Lipid soluble vitamin A (retinol) interacts with retinol binding protein 4 (RBP4) upon ingestion for uptake into tissues (**Figure 1.7**). Alcohol and retinol dehydrogenases (ADH and RDH) oxidise retinol to generate retinal, which is further oxidised by retinaldehyde

dehydrogenases (Aldh1a1, Aldh1a2 and Aldh1a3). Consequently, RA is produced and elicits an effect by binding to an appropriate receptor which comprises of a complementary ligand binding pocket (reviewed by Rhinn and Dollé, 2012).

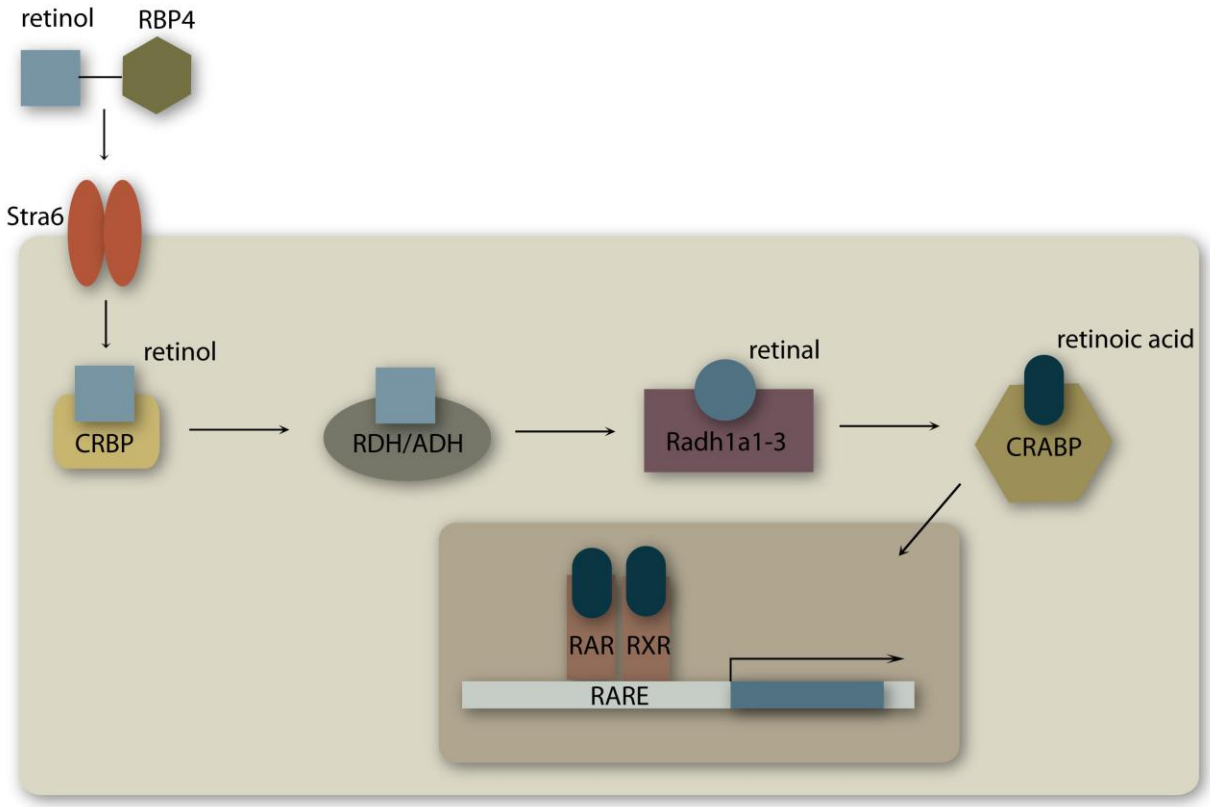



**Figure 1.7 Retinoic acid signalling pathway** While in circulation, retinol binds to retinol binding protein (RBP4) and is transported across the cell membrane via Stra6. While in the cytoplasm, retinol binds to cellular retinol binding protein (CRBP) until oxidation is facilitated through the action of retinol or alcohol dehydrogenase (RDH/ ADH). This generates retinal which is further oxidised by retinal dehydrogenase (Radh1) to form retinoic acid (RA). As the active metabolite, RA associates with cellular retinoic acid binding protein (CRABP) and is transported into the nucleus where it is recognised and bound to retinoic acid receptors (RARs). By influencing Retinoic Acid Response Elements (RAREs), this association results in the initiation of transcriptional events and triggers a cascade of signalling events.

### 1.5.1 Retinoic Acid Receptors

The isolation of the first receptor that interacts with RA came as a fortuitous event (Giguere et al., 1987). While exploring hepatitis B viral integration events within human hepatocellular carcinoma (Dejean et al., 1986), a 462 amino acid open reading frame, of which 66 amino acids shared close homology to the DNA binding domain (DBD) within glucocorticoid and thyroid hormone receptors, was identified. By substituting its DBD with an equivalent from a human glucocorticoid receptor, the ligand corresponding to the novel protein was revealed to be RA. This novel protein was aptly termed as Retinoic Acid Receptor alpha (RARα).

Additional receptors that interact with RA were found in quick succession (Petkovich et al., 1987; Benbrook et al., 1988; Brand et al., 1988; Krust et al., 1989; Zelent et al., 1989). Overall, three receptors were identified in the mouse and human, designated as Retinoic Acid Receptor alpha (RARα), Retinoic Acid Receptor beta (RARβ) and Retinoic Acid Receptor gamma (RARγ). These receptors were classified under the nuclear receptor superfamily and show high resemblance in their DBDs and ligand binding domains (LBDs) (reviewed by Chambon, 1996). The DBD integral to RARs provides specificity when tethering to cognate response elements. The recognition motif (A/G)G(G/T)TC(A/G) and its degenerate variations, have been designated as Retinoic Acid Response Elements (RAREs). Conventionally, direct repetition (DR) of the RARE is commonly found at binding sites, where an interval of 1, 2 or 5 base pairs (DR1, DR2 or DR5) is apparent (Balmer and Blomhoff, 2002). Recent chromatin immunoprecipitation experiments targeted at RARα and RARγ in ESCs and MEFs indicate otherwise (Delacroix et al., 2010), where single and anomalously spaced RARE half sites were also detected.

The finding that purified RARs are unable to bind to RAREs *in vitro* led to the identification of the partnership between RARs and Retinoic X Receptors (RXRs) in forming heterodimers (Yu et al., 1991). Similar to RARs, RXRs are made up of three family members, RXRα, RXRβ and RXRγ. Although RXRs are unable to bind to endogenous all-trans RA ligands, it is purported to co-operate with RARs to elicit transcriptional activation (Leid et al., 1992; Kliewer et al., 1992; Zhang et al., 1992).

The canonical pathway involving RARs encompasses the presence of the ligand, its association to the receptor and the trigger of transcriptional events (**Figure 1.8**). However, unconventional modes of action have been described for RARs (Huggenvik et al., 1993; Matkovits and Christakos, 1995; Alsayed et al., 2001; Gupta et al., 2008; Laursen et al., 2012). This has been described through the rapid activation of mitogen-associated protein kinases (MAPKs), phosphoinositide 3-kinase (PI3K) and Src kinases.

In the absence of its natural ligand, RARs have also been reported to act as repressors. This can occur in two folds. First, attachment of the receptor to its recognition motif can proceed independent of the ligand, in condition that the site is not impeded by nucleosomes (Narlikar et al., 2002). Second, RAR target genes are capable of interacting with repressors such as Topoisomerase IIβ (McNamara et al., 2008), Polycomb group proteins (Gillespie and Gudas, 2007a; Gillespie and Gudas, 2007b) and calmodulin kinase IIγ (Si et al., 2007). Specific to RARα, absence of RA results in the exposure of helix 12 and recruitment of co-repressor complexes (Hu and Lazar, 1999)

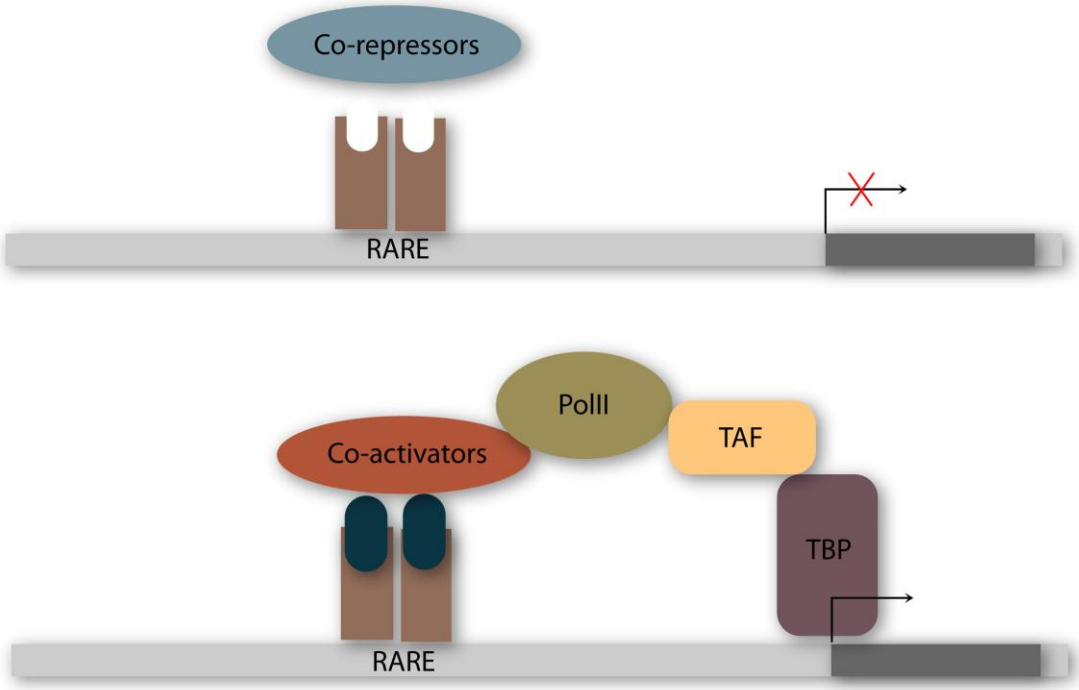



**Figure 1.8 Ligand dependent transcriptional control** (Top) In the absence of its ligand, Retinoic Acid Receptors (RARs) are able to recognise and bind to Retinoic Acid Response Elements (RAREs). Through its association to co-repressor complexes, transcription at the target gene is inhibited. (Bottom) In the presence of Retinoic Acid (RA), binding to RARs occurs and co-repressor complexes dissociate. Co-activators are attracted to the ligand bound receptors and recruit RNA polymerase (PolII), TBP associated factors (TAF) and TATA-binding protein (TBP). These factors initiate transcription and drive the expression of the target gene.

### 1.5.2 Retinoic acid signalling during development and reprogramming

Members within the RAR family exhibit redundancy, where single ablation of RARα, RARβ or RARγ yields viable mice (Lufkin et al., 1993; Ghyselinck et al., 1997; Lohnes et al., 1993). Surviving mice display a few congenital abnormalties which resemble vitamin A deficiency syndrome, but the extent of physical aberration does not apply to the entire spectrum of tissues which express the receptor (reviewed by Mark et al., 2009). In contrast, double mutants display overt developmental defects and die *in utero* (Ghyselinck et al., 1997; Lohnes et al., 1994; Kastner et al., 1994; Mendelsohn et al., 1994). These abnormalities were not previously described in vitamin A deficiency syndromes seen in murine and porcine models (Hale, 1933; Wilson et al., 1953) and included aberrations in the Harderian glands, skull, face, vertebrae, limbs and forebrain (Ghyselinck et al., 1997; Lohnes et al., 1993; Lohnes et al., 1994; Luo et al., 1996; Mendelsohn et al., 1994; Subbarayan et al., 1997).

RARα and RARγ are expressed both maternally and zygotically at pregastrulation stages in *Xenopus Laevis* (Ellinger-Ziegelbauer and Dreyer, 1991; Shiotsugu et al., 2004), but not RARβ. Similar patterns were observed in bovine and murine embryos (Mohan et al., 2001; Ang and Duester, 1997). During gastrulation in murine embryos, expression levels of RARα and RARγ are sustained and RARβ is detected at headfold, posterior midline tissues and lateral regions of the egg-cylinder (Ruberte et al., 1991). As development progresses to the somitic stages through to adulthood, the localisation of the RARs becomes confined (reviewed by Dolle, 2009) and RARα is expressed in a widespread manner, but RARβ and RARγ have more specific distribution patterns.

Initial studies of RA postulated its role as a morphogen, eliciting a concentration gradient across the anterior-posterior axis during embryonic development (Casci, 2008). The versatility of developmental potential induced by stratification of RA dosage is exemplified in ESCs where adjustment of RA concentration results in the acquisition of dorsal or ventral phenotypes in neural derivatives (Okada et al., 2004) (**Figure 1.9**). In addition, varied concentrations of RA and a permutation of co-operating chemicals are able to convert ESCs into neuronal derivatives (Schuldiner et al., 2001), insulin secreting

cells (Jiang et al., 2007), osteoblasts (Yamashita et al., 2005) and epithelial cells (Metallo et al., 2008). Exposure of EC cells to RA is capable to inducing differentiation (Strickland and Mahdavi, 1978). Interestingly, exposure of RA to various somatic cell types incites distinct cellular responses, where leukemic cells in bone marrow aspirates were committed to differentiate (Flynn et al., 1983) whereas hepatocytes were stimulated to proliferate (Ledda-Columbano et al., 2004).

Comparison of studies which have delineated genomic regions recognised by RARs and transcriptional changes stimulated by the addition of RA (Delacroix et al., 2010; Mahony et al., 2011; Simandi et al., 2010) reveals common gene targets involved in a spectrum of developmental processes, reinforcing the notion that RA signalling plays a pivotal role in embryonic development. Furthermore, a recent study has implicated the role of RARγ in the efficient and rapid generation of murine iPSCs (Wang et al., 2011c). This was attributed to its synergistic co-operation with Nr5a2 to transcriptionally activate Oct4 by binding to an RARE at its proximal promoter (RAREoct). These studies shed light on the developmental role of RA signalling. Dissecting these pathways will enlighten us about the molecular intricacies during development and reprogramming.

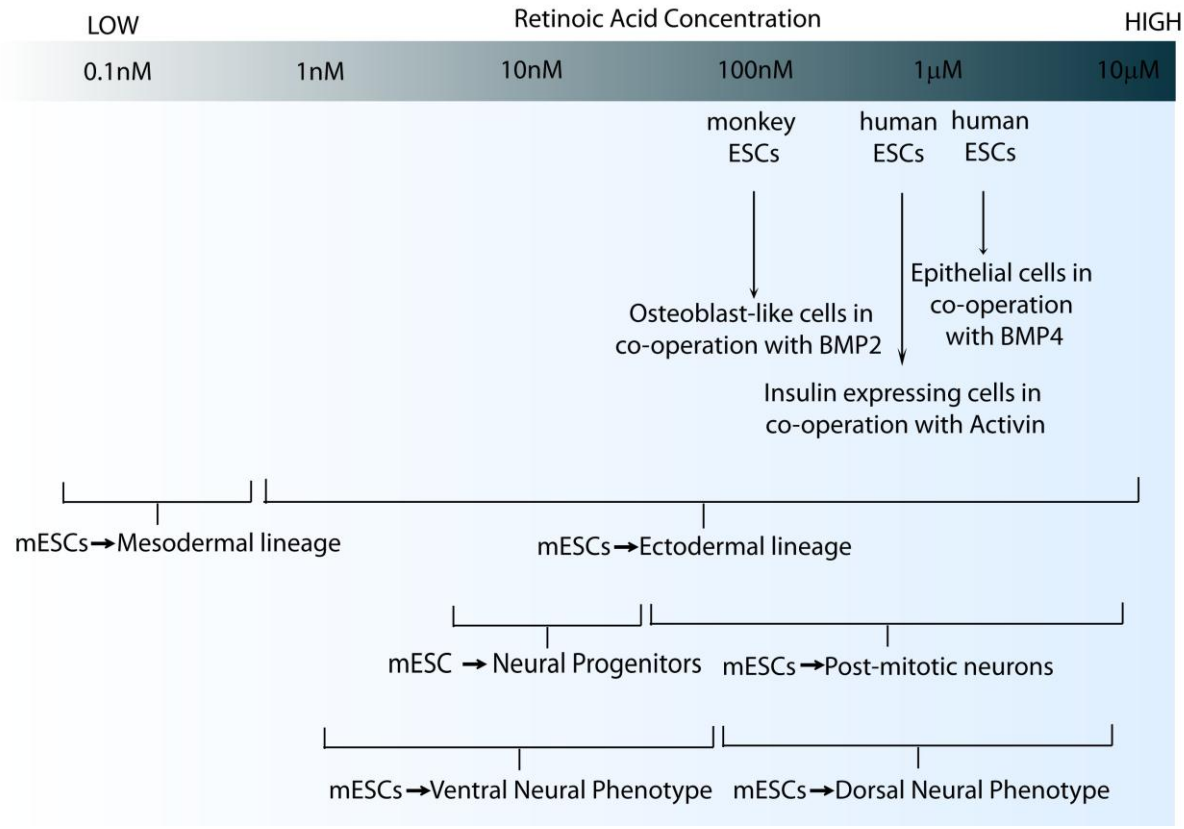



**Figure 1.9 Dosage dependence in RA-stimulated differentiation** Distinct concentrations of RA can lead to differentiation into different cell types. Overall, introduction of high concentrations of RA to mouse ESCs (mESCs) gives rise to neuronal lineages whereas low concentrations ($<10^{-10}$M RA) leads to mesodermal lineages. Differing neuronal phenotypes stemming from mESCs can also be achieved by varying the concentrations of RA. In addition, co-operating stimuli such as Bone Morphogenic Proteins (BMPs) and Activin result in directed differentiation of various ESCs into insulin-expressing, osteoblast-like and epithelial cells.

## 1.6 Genetic Screens

The law of hereditary as proposed by Gregor Mendel (Castle, 1903) ignited our understanding of the genetic constitution within organisms. In combination with the discovery of the molecular structure of DNA (Watson and Crick, 1953), these findings sparked a surge in interest to uncover the genetic content of organisms. After the establishment of a robust sequencing protocol in 1977 (Sanger et al., 1977), an explosion of efforts led to the elucidation of genomic components within viruses and viroids, naturally occurring plasmids, organelles, eubacteria, archaea, fungi, plants and animals. The relentless drive for genomic compositions has generated blank canvases for the comprehension of unknown elements in the genome.

A range of tools has facilitated our understanding of the genome, spanning from chromatin immunoprecipitations to cDNA libraries and genome wide association studies. Retroviruses and transposons comprise of a subset of gene discovery tools. In the 1930s, avian myeloblastosis viruses were identified to infect chickens and cause a disease similar to acute myelogenous leukemia in humans. Using sequencing techniques, this disease causing mutation was narrowed down to N-Myb (Klempnauer et al., 1982; Rushlow et al., 1982; Souza et al., 1980). Analogous discoveries were made by analyzing the Rous sarcoma virus and Abelson Murine leukemia virus, where oncogenes v-Src and v-Abl were unearthed (Brugge and Erikson, 1977; Witte et al., 1978; Goff et al., 1981). The discovery of the tumorigenic potential of a subset of slow transforming viruses opened avenues for the identification of oncogenes and tumour suppressor genes. These viruses, such as avian leucosis viruses and murine leukemia viruses, have a long latent period of 4-12 months and do not have inherent transforming genes. Instead, they integrate into the host genome and elicit transcriptional activation at the site of integration. Deregulated expression of genes critical for the maintenance of cellular homeostasis results in the formation of tumours (Hayward et al., 1981; Neel et al., 1981). Several oncogenes have been discovered through this avenue, such as Pim1, a serine protein kinase associated with MoLV transformation of T cells (Selten et al., 1986; Meeker et al., 1987) and Evi1, a multifunctional transcription factor identified from AKXD-23 myeloid tumors (Mucenski et al., 1988; Morishita et al., 1988; Bard-Chapeau et al., 2012).

Extending these findings, infection of organisms with retroviruses has been employed as an approach in forward genetics to dissect genetic elements associated to defined phenotypes (Jaenisch, 1976; Lin et al., 1994). Performed in mice and zebrafish, insertional mutagenesis provided a separate dimension to study development and tumorigenesis (Zambrowicz et al., 1998; Golling et al., 2002; Amsterdam et al., 2004; Li et al., 1999; Suzuki et al., 2006).

N-ethyl-N-nitrosourea (ENU) mediated mutagenesis poses as a distinct approach to perform functional genetic analysis. ENU is an alkylating agent that predominantly induces base substitutions in the host genome, thereby creating a mutagenised library when introduced into target cells. Exemplified in ESCs, allelic mutations in Smad2 and Smad4 were observed after exposure to ENU (Chen et al., 2000). Introduction of Smad mutant ESCs into mouse blastocysts led to developmental aberrations in resultant chimeras, recapitulating the significance of Smad proteins during embryogenesis (Vivian et al., 2002). A limitation reflected by ENU mutagenesis is the diploid nature of the genome, where the isolation of candidate genes responsible for a given phenotype is reliant on biallelic disruption of genes. To circumvent this bottleneck, Bloom-deficient ESCs that exhibit increased rates of loss of heterozygosity were utilised (Yusa et al., 2004). The recent derivation of haploid ESCs presents as an additional source of starting material which overcomes the requirement for biallelic disruption of genes, thus enabling convenient detection of phenotypic mutants (Leeb and Wutz, 2011).

The discovery of transposons exemplifies another source of understanding the genome (McClintock, 1950). Recognised in maize, McClintock noticed that the short arm of chromosome 9 was unstable and mutable, and attributed it to the co-operation of Dissociator (Ds) and Activator (Ac) complexes. These elements were subsequently cloned and termed as transposon and transposase respectively. As potent mutagenic tools, the transposon system is now widely used to perform genetic screens.

### 1.6.1 Transposons

Transposons are mobile genetic elements which can be mobilised in the presence of a transposase. A spectrum of transposon systems, such as Sleeping Beauty, Tol2 and piggyBac, have been isolated from Salmon, Medaka fish and Cabbage Looper moth respectively (Ivics et al., 1997; Kawakami et al., 2000; Fraser et al., 1996; Ding et al., 2005; Cadinanos and Bradley, 2007; Wang et al., 2008). Applications of these transposons include gene delivery and genetic screens (**Figure 1.10**). To illustrate the potential of transposon mediated screens, Sleeping Beauty was described as the first nonviral insertional mutagen used for cancer gene identification (Dupuy et al., 2005). Succeeding studies have employed this strategy in identifying common insertion sites in a range of cancer types, spanning from hepatocellular carcinoma to leukemia to colorectal cancer and pancreatic adenocarcinoma (Starr et al., 2009; Dupuy et al., 2009; Keng et al., 2009; Rad et al., 2010; Mann et al., 2012). A closer look at these genes reveals a correlation to known cancer genes and uncovers potential oncogenes and tumour suppressors. Moreover, the application of transposon-mediated mutagenesis has been stretched to prove effective in the validation of oncogenes (Su et al., 2008).

In comparison to retroviral insertional mutagenesis, transposon mediated mutagenesis holds several advantages. It does not preferentially occur in the hematopoietic system and is relatively unbiased (Dupuy et al., 2005), hence it has gained in favour to facilitate genome wide screens.

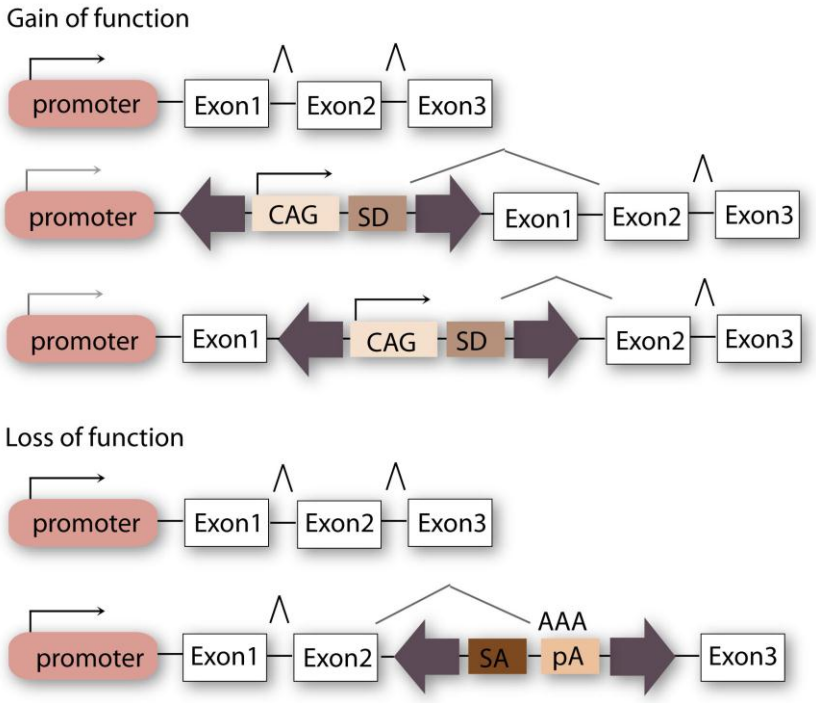



**Figure 1.10 Principle behind the execution of transposon mediated genetic screens** (Top) Gain of function genetic screens can be conducted through the use of transposons encompassing a strong constitutive promoter, such as CMV early enhancer/chicken beta actin (CAG), preceding a splice donor (SD) site. Insertion of the trapping cassette into the genome results in the forced transcription of exons located downstream of the insertion site. (Bottom) Loss of function genetic screens can be performed through the application of transposons flanking a splice acceptor (SA) site adjacent to a poly A signal. Integration of the transposon cassette results in the premature termination of the inserted gene, creating a truncation mutant.

### 1.6.2 piggyBac

Within the spectrum of transposon systems, piggyBac has been demonstrated to display high transposition efficiency and competence in the delivery of large cargo size (Wu et al., 2006; Wang et al., 2008; Li et al., 2011b; Liang et al., 2009). Targeted at TTAA sites, piggyBac assisted transposition events abolish the presence of genetic footprints upon its excision from the genome (Wang et al., 2008). This unique quality has led to its application as a gene delivery tool in the generation of iPSCs (Woltjen et al., 2009; Yusa et al., 2009), with the notion of excising the reprogramming factors upon the attainment of pluripotency to obtain transgene-free iPSCs. Other properties of piggyBac which have been observed include reduced excision efficiencies of methylated transposon cassettes and no clear indication of local hopping, where the excised transposon reintegrates into a neighbouring genomic location (Wang et al., 2008). These descriptions of piggyBac portray its usefulness as a genetic tool for genome wide screens.

### 1.6.3 piggyBac transposons as a tool to conduct genetic screens in cellular assays

piggyBac transposons exhibit qualities in performing *in vitro*, as well as *in vivo* screens. Competent in dissecting the molecular complexities of tumorigenesis (Rad et al., 2010), the transposon system is also a versatile genetic tool to modifying cellular phenotypes. A recent study employed the use of piggyBac transposition to create a mutagenic library of EpiSCs and effectively isolated genes that were capable of transforming EpiSCs to a naïve pluripotent state (Guo et al., 2010). The application of piggyBac assisted genetic screens was subsequently extended in co-operation with Bloom-deficiency, to identify genes capable of conferring resistance to ricin (Wang et al., 2011b) and differentiation (Guo et al., 2011).

With little bias to specific genetic loci, it is evident that piggyBac transposition represents a powerful tool to explore the host genome and identify genes involved in a phenotype of interest. Due to the capacity to transport large cargo (Li et al., 2011b), various genetic modifications, such as the inclusion of selection markers or genes that improve the sensitivity of the screen, can be made to the transposon to facilitate the screen. The incorporation of splice donor and acceptor sites into the transposon cassette confers gene

trapping capacity and allows convenient identification of candidate genes. Together, these modifications enable us to explore the genetic basis of various cellular phenotypes.

### 1.7 Project Aim

Reprogramming of somatic cells to iPSCs involves the co-operation of four transcription factors, Oct4, c-Myc, Klf4 and Sox2. Although the reprogramming capacity of the four genes has been recapitulated across cell types and species, the molecular intricacies involved in the reprogramming process are not fully delineated. To address this, genetic analyses have been conducted to identify factors that can replace or enhance the defined set of four genes. Oct4 has been most recalcitrant to replacement and its importance is highlighted in the obligatory need of Oct4 during development. To decipher the role of Oct4 during the establishment of pluripotency, this project aims to identify genetic replacements of exogenous Oct4 during the generation of iPSCs using a genome-wide piggyBac transposon-assisted mutagenesis approach.

The experimental strategy behind the execution of the screen and the dissection of results obtained from the screen will be described in **Chapter 3**. Validation of isolated candidate genes will be explored in **Chapter 4**. Focusing on one candidate gene, **Chapter 5** will delineate the mechanisms behind its ability to act as a substitute for the ectopic expression of Oct4. Ultimately, these results offer a broad perspective on alterations in the epigenetic landscape elicited by reprogramming events, and heighten our awareness on the molecular changes that accompany the acquisition of pluripotency.

# CHAPTER 2 : MATERIALS AND METHODS

## 2.1 Cell Culture

### 2.1.1 Embryonic Stem Cells

#### 2.1.1.1 AB2.2

AB2.2 was derived from a 129S6 blastocyst (McMahon and Bradley, 1990). ESCs were maintained in Knockout$^{TM}$ DMEM (Invitrogen), supplemented with 15% serum, 1x Penicillin-L-glutamine-Streptomycin, 100 μM β-mercaptoethanol and 100 U/ml leukaemia inhibitory factor (LIF) (Millipore) (M15-LIF medium) on a layer of irradiated SNL76/7 feeder fibroblasts as previously described (Ramirez-Solis et al., 1993). For the purpose of maintenance, medium was changed daily. To perform routine passaging, cells were washed once with phosphate buffered saline (PBS) before being treated with 0.1% trypsin-EDTA in PBS at 37°C for 5 minutes. Trypsin activity was quenched with an equal volume of M15-LIF medium and clumps were mechanically disrupted by pipetting before centrifugation. Supernatant is aspirated and the resuspended pellet is transferred at desired proportions into a fresh plate pre-fed with M15-LIF.

#### 2.1.1.2 ZHBTc4 ESCs

ZHBTc4 ESCs were derived from CGR8 ESCs, where genetic manipulations (i) ablated the remaining endogenous copy of Pou5f1 (Niwa et al., 2000) and (ii) inserted a transgene expressing Oct4 under a tetracycline repressible system. These cells were a kind gift from Dr. Austin Smith (Centre for Stem Cell Research, Cambridge). These ESCs were maintained in GMEM (Sigma), supplemented with 10% serum (Hyclone), 1x Penicillin-L-glutamine-Streptomycin (Invitrogen), 1x non-essential amino acids (Invitrogen), 1x sodium pyruvate (Invitrogen), 100 μM β-mercaptoethanol (Sigma) and 100 U/ml LIF (Millipore), on tissue culture plates coated with 0.1% Gelatin (Sigma). Cells were maintained in a similar manner to AB2.2 ESCs.

### 2.1.2 Mouse embryonic fibroblasts

Mouse embryonic fibroblasts (MEFs) were maintained in Knockout$^{TM}$ DMEM (Invitrogen), supplemented with 10% serum, 1x Penicillin-L-glutamine-Streptomycin and 1x non-essential amino acids (M10). To obtain MEFs, mice of desired genotypes were set as mating pairs. Females were inspected for the formation of plugs daily. Upon the observation of a plug, the pregnant female was sacrificed 13 days later, such that the embryos have reached developmental stage E13.5. The embryos were separated from the female and placed into cold PBS. On a Petri dish, the embryos were dissected to remove the surrounding membranes and placenta. The head and visceral organs were discarded and the remaining tissues were washed with fresh PBS. If genotyping is required, a small portion of the tail is extracted for DNA. Using a pair of sterile scissors, the tissues were finely minced and transferred into 3ml of 0.1% trypsin-EDTA in PBS at 37°C for 20 minutes. An equal volume of trypsin was added and incubated at 37°C for another 20 minutes. To neutralise trypsin, 6ml of M10 was added to the mixture and centrifuged. The cell pellet obtained from one embryo was distributed into one 15-cm 0.1% gelatin coated plate. After 3 days, the cells had reached confluency and were dissociated using trypsin treatment. The cells were counted and frozen into vials at a density of 5 million cells.

### 2.1.3 iPSCs

#### 2.1.3.1 Generation of iPSCs

To generate iPSCs, frozen MEFs were thawed and 5 million cells were seeded into three 15-cm gelatin-coated plates. After 4 days or when the cells attain 70% confluency, the cells were washed once with PBS and dissociated in 4ml of trypsin-EDTA at 37°C for 5 minutes. An equal volume of M10 was added to neutralise trypsin and the cell clumps were disrupted mechanically through the use of pipettes. 10µl of the cell suspension was used for counting, while the remaining cells were collected and centrifuged at 1000rpm for 3 minutes. The cell pellet was resuspended in OptiMEM (Invitrogen) at a density of 5 million cells/ml. Plasmids used for reprogramming were aliquoted at appropriate amounts (**Table 1**) into 1.5ml eppendorf tubes and mixed with 200µl of cell suspension and transferred into an Amaxa nucleofection cuvette. The cuvette was inserted into the Amaxa

nucleofector and electroporation was initiated using program A-023. Once electroporation occurred, the cuvette was quickly removed and 1ml of M10 was added to resuscitate the cells. The cell mixture was then transferred into a fresh 10-cm plate pre-fed with M15-LIF. To increase reprogramming efficiency, irradiated SNL76/7 feeder fibroblasts were employed to support the transfected cells. In the event that the doxycycline inducible platform was applied, 1.0μg/ml doxycycline (Clontech) was introduced immediately after transfection. Media for these transfected cells were changed regularly and the plates were maintained over a span of three weeks to determine reprogramming efficiencies.

| PLASMID | AMOUNT TRANSFECTED |
|---|---|
| piggyBac transposase (PL623) | 2μg |
| PB-CAG-cDNA | 1μg |
| PB-CAG-rtTA | 1μg |
| PB-TRE-cDNA | 1μg |

**Table 2.1 Amounts of plasmid DNA transfected into MEFs to obtain iPSCs** In a typical reprogramming experiment, 1 x $10^6$ MEFs were transfected with 2μg of PL623 and 1μg of each transposon construct consisting of reprogramming factors, unless otherwise stated. If doxycycline inducible platform was employed, 1μg of PB-CAG-rtTA was co-introduced into the cells.

### 2.1.3.2 Establishment of iPSC colonies

Upon transfection of MEFs, the cells are maintained for 14 days or until the colonies are visible by eye. Prior to the picking of colonies, 50µl of trypsin-EDTA was added into each well of a 96-well round bottom plate. The reprogrammed cells were washed once with PBS and 10ml of fresh PBS was added into each 10-cm plate. Viewed under the microscope, colonies were picked and transferred into 50µl of trypsin-EDTA. After picking, an equal volume of M15-LIF was added and the colony was disrupted by pipetting the mixture approximately 10-30 times, depending on the compactness of the colony. The cell suspension was then transferred into a 96-well feeder coated plate pre-fed with fresh M15-LIF. The cells are maintained by a change of fresh media daily. Once confluent, the cells were dissociated using trypsin-EDTA and the cell pellet obtained after neutralisation and centrifugation was seeded onto a feeder coated 24-well plate.

To obtain a homogenous population of ground state naïve iPSCs, the colonies are maintained in NDiff $N_2B_{27}$ (Stem Cells, Inc) supplemented with 1µM PD 0325901 (Axon Medchem), 3µM CHIR 99021 (Axon Medchem) and 100 U/ml leukaemia inhibitory factor (LIF) (Millipore) (2i-LIF).

### 2.1.3.1 4F-iPSCs and 6F-iPSCs

To generate 4F iPSCs, 2µg PL623, 1µg PB-TRE-OCKS and 1µg PB-CAG-rtTA were introduced into 1 x $10^6$ Rex1::EGFP-IRES-Puro MEFs. 6F iPSCs were obtained using 2µg PL623, 1µg PB-TRE-OCKS, 1µg PB-TRE-RL and 1µg PB-CAG-rtTA. Primary colonies in both conditions were picked and maintained in 2i-LIF on gelatin coated plates.

## 2.2 Genetic Screen

### 2.2.1 Cloning of PB-CAG-CKS-CAG-SD

A transposon cassette containing CAG promoter and the four murine factors (mFx), namely OCT4, C-MYC, KLF4 and SOX2, was obtained from Wei Wang, Wellcome Trust Sanger Institute. To generate PB-CAG-C-MYC-KLF4-SOX2, OCT4 was excised using EcoRI and SbfI. As C-MYC contains an internal SbfI recognition site, PB-CAG-C-MYC (also obtained from Wei Wang) was simulataneously digested with EcoRI and SbfI (New England Biolabs) and ligated to the products obtained from the pre-described restriction digestion step. To generate PB-CAG-C-MYC-KLF4-SOX2-CAG-SD, PB-CAG-C-MYC-KLF4-SOX2 and PB-CAG-SD were both digested with XhoI and NheI (New England Biolabs) and ligated using T4 ligase (New England Biolabs).

### 2.2.2 Transduction of PB-CAG-CKS-CAG-SD into MEFs

As previously described, MEFs at 70% confluency were treated with 0.05% trypsin (Invitrogen), pelleted and resuspended in OptiMEM. 10μg of PB-CAG-CKS-CAG-SD was mixed with 1 x $10^6$ cells. 10 x $10^6$ cells were used in ten separate electroporations. Amaxa Nucleofector program A-023 was used to electroporate the mixtures. Modified cells from each electroporation were transferred to 10-cm feeder coated plates and maintained regularly in M15-LIF. Colonies were observed after 20 days and 2μg/mL puromycin was introduced to positively select for fully reprogrammed colonies.

### 2.2.3 Analysis of integration sites

#### 2.2.3.1 Extraction of DNA

10cm plates of iPS colonies were treated with 0.05% trypsin (Invitrogen) and dislodged cells were pelleted and washed with PBS. The cells were pelleted by centrifugation and resuspended in 1mL of lysis buffer (50mM Tris-HCl pH7.5, 50mM EDTA pH8.0, 100mM NaCl, 1% SDS, 1mg/mL proteinase K) The cells were incubated overnight in 65°C and double the amount of 100% ethanol was added. Precipitation was allowed to occur at room

temperature for 2 hours and DNA was pelleted by centrifuging at 1000rpm for 10 minutes. Supernatant was aspirated and 70% ethanol was used to wash the DNA pellet twice. After another round of centrifugation, the precipitate was resuspended in Tris-EDTA buffer overnight at 65°C.

#### 2.2.3.2 Splinkerette PCR and cloning

4μg of DNA from each 10cm plate was digested with BfuCI (New England Biolabs) at 37°C for 3 hours and ligated with splinkerettes (**Table 2.2**) at 16°C for 15 hours. Upon ligation, splinkerette PCR was conducted using primers against the splinkerette and the left or right transposon arms. A small aliquot of the PCR products were separated on 1.5% agarose to visualise a ladder of various band sizes. The remaining PCR products were purified using the QIAGEN PCR Purification Kit and eluted in 30μL of the provided elution buffer. 3μL of the purified PCR products was cloned into pGEMT-EASY (Promega). Subsequently, 96 colonies were picked for DNA obtained from each 10cm plate and grown in LB with 100μg/mL ampicillin. The resultant bacterial cultures were sequenced using capilliary sequencing. Sequences were analysed with the help of Stephen Rice (WTSI) and categorised according to the exact integration site, query coverage and quality, neighbouring annotated genes, percentage of match, and the orientation of the integration.

| PRIMER/OLIGO NAME | SEQUENCE |
|---|---|
| Splinkerette 1 | CGAAGAGTAACCGTTGCTAGGAGAGACCGTGGCTGAATGAGACTGGTGTCGACACTAGTGG |
| Splinkerette 2 | GATCCCACTAGTGTCGACACCAGTCTCTAATTTTTTTTTTCAAAAAAA |
| Splinkerette primer 1 | CGAAGAGTAACCGTTGCTAGGAGAGACC |
| Splinkerette primer 2 | GTGGCTGAATGAGACTGGTGTCGAC |
| piggyBac left primer 1 | CAGTGACACTTACCGCATTGACAAGCACGC |
| piggyBac left primer 2 | GAGAGAGCAATATTTCAAGAATGCATGCGT |
| piggyBac right primer 1 | CCTCGATATACAGACCGATAAAACACATGC |
| piggyBac right primer 2 | ACGCATGATTATCTTTAACGTACGTCACAA |

**Table 2.2 Sequences of splinkerettes and primers used during splinkerette PCR**

### 2.2.3.3 Validation of integration sites within BNC2

Collection of DNA from pools of iPSCs obtained from the genetic screen was described in 2.2.3.1. 1µl of DNA was used to amplify BNC2 genomic regions comprising of transposon integrations. Primers employed are listed in **Table 2.3** and Extensor Hi-Fidelity PCR master mix 2 (ABgene) was used to perform PCR using the following settings: 94°C for 4 min, followed by 30 cycles of 94°C for 10 sec, 55°C for 30 sec and 68°C for 30 sec. Final incubation was at 68°C for 10 min.

| Name | Primer Sequence | Position of integration | Position of Primer | Distance from Integration Site |
|---|---|---|---|---|
| **Forward-1** | CAGGCAATCACGGCTATTTT | 84147715-84147827 | 84147534 | -181 |
| **Reverse-1** | GGGAGGCCACACACATAATC | 84147715-84147827 | 84148031 | 316 |
| **Forward-2** | TGATTCAGAAAGCCCAACCT | 84304000-84304026 | 84303769 | -231 |
| **Reverse-2** | CCCACATCACAACCAAAACA | 84304000-84304026 | 84304183 | 183 |
| **Forward-3** | AAAGGGTGAACAAGCAAAGC | 84193364-84193694 | 84193024 | -340 |
| **Reverse-3** | CGCTTGCTCATCTTTTCTCA | 84193364-84193694 | 84193781 | 417 |
| **Forward-4** | CAGCCACAGTCCTCTGACAA | 84195187-84195313 | 84194988 | -199 |
| **Reverse-4** | GGTGCCTTGGTTCTGTTGTT | 84195187-84195313 | 84195515 | 328 |

**Table 2.3 Sequences of primers which correspond to the integrations within BNC2 observed in the genetic screen**

## 2.3 Constructs generated for validation

### 2.3.1 PB-CAG-BNC2

A transposon cassette containing CAG promoter and RARγ and LRH1, was obtained from Wei Wang, Wellcome Trust Sanger Institute (Wang et al., 2011c). Using EcoRI and NotI, cDNAs encoding RARγ and LRH1 were excised. BNC2 cDNA was obtained from GeneService and primers flanking the cDNA were designed. The forward and reverse primers were tethered to EcoRI and NotI recognition sites respectively (**Table 2.4**). BNC2 cDNA was amplified using Pwo Master (Roche) according to the manufacturer's instructions. A portion of the PCR product was visualised on an agarose gel to determine its purity and size. The remaining PCR product was purified using QIAGEN PCR Purification Kit and digested using NotI (New England Biolabs) at 37°C for an hour. As there are internal EcoRI restriction sites within the coding sequence of BNC2, 0.05µl of EcoRI (New England Biolabs) was added to the digested mixture and incubated at 37°C for 15 minutes. The digestion mixture was immediately run on an agarose gel. The partial digestion gave rise to a series of bands and the one of correct size was extracted and purified according to the manufacturer's instructions (QIAGEN Gel Purification Kit). The purified fragments were ligated to a pre-digested PB-CAG-bpA plasmid backbone described earlier, using T4 Ligase (New England Biolabs) in a 10µl reaction at room temperature for 1 hour. 1.5µl of the ligation mix was introduced to a vial of MegaX DH10β electrocompetent cells (Invitrogen) and transferred to an electroporation cuvette. The cuvette was placed into the BioRad GenePulser X cell and electroporation took place at a setting of 1800V, 25µF, 200Ω. Immediately after electroporation, 300µl of Lysogeny Broth (LB) was added to the bacterial culture to resuscitate the cells and the entire mixture was subsequently incubated at 37°C, 200rpm, for 1 hour. The cells were then plated onto a pre-warmed LB Agar plate supplemented with 100µg/mL ampicillin and incubated overnight at 37°C. Bacterial colonies were picked the next day and grown in LB supplemented with 100µg/mL ampicillin and incubated for 15 hours at 37°C. Following the manufacturer's protocol, the cloned plasmids were retrieved from the bacterial cultures using QIAGEN Miniprep Kit.

| PRIMER NAME | RESTRICTION SITE | SEQUENCE |
|---|---|---|
| mBNC2 Forward | EcoRI | AATTGAATTCTGACAACTCCATGCAGTTCG |
| mBNC2 Reverse | NotI | AATTGCGGCCGCCTAATCTATTGAAGTGAAGGGA |
| mRARβ Forward | EcoRI | AATTGAATTCAGGATAAGCACTTTTGCAGAG |
| mRARβ Reverse | NotI | AATTGCGGCCGCCTGCAGCAGTGGTGACTGA |
| mRORα Forward | BamHI | AATTAAGCTTCGCGGCGTAAAGGATGTATTT |
| mRORα Reverse | XbaI | AATTCTAGATTACCCATCGATTTGCATGGC |

**Table 2.4 Sequences of primers used for cloning (Chapters 2.3.1- 2.3.6)**

### 2.3.2 PB-CAG-RARβ

A similar approach was taken to generate PB-CAG-RARβ. cDNA corresponding to full length RARβ was obtained from GeneService and primers flanking the cDNA were designed. EcoRI and NotI sites were included in the forward and reverse primers respectively (**Table 2.4**). RARβ cDNA was amplified using Pwo Master (Roche) according to the manufacturer's instructions and purified. As RARβ contains internal EcoRI recognition sites, the PCR product was first digested using NotI and partially digested by EcoRI. The correct product size was identified and purified, before ligating to a pre-digested PB-CAG-bpA plasmid backbone.

### 2.3.3 PB-CAG-RORα

An identical approach was taken to generate PB-CAG-RORα. cDNA corresponding to full length PB-CAG-RORα was obtained from GeneService and primers flanking the cDNA were designed. BamHI and XbaI sites were included in the forward and reverse primers respectively (**Table 2.4**). PB-CAG-RORα cDNA was amplified using Pwo Master (Roche) according to the manufacturer's instructions and purified. The correct product size was identified and purified, before ligating to a pre-digested PB-CAG-bpA plasmid backbone.

### 2.3.4 PB-TRE-BNC2

A similar approach was taken to generate PB-TRE-BNC2. PB-TRE-bpA plasmid backbone was achieved through the digestion of PB-TRE-RARγ-2A-LRH1 using EcoRI and NotI (**Table 2.4**). This plasmid was obtained from Wei Wang, Wellcome Trust Sanger Institute (Wang et al., 2011c). Using PCR products of BNC2 which have been digested using EcoRI and NotI, as described above, both the insert and plasmid backbone were ligated and transformed into MegaX DH10β electro-competent cells.

### 2.3.5 PB-TRE-RARβ

A similar approach was taken to generate PB-TRE-RARβ. PB-TRE-bpA plasmid backbone was achieved through the digestion of PB-TRE-RL using EcoRI and NotI (**Table 2.4**). Using PCR products of RARβ which have been digested using EcoRI and NotI, as described above, both the insert and plasmid backbone were ligated and transformed into MegaX DH10β electro-competent cells.

### 2.3.6 PB-TRE-RORα

A similar approach was taken to generate PB-TRE-RORα. PB-TRE-bpA plasmid backbone was achieved through the digestion of PB-TRE-RL using BamHI and XbaI (**Table 2.4**). Using PCR products of RORα which have been digested using BamHI and XbaI, as described above, both the insert and plasmid backbone were ligated and transformed into MegaX DH10β electro-competent cells.

### 2.3.7 PB-TRE-CKS

A similar approach was taken to generate PB-TRE-CKS. PB-TRE-bpA plasmid backbone was achieved through the digestion of PB-TRE-RL using EcoRI and NotI. PB-CAG-CKS (as described in 2.2.1) was digested using EcoRI and NotI to excise c-Myc-2A-Klf4-2A-Sox2. This fragment was ligated with a pre-digested plasmid backbone containing PB-TRE-bpA and transformed into MegaX DH10β electro-competent cells.

### 2.3.8 Other plasmids used in the study

PB-CAG-rtTA, PB-CAG-LRH1, PB-CAG-RARγ, PB-CAG-RARα, PB-CAG-OCKS, PB-CAG-RL, PB-TRE-LRH1, PB-TRE-RARγ, PB-TRE-RARα, PB-TRE–OCKS and PB-TRE-RL were obtained from Wei Wang, Wellcome Trust Sanger Institute (Wang et al., 2011c).

## 2.4 Analysis of iPSC colonies

### 2.4.1 Reverse Transcription Polymerase Chain Reaction (RT-PCR)

RNA was first extracted using RNeasy Mini Kit (Qiagen) following manufacturer's guidelines. Subsequently, reverse transcription reactions were performed using QuantiTect Reverse Transcription Kit (Qiagen). PCR was then carried out using primers against pluripotency marker genes (Sigma) (**Table 2.5**) and an Extensor Hi-Fidelity PCR master mix 2 (ABgene). PCR was performed using the following settings: 94°C for 4 min, followed by 30 cycles of 94°C for 10 sec, 55°C for 30 sec and 68°C for 30 sec. Final incubation was at 68°C for 10 min.

| PRIMER NAME | SEQUENCE |
|---|---|
| **mOct4-RT-En-Forward** | TCTTTCCACCAGGCCCCCGGCTC |
| **mOct4-RT-En-Reverse** | TGCGGGCGGACATGGGGAGATCC |
| **mSox2-RT-En-Forward** | TTGCCTTAAACAAGACCACGAAA |
| **mSox2-RT-En-Reverse** | TAGAGCTAGACTCCGGGCGATGA |
| **mKlf4-RT-En-F orward** | GCGAACTCACACAGGCGAGAAACC |
| **mKlf4-RT-En-Reverse** | TCGCTTCCTCTTCCTCCGACACA |
| **mNanog-RT-En-F orward** | CAGGTGTTTGAGGGTAGCTC |
| **mNanog-RT-En-Reverse** | CGGTTCATCATGGTACAGTC |
| **mRex1-RT-F orward** | ACGAGTGGCAGTTTCTTCTTGGGA |
| **mRex1-RT-Reverse** | TATGACTCACTTCCAGGGGGCACT |
| **mEsg1-RT-F orward** | GAAGTCTGGTTCCTTGGCAGGATG |
| **mEsg1-RT-Reverse** | ACTCGATACACTGGCCTAGC |

| | |
|---|---|
| **mFgf4-RT-Forward** | CGTGGTGAGCATCTTCGGAGTGG |
| **mFgf4-RT-Reverse** | CCTTCTTGGTCCGCCCGTTCTTA |
| **mDax1-RT-Forward** | TGCTGCGGTCCAGGCCATCAAGAG |
| **mDax1-RT-Reverse** | GGGCACTGTTCAGTTCAGCGGATC |
| **mGdf3-RT-F orward** | GTTCCAACCTGTGCCTCGCGTCTT |
| **mGdf3-RT-Reverse** | AGCGAGGCATGGAGAGAGCGGAGCAG |
| **mUtf1-RT-F orward** | GGATGTCCCGGTGACTACGTCTG |
| **mUtf1-RT-Reverse** | GGCGGATCTGGTTATCGAAGGGT |
| **mBactin-RT-Forward** | GTTTGAGACCTTCAACACCCC |
| **mBactin-RT-Reverse** | GTGGCCATCTCCTGCTCGAAGTC |

**Table 2.5 Sequences of primers used for RT-PCR of pluripotency markers**

### 2.4.2 Quantitative real time PCR

In a probe-based assay (**Table 2.6**), quantitative real time PCR was performed using Absolute Fast qPCR Master Mix (ABgene) according to manufacturer's instructions. Data were normalised with GAPDH. The assay was performed using the following settings: 95°C for 5 min, followed by 40 cycles of 95°C for 1 sec and 60°C for 20 sec.

| ASSAY ID | TARGET | APPLIED BIOSCIENCE GENE NAME |
|---|---|---|
| **Mm00658129_gH** | mOct4 | POU domain, class 5, homeobox 1 |
| **Mm00516104_m1** | mKlf4 | Kruppel-like factor 4 |
| **Mm02384862_g1** | mNanog | Nanog homeobox |
| **Mm03053975_g1** | mRex1 | Zinc finger protein 42 |
| **4352341E** | β-actin | Mouse ACTB Endogenous Control |

**Table 2.6 Applied Bioscience pre-designed TaqMan probes to detect quantitative levels of pluripotency markers**

Using a primer-based approach, $RT^2$ qPCR SYBR Green/ROX mastermix (Qiagen) was applied following the manufacturer's instructions. PCR was performed at 95 °C for 10 min followed by 40 cycles of 95 °C for 15 sec and 60 °C for 60 sec. Data was normalised to GAPDH.

### 2.4.3 Immunostaining

iPSCs were pre-cultured in gelatin coated 24-well plates and fixed with 4% paraformaldehyde in PBS for 10 min at room temperature and permeabilised with 0.3% Triton X100 in PBS. The cells were then blocked with blocking buffer (1% BSA, 3% goat serum in 0.3% Triton in PBS) for 10 min at room temperature. Primary antibodies were made up using the blocking buffer with the following dilutions: Nanog (1:150, Rabbit

polyclonal; ab21603, Abcam), SSEA1 (1:10, mouse monoclonal, gift from Sheffield University) and Oct4 (1:150, Mouse monoclonal; sc-5279, Santa Cruz). After overnight incubation at 4°C, cells were washed 4 times with 0.3% Triton X100 in PBS and labelled with Alexa 488- or Alexa 594-conjugated secondary antibodies (Invitrogen). After a final wash with 0.3% Triton X100 in PBS, the cells were incubated with DAPI and visualised using Olympus IX81.

### 2.4.4 Bisulfite Sequencing

DNA from iPSCs was extracted as described above. Bisulphite conversion was performed using EpiTect Bisulfite Kit (Qiagen), according to manufacturer's recommendations. Approximately 400bp of the promoter regions of Rex1, Oct4 and Nanog was amplified using the following primers:

The PCR fragments were cloned into pGEMT-EASY (Promega) and sequenced. The sequences obtained were aligned to the following templates and CpG dinucleotides which were analysed are underlined.

Oct4

GGTTTTTTAGAGGATGGTTGAGTGGGTTGTAAGGATAGGTCGAGAGGGTGTAG
TGTTAATAGGTTTTGTGGTGCGATGGGGTATTCGAGTAATTGGTTTGTGAGGTG
TTCGGTGATTTAAGGTAGGGGTGAGAGGATTTTGAAGGTTGAAAATGAAGGTT
TTTTGGGGTTTCGTTTTAAGGGTTGTTTTGTTTAGACGTTTTTAATTTTCGTTTG
GAAGATATAGGTAGATAGCGTTCGTTTTAGTTTTTTTTATTTTTATAGTTTTGTT
TTTTTATTTATTTAGGGGGCGGGGTTAGAGGTTAAGGTTAGAGGGTGGGATTG
GGGAGGGAGAGGTGAAATCGTTTTTAGGTGAGTCGTTTTTTTATTAGGTTTTCG
GTTCGGGGTGTTTATTTTTTTTATGGTTGGATATTTGGTTTT

Nanog

GATTTTGTAGGTGGGATTAATTGTGAATTTATAGGGTTGGTGGGGCGTGGGTG
TCGTTTGGGTGTTTGGGAGAATAGGGGGTGGGTAGGGTAGGAGGTTTGAGGG
GGGAGGAGTAGGATTTATTTTTTAAATTTATCGTTTTGAGTCGTTGGTTTTTAG

ATAGGTTGATTTGGTTGGTGTTTTGTTTTTTTTGTGGGAAGGTTGCGGTTTATTT
TTTTTTGATTTTTTGATAATTTTGTATTAGATATTTAATTTTTTTTTTTTATGATTT
TTTTTTTTAGATATTGAGTTTTTTGGTTGTTGTTTAAAATTTTTTTAGAAATTTTT
TTTTTCGTTATTATATTGATATGAGTGTGGGTTTTTTTGGT

Rex1

TATATTAATGTTGGAAAAAGTTTAGGTAATTAGTGTATTTTGTAGCGGGGTTCG
GGAGAGGTTGGGGTTTAGAGTGGCGATGGGACGAAAGTGTAAAGAAAGTTTT
CGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAGGTAGGTAG
GTAGAAATTGTATTTTTTGTTTGTGTAAATTCGGTTATAAGTTTGTAATAGAGG
TATTGAGATGTGATTGAGTTTTAAGGTTAGGCGATCGGGATTTTGAAGAGGTA
TTTGTATAATTGAGTAAGAGTTTTTGTTTTATTTTTTTACGTCGATTTAAGATAA
GGGGATAGTTTTGGGTGGGTTATTTGAAGGGTTAGGGGTTCGTTTATATTTTCG
TTTATATTTATTTTGAGCGTTTTTTATTGGGTATTTTAGTTTATTTAGTTCGGAG
TTAGTTATTTCGTAAGTGTGGTCGGAATAGAGTTCGTTTATTTAAAAAGGGAG
GGGTTTAAGGAGTT

| PRIMER NAME | SEQUENCE |
|---|---|
| Oct4-DMR-Forward | GGCCCCAGAAATAATTGG |
| Oct4-DMR-Reverse | TTTTGGTTTATAGGTTGG |
| Rex1-DMR-Forward | CACATTAATGCTGGAAAAAGTTCAGGCAA |
| Rex1-DMR-Reverse | TTGAGGAATTTGGGGAGGGAAAAATTTATT |
| Nanog-DMR-Forward | GACCCTGCAGGTGGGATTAACTGTGAATTC |
| Nanog-DMR-Reverse | TGGTTTTTTTGGGTGTGAGTATAGTTATAT |

**Table 2.7 Sequences of primers used to amplify promoter regions of pluripotency markers for bisulfite sequencing**

### 2.4.5 *In vitro* differentiation

For differentiation into the ectoderm lineage, iPSCs were seeded at a density of $2x10^5$ cells / well in a 6-well plate. The cells were cultured in NDiff® N2B27 (Stem Cells, Inc).

For differentiation into the endoderm and mesoderm lineage, iPSCs were trypsinised and seeded onto an uncoated 6-cm dish to form embryoid bodies. After 3 days, the embryoid bodies were seeded onto a gelatin coated 6-cm dish and media was switched to M10 for 3 days before staining.

### 2.4.6 Teratoma Formation

iPSCs were resuspended at a concentration of $5 \times 10^6$ cells/ml in PBS. 200µl of the cell suspension was administered subcutaneously into the dorsal flank. 3 to 5 weeks after the injection, tumours were observed and surgically dissected from the mice. Samples were fixed in 10% formalin, and embedded in paraffin. Sections were prepared and stained with hematoxylin and eosin.

### 2.4.7 Chimera Production

iPSCs were injected into blastocysts of C57BL/6 (albino) and implanted to C57BL/6 (albino) foster females. Chimeric males were chosen, based on iPSC coat colour contribution, to breed with C57BL/6 (albino) females in order to test for germline transmission.

### 2.4.8 Alkaline Phosphatase Staining

Staining was performed according to manufacturer's instructions (Alkaline Phosphatase Kit, Sigma). iPSCs were washed once with fresh PBS and incubated in 4% PFA in PBS at room temperature for 20 minutes. The cells were then washed once with PBS before

freshly prepared alkaline-dye solution was added. The plates were incubated in the dark for 20 minutes before they were rinsed with distilled water and left to dry.

### 2.4.9 Karyotyping

Cells were treated with 0.1 µg/ml colcemid for 50 minutes before trypsin treatment for collection. 10ml of 75mM KCl was added in a dropwise fashion using a Pasteur pipette. The solution was incubated at room temperature for 10 minutes before centrifugation at 1600rpm for 8 minutes. 1ml of fixing solution which consists of Glacial Acetic Acid and Methanol at a ratio of 3:1 was added to the disturbed cell pellet and centrifuged at 400g for 8 minutes. The pelleted cells were then resuspended in 2ml of fixing solution.

To prepare slides, they were sonicated and soaked in 96% ethanol before drying. 20µl of cell suspension obtained in the previous step was dropped onto the slide and placed in a 50°C water bath for 1 minute. The spread of the chromosomes were inspected under the microscope.

After aging the slides overnight, the slides were placed in 10mM HCl with 0.01% pepsin for 3 minutes. The slides were then washed twice with 2xSSC for 5 minutes each. For dehydration, the slide was passed through an ethanol dehydration series: 70%, 90% and 100% ethanol for 5 min each. Denaturation of DNA was next performed by incubating the slides in 70% formamide and 30% 2xSSC and the slides were put through another series of dehydration as previously described. The slides were air-dried and baked at 65°C for 1 hour.

Formamide was added to the slides at 61°C for 1.5 minutes before quenching in ice cold 70% ethanol. The slides then had to undergo dehydration from 70%, 90% to 100%. The slides were then air-dried and pre-aliquoted probes were added before coverslips were placed.

For detection, the slides were placed into a 43°C water bath for 5 minutes to remove the coverslip. Probes were washed using 50% Formamide and 50% 2xSSC twice for 5 minutes each. Pre-aliquoted antibodies were added onto the slides and incubated at 37°C. The slides were then visualised under the microscope.

### 2.4.10 Analysis of transgene expression

RNA was first extracted using RNeasy Mini Kit (Qiagen) following manufacturer's guidelines. Subsequently, reverse transcription reactions were performed using QuantiTect Reverse Transcription Kit (Qiagen). PCR was then carried out using primers against junctions between transgenes (Sigma) (**Table 2.8**) and an Extensor Hi-Fidelity PCR master mix 2 (ABgene). PCR was performed using the following settings: 94°C for 4 min, followed by 30 cycles of 94°C for 10 sec, 55°C for 30 sec and 68°C for 30 sec. Final incubation was at 68°C for 10 min.

| PRIMER NAME | SEQUENCE | APPLICATION |
|---|---|---|
| **Oct4-Forward** | CAGAAGGGCAAAAGATCAAG | Amplification of Oct4-c-Myc junction |
| **c-Myc-Reverse** | TACGGAGTCGTAGTCGAGGT | |
| **c-Myc-Forward** | CGCCTACATCCTGTCCATTC | Amplification of c-Myc-Klf4 junction |
| **Klf4-Reverse** | GCCGGGCCGGACGCGAACGT | |
| **LRH1-Forward** | ATCCGGGCAATCAGCAAGCA | Amplification of LRH1-bpA junction |
| **bpA-Reverse** | GGCACAGTCGAGGCTGATCAGC | |
| rtTA-Forward | GGACGAGCTCCACTTAGACG | Amplification of rtTA |
| rtTA-Reverse | AGGGCATCGGTAAACATCTG | |

**Table 2.8 Sequences of primers used to detect the presence of reprogramming transgenes**

## 2.5 Chromatin Immunoprecipitation (ChIP)

### 2.5.1 RARβ-directed ChIP in MEFs

Frozen MEFs (as described earlier) derived from wildtype C57BL/6J mice were thawed and 5 million cells were seeded into three 15-cm gelatin-coated plates. After one week, the cells attained 100% confluency. Harvest of the cells was conducted using ChIP-IT Express (Active Motif), according to the manufacturer's instructions. Media was aspirated and 20ml of fresh serum-free media was added. 0.54ml of 37% Paraformaldehyde was added to the media and the plate was rocked gently at room temperature for 20 minutes. The cells were washed once with ice cold PBS and 10ml of Glycine was added to quench the action of formaldehyde. The mixture was rocked gently for 10 minutes at room temperature and washed with ice cold PBS. 2ml of fresh ice cold PBS supplemented with 0.5mM PMSF was added onto the plate for the scraping of cells. A cell scraper was used to collect the cells and the cells were pelleted through centrifugation at 4°C, 2500rpm for 10 minutes. The cell pellet was resuspended in 1ml of supplied lysis buffer supplemented with 1mM PMSF and protease inhibitor cocktail. Lysing of the cells took place on ice for an hour. Cells were pelleted at 5000rpm for 10 minutes and the pellet was resuspended in 600µl of the provided shearing buffer. The mixture was divided into two Lo-Bind eppendorf tubes (Eppendorf), placed into a BioRuptor (Diagenode) and pulsed with 15 cycles of 30 seconds sonication and 30 seconds rest. The resultant mixture was centrifuged at 13,500rpm for 20 minutes and 100µl of the supernatant was used for the execution of ChIP. 20µl of the lysate was used to examine the quality of shearing, as stated in the manufacturer's protocol.

To conduct ChIP, 100µl of the lysate was incubated with ChIP buffer 1, 3µg of RARβ (Santa Cruz) or IgG (Cell signalling) antibody, protease inhibitor cocktail and 25µl of provided magnetic Protein G beads in a total volume of 400µl. The tubes were placed on a rotator in 4°C overnight.

After an overnight incubation with the respective antibodies, the tubes were placed in a rack with a magnetic strip to pellet the beads. Supernatant was removed and 800µl of ChIP buffer 1 was used to wash the beads. The beads were subsequently washed twice with ChIP buffer 2. 50µl of elution buffer was added to the beads and rotated at room temperature for 20 minutes. 50µl of reverse cross-linking buffer was then added and the beads were pelleted. Supernatant was collected and purified using QIAGEN PCR Purification Kit. For quantitative analysis, 1µl of the eluate was used for real-time PCR (**Table 2.9**).

| PRIMER NAME | SEQUENCE |
|---|---|
| 8873bp Forward | AAGGTCAGATGAGGGCATTG |
| 8873bp Reverse | CAGCACCTGAATGATGGATG |
| 8244bp Forward | GGTGCTCTTACCCACTGAGC |
| 8244bp Reverse | AGGCCAAACACTCCAATGAC |
| 6932bp Forward | GGTCTTTTGAGCCACCAGAA |
| 6932bp Reverse | GATTTGCCTCTCTGGGTCAG |
| 4836bp Forward | AGCCAGGGCTACACAGAGAA |
| 4836bp Reverse | GTGGGGAGACAGGAATGAGA |
| 2610bp Forward | CCAAGTGCTGGGATTAAAGC |
| 2610bp Reverse | TCTGCCCCCTTTAAGAGTCA |
| 2170bp Forward | GGCTGCAGGCATACTTGAAC |
| 2170bp Reverse | GCTACAACCTCCCCACACC |
| 1940bp Forward | CTCTCGTCCTAGCCCTTCCT |
| 1940bp Reverse | CCTCCACTCTGTCATGCTCA |
| 1611bp Forward | TGGTGAAGTCGATGAAGCTG |
| 1611bp Reverse | GAGCTGTTGGCTAGGGTCAG |
| 1130bp Forward | TCCTCCTAATCCCGTCTCCT |
| 1130bp Reverse | ATACCCTGCTTCCCTTCCTC |
| 275bp Forward | AATTGGCACACGAACATTCA |
| 275bp Reverse | GTCCTTACAGCCCACTCAGC |
| 100bp Forward | CCTAAGGGTTGTCCTGTCCA |
| 100bp Reverse | AGCGCTATCTGCCTGTGTCT |

**Table 2.9 Sequences of primers used for ChIP-qPCR of putative RAREs upstream of Oct4**

### 2.5.1 H3K4me3-directed ChIP in differentiated iPSCs

4F and 6F iPSCs were maintained on 10-cm gelatin coated plates in 2i-LIF. When confluent, the cells were collected using trypsin and seeded onto three 15-cm gelatin coated plates. The cells were maintained in M10 supplemented with 0.1μM all-trans retinoic acid for 6 days to drive differentiation. After 6 days, the cells were treated with doxycycline as described and cross-linked and harvested as earlier described. After lysis and sonication, the lysate was incubated at 4°C overnight with 1μl of α-H3K4me3 antibody (Cell signalling). The pull down was conducted as earlier described.

## 2.6 Luciferase Assay

### 2.6.1 Cloning of luciferase reporter constructs

#### 2.6.1.1 Constructs which test promoter function

To generate fragments consisting of various elements within the upstream region of Pou5f1, forward primers were designed at 2190bp, 1393bp, and 477bp upstream of the transcriptional start site (TSS). The reverse primer was designed 11bp before the TSS. The forward and reverse primers were flanked with XhoI and KpnI restriction sites respectively (**Table 2.10**).

PCR was conducted using mouse genomic DNA as a template. Pwo master mix (Roche) was employed to execute the PCR. An aliquot of the reaction mix was visualised on an agarose gel to ascertain the purity and size of PCR product. The remaining mixture was purified using QIAGEN PCR Purification Kit and digested using XhoI and KpnI (New England Biolabs) at 37°C for 1 hour. The digested products were purified using the QIAGEN PCR Purification Kit. pGL3-basic (Promega) was also digested using XhoI and KpnI at 37°C for 1 hour. The inserts and plasmid backbone were ligated using T4 ligase (New England Biolabs) and transformed into One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen) following the manufacturer's instructions. Once recovered in LB, the

bacterial cells were plated onto LB agar plates supplemented with 100μg/ml ampicillin and picked the next day to inspect for correct insertions.

| PRIMER NAME | RESTRICTION SITE | SEQUENCE |
|---|---|---|
| **DE-PE-RAREoct Forward** | XhoI | AATTCTCGAGAGGGCACATCTGTTTCAAGC |
| **DE-PE-RAREoct Reverse** | KpnI | GGTACCCTCTAGACGGGTGGGTAAGCAAGA ACTG |
| **PE-RAREoct Forward** | XhoI | AATTCTCGAGAAACACCATCCCTTGCAGAC |
| **PE-RAREoct Reverse** | KpnI | GGTACCCTCTAGACGGGTGGGTAAGCAAGA ACTG |
| **RAREoct Forward** | XhoI | CTCGAGCTCTAGACGGGTGGGTAAGCAAGA ACTG |
| **RAREoct Reverse** | KpnI | GGTACCCTCTAGACGGGTGGGTAAGCAAGA ACTG |
| **DE Forward** | XhoI | AATTCTCGAGAGGGCACATCTGTTTCAAGC |
| **DE Reverse** | KpnI | AATTGGTACCAGGAAGGGCTAGGACGAGAG |
| **DE-PE Forward** | XhoI | AATTCTCGAGAGGGCACATCTGTTTCAAGC |
| **DE-PE Reverse** | KpnI | AATTGGTACCGTCCTTACAGCCCACTCAGC |

**Table 2.10 Primers used for the cloning of Oct4 regulatory elements into luciferase reporter vectors**

### 2.6.1.2 Constructs which test enhancer function

To generate fragments consisting of various elements within the upstream region of Pou5f1, a forward primer was designed at 2190bp upstream of the transcriptional start site (TSS). The reverse primers were designed at 1985bp, 383bp and 11bp before the TSS. The forward and reverse primers were flanked with XhoI and KpnI restriction sites respectively (**Table 2.10**).

The steps for cloning are similar to 2.6.1.1, except the plasmid employed was pGL4.23 (Promega) which comprised of a minimal promoter sequence before the luciferase reporter.

### 2.6.2 Transfection of constructs

To perform luciferase assays, MEFs were transfected with the described luciferase constructs. MEFs were thawed and expanded as previously described. When 70% confluent, the cells were trypsinised and collected as described and 1 million cells were used for each transfection. To explore the effect of cDNAs on luciferase activity, 10μg of each cDNA and 10μg of luciferase reporters were used for transfection. 1μg of TK-Renilla was co-introduced for the normalization of firefly luciferase readings. Transfection was performed as described earlier and resuscitated cells were transferred into a fresh 6-well gelatin coated plate pre-fed with M10. The experiment was conducted in triplicates.

### 2.6.3 Performing the assay

Luciferase assays were conducted according to manufacturer's instructions. Dual-Luciferase® Reporter Assay System (Promega) was used for this purpose. After 48 hours, unless otherwise specified, the cells were washed with PBS twice and lysed using the buffer provided. The cells were rocked at room temperature for 20 minutes and 20μl of the lysate was transferred to a 96-well luminometry plate. To obtain luciferase readings, 100μl of Luciferase Assay Reagent was added to each well and the plate was placed in

MicroLumat Plus LB 96V (Berthold Technologies). WinGlow was the program used to read the luciferase activity. After the firefly luciferase readings were taken, 100μl of Stop&Glo Reagent was added to quench the firefly luciferase activity and trigger Renilla activity. WinGlow was also used to determine the level of Renilla activity.

## 2.7 Manipulation of ZHBTc4 ESCs

### 2.7.1 Differentiation of ZHBTc4 ESCs

Differentiation of ZHBTc4 ESCs was performed under similar conditions to 4F and 6F iPSCs. ZHBTc4 ESCs were maintained in 6-well plates. Once confluent, the cells from one 6-well were trypsinised and seeded onto one gelatin coated 10-cm plate. The cells were differentiated in M10, supplemented with 0.1μM all-trans retinoic acid and 1μg/ml doxycycline, for 6 days.

### 2.7.2 Transfection of potential rescue constructs

After 6 days of differentiation, ZHBTc4 cells were trypsinised and collected as previously described. cDNAs were analysed for their ability to reprogram these differentiated cells in the presence or absence of doxycycline. To address this, cDNAs were introduced at 1μg per construct and appropriate cDNA combinations were transfected into 1 million ZHBTc4 cells. Transfected cells were divided into two 6-wells that were lined with SNL76/7 feeder cells. In one well, doxycycline was added at 1μg/ml, whereas the other well was left doxycycline-free.

# CHAPTER 3 : PIGGYBAC-ASSISTED SCREEN TO IDENTIFY GENETIC REPLACEMENTS OF EXOGENOUS OCT4 DURING REPROGRAMMING

## 3.1 Introduction

Introduction of four genetic factors, Oct4, c-Myc, Klf4 and Sox2, into somatic cells has been demonstrated to implicate the loss of differentiation, as triggered by broad epigenomic changes, and result in a gain in pluripotency. Resultant pluripotent derivatives resemble ESCs and are termed as induced pluripotent stem (iPS) cells (Takahashi and Yamanaka, 2006). This phenomenon is reproducible in a spectrum of cell types and animal species, but we are only beginning to elucidate the molecular mechanisms necessary to trigger this dedifferentiation process.

### 3.1.1 Methods to extricate pathways and factors involved in the generation of iPSCs

To illuminate the molecular circuitry behind the acquisition of pluripotency, an array of approaches has been employed. Examination of proteomic and transcriptomic profiles over the course of reprogramming has identified molecular pathways, such as chromatin remodelling and mesenchymal-to-epithelial transition, to play significant roles in the initiation of reprogramming events (Samavarchi-Tehrani et al., 2010; Li et al., 2010). Extending the initial strategy employed by Takahashi and Yamanaka, screens which encompass a wider set of transcription factors have been performed to identify genes proficient at replacing or enhancing the reprogramming potential exhibited by the established set of four reprogramming factors (Maekawa et al., 2011; Han et al., 2010b; Nagamatsu et al., 2011). However, these screens focussed on an exhaustive list of transcription factors, where genes-of-interest were co-introduced with the conventional cocktail of reprogramming factors to determine differences in reprogramming efficiencies.

To explore beyond the realm of transcription factors, genome-wide screens have been executed to provide a different perspective on the reprogramming process (Guo et al., 2010). Using a piggyBac assisted mutagenesis in EpiSCs, Nr5a2 was identified to possess the capacity to drive cells from primed to naïve pluripotency. A separate siRNA screen

was performed in secondary MEFs (Samavarchi-Tehrani et al., 2010) to observe for genes that reduce reprogramming efficiency. Using AP expression as a selection criterion, 4010 genes were assessed for their ability to result in a reprogramming deficit. These genetic targets consist of transcription factors, chromatin regulators and signalling elements, which represent a small subset of all annotated genes within the genome. In a bid to increase the breadth of genes studied during the acquisition of pluripotency, this Chapter describes an unbiased genome-wide screens to identify Oct4 replacements for exogenous Oct4 in reprogramming.

### 3.1.2 Identification of Oct4 substitutions that can elicit reprogramming events

Oct4 (Octamer-binding transcription factor-4) belongs to class V of the POU transcription factor family (Okamoto et al., 1990; Schöler et al., 1990a; Schöler et al., 1990b; Rosner et al., 1990) that recognises a consensus octamer motif ATGCAAAT (Klemm et al., 1994) within cis-acting elements and elicits transcriptional activation. Oct4 was the first transcription factor identified to play a role in pluripotency where maternal transcripts of Oct4 are present within oocytes and persists until fertilization (Rosner et al., 1990; Pesce et al., 1998). It was subsequently uncovered that Oct4 functions as a master regulator of the transcriptional network within ESCs, and acts in coordination with Sox2 and Nanog to regulate a plethora of genes (Boyer et al., 2005; Kim et al., 2008; Masui et al., 2007). Embracing a pivotal role in the maintenance of pluripotency, Oct4 expression levels are tightly regulated. Absence of Oct4 results in the loss of pluripotency and differentiation into trophoblast cells, whereas ectopic expression of Oct4 results in differentiation into the primitive endoderm (Nichols et al., 1998; Niwa et al., 2000). In addition, ablation of Oct4 leads to embryonic lethality at E3.5, underpinning its imperative requirement during the formation of the mouse epiblast (Nichols et al., 1998). Apart from development, the obligatory need for Oct4 is demonstrated in reprogramming. Several studies have attempted to negotiate the need for the complete set of reprogramming factors. This has been successfully done but Oct4 remains to be essential for the acquisition of pluripotency (Tsai et al., 2010; Kim et al., 2009b; Kim et al., 2009c; Giorgetti et al., 2009). Efforts have also been made to substitute each of the four reprogramming factors with their family members (Nakagawa et al., 2008). Although Sox2, c-Myc, Klf4 and their respective

relatives display functional redundancy, Oct4 was not readily substituted by Oct1 and Oct6. Furthermore, a recent finding demonstrates the correlation between amounts of exogenous Oct4 and the resetting of the epigenetic landscape in reprogrammed cells, reinforcing the importance of Oct4 (Carey et al., 2011).

The importance of Oct4 in the maintenance and acquisition of pluripotency has implored several groups to scout for genes that can replace the function of exogenous Oct4 during reprogramming. Two studies have successfully identified Nr5a2 and E-cadherin (Heng et al., 2010; Redmer et al., 2011) as competent substituents. However, as previously mentioned, their methods for identification were candidate-gene approaches and a large proportion of coding genes were neglected. This project aims to overcome this bottleneck by determining replacements of exogenous Oct4 through an unbiased genome wide transposon mediated mutagenesis approach.

### 3.1.3 Transposon-assisted mutagenesis

Transposons are mobile genetic elements that integrate into the host genome in the presence of its accompanying transposase. A spectrum of transposon systems, such as Sleeping Beauty, Tol2 and piggyBac, have been isolated from salmon, medaka fish and cabbage looper moth respectively and demonstrated to display transposition capacities in mammalian systems (Ivics et al., 1997; Kawakami et al., 2000; Fraser et al., 1996; Ding et al., 2005; Cadinanos and Bradley, 2007; Wang et al., 2008). Additional transposons have been described to function in disparate organisms (Bingham et al., 1982; Rubin et al., 1982; Bessereau et al., 2001). In particular, the P-element which was discovered in natural populations of *Drosophila melanogaster* but not laboratory stock, has been widely used as a molecular gene manipulation tool (Cooley et al., 1988; Gloor et al., 1991). To this end, P-elements have successfully been used as a mutagen in *Drosophila melanogaster* through excision and subsequent gap repair of DNA double strand breaks (Gloor et al., 1991).

The ability of transposons to integrate into host genomes has been utilised to perform genetic screens in the search of genes that participate in tumorigenesis and the acquisition of pluripotency and antibiotic resistance (Starr et al., 2009; Dupuy et al., 2009; Keng et al., 2009; Rad et al., 2010; Mann et al., 2012; Guo et al., 2010; Guo et al., 2011; Wang et al., 2011b). Two transposon systems, Sleeping Beauty and piggyBac, display competent transposition capacities in mammalian systems and serve as popular choices to conduct genetic screens. These genetic screens identified an assortment of expected and novel candidate genes, fortifying its ability as a tool to unravel genetic components within a cellular phenotype of choice.

As genomic coverage of mobilised transposons varies across systems, it is important to be aware of their inherent flaws and capabilities while designing a genetic screen. Sleeping Beauty has been demonstrated to exhibit local hopping where half of excised transposons re-integrate within the same chromosome (Luo et al., 1998). In contrast, a smaller fraction (18%) of piggyBac transposons was reported to re-integrate into the original chromosome (Wang et al., 2008). As local hopping may result in local deletions and cell death (Geurts et al., 2006), transposition frequencies can be elevated to an extent where incidences of local hopping are insignificant (Dupuy et al., 2005).

Another property exhibited by transposons is the tendency to integrate into recurrent genomic regions. These sites are labelled as hotspots and introduce a bias during the performance of a genetic screen. Analysis of Sleeping Beauty integration sites delineated a statistical bias toward transcriptional units and their upstream regulatory sequences, whereas P-elements display preference toward the 5′ ends of transcriptional units in flies (Yant et al., 2005; Tsubota et al., 1985; Kelley et al., 1987). Similarly, piggyBac has been described to harbour preferential integration sites in long terminal repeats, CpG islands and DNaseI hypersensitive sites when introduced into human cells (Meir et al., 2011; Wilson et al., 2007; Huang et al., 2010). In contrast, delivery of piggyBac into mouse cells leads to even distribution across the genome where 43% of insertion sites were observed in annotated genes (Wang et al., 2008). Taking into consideration that (i) this thesis employs

the use of mouse as the study organism and (ii) piggyBac portrays enhanced efficiencies when compared to alternative transposon systems (Wang et al., 2008; Liang et al., 2009), piggyBac was chosen to conduct the genetic screen.

In order to gain a clearer perspective on the properties of piggyBac transposon mediated mutagenesis, a comprehensive study on transposition efficiencies of piggyBac in mammalian cells has been conducted (Wang et al., 2008). It was demonstrated that transposition efficiencies is proportional to the amounts of helper (transposase) and donor (transposon) plasmids present. Increasing proportions of donor plasmids from 1-50 μg augments transposition efficiencies by 14-fold, whereas increasing helper plasmid proportions in a similar fashion augments transposition efficiencies by 4-fold (Wang et al., 2008). In addition, increasing amounts of donor plasmids favour higher incidences of integrations. However, escalated volumes of transfected DNA may lead to cell death and induce cellular stress which could invoke debilitating consequences (Felgner et al., 1987).

As mentioned above, the piggyBac transposition-mediated mutagenesis poses as a valuable tool to trap genes and influence their expression levels. This can be executed at two levels. First, donor plasmids may encompass strong constitutive promoters preceding a splice donor site, thus resulting in forced expression of the integrated gene (Rad et al., 2010). Second, the presence of splice acceptor sites adjacent to poly-A signals may lead to disruption of gene transcription (**Figure 1.10**) (Guo et al., 2011). Both methods deregulate gene expression levels and create truncation products. If the mutation results in a phenotypic change that can be scored, phenotypic mutants and their corresponding genetic disruptions can be identified and validated.

### 3.1.4 Chapter Aim

In this chapter, I aim to employ piggyBac transposons as a tool to create a mutagenised library of MEFs. The introduced mutagen is linked to an incomplete set of reprogramming factors, c-Myc, Klf4 and Sox2, driven by a composite CMV/ chicken β-actin (CAG)

promoter (Niwa et al., 1991). In addition, the donor plasmid comprises of a second CAG promoter followed by a splice donor site or not. To recapitulate reprogramming events, only cells which exhibit integrations within genes that readily replace the function of exogenous Oct4 would trigger dedifferentiation. The ability to undergo reprogramming was scored and integrations responsible for the phenotype were analysed.

The analysis may culminate in the unravelling of possible substitutes of Oct4, comprising of (i) upstream regulators of Oct4, (ii) downstream targets of Oct4 and (iii) participants of alternative pathways regulating gene targets of Oct4. These gene targets may shed light on the intricacies behind the molecular circuitry involving Oct4 and perhaps uncover novel pathways in the maintenance of pluripotency, hence deepening our understanding of the reprogramming process. As such, this chapter provides a stepping block to address these pertinent questions and further validation will be carried out in subsequent chapters.

## 3.2 Experimental Design

### 3.2.1 Factors taken into consideration for a saturated genetic study

The aim of this study encompasses the use of piggyBac mediated mutagenesis to identify factors that can act as replacements to exogenous Oct4 during the acquisition of pluripotency. In order to conduct a comprehensive genome-wide search, the experiment was designed such that the genome could be saturated with transposon integration events and most genes would be assessed for their potential to act as an Oct4 substitute. To design the screen, experimental parameters were adjusted according to known technical capacities of piggyBac mediated mutagenesis and are illustrated below.

First, the delivery method employed to introduce the transposon cassettes was established. Due to high cell viability and transfection efficiency initiated by Nucleofection$^{TM}$ (Amaxa, Lonza), this transfection method was selected. In addition, Nucleofection$^{TM}$ shuttles

plasmids directly into the nucleus instead of the cytosol (Gresch et al., 2004). As MEFs are slow cycling cells, delivery of exogenous material into the nucleus is beneficial.

Second, the transposition efficiencies of piggyBac were considered to determine the amount of donor plasmids and number of MEFs necessary to prompt a saturated genome-wide screen. To gain a perspective on the efficiencies of piggyBac assisted transposition, references were made to a comprehensive analysis of piggyBac that had been previously conducted using ESCs (Wang et al., 2008). ESCs were employed in the study due to their ability to generate colonies from single cells, posing as an attractive tool for the quantification of transfection efficiency. As this experimental study starts with the application of MEFs, ESCs can only serve as a guideline to design experimental constraints. The introduction of 10μg of donor plasmids results in transposition efficiencies up to 10%, depending on transposase amounts. In this study, transposase activity was introduced through two different avenues, as determined by the employment of two variations of MEFs. The first comprises of piggyBac transposase (PBase) constitutively expressed under the endogenous ROSA26 promoter (Zambrowicz et al., 1997). The ROSA26 locus was originally discovered in a gene-trap screen in murine ESCs, and was subsequently described to be ubiquitously expressed during embryogenesis and in adult mice (Friedrich and Soriano, 1991; Zambrowicz et al., 1997). The employment of ROSA26 allows constitutive expression of PBase (Wang et al., 2008) and ensures effective mobilisation of transposon cassettes. The second variation of MEFs lacks the presence of PBase under the endogenous ROSA26 promoter and relies on the transient ectopic expression of PBase (Wang et al., 2008). Despite considerable differences in PBase activities in both cell populations, calculations were based on a conservative estimate implicating the exogenous introduction of 2μg of PBase in 1 x $10^6$ cells. In these conditions, introduction of 10μg of transposon plasmids results in approximately 1% of surviving cells successfully integrating at least one copy of the transposon cassette (Wang et al., 2008). Moreover, as the reference study employed conventional electroporation methods, the predicted transfection efficiency is an underestimate.

Armed with knowledge that 1% of transfected cells receive transposon induced mutations, the next step was to resolve the number of integration sites necessary to saturate the genome. The mouse genome has been successfully sequenced and analysed within the past decade (Waterston et al., 2002). The size of the euchromatic mouse genome was estimated to be approximately 2.5Gb, through the comparison of sizes of ultracontigs and their corresponding gaps. As each cell contains two sets of chromosomes, the total size of euchromatin within each cell is 5Gb. On the basis of evidence-based analysis and *de novo* gene predictions, there are estimated to be 30,000 protein coding genes. Using the human genome as a reference and assuming that the mouse genome contains a similar number of protein coding genes, it is predicted that one gene exists at every 83kb interval within the mouse genome. As a conservative estimate to obtain a saturated genome in this study, the acquisition of an integration event every 20kb was decided to be sufficient to incorporate mutations into each coding gene. To be able to attain an integration event at every 20kb interval, there should ideally be $1.25 \times 10^5$ genomic insertions in the entire screen, on the premise that integrations take place in a random fashion.

As 1% of surviving cells comprise of at least one integration site, $1.25 \times 10^7$ cells have to be transfected in order to generate $1.25 \times 10^5$ integrations. Keeping in mind that the use of MEFs that constitutively express PBase would result in transposition efficiencies exceeding the conservative estimate of 1% (Wang et al., 2008), $1 \times 10^7$ cells were employed at the beginning of the genetic screen.

### 3.2.2 Constructs generated for the execution of the genetic screen

In a bid to conduct a transposon based screen to identify replacement factors for exogenous Oct4 during reprogramming, transposon cassettes were generated to serve our purpose (**Figure 3.1**). Flanked by piggyBac transposon arms, c-Myc, Klf4 and Sox2 were incorporated into a CAG-driven expression cassette (**Figure 3.1**). Unpublished data from previous work has demonstrated the ability of CAG promoters to drive gene expression at significantly higher levels than promoters such as Murine Stem Cell Virus (MSCV), Cytomegalovirus (CMV) and Phosphoglycerate kinase (PGK), representing a strong

impetus to trigger reprogramming events. Another feature of the transposon cassette was the presence of 2A peptides interspersed between the three Yamanaka factors. 2A peptides allow the expression of several gene products from a single transcript, hence allowing the design of one vector carrying all three Yamanaka factors. 2A peptides are self-cleaving peptides that originated from foot-and-mouth disease virus (F2A) and comprise of approximately 20–28 amino acids (Szymczak et al., 2004). To confer gene-trapping capacity, an additional CAG-driven expression cassette consisting of a splice donor site was included. The presence of this site allows the transcription machinery to search for the next splice acceptor site and as a result, drive the expression of the gene at the site of integration. In order to allow simultaneous expression of c-Myc, Klf4, Sox2 and the integrated gene, the reprogramming factors and splice donor site were cloned into a single transposon cassette. For labelling purposes, this gene trap cassette is designated as PB-CAG-CKS-CAG-SD **(Figure 3.1)**.

The utility of the splice donor site implicates the creation of truncation products, discounting the first and possibly more exons. To overcome this bottleneck, an additional trapping construct was generated with the omission of the splice donor site. This modification results in the reliance on the inherent ability of a strong constitutive promoter like CAG to drive transcription of genes neighbouring the integration site. As the CAG promoter possesses intrinsic enhancer properties, it can promote the expression of the inserted gene from a distance. This transposon construct will hereby be referred to as PB-CAG-CKS-CAG. Both plasmids are described in detail in **Figure 3.1**.

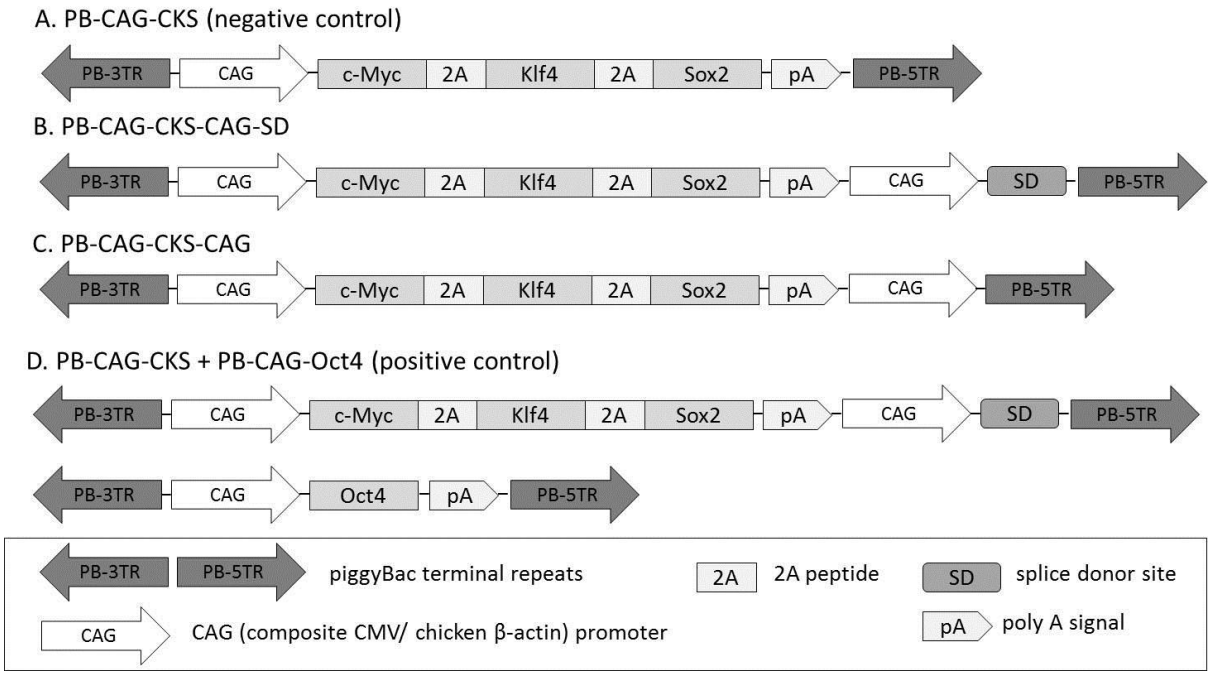



**Figure 3.1 Schematic diagram of piggyBac transposon constructs (**A) PB-CAG-CKS cassette. The dark arrows flanking both ends depict piggyBac terminal repeats. Between the two boundaries lies a constitutively active CAG promoter driving the expression of c-Myc, Klf4 and Sox2. (B) PB-CAG-CKS-CAG-SD cassette. The structure of this plasmid is similar to the top. The only difference lies in the inclusion of a second CAG promoter which precedes a splice donor site. In combination, the CAG promoter and splice donor site would recruit transcriptional machinery and drive transcription starting from the nearest splice acceptor site. (C) PB-CAG-CKS-CAG cassette. The structure of this plasmid is similar to (B), but the splice donor site is lacking. The second CAG promoter will act as an enhancer or promoter to the integrated gene. (D) PB-CAG-CKS-CAG-SD + PB-CAG-Oct4. Positive control included into the genetic screen, encompassing of PB-CAG-CKS-CAG-SD as depicted in (B) and a piggyBac transposon flanking CAG-promoter and cDNA encoding for Oct4, PB-CAG-Oct4.

## 3.3 Results

### 3.3.1 Generation of mutagenised MEF library

To identify potential substitutes for the requirement of exogenous Oct4 in the creation of iPSCs, a library of mutagenised MEFs was generated. piggyBac mediated transposition was used for this purpose as it is a potent mutagen and provides an unbiased coverage of the mouse genome. **Figure 3.2** shows a schematic diagram of the experimental strategy employed. 10μg of donor plasmids (PB-CAG-CKS-CAG-SD or PB-CAG-CKS-CAG) was introduced into 1 x $10^6$ MEFs via Nucleofection$^{TM}$. In order to saturate the genome, transfection was repeated nine times such that a total of 1 x $10^7$ cells were utilised. As a negative control, a construct which comprises only of CAG promoter-driven c-Myc, Klf4 and Sox2, PB-CAG-CKS (**Figure 3.1**), was transfected into 1 x $10^6$ cells. To ensure that the mutagenic construct was functional, PB-CAG-CKS-CAG-SD was co-transfected with a separate construct containing CAG promoter-driven Oct4 into 1 x $10^6$ cells (**Figure 3.1**).

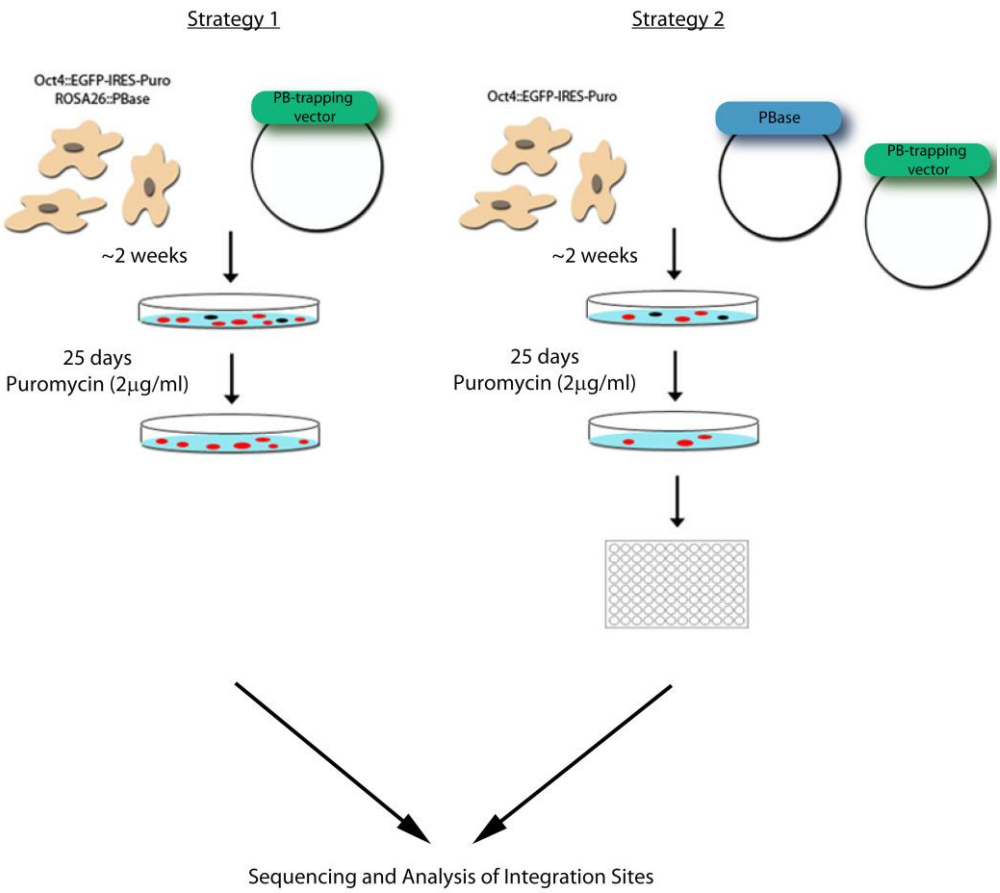



**Figure 3.2 Experimental Strategy of Transposon-assisted Genetic Screen** On the left: Strategy 1 involving PB-CAG-CKS-CAG-SD. MEFs that harbour two modifications were used as starting material: (i) EGFP expression and Puromycin resistance driven by endogenous Oct4 locus, and (ii) PBase expression controlled by the ROSA26 locus. 10µg of PB-CAG-CKS-CAG-SD was introduced into one million MEFs and the process was repeated nine times to create a mutagenised library of 10 million cells. Colonies were observed after approximately 2 weeks and puromycin (2µg/ml) was introduced at 25 days post transfection. Resistant colonies were pooled and DNA was extracted. Splinkerette PCR and capillary sequencing were performed to analyse the integration sites. On the right: Strategy 2 involving PB-CAG-CKS-CAG. MEFs that possess one genetic alteration (EGFP expression and Puromycin resistance driven by endogenous Oct4 locus) were used as starting material. 10µg of PB-CAG-CKS-CAG and 2µg of CAG-PBase were introduced into one million MEFs and the process was repeated nine times to create a mutagenised library of 10 million cells. Colonies were observed after approximately 2 weeks and puromycin (2µg/ml) was introduced at 25 days post transfection. Seven puromycin resistant colonies were picked and expanded. DNA extracted from these colonies was used for splinkerette PCR and amplified products were sequenced to identify integration sites.

#### 3.3.1.1 The PB-CAG-CKS-CAG-SD screen

To facilitate the execution of the screen, MEFs that consisted of two components were utilised: (i) a selection marker to identify reprogrammed colonies and (ii) constitutive expression of PBase to allow efficient transposition of the mutagenic cassette during the genetic screen. To obtain MEFs that comprised of both components, MEFs were obtained from E13.5 embryos generated by a cross between two mouse lines. One mouse line harboured Puromycin resistance cassette and cDNA encoding EGFP under the control of the endogenous Oct4 promoter (Ying et al., 2002), whereas the other mouse line controlled PBase expression under the endogenous ROSA26 locus (Wang et al., 2008). Both mouse lines were derived from a mixed background between 129S5 and C57BL/6. As the activity of the endogenous Oct4 locus corresponds to the pluripotent nature of reprogrammed MEFs, the Puromycin resistance selection permits convenient detection and isolation of cells which have successfully acquired pluripotency.

#### 3.3.1.2 The PB-CAG-CKS-CAG screen

MEFs employed in this study were of a similar background to **Chapter 3.3.1.1**. However, for the purpose of this screen, the expression of PBase was not controlled under the ROSA26 locus. Instead, transient expression of the helper plasmid was driven by electroporation of a vector containing CAG promoter preceding cDNA encoding PBase. Constitutive expression of transposase leads to the continuous mobilisation of donor plasmids. As a result, integrations within ES cell-like colonies may not faithfully reflect the mutations that participated in the reprogramming event. To circumvent this, fleeting levels of the helper plasmid were introduced at primary stages of the experiment. This allows the trapping of genes essential for the trigger of reprogramming events and these modifications remain intact for detection at later stages. The major disadvantage of this approach is its low efficiency.

### 3.3.1.3 Reprogramming MEFs to iPSCs during the genetic screens

Upon electroporation of the transposon cassettes, cells were seeded onto a layer of feeder cells, SNL76/7, that have been pre-treated with mitomycin-C (McMahon and Bradley, 1990). SNL76/7 are mouse fibroblasts which were clonally derived from STO cell line and transformed to express genes encoding for neomycin resistance and murine LIF. These fibroblasts act as a feeder layer to support the growth of mouse ESCs and iPSCs. To facilitate the reprogramming of MEFs which possess advantageous mutations, cells were maintained in ES cell growth media. The use of both transposon donor plasmids, PB-CAG-CKS-CAG-SD and PB-CAG-CKS-CAG, resulted in the observation of colonies under the microscope after 14 days. Colonies were allowed to expand over a week. There was an assortment in the rate at which colonies became visible, spanning from 14 to 20 days. This could be attributed to the innate properties of the disrupted genes, as gene dosage and the hierarchical positioning of the gene in the reprogramming pathway could affect both the efficacy and efficiency of Oct4 substitution. Furthermore, in situations where constitutive expression of transposase was employed, continuous mobilisation of the transposon cassettes could dictate the time taken for an insertion in a critical gene to occur, potentially instigating the heterogeneous reprogramming rate.

To serve as a control, an independent experiment was conducted using transposon cassettes that were only comprised of the three reprogramming factors, c-Myc, Klf4 and Sox2, regulated by a CAG-promoter (**Figure 3.1**). The three transcription factors alone did not yield any colonies, reaffirming that observations of reprogrammed fibroblasts were attributable to the mutagenic ability of the splice donor site and/or CAG promoter. To verify the pluripotent nature of the ES cell-like colonies, puromycin selection was introduced 25 days after the start of the experiment. As puromycin resistance is encoded under the endogenous Oct4 promoter, colonies which have successfully activated endogenous Oct4 transcription will exhibit resistance.

Not surprisingly, the amount of helper plasmids played a significant role in dictating the transposition efficiencies of piggyBac (Wang et al., 2008). Electroporation of $1 \times 10^7$

MEFs which expressed PBase constitutively under the endogenous ROSA26 locus generated thousands of ES cell-like colonies, whereas MEFs dependent on the fleeting ectopic expression of PBase gave rise to seven colonies.

Five days after puromycin selection, genetic material from reprogrammed colonies was extracted and splinkerette PCR was performed. As numerous colonies were observed when PB-CAG-CKS-CAG-SD was employed as a mutagen, DNA pools from each of the ten electroporations were collected for further analysis. In contrast, transient expression of PBase only resulted in the formation of seven colonies, hence each colony was analysed individually. **Figure 3.3** shows a gel electrophoresis image of splinkerette PCR products obtained from ten pools of DNA originating from the application of PB-CAG-CKS-CAG-SD. The bands observed correspond to transposon integration sites. After purification, PCR products were cloned into pGEMT-EASY plasmids (Promega). 96 colonies were picked from each DNA pool or colony and sequenced.

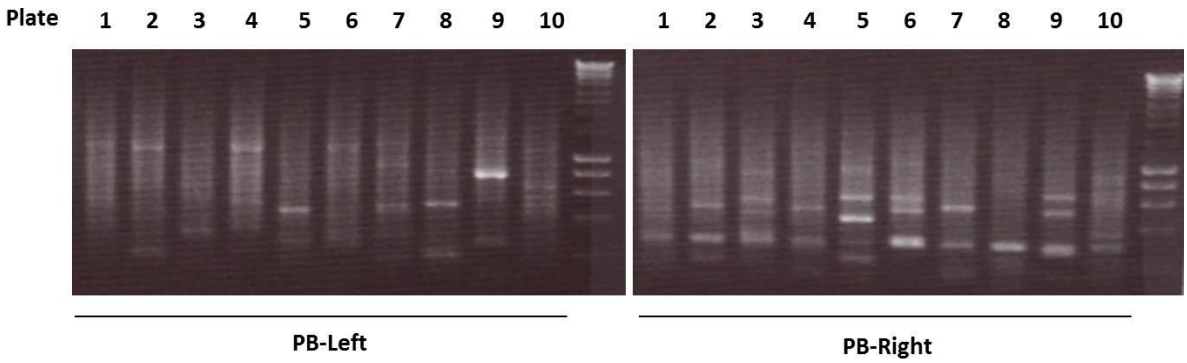



**Figure 3.3 Splinkerette PCR of integration sites** Gel electrophoresis image of splinkerette PCR products obtained from each of ten 10cm plates with induced pluripotent stem cell colonies as instigated by the presence of PB-CAG-CKS-CAG-SD. Each band represents the presence of an integration site. PB-Left and PB-Right refer to primers used to obtain these PCR products. PB-Left employs the use of primers against the left transposon arm and PB-Right employs the use of primers against the right transposon arm.

### 3.3.2 Analysis of the PB Integration sites

Bioinformatics analysis was performed on the sequencing outcomes corresponding to the splinkerette PCR products. This analysis involved the mapping of the sequence to an annotated high coverage mouse assembly, NCBI m37, allowing both the site of integration and the orientation of the inserted transposon to be ascertained. To cleanse the dataset of unwanted reads, genome matches of less than 50 base pairs and insignificant genome matches of less than 90% were ignored. As the two genetic screens exhibited varied dynamics and portrayed dissimilar outcomes, distinct methodology was implemented to trim the list of integration sites and will be described below.

#### 3.3.2.1 Identification of candidate genes in the PB-CAG-CKS-CAG-SD screen

In this screen, PB-CAG-CKS-CAG-SD was employed to mutagenise MEFs. As PB-CAG-CKS-CAG-SD possesses a splice donor site, components of transcriptional machinery will be attracted to the splice donor site upon integration into the host genome and force the recognition of a neighbouring splice acceptor site to initiate transcription. In addition, MEFs that were transfected harboured genetic modifications, where PBase was constitutively expressed from the endogenous ROSA26 locus. A combination of the splice donor site and constitutive expression of PBase led to the formation of a multitude of ES cell-like colonies. The sheer number of reprogrammed colonies made it impossible to analyse individual colonies. Instead, integration sites attained were reflective of a pool of cells. As copious amounts of integration sites were identified, careful reduction in list of integration sites was pertinent. As a result, intergenic regions and transposon cassettes which land in the opposing orientation to the integrated gene were disregarded. In summary, from an initial screen involving $1 \times 10^7$ MEFs, 275 hits were obtained, of which 179 were within intergenic regions and 96 were within annotated genes. After subtracting transposon cassettes which were integrated in an opposite orientation to the targeted gene, 44 gene targets remained. Each of these 44 genes was carefully investigated and **Figure 3.4** shows a summary of the genes and their known functions. Pathways highlighted in the screen may provide clues to the molecular circuitry behind the function of Oct4 in reprogramming or suggest alternative pathways which play crucial roles in driving

pluripotency. Hence a closer look is needed to carefully select potential targets which may replace Oct4.

| Gene | Function/Pathway |
|---|---|
| Gmds | metabolic pathway |
| Mlh3 | DNA damage response |
| Nup210l | |
| Slc44a5 | Poorly characterised |
| Dlgap1 | |
| 2310045A20Rik | |
| AL772224.2 | |
| Atp9a | |
| BC099486 | |
| Bzrap1 | |
| Ccdc120 | |
| Cobll1 | |
| Dnahc5 | |
| RP23-144M4.3 | |
| RP23-309K20.1 | |
| Pkhd1 | |
| Bnc2 | |
| Rasa2 | Cell signaling components |
| Fyn | |
| Ankrd12 | |
| Gipc2 | |
| Nomo1 | |
| Gab1 | |
| Siah1a | ubiquitination |
| Usp34 | |
| Trp63 | Transcriptional regulators involved in developmental processes |
| Sox5 | |
| Rarb | |
| Zbtb2 | |
| Rexo4 | |
| Glis3 | |
| Foxf2 | |
| Cnot4 | |
| Ctnnd2 | cell-cell interaction/communication |
| Spon1 | |
| Fat3 | |
| Palm2 | |
| Galnt1 | |
| Mpp3 | |
| Tsga10 | |
| Nsfl1c | cellular transport |
| Trpc4 | |
| Syt9 | |
| Dennd1a | |

**Figure 3.4 List of candidate genes identified from the PB-CAG-CKS-CAG-SD driven genetic screen** List of factors obtained from PB-CAG-CKS-CAG-SD-assisted genetic screen. 44 genes were isolated after elimination of poor, short reads and intergenic regions. Genes listed are annotated in Ensembl and harbour transposon integrations in the same orientation. Potential targets were grouped according to their functions and displayed in the table.

Classification of 44 genes according to their functional categories reveals a diverse ontological representation (**Figure 3.5**). The largest group consists of genes which have been poorly characterised. This strengthens the notion of an unbiased screen, where genes of unknown functions were unravelled. The sector of following magnitude comprises of transcription factors involved in developmental processes. Exemplified by the potency of the reprogramming cocktail, this segment of candidate factors was anticipated and reaffirmed the reliability of the genetic screen. Genes involved in cell-adhesion/communication were also significantly represented in the list of potential factors. Intriguingly, it was recently reported that E-cadherin expression was indicative of fully reprogrammed iPSCs and its ablation led to reduced reprogramming efficiencies (Redmer et al., 2011). Moreover, E-cadherin was a competent substitute for exogenous Oct4 during reprogramming (Redmer et al., 2011). This offers a separate dimension to highlight the dependability of the genetic screen. Although E-cadherin was not identified from the genetic screen attributable to an unsaturated genome, its co-operating partners CTNND1 and CTNND2 were identified, reinforcing the notion that changes in cell adhesion are integral during reprogramming.

Among the 44 genes which were identified in the screen, two were of particular interest as integration events were witnessed in more than one independent DNA pools, decreasing the likelihood that both genes were observed by chance. These two genes have been annotated as Retinoic Acid Receptor beta (RARβ) and Basonuclin-2 (BNC2). The integration sites identified within these genes are indicated in **Figure 3.6**.

Between the two candidate genes, BNC2 was portrayed to harbour four integration sites, whereas RARβ consisted of two. As a form of validation, BNC2 was employed as a reference and primers flanking each of the four insertion sites were designed. PCR was conducted using these primers to verify the presence of transposon cassettes within the genomic content of the reprogrammed cells (**Figure 3.7**). Three of four predicted sites were validated through PCR. Failure to amplify the remaining site could be attributed to

technical limitations such as poor primer design. Overall, these results reaffirm the accurate and dependable nature of the sequencing results.

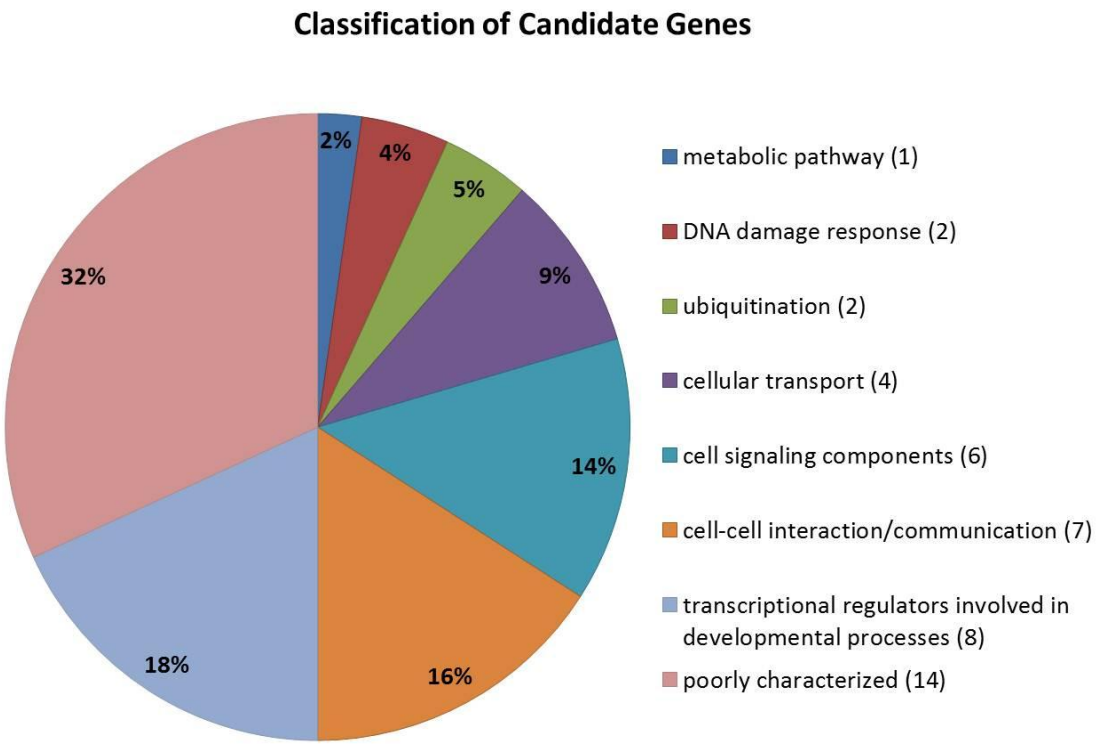



**Figure 3.5 Overview of functional classification of candidate genes identified from the PB-CAG-CKS-CAG-SD driven genetic screen** Pie chart describing the diverse functions of candidate genes. The colours depict distinct categories, as labelled in the legend. The actual number of genes associated to each category is shown in brackets. Percentage representation of each group is displayed on the pie chart fractions.

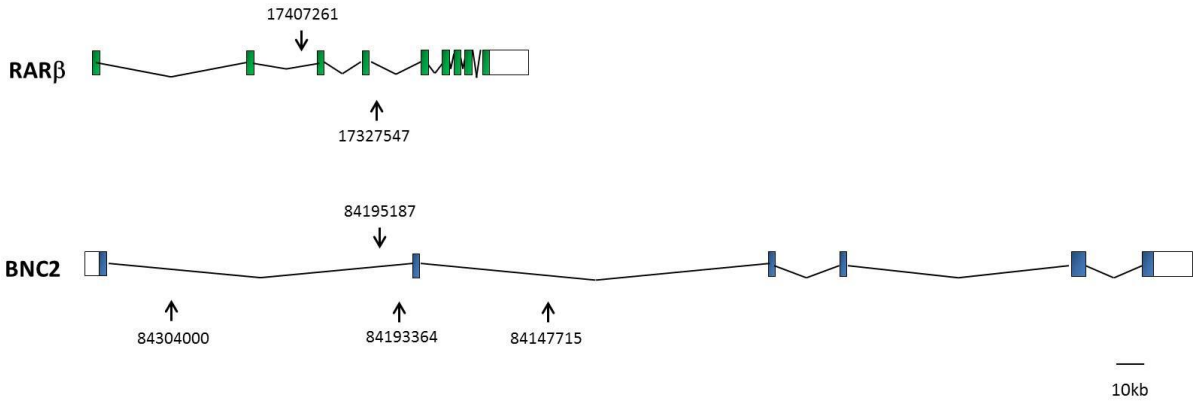



**Figure 3.6 Schematic diagram showing transposon integration sites within two candidate genes** Integration sites obtained from the PB-CAG-CKS-CAG-SD-assisted genetic screen. Coloured boxes indicate the exons corresponding to the two genomic regions. Arrows depict the location of the integration sites, and are conjugated to the exact genomic position of the insertions. The scale bar represents a genomic size of 10kb.

| Plate Number | Primers |
|---|---|
| 1 | 1-1, 1-2 |
| 2 | 4-1, 4-2 |
| 6 | 3-1, 3-2 |
| 8 | 2-1, 2-2 |

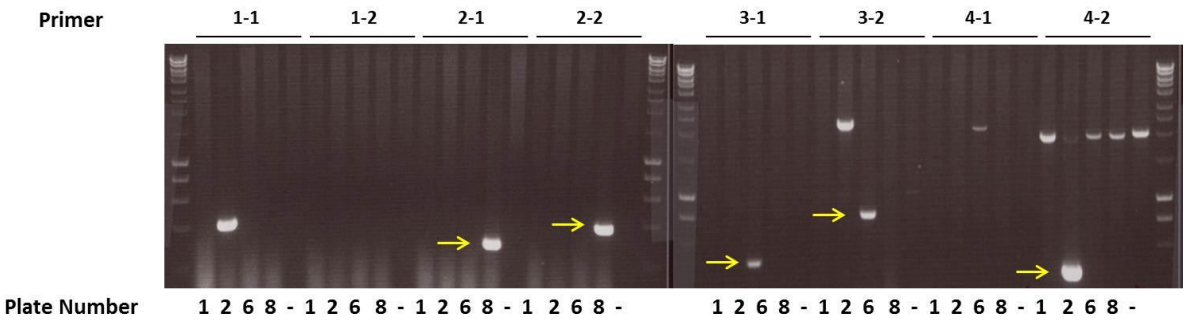



**Figure 3.7 Validation of BNC2 insertion sites** A pair of primers were designed for each of the 4 insertion sites. Genomic DNA was obtained from each of ten 10cm plates with iPSC colonies. DNA from plates 1, 2, 6, 8 and 10 were used. Insertions in the genomic locus corresponding to BNC2 were observed in plates 1, 2, 6 and 8. As plate 10 did not possess insertions in BNC2, it was employed as the negative control and is designated as a dash. Primers that recognise the four integration sites are listed in the table. PCR was conducted using all 8 primer pairs and the image represents the visualization of the PCR products on an agarose gel, as indicated by the yellow arrows.

### 3.3.2.2 Identification of candidate genes in the PB-CAG-CKS-CAG screen

The use of PB-CAG-CKS-CAG as a mutagen provides a rigorous platform to deregulate gene expression levels of factors that are competent at replacing the ectopic requirement of Oct4 during reprogramming. Due to the stringent nature of this screen in the absence of a splice donor site, a total of seven iPSC colonies were obtained and individually analysed. This is a stark contrast to the first screen where large numbers of colonies were seen and analysis on pools of reprogrammed cells was performed. Instead, it was possible to examine integration sites within individual iPSC colonies obtained from the screen.

Due to the absence of a splice donor site in the mutagenic cassette, it is essential that the promoter and enhancer attributes of CAG drive the expression of the inserted gene. The resultant gene product would not be truncated or modified, unlike the previously mentioned screen. With this in mind, all integration sites were analysed and intergenic regions were taken into consideration. In order to score intergenic regions, the nearest neighbouring gene in the direction akin to the CAG promoter was taken into account.

**Figure 3.8** provides an overview of the integration sites within the seven clones. Examination of the orientation of insertion sites revealed no bias, such that both orientations were observed at equal frequencies. An average of 20 integration sites was detected in each iPSC colony. This is a huge discrepancy to the calculations performed in **Chapter 3.2.1** which predicts approximately one integration site per cell that has been successfully transfected with both donor and helper plasmids. However, it is plausible that stringency of the experiment resulted in the isolation of rare cells that exhibit increased frequencies of transposon insertions. Moreover, the ability to replace the exogenous requirement of Oct4 is challenging, hence cells which are highly mutated are inherently more likely to possess a critical aberration to overcome the selection process. Lastly, it is possible that several genes work in synergy to surmount the absence of exogenous Oct4, explaining the high number of integrations per established iPSC colony.

In total, 138 integrations were uncovered. Assembling these insertion sites to the mouse genome reveals a total of 133 genes. Ontological classification of these genes is described in **Figure 3.9**. Analagous to the results obtained from the first strategy, a large proportion of 47 genes (35%) were poorly characterised, reiterating the potential of the genetic screen to uncover novel candidates and pathways associated to reprogramming. Genes involved in cell signalling, transcriptional regulation and mRNA processing were also over-represented in the analysis. All seven iPSC colonies displayed deregulation in one or more of these pathways (**Figure 3.10**).

| Clone 1 | Clone 2 | Clone 3 | Clone 4 | Clone 5 | Clone 6 | Clone 7 |
|---|---|---|---|---|---|---|
| Lrig2 | Vcpip1 | U6 | Mllt10 | U6 | AC157923.4 | Xrcc5 |
| Dnntip2 | Dbi | Dynlrb | 46341642Rik | Prg4 | Nmnat2 | Orc4l |
| Crocc | Prpf18 | Tbx3 | Faf1 | Rpl18-rs8 | RP23-350C1.5 | Slc7a11 |
| KIAA1530 | Trim33 | Rab28 | Akap13 | mmu-mir-29c | U4 | AC127339.2 |
| BC064033 | D3Ertd300e | AC164700.3 | Arglu1 | Rnf13 | AL669981.11 | Sts1 |
| Plxnc1 | Cmpk | Cald1 | Pknox2 | Olfr151 | Sdk1 | Rora |
| 1700024P04Rik | U6 | Adamts18 | Nkiras2 | Dbr1 | Pftk1 | Stag1 |
| Dedd | Eya4 | Mcph1 | Aoah | Tspan10 | Cald1 | Gam14 |
| Prkcq | Zdhhc17 | Dctd | 1600140C23Rik | SNORA17 | N4bp1 | Cox16 |
| Act16a | Olfr804 | Casp1 | AC116999.2 | Pgr4 | Nav3 | Slc38a9 |
| RP-23-140F18.5 | Akap7 | 22104415F13Rik | Gpr177 | Ifih1 | AC158122.6 | Zfp759 |
| Rora | Ercc6 | UPF0528 | Fzd2 | Pdgfc | Fev | Douson1 |
| Insig2 | Nalcn | Rcqd1 | Vrk3 | Mrps21 | Wfdc3 | Prune2 |
| AC113507.12 | Diap1 | Chchd7 | RP23-146A17 | Cdkal1 | RP23-103G12.1 | Unc5c |
| AC138594.10 | Xiap | AC116591.4 | SNORA79 | Col11a1 | Ctnnd1 | Gclc |
| Tshz3 | mmu-mir-222 | Slc24a | Ckn5 | Nop16 | Col4a1 | Cd276 |
| Cand1 | Sephs2 | Rnf121 | Fga | Tdrd7 | Wdr22 | Capn13 |
| RP23-436D9.2 | AC105256.7-2 | Tspo | Rpl29 | Stk40 | AC164093.2-1 | Rapgef5 |
| Nin | AC127568.4 | Nck2 | 4930578I06Rik | Rpn1 | 9530020O07Rik | Rnf141 |
| D130079A08Rik | AC153658.4-1 | Rbm43 | Phf21b | Msl2 | Zfp608 | RP23-198C2.5 |
| U2 | Pik3cg | Pdha2 | Mc4r | Marcks | Aqp4 | Smek2 |
| Rbm3 | EG432723 | AC112688.10 | Prl8a2 | Atxn1 | B230399E16Rik | |
| RP23-149H3.2 | AC164883.5-1 | Far2 | Cpt1a | AC101970.11 | | |
| Bbs9 | Etv6 | Antxr1 | Fhdc1 | Col11a1 | | |
| Fryl | Fgf14 | AC116499.9 | Utp6 | Manea | | |
| Gcap14 | Lass6 | Setd2 | | SNORA17 | | |
| Lancl3 | Sfmbt2 | B3galnt2 | | Med13l | | |
| Tmcc3 | Syn3 | Rnase9 | | Cntn6 | | |
| | Tsc2 | CT025601.7 | | Cited2 | | |
| | | 2010321M09Rik | | AC135667.5 | | |
| | | Ak5 | | Ccdc125 | | |
| | | Etl4 | | Ephb1 | | |
| | | Plag1 | | | | |
| | | Wdr22 | | | | |

**Figure 3.8 List of candidate genes identified from the PB-CAG-CKS-CAG driven genetic screen** List of factors obtained from PB-CAG-CKS-CAG-assisted genetic screen. Seven individual colonies were picked and sequenced. Candidate genes from each of the seven clones are displayed in the table. Black and blue fonts highlight genes in the same and opposite orientation as the insertion cassettes respectively.

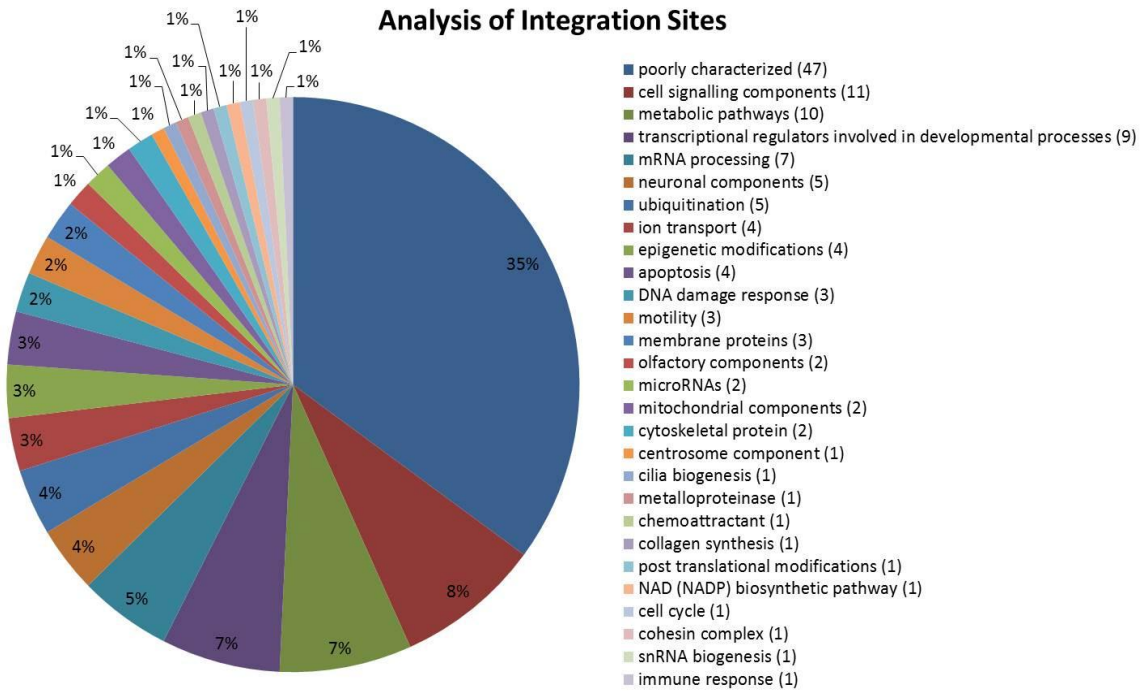



**Figure 3.9 Overview of functional classification of candidate genes identified from the PB-CAG-CKS-CAG driven genetic screen** Pie chart describing the diverse functions of candidate genes. The colours depict distinct categories, as labelled in the legend. The actual number of genes associated to each category is shown in brackets. Percentage representation of each group is displayed on the pie chart fractions.

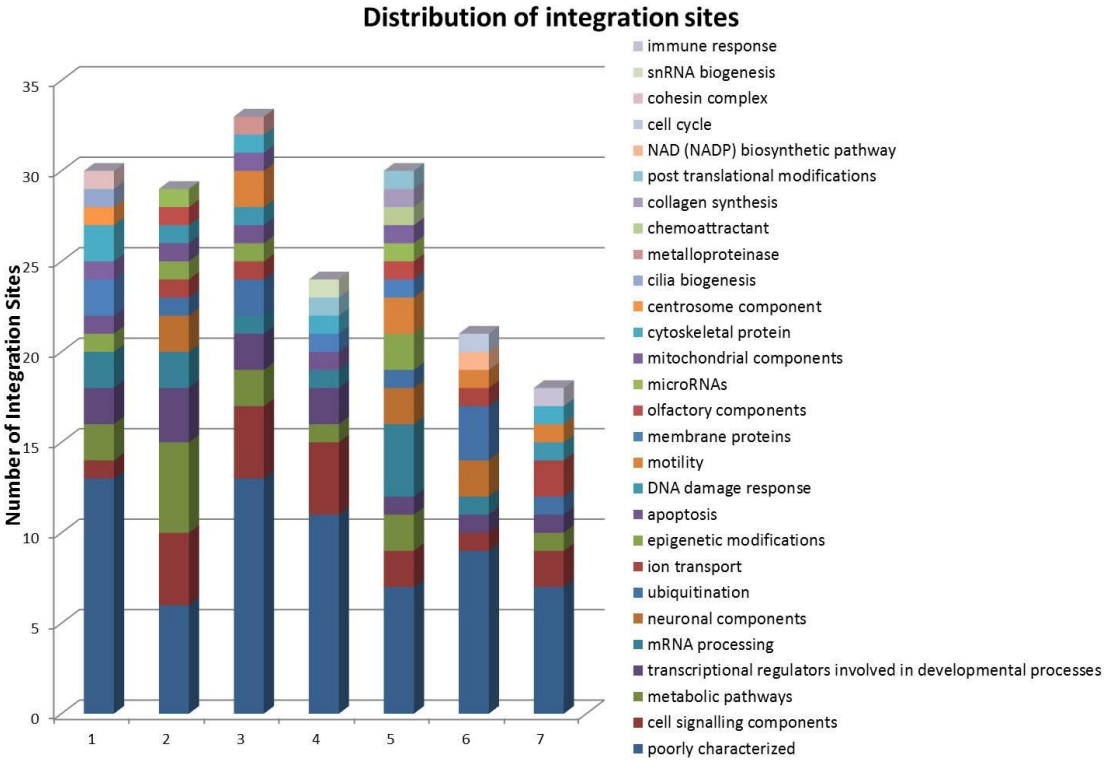



**Figure 3.10 Graphical representation of candidate genes identified from the PB-CAG-CKS-CAG driven genetic screen** Bar chart displaying the range of mutations in functional groups in each of the seven iPSC colonies. Array of colours depict different categories as listed on the right.

Interestingly, additional pathways such as ubiquitination and apoptosis also exhibited aberrance in these reprogrammed cells. Ablation of p53 and the consequent increase in cell proliferation has been reported to increase reprogramming efficiency (Zhao et al., 2008; Hanna et al., 2009; Utikal et al., 2009; Marión et al., 2009; Li et al., 2009; Kawamura et al., 2009; Hong et al., 2009), establishing the possibility that overcoming protective mechanisms such as apoptosis and protein degradation could alleviate the reprogramming process. However, it is necessary to keep in mind that each iPSC colony received an average of 20 integration sites, and a mere subset of them contributed to the initiation of reprogramming events in the absence of Oct4. As such, the analysis must be performed with caution and future validation steps are critical.

A similar criterion to the abovementioned (**Chapter 3.3.2.1**) was employed to taper the list of candidate factors. Among 133 candidate genes, three genes were identified to harbour more than one transposon integrations in independent iPSC colonies, hence reducing the probability that these integrations occurred by chance. These genes were Retinoic acid receptor-related Orphan Receptor alpha (RORα), Caldesmon (Cald1) and DDB1–CUL4A-associated factor-5 (DCAF5). **Figure 3.11** provides a schematic overview of the location of insertion sites.

Having identified a handful of candidate genes, it is essential to validate their abilities to reprogram MEFs in the absence of exogenous Oct4. These experiments will be described in the following chapter.

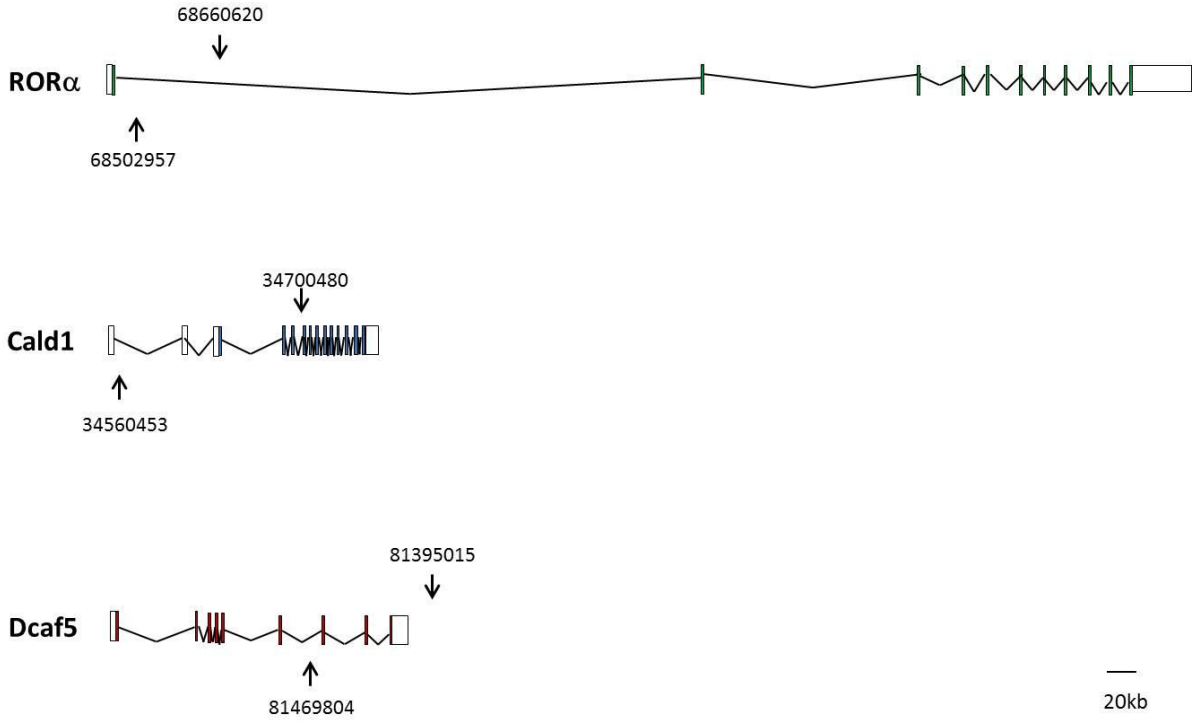



**Figure 3.11 Schematic diagram showing transposon integration sites within three candidate genes** Integration sites obtained from the PB-CAG-CKS-CAG-assisted genetic screen. Coloured boxes indicate the exons corresponding to the three genomic regions. Arrows depict the location of the integration sites, and are conjugated to the exact genomic position of the insertions. The scale bar represents a genomic size of 20kb.

## 3.4 Discussion

This chapter attempts to identify candidate genes which can co-operate with ectopic expression of c-Myc, Klf4 and Sox2, to drive the reversion of somatic signatures in MEFs. A transposon assisted mutagenesis screen was conducted to address this objective. Two approaches differing in the expression of PBase and gene trapping strategies were employed and iPSCs were successfully generated. This highlights the possibility to overcome the exogenous requirement for Oct4 in order to trigger dedifferentiation.

### 3.4.1 Analysis of the ontological classification of candidate genes

Sequencing of the integration sites generated from both screening strategies yielded an immense amount of information. Categorisation of genes that exhibited insertion sites provided a broad insight on the cellular processes that were ignited during the acquisition of pluripotency. By comparing both experimental strategies, it is evident that there are several segments of genes which were recurrent. These include transcriptional regulators that affect development, components of metabolic and cell signalling pathways, and mRNA processing. To comprehend the repercussions of identifying these gene segments, each gene segment is elaborated below.

Transcriptional regulators have been demonstrated to play critical roles in the generation of iPSCs from its conception. Four transcription factors specifically expressed in ESCs were able to convert somatic cell types to ES cell-like states (Takahashi and Yamanaka, 2006) and an expanded list of pluripotency-associated transcription factors have been demonstrated to augment this dedifferentiation process (Han et al., 2010b; Heng et al., 2010; Feng et al., 2009; Nagamatsu et al., 2011; Maekawa et al., 2011). In addition, epigenetic modifiers have been described to affect reprogramming efficiencies (Mikkelsen et al., 2008; Huangfu et al., 2008a).

Metabolism has been delineated to participate in the transformation of cells during tumorigenesis by fuelling proliferation (Vander Heiden et al., 2009; reviewed by Hamanaka and Chandel, 2012). In addition, pluripotent cells have been described to possess disparate metabolic signatures to somatic cell types and the acquisition of pluripotency incurs a shift in these metabolic elements (Varum et al., 2011; Prigione et al., 2010). As such, overriding metabolic pathways may surmount obstacles that hinder the initiation of dedifferentiation.

Cell signalling pathways have been well-characterised in pluripotent stem cells (Yoshida et al., 1994; Niwa et al., 1998; Ying et al., 2003). The use of small chemicals such as MEK and GSK3β inhibitors are routinely performed to sustain ground state pluripotency (Ying et al., 2008). Recently, it has been established that promotion of Wnt signalling increases reprogramming efficiencies (Marson et al., 2008a; Lluis et al., 2011). Extending this, two receptors of Wnt ligands, Gpr177 and Fzd2, were discovered to harbour integration sites from the genetic screen and may provide closer insight to the role of Wnt signalling during the generation of iPSCs.

mRNA processing is an evolving field and evidence of its role in the execution of nuclear reprogramming is budding (Ji and Tian, 2009). Comparisons between proteomic and transcriptional expression levels in ESCs reveal discrepancies, where protein and mRNA levels of pluripotency-associated factors change non-synonymously during differentiation, implicating an interruption in mRNA processing (O'Brien et al., 2010). Moreover, transcriptomic comparisons between oocytes and 2-cell embryos revealed differences in expression levels of components that participate in mRNA processing (Macfarlan et al., 2012). Overall, these results surmise the significance of mRNA processing in development and nuclear reprogramming of somatic cells, leaving much to be uncovered.

Mutagenic diversity within a single iPS clone (**Figure 3.10**) proposes that deregulation of a spectrum of cellular processes is necessary to attain pluripotency. Close inspection of the

breadth of genetic aberrations in each of the seven iPSC colonies may shed light on potential genetic co-operations. On a similar note, previously described transposon mediated genetic screens revealed not only mutually exclusive oncogenic roles of Ikaros and Notch, but also putative co-operating partners of Notch (Dupuy et al., 2005).

Although the classification of the candidate genes sheds light on critical processes in the initiation of reprogramming, it is key to note that a subset of genes may possess "passenger mutations" which are not causal to the appearance of iPSCs. As the conversion of somatic cells to pluripotent derivatives is accompanied by alterations in chromatin conformation, transposons may gain access to remote regions of the genome. This leads to increased incidences of insertion events at genomic loci which are exposed during reprogramming, irrespective to their contribution to the reprogramming process. Hence, it is necessary to look at the list with scepticism and conduct further validation experiments.

### 3.4.2 Technical considerations of the screen

To assess the coverage of the transposition events, uncurated insertion sites across each chromosome were tallied (**Figure 3.12**). As piggyBac has been described to not exhibit a selection bias (Wang et al., 2008), the number of integration sites per chromosome should draw a parallel to its size. However, an unequal distribution of insertion sites was observed. By comparing both genetic screens, transposition events in chromosomes 12, 13 and 18 were over-represented, whereas insertion sites in chromosomes 11, 16 and 19 were infrequent. Interestingly, among the candidate genes identified, only DCAF5 is situated in genomic regions which were over-represented, eliminating the possibility that candidate genes were identified due to their presence at preferential integration sites.

Although the numbers are small and lacks replicates, it is important to keep these figures in mind when designing future screens. Increasing the scope of the experiment may result in a more balanced distribution of integration sites. On the other hand, it is also plausible that genes critical in the reprogramming process are dominantly located on mouse

chromosomes 12, 13 and 18. From a different perspective, PB-CAG-CKS-CAG-SD and PB-CAG-CKS-CAG- assisted mutagenesis gave rise to an average of 1.15 and 0.96 integration sites per 10Mb of the mouse genome respectively. As these values were obtained from a small population of transfected cells that survived the selection process, it gives us confidence that the genome was saturated (as denoted in **Chapter 3.2**) and more than one integration site was achieved per 20kb of the genome.

To ascertain the saturation of the genetic screen, integrations within Oct4 should have been observed. However, Oct4 was not in the list of candidate genes. This could be attributed to the relatively small genomic size of the Oct4 locus. Spanning a mere 4739bp, it is unlikely that an insertion would occur in its vicinity. In addition, Oct4 is specifically expressed in pluripotent stem cells and not in MEFs, possibly exhibiting steric hindrance at its genomic locus. Furthermore, the genomic locus corresponding to Oct4 is laden with regulatory sites, hence integrations that affect these regulatory elements may abrogate Oct4 expression. On another note, application of a splice donor site in the first experimental strategy would have resulted in a truncated transcript if an insertion had occurred within the Oct4 locus. This would have led to a loss of function, and reprogramming would have been stalled. Ultimately, this Chapter describes a minimal coverage screen. To this end, augmenting the coverage through increasing cell numbers or increased transposon amounts may allow the detection of integration sites within Oct4. Reprogramming of MEFs to iPSCs is a demanding and stochastic process which displays low efficiencies between 0.01-0.2% (reviewed by Hochedlinger and Plath, 2009; Hanna et al., 2009). As calculations performed in **Chapter 3.2.1** did not consider low reprogramming efficiencies, it is possible that although the transposon cassette landed in the Oct4 locus, it was unable to elicit reprogramming events within 25 days. As such, increasing starting amounts of MEFs may overcome this bottleneck.

A separate limitation of the genetic screen is the detection of insertion sites. First, the detection of insertion sites relies on amplification via polymerase chain reaction. According to amplification conditions, only medium sized products of approximately

500bp were generated, whereas the production of short and long products were challenging. Another drawback is evident in the bias faced during the ligation of amplified products into pGEMT-Easy. A similar phenomenon was experienced where medium-sized products were ligated into pGEMT-Easy with more ease than short and long fragments. In addition, of the fragments that were successfully ligated and transformed, only a subset of bacterial colonies was ultimately sequenced. This selection process could potentially lead to the omission of critical genes. Lastly, in addition to increasing cell numbers to further saturate the genome, a different subset of candidate genes may be illuminated through the application of a separate restriction enzyme during the execution of splinkerette PCR.

| Chromosome number | Number of integration sites | Number of integration sites per 10Mb | |
|---|---|---|---|
| 1 | 31 | 1.57 | Strategy 1: PB-CAG-CKS-CAG-SD |
| 2 | 18 | 0.99 | |
| 3 | 23 | 1.44 | |
| 4 | 16 | 1.03 | |
| 5 | 15 | 0.98 | |
| 6 | 20 | 1.33 | |
| 7 | 18 | 1.18 | |
| 8 | 18 | 1.36 | |
| 9 | 11 | 0.89 | |
| 10 | 11 | 0.85 | |
| 11 | 7 | 0.57 | |
| 12 | 23 | 1.90 | |
| 13 | 23 | 1.92 | |
| 14 | 10 | 0.80 | |
| 15 | 11 | 1.07 | |
| 16 | 4 | 0.41 | |
| 17 | 12 | 1.26 | |
| 18 | 15 | 1.65 | |
| 19 | 4 | 0.66 | |

| Chromosome number | Number of integration sites | Number of integration sites per 10Mb | |
|---|---|---|---|
| 1 | 23 | 1.17 | Strategy 2: PB-CAG-CKS-CAG |
| 2 | 17 | 0.93 | |
| 3 | 17 | 1.06 | |
| 4 | 13 | 0.83 | |
| 5 | 11 | 0.72 | |
| 6 | 14 | 0.93 | |
| 7 | 16 | 1.05 | |
| 8 | 15 | 1.14 | |
| 9 | 8 | 0.65 | |
| 10 | 10 | 0.77 | |
| 11 | 7 | 0.57 | |
| 12 | 20 | 1.65 | |
| 13 | 21 | 1.75 | |
| 14 | 8 | 0.64 | |
| 15 | 9 | 0.87 | |
| 16 | 4 | 0.41 | |
| 17 | 10 | 1.05 | |
| 18 | 13 | 1.43 | |
| 19 | 4 | 0.66 | |

**Figure 3.12 Table displaying spread of transposition events** Uncurated insertion sites were tallied according to their chromosomal positions. Using the length of each chromosome, as stated in the Mouse Genome Informatics, Jackson Laboratories, the frequency of insertion sites per 10Mb of each chromosome was derived. As two genetic screens were calculated, the values corresponding to the first strategy using PB-CAG-CKS-CAG-SD are shown on the left. Similarly, values associated to PB-CAG-CKS-CAG-assisted mutagenesis are revealed on the right.

## 3.5 Conclusion

This chapter illustrates the possibility of conducting a transposon-assisted mutagenesis screen to identify genes involved in reprogramming. Moreover, these results demonstrate the ability to replace the exogenous requirement of Oct4 during reprogramming. The candidate genes obtained from the screen are mostly made up of poorly characterised genes, allowing room for validation. In addition, a large proportion of genes identified are associated to cellular pathways critical in the maintenance of pluripotency and growth. Due to the extended list of candidate genes, it is challenging to isolate factors for further characterization. To this end, the list was narrowed to five genes which were identified from independent pools of genetic material.

# CHAPTER 4 : VALIDATION OF CANDIDATE GENES

## 4.1 Introduction

In this chapter, I aim to validate the candidate genes identified in **Chapter 3**, by assessing their ability to replace the exogenous requirement of Oct4 during reprogramming. These candidate genes were discovered through a piggyBac-assisted mutagenesis screen where integration sites within these genes were observed in more than one occasion, diminishing the possibility that these occurrences were dependent on chance. Validation was conducted in a similar approach to the screen where piggyBac-mediated transposition was employed to deliver the cDNAs. To this end, transposon cassettes comprising of the candidate genes were generated. A total of five genes identified from the screens, RARβ, RORα, BNC2, Cald1 and DCAF5, were subjected to validation assays.

### 4.1.1 The piggyBac transposon as a reprogramming tool

In 2006, a milestone in nuclear reprogramming was established. Retroviral delivery of four transcription factors, Oct4, c-Myc, Klf4 and Sox2, into mouse fibroblasts was sufficient to revert transfected cells into a primitive ES cell-like state (Takahashi and Yamanaka, 2006). Reprogrammed cells resembled ESCs in transcriptional and functional properties, and were termed as iPSCs. The inception of iPSC technology evoked a multitude of studies attempting to replicate and enhance the original findings. One such advancement was the refinement of delivery methods used to introduce the four reprogramming factors into desired cell types. As the conventional cocktail of reprogramming genes is made up of critical regulators of cell fate and development, they also serve as potent oncogenes. The generation of mice from iPSCs originally obtained through retroviral infections gave rise neck and other tumours at a frequency of 20%, attributable to the reactivation of c-Myc (Okita et al., 2007).

To circumvent this problem, a spectrum of alternative gene delivery tools was employed. Non-integrating approaches such as adenoviruses, Sendai virus, episomal DNA, repeated transfections of mRNA or plasmid DNA and purified proteins have been described to

generate iPSCs (Fusaki et al., 2009; Stadtfeld et al., 2008; Yu et al., 2009; Warren et al., 2010; Okita et al., 2008; Kim et al., 2009a; Zhou et al., 2009). Although a portion of these methods do not implicate the use of genetic material, certain protocols are labour-intensive and challenging to be faithfully reproduced. Ultimately, reprogramming efficiencies instigated by these methods are low and create a bottleneck in the creation of iPSCs, where non-integrating viruses generate iPSCs at efficiencies between 0.0001% and 0.0018%, approximately a third of reprogramming efficiencies achieved by retroviruses (reviewed by González et al., 2011).

Apart from non-integrating protocols, transposon-mediated mutagenesis exemplifies an efficient tool to deliver the reprogramming cocktail into cells (Woltjen et al., 2009; Kaji et al., 2009; Yusa et al., 2009). Despite its integrative nature, expression of an appropriate transposase can allow the excision of the transposon. In particular, the piggyBac transposon exhibit the absence of a genetic footprint upon mobilisation (Fraser et al., 1996), producing transgene-free iPSCs (Woltjen et al., 2009; Kaji et al., 2009; Yusa et al., 2009).

Analogous to **Chapter 3**, the capabilities of piggyBac-assisted transposition have also been employed to isolate genes capable of reverting primed EpiSCs to naïve ESCs (Guo et al., 2010). Nr5a2 was isolated from the screen and its ability to reprogram EpiSCs was validated when introduced into EpiSCs in a piggyBac-assisted manner. Parallel to the published finding, this Chapter describes the use of transposition to validate the candidate genes obtained from **Chapter 3**.

### 4.1.2 Candidate Genes

#### 4.1.2.1 Retinoic Acid Receptor beta (RARβ)

Retinoic Acid signalling plays an integral role during development (Thaller and Eichele, 1987; Dickman et al., 1997; White et al., 1998). The cascade of molecular events elicited

by Retinoic Acid signalling relies on the presence of three receptors which recognise and bind to Retinoic Acid, namely Retinoic Acid Receptor alpha (RARα), Retinoic Acid Receptor beta (RARβ) and Retinoic Acid Receptor gamma (RARγ) (Giguere et al., 1987; Dejean et al., 1986; Petkovich et al., 1987; Benbrook et al., 1988; Brand et al., 1988; Krust et al., 1989; Zelent et al., 1989). These receptors were classified under the nuclear receptor superfamily and show high resemblance in their DNA binding domains (DBDs) and ligand binding domains (LBDs) (reviewed by Chambon, 1996).

RARβ represents one family member and the human isoform was originally discovered in 1988 by two independent laboratories using distinct approaches. One study detected homology to RARα (Brand et al., 1988), whereas the other determined integration sites in Hepatitis B virus (HBV) infected hepatocellular carcinoma (Benbrook et al., 1988). To explore its function, the ligand binding domain of RARβ was tethered to the DNA binding domain of human estrogen receptor. Exposure of RA at physiological concentrations led to the activation of estrogen responsive genes, reflecting ligand dependent transcriptional activity. Sharing close homology, mouse RARβ was discovered soon after (Zelent et al., 1989).

Mouse RARβ exists in three isoforms, RARβ1, RARβ2 and RARβ3, which are identical except for the hypervariable domain at the $NH_2$-terminal end (Zelent et al., 1991). Exhibiting distinct 5' untranslated regions (UTRs), it was proposed that the three isoforms are regulated by multiple promoters and undergo alternative splicing. RARβ1 and RARβ3, observed at lower frequencies than RARβ2, are primarily expressed in the brain, whereas RARβ2 is highly expressed in EC cells and is postulated to function in developmental pathways. Interestingly, loss of RARβ2 has been observed in an array of cancers, highlighting its role in development (Picard et al., 1999; Xu et al., 1999; Xu et al., 1997; Widschwendter et al., 1997).

Recent studies have implicated another RAR family member, RARγ, in the promotion of reprogramming events (Wang et al., 2010). This phenomenon was observed in the accompaniment of LRH1 and was attributed to activation of the proximal promoter of endogenous Oct4. Having delineated that RARγ can enhance the generation of iPSCs through the activation of Oct4, it suggests that RARβ may play a similar role in the derivation of iPSCs.

#### 4.1.2.3 Retinoic Acid Receptor-related Orphan Receptor alpha (RORα)

RORα belongs to a family of orphan receptors (RORs) which shares homology to RARs. RORs have been described to play pleiotropic functions in development, inflammation and metabolism (reviewed by Jetten, 2009). Analagous to other steroid hormone receptors, RORs function as ligand activated transcription factors. Structural and functional studies have demonstrated that cholesterol has the capacity to bind to RORs and consequently enhances transcription (Kallen et al., 2004), insinuating its capacity as a natural ligand.

The discovery of RORs was spurred by the cloning of several steroid hormone receptors in the 1980s. This brought about the identification of three family members, namely RORα (Becker-André et al., 1993; Giguère et al., 1993; Giguère et al., 1994), RORβ (André et al., 1998; Calberg et al., 1994; Schaeren-Wiemers et al., 1997), and RORγ (He et al., 1998; Hirose et al., 1994; Medvedev et al., 1996). RORs comprise of an N-terminal domain, DNA-binding domain, a hinge domain, and a C-terminal ligand-binding domain, where the DNA-binding domain displays close identity and the ligand binding domain exhibits moderate resemblance among family members (reviewed by Jetten et al., 2009).

RORα is expressed in a variety of tissues, including testis, kidney, adipose and liver, with significant levels in the brain (Becker-André et al., 1993; Carlberg et al., 1994; Nakagawa et al., 1997; Nakagawa and O'Leary, 2003). Disruption of RORα through the deletion of its fifth exon or complete ablation results in thin long bones (Meyer et al., 2000), ataxia

and severe cerebellar atrophy (Hamilton et al., 1996; Doulazmi et al., 2001; Doulazmi et al., 1999; Dussault et al., 1998). RORα has also been implicated in the study of cancer, suggesting its importance in developmental processes (Lee et al., 2010; Xiong et al., 2012). As links between nuclear reprogramming and RORα remain elusive, it will be interesting to determine if RORα can act as a substitute for ectopic Oct4 during the acquisition of pluripotency.

#### 4.1.2.4 Basonuclin-2 (BNC2)

Basonuclins were first discovered in an incidental manner through a search for regulators that possess helix-loop-helix (HLH) domains in human keratinocytes (Tseng and Green, 1992). Instead, they identified an atypical zinc finger motif that consists of six $C_2H_2$ zinc fingers arranged in three interspersed pairs. Designated as basonuclin-1 (BNC1), it also possesses a serine stripe and nuclear localisation signal between the second and third zinc fingers, delineating its ability to shuttle between the nucleus and cytoplasm in keratinocytes (Iuchi and Green, 1997; Iuchi et al., 2000).

BNC2 was subsequently discovered by searching chicken and mouse expressed sequence tag (EST) databases for homology to BNC1 (Romano et al., 2004; Vanhoutteghem and Djian, 2004). Although BNC2 comprises of similar elements to BNC1, comparison of both human isoforms reveal 58% homology at sequence level and 44% homology at amino acid level. Across species, mouse BNC1 is 85.6% similar to its human counterpart, whereas mouse and human BNC2 exhibit 97.0% homology.

In resemblance to BNC1, BNC2 is expressed in the epidermis, the testis and the ovary. However, arising pieces of evidence have suggested a wider expression pattern of BNC2 in the kidney, intestine, uterus, palate, connective tissue surrounding the brain, cartilage and bones (Romano et al., 2004; Vanhoutteghem and Djian, 2004; Vanhoutteghem et al., 2011).

Disruption of BNC2 poses as a tool to elucidate its function and has been exemplified across species. Fusion between the first exon of BNC2 and LacZ in mice leads to a cleft palate, abnormalities in craniofacial bones and tongue, resulting in death within 24 hours of birth (Vanhoutteghem et al., 2009). Nitrosonurea incurred BNC2 mutations have been demonstrated in a zebrafish mutant known as bonparte. These zebrafish exhibit infertility and lack adult stripe patterns, attributable to the presence of a premature stop codon and the creation of a truncation product which lacks all six zinc fingers (Lang et al., 2009). However, these observations may be a result of genetic redundancy by its family member, BNC1, thus shadowing the true function of BNC2. As such, further elucidation of the mechanisms related to BNC2 is necessary.

In humans, computational analysis of the genomic locus corresponding to BNC2 predicts the presence of six promoters, seven major and sixteen minor exons, and the potential to generate 90,000 mRNA spliced isoforms (Vanhoutteghem et al., 2007). It extends over 461 kb of the genome and has been implicated in ovarian and esophageal cancer (Song et al., 2009; Akagi et al., 2009; Sundqvist et al., 2011). Although the function of BNC2 has not been clearly delineated, both members of the basonuclin family depict critical roles in developmental pathways. As such, it will be intriguing to explore the role of BNC2 in the generation of iPSCs.

#### 4.1.2.5 Caldesmon-1 (Cald1)

Cald1 encodes for Caldesmon, a multimodular protein which plays significant roles in actomyosin contraction and actin cytoskeleton remodelling (reviewed by Lin et al., 2009). It was first identified in the chicken gizzard through its interaction with Calmodulin (Sobue et al., 1981). Two isoforms have subsequently been discovered: high molecular mass Cald1 (h-Cald1) and low molecular mass Cald1 (l-Cald1). L-Cald1 lacks a single helical region that separates the N- and C- terminal domains (Owada et al., 1984; Sobue et al., 1985). Although both isoforms exhibit similar biochemical properties *in vitro*, they

have been observed in mutually exclusive compartments; h-Cald1 is primarily expressed in smooth muscle cells (Marston and Lehman, 1985; Fürst et al., 1986; Bretscher and Lynch, 1985), whereas l-Cald1 is found in non-smooth muscle cells (Sobue et al., 1985).

Disruption of h-Cald1 through the excision of exon 3 has been performed in mice (Guo and Wang, 2005). Although some homozygotes survived until adulthood, most F2 newborns developed umbilical hernia and died within 5-7 hours following birth. Cald1 has also been deregulated in zebrafish, via the introduction of morpholinos to knockdown gene expression levels, influencing its capacity to modulate cardiac morphogenesis, muscularisation and function (Zheng et al., 2009a; Zheng et al., 2009b). In combination with observations that portray its influence on cellular transformation and cancer (Novy et al., 1991; Owada et al., 1984; Ross et al., 2000; Tanaka et al., 1993), it is evident that Cald1 plays a pivotal role in vertebrate development. As such, implications that Cald1 may participate in the process of nuclear reprogramming will assist in the elucidation of its role during development.

#### 4.1.2.6 DDB1 and CUL4 associated factor 5 (DCAF5)

DCAF5 belongs to a family of DDB1 and CUL4 interactors. CUL4 is an example of a cullin which acts as a scaffold around the assembly unit associated with ubiquitin ligases (reviewed by Petroski and Deshaies, 2005). As the molecular mechanisms behind the action of CUL4 remain unclear, binding partners were identified using immunoprecipitation of DDB1 (Jin et al., 2006). Using proteomic approaches, DDB1 interating partners were delineated and 18 DCAFs were identified. Among these candidates, 14 comprised of WD40 repeats, facilitating their interaction with DDB1. Although initial studies implicate DCAF5 in process of ubiquitination, little research has been conducted to reinforce these findings. Identification of DCAF5 as a candidate substitute for ectopic expression of Oct4 during reprogramming may shed light on the influence of ubiquitination on reprogramming, and may delineate the molecular mechanisms associated to DCAF5. Moreover, Oct4 expression levels has been described

to be regulated through ubiquitination (Saxe et al., 2009), inferring that DCAF5 may play a role in reprogramming through direct modification of Oct4 protein.

### 4.1.3 Chapter Aim

This chapter aims to address the results achieved from the piggyBac assisted mutagenesis screen described in **Chapter 3**. The genetic screen was adopted to identify factors which display competence in negotiating the exogenous requirement of Oct4 during the acquisition of pluripotency. Five candidate genes, as described in detail, were identified and will be addressed in this chapter. Candidate genes will be introduced into mouse fibroblasts, in combination with c-Myc, Klf4 and Sox2, using a piggyBac mediated approach. The emergence of iPSCs will reflect the ability of the candidate gene to trigger reprogramming events, allowing the assessment of the dependability of the genetic screen. In the event that these genes competently negate the need for Oct4 over-expression, iPSCs derived using these transgenes will undergo an array of tests to determine their pluripotent potential. As depicted in **Figure 4.1**, this array of tests includes transcriptional, proteomic and epigenetic analysis, culminating in the test for germline contribution.

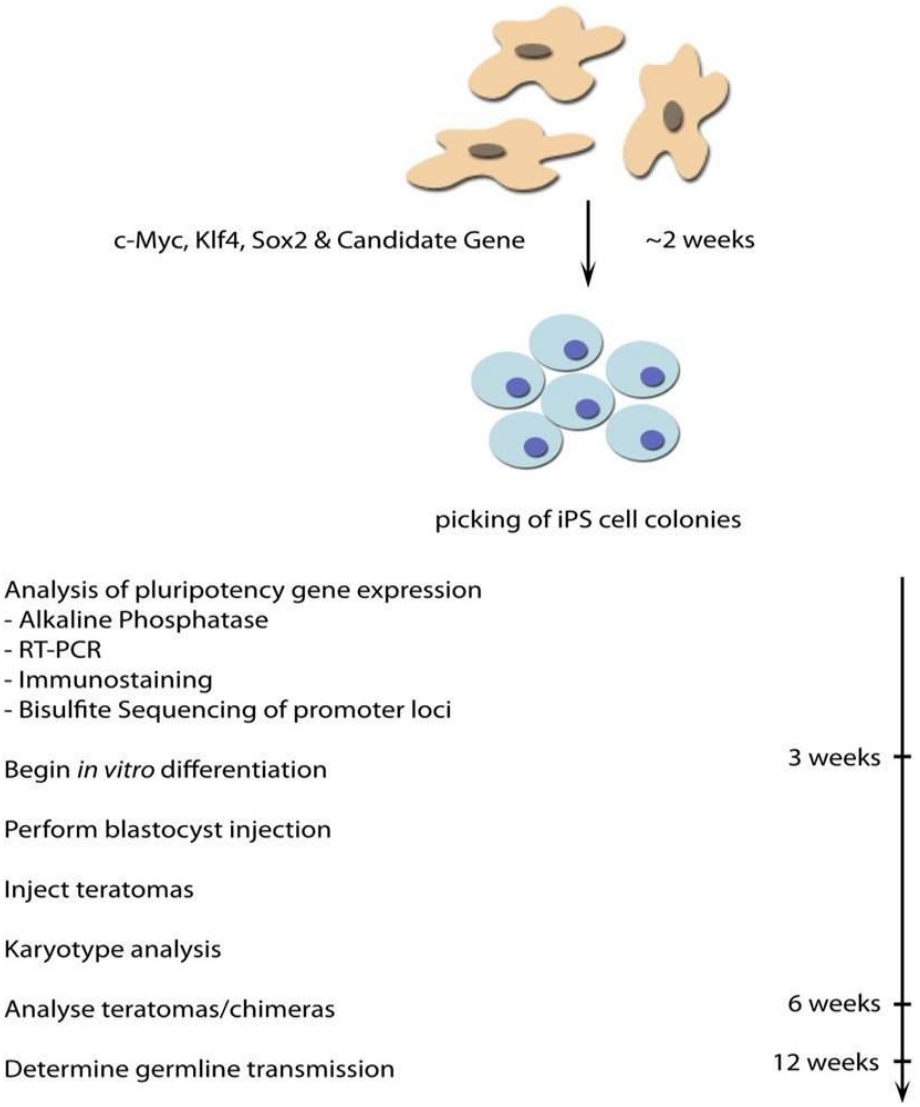



**Figure 4.1 Schematic diagram showing strategy to validate pluripotency of iPSCs** MEFs transfected with the desired reprogramming cocktail established iPSC colonies in an average of 2 weeks. These colonies were picked and expanded. As the iPSCs were reprogrammed in doxycycline inducible manner, withdrawal of the drug was necessary before proceeding to following steps. The cells were put through a battery of tests as described on the left. To the right, an estimated time line describes the amount of time required to fulfil each task.

## 4.2 Experimental Strategy

In order to validate the efficacy of the candidate genes in replacing the ectopic requirement of Oct4 during reprogramming, piggyBac mediated transposition was applied, mirroring the genetic screen performed in **Chapter 3**. **Figure 4.2** provides an illustration of the experimental approach employed. piggyBac transposons encompassing CAG-promoters were used to drive the expression of candidate genes and their co-operating factors. c-Myc, Klf4 and Sox2 were inserted into one transposon plasmid, and the candidate genes were inserted into separate transposon plasmids. This recapitulates the conditions experienced in the genetic screen where CAG-promoter was chosen to drive the three reprogramming factors, c-Myc, Klf4 and Sox2, synonymously with the inserted genetic locus. Plasmids that correspond to each candidate gene and its co-operating factors were introduced into MEFs by nucleofection$^{TM}$, seeded onto a feeder layer and cultivated in media containing LIF and serum to support the growth of iPSCs. These cells were maintained for 3 weeks and inspected for the appearance of iPSC colonies.

If colonies were observed using a CAG promoter driven approach, an additional experiment was carried out using the doxycycline inducible expression system. The ability to use tetracycline as a gene expression switch was first demonstrated through the modification of the prokaryotic tet repressor (tetR) complex (Gossen and Bujard, 1992). Fusion between tetR and a transcriptional activator from the herpes simplex virus (HSV-VP16) resulted in the formation of tetracycline transactivator (tTA). In the absence of tetracycline, tTA associates with tetracycline responsive elements (TRE) and induces gene expression levels. Conversely, in the presence of tetracycline, tTA binds to tetracycline and is displaced from TRE, consequently repressing gene expression levels. The temporal control of gene expression is rapid where levels fluctuate >50% within 24 hours. It was subsequently discovered that four amino acid substitutions within tTA resulted in the generation of reverse tetracycline transactivator (rtTA), and initiated transcription when bound to tetracycline (Gossen et al., 1995). Doxycycline, a derivative of tetracycline, was observed to be the most potent effector and triggered full expression within 24 hours. With the capability to gain temporal control over gene expression levels, this system was employed in our second stage of validation. Candidate genes which succeeded in the first

round of assessment were re-introduced into fibroblasts in the presence of doxycycline. If the gene combinations elicited reprogramming events in a doxycycline inducible manner, doxycycline would be withdrawn to investigate if iPSCs were independent of transgene expression, alluding to the attainment of true pluripotency.

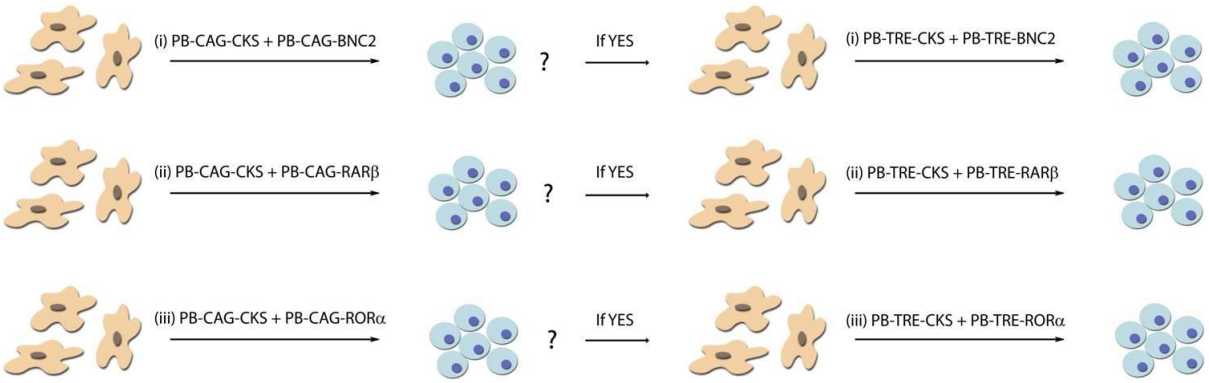



**Figure 4.2 Schematic diagram displaying validation strategy** (Left) piggyBac transposase was electroporated into MEFs with PB-CAG-CKS and an individual candidate gene. Transfected MEFs were seeded onto a mitotically inactive feeder layer and maintained for 3 weeks. (Right) If colonies were seen, a doxycycline inducible system was explored. PB-TRE-CKS and the validated candidate gene were introduced into MEFs, in combination with the helper vector and PB-CAG-rtTA. If colonies were observed, they were picked and expanded. Determination of their pluripotent potential is described in **Figure 4.1**.

### 4.2.1 Selection of validation candidates

The genetic screen conducted in **Chapter 3** identified five candidate genes which may exhibit properties that replace the requirement for exogenous Oct4 during reprogramming. As a first step, I focussed my efforts on BNC2, RARβ and RORα.

Although little information has been revealed about BNC2, it has been postulated to participate in mRNA processing (Vanhoutteghem and Dijan, 2004; Vanhoutteghem et al., 2006), which may perpetrate a global effect on the cellular machinery during reprogramming. In addition, mRNA processing has amassed interest due to its effects on pluripotency (Ji and Tian, 2009; O'Brien et al., 2010; Macfarlan et al., 2012), highlighting the potential effect of BNC2 in the acquisition of pluripotency. On the other hand, the RAR family has been depicted to influence the promoter and enhancer regions of Oct4 (Ben-Shushan et al., 1995; Delacroix et al., 2010) and the rate of reprogramming (Wang et al., 2011c). In addition, it is intriguing that two steroid hormone receptors were independently isolated from the genetic screen. With these reasonings, BNC2, RARβ and RORα were investigated for their capacities to influence reprogramming events.

## 4.3 Results

### 4.3.1 Generation of transposons expressing candidate genes

To obtain expression vectors that allow the ectopic expression of candidate genes in MEFs, eight transposon cassettes were generated: four CAG promoter-driven cassettes PB-CAG-CKS, PB-CAG-BNC2, PB-CAG-RARβ, PB-CAG-RORα, and four TRE-driven cassettes PB-TRE-CKS, PB-TRE-BNC2, PB-TRE-RARβ, PB-TRE-RORα (**Figure 4.3**). The strategies employed to create these constructs are described in detail in **Chapter 2**.

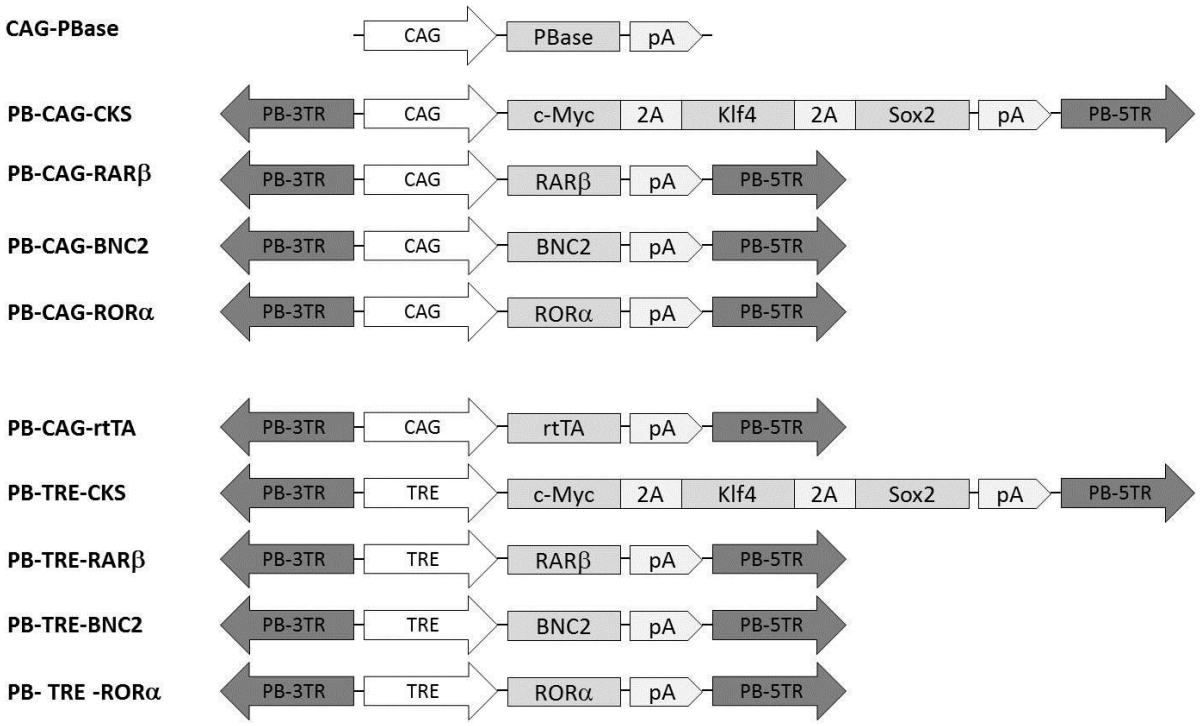



**Figure 4.3 Schematic diagram of transgenes employed in the validation process** (Top) The expression construct comprised of piggyBac transposase driven by a CAG promoter. (Middle) PB-CAG cassettes. The dark arrows flanking both ends depict piggyBac terminal repeats. Between the two boundaries lies a constitutively active CAG promoter driving the expression of (i) c-Myc, Klf4 and Sox2, or (ii) the candidate gene. (Bottom) PB-TRE cassettes. In a similar layout to the above, except the presence of a tetracycline responsive element controlling the expression of (i) c-Myc, Klf4 and Sox2, or (ii) the candidate gene. In addition, PB-CAG-rtTA comprises of a CAG promoter driving the expression of tetracycline reverse transcriptional activator (rtTA).

### 4.3.2 BNC2 and RARβ can function as Oct4 substitutes

To score the credibility of the three candidate genes, PB transposons carrying full length cDNAs encoding BNC2, RARβ and RORα were individually transfected with c-Myc, Klf4 and Sox2 to reprogram MEFs into iPSCs. Using a CAG-promoter driven approach, colonies were observed in reprogramming cocktails inclusive of BNC2 (combination CKS-B) (**Figure 4.4a**) and RARβ (combination CKS-Rβ) (**Figure 4.4b**). The appearance of CKS-B colonies arose 7 days after transfection, similar to the defined set of Yamanaka factors including Oct4 (combination CKS-O) (**Figure 4.4c**), whereas CKS-Rβ colonies emerged 10 days after transfection.

In contrast, RORα did not produce any iPSC colonies in co-operation with c-Myc, Klf4 and Sox2 (**Figure 4.5b**). The absence of reprogramming impetus can be attributed to the expression levels necessary to trigger pluripotency, or the window of expression required to initiate reprogramming events. The use of various promoters and the titration of DNA amounts can be applied to address this concern. Although the strategy behind the genetic screen employed CAG-driven transposon cassettes, the complex cellular milieu due to excessive amounts of transposition and forced transcription may confound the actual effect of RORα. In addition, co-operating aberrations may have assisted the effect of RORα. To overcome this, closer inspection of the results obtained from the genetic screen is necessary and this can be executed through deep sequencing of DNA acquired from iPS clones derived via the genetic screen. Alternatively, RORα agonists (Wang et al., 2010) or antagonists (Solt et al., 2011) can be introduced into the conventional reprogramming milieu to determine the effect of RORα on the acquisition of pluripotency.

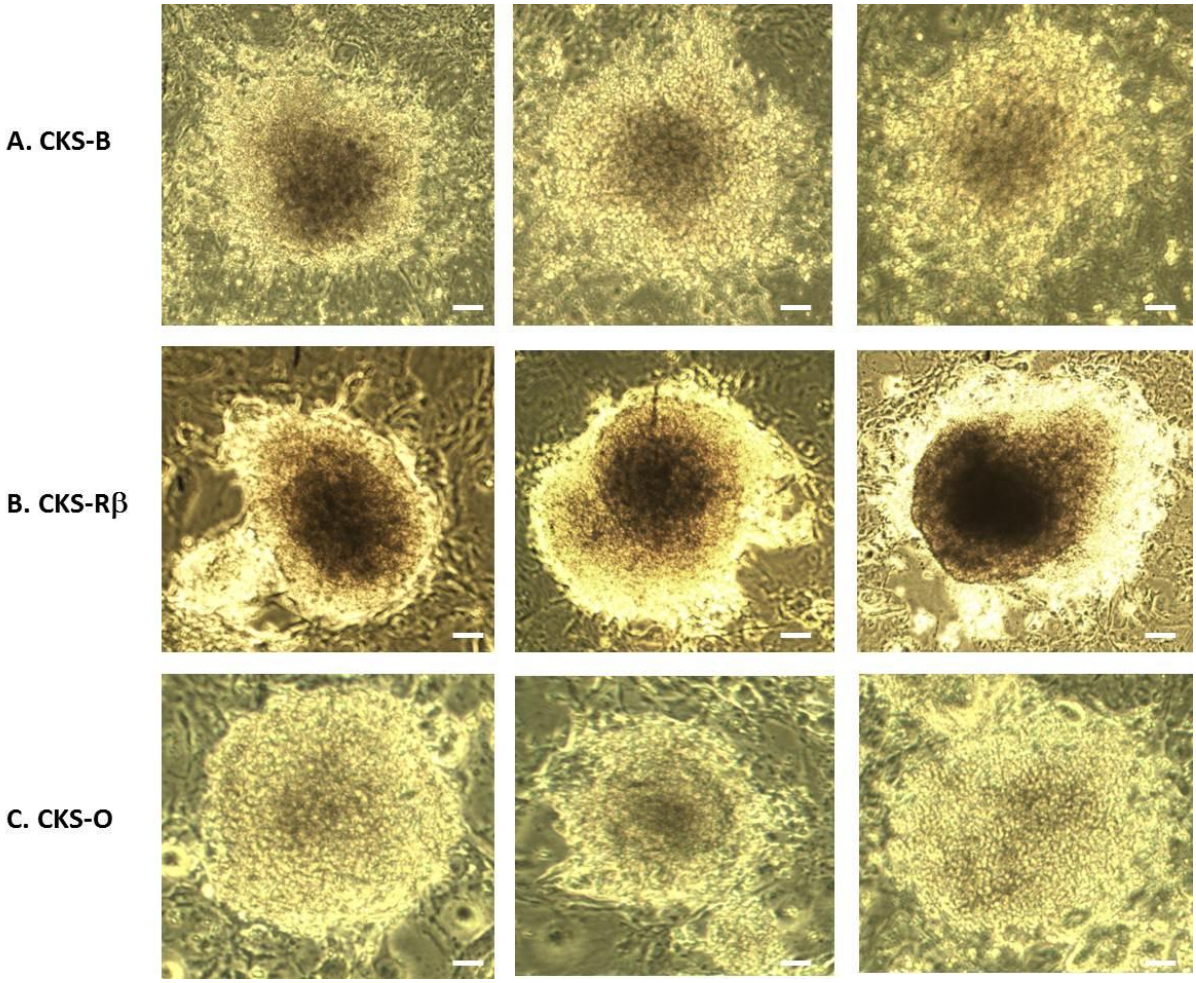



**Figure 4.4 Microscope pictures of primary colonies obtained from candidate genes, BNC2 and RARβ** MEFs were transfected with 2μg of CAG-PBase, 1μg of PB-CAG-CKS and 1μg of PB-CAG-mFx, where mFx refers to BNC2 (A), RARβ (B) or Oct4 (C). Transfected MEFs were seeded onto a mitotically inactive feeder layer and maintained for 18 days. Scale bar: 50μm.

#### 4.3.2.1 Comparison of efficiencies between BNC2, RARβ and Oct4

The abilities of RARβ and BNC2 in the substitution of Oct4 were measured through the assessment of reprogramming efficiencies (**Figure 4.5**). Using a fixed stoichiometric ratio, combinations of reprogramming factors were transfected into MEFs such that PB-CAG-CKS:PB-CAG-mFx, where Fx refers to Oct4, BNC2 or RARβ, was 1:1. Transfected MEFs were maintained on a feeder layer for 21 days and resultant colonies were stained for the expression of alkaline phosphatase. As iPSCs and ESCs have been demonstrated to express AP (Pease et al., 1990), its expression provides an indication of the reprogramming efficiencies elicited by various gene combinations. In summary, CKS-O gave rise to an average of 3500 colonies in three independent transfections (**Figure 4.5e**), whereas CKS-B (**Figure 4.5d**) and CKS-Rβ (**Figure 4.5c**) generated 2800 and 1200 colonies respectively. Although both replacement factors could not attain reprogramming efficiencies similar to Oct4 itself, this could be attributed to sub-optimal stoichiometric ratios of the transgenes. It has been established that gene dosage is critical during reprogramming (Carey et al., 2011; Lohle et al., 2012), hence modifications in transfected amounts may promote reprogramming efficiencies of BNC2 and RARβ.

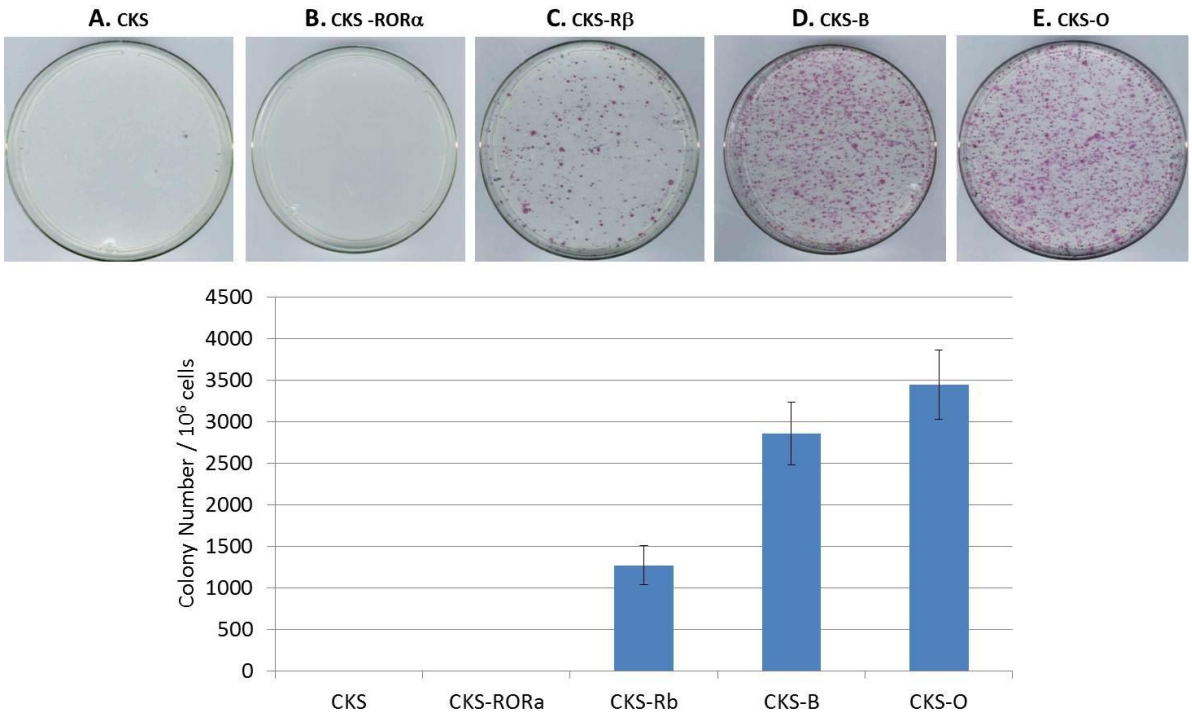



**Figure 4.5 Comparison of reprogramming efficiencies between Oct4, BNC2 and RARβ** MEFs were transfected with 2μg of CAG-PBase, 1μg of PB-CAG-CKS and 1μg of PB-CAG-mFx, where mFx refers to an empty vector (A), RORα (B), RARβ (C), BNC2 (D), or Oct4 (E). Transfected MEFs were seeded onto a mitotically inactive feeder layer and maintained for 21 days. The plates were assessed for the presence of AP, as depicted by the red stain (Top). The number of AP-expressing colonies were counted and shown in the bar chart (Bottom). Triplicates were performed, accounting for the error bars.

#### 4.3.2.2 Comparison of extent of reprogramming between BNC2, RARβ and Oct4

Although it has been described that CKS-B and CKS-Rβ lead to lower reprogramming efficiencies than CKS-O, the extent of reprogramming was not assessed. By using the activation of the endogenous Rex1 locus as a surrogate tool to signify pluripotency, determination of the extent and level of endogenous Rex1 expression will provide insight to the reprogramming potential of BNC2 and RARβ as replacements to ectopic Oct4. Initial reprogramming studies applied Fbx15 as a selection marker to identify reprogrammed cells (Takahashi and Yamanaka, 2006). However, selected cells were unable to contribute to the mouse germline. As Rex1 and Nanog expression mark ground state pluripotency (Silva et al., 2009; Guo and Smith, 2010), this bottleneck has been overcome through the stringent selection of reprogrammed cells using activated Nanog and Rex1 loci (Okita et al., 2007; Wang et al., 2011c). In this study, to enable strict assessment of the attainment of pluripotency, MEFs derived from Rex1::EGFP-IRES-Puro mice were utilised. Cells from these mice were genetically manipulated at the endogenous Rex1 locus, where the open reading frame of Rex1 was replaced by a cassette consisting of cDNAs encoding for enhanced green fluorescence protein (EGFP) and puromycin resistance. As such, green fluorescence and resistance to puromycin are indicative of the acquisition of ground state pluripotency. To determine if BNC2 and RARβ, in co-operation with c-Myc, Klf4 and Sox2, could lead to the attainment of naïve pluripotency, CKS-B and CKS-Rβ were independently transfected into Rex1::EGFP-IRES-Puro MEFs. Twenty-one days after transfection, the iPSC colonies were examined under the microscope for the presence of EGFP expression (**Figure 4.6**). CKS-O led to the formation of colonies that exhibited EGFP expression, depicting activation of the endogenous Rex1 locus (**Figure 4.6c**). Close inspection of the colonies reveal heterogenous distribution of EGFP signals, representing heterogeneity within the colony. This was reiterated when CKS-B was introduced into MEFs (**Figure 4.6a**). These findings reinforce prior observations of extensive heterogeneity within single iPSCs and between colonies (Masaki et al., 2008; Narsinh et al., 2011). Furthermore, colonies obtained from CKS-O and CKS-B were not compact and exhibited differentiation around the edges of the colonies. In contrast, CKS-Rβ gave rise to compact, three-dimensional colonies that expressed Rex1 homogenously (**Figure 4.6b**). These results are reminiscent of a recent study, where the

addition of RARγ and LRH1 to the conventional reprogramming mix instigated homogenous activation of the endogenous Rex1 locus within 4 days, whereas the traditional four transcription factors triggered Rex1 expression 12 days after transfection, in a disproportionate manner (Wang et al., 2011c).

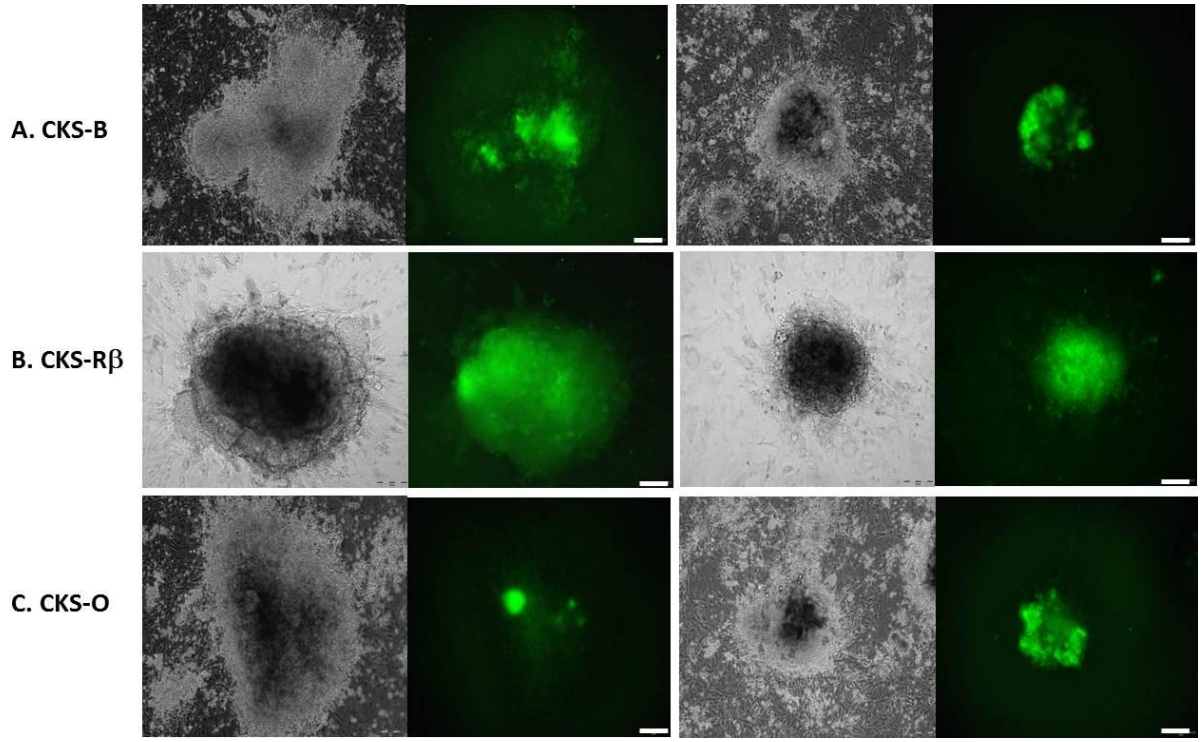



**Figure 4.6 Rex1 activation in primary colonies obtained from candidate genes, BNC2 and RARβ** Rex1::EGFP-IRES-Puro$^{R}$ MEFs were transfected with 2μg of CAG-PBase, 1μg of PB-CAG-CKS and 1μg of PB-CAG-mFx, where mFx refers to BNC2 (A), RARβ (B) or Oct4 (C). Transfected MEFs were seeded onto a mitotically inactive feeder layer and maintained for 18 days. The primary colonies were observed for Rex1 activation as depicted by EGFP expression and are shown at magnification of 100x. Scale bar: 50μm.

#### 4.3.2.3 Additive effect of BNC2 and RARβ on conventional reprogramming cocktail

Having demonstrated that BNC2 and RARβ exhibit properties that abolish the requirement for exogenous Oct4 during the reprogramming process, it is not clear if these inherent abilities of BNC2 and RARβ could augment the reprogramming efficiencies elicited by the conventional reprogramming mix, Oct4, c-Myc, Klf4 and Sox2 (OCKS). To address this concern, PB-CAG-BNC2 or PB-CAG-RARβ was co-electroporated with PB-CAG-OCKS and assessed for colony numbers after 10 and 14 days respectively.

Interestingly, co-expression of BNC2 with OCKS (combination OCKS-B) heightened its reprogramming efficiency by 3.7 fold (**Figure 4.7a**), evident by day 10, suggesting that BNC2 plays an additive role in the presence of Oct4 during the generation of iPSCs. BNC2 has been postulated to play a role in mRNA processing due to its localisation in nuclear speckles (Vanhoutteghem and Dijan, 2004; Vanhoutteghem et al., 2006), hence its over-expression may elicit a global transformation in transcripts, creating a conducive environment for cellular remodelling events. On the contrary, RARβ did not enhance reprogramming efficiencies of OCKS (combination OCKS-Rβ) to a large extent (**Figure 4.7b**). At 14 days post transfection, supplement of RARβ led to a marginal increase of 1.4 fold in the quantity of reprogrammed colonies, significantly lower than the difference elicited by BNC2. This suggests that the pathways triggered by RARβ are shadowed by ectopic Oct4. Alternatively, RARβ may primarily function through the regulation of Oct4, explaining the lack of an additive effect when RARβ and Oct4 are co-expressed in a reprogramming cocktail. From a different perspective, it is also plausible that stoichiometric ratios of RARβ and OCKS were suboptimal, and titration of transfected DNA amounts may be required to observe an additive effect. Overall, although RARβ did not elicit an improvement in reprogramming efficiencies when co-introduced with the conventional reprogramming cocktail, it will be intriguing to determine if RARβ affects the speed and homogeneity of reprogramming events.

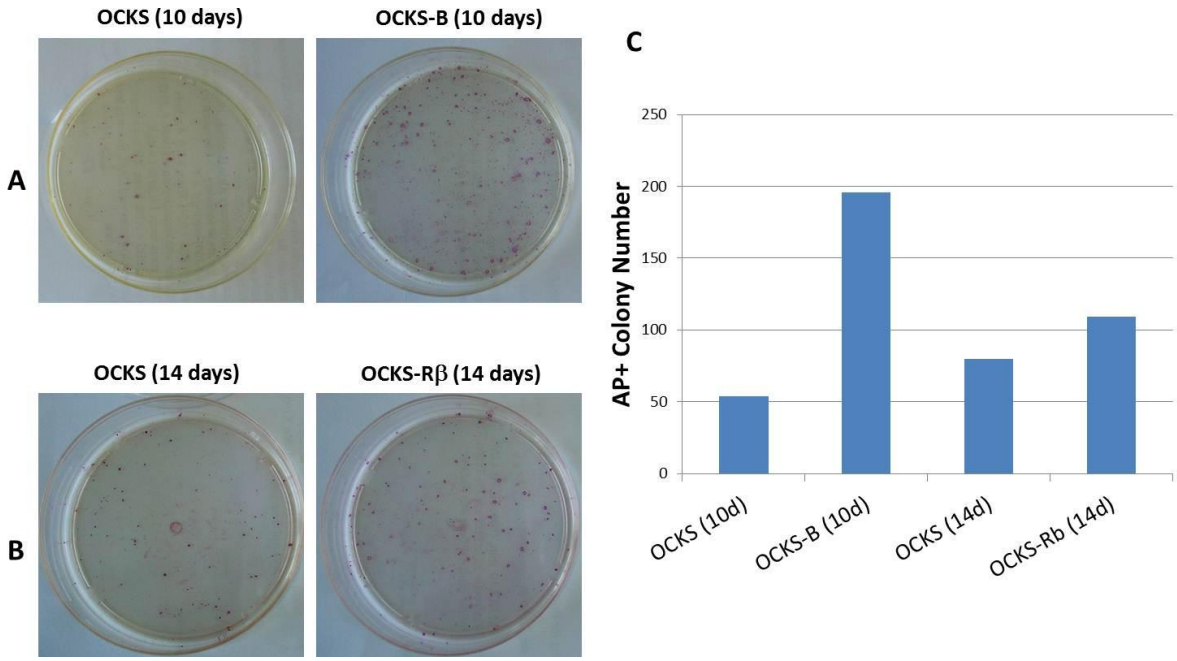



**Figure 4.7 Investigation of reprogramming efficiencies in the presence of Oct4 and BNC2 or RARβ** MEFs were transfected with 2μg of CAG-PBase, 1μg of PB-CAG-OCKS and 1μg of PB-CAG-mFx, where mFx refers to (A) BNC2 or (B) RARβ, or an empty vector. Transfected MEFs were seeded onto a mitotically inactive feeder layer and maintained for either 10 or 14 days (as indicated). The plates were assessed for the presence of AP, as depicted by the red stain. (C) Graphical representation of $AP^+$ colony numbers observed in (A) and (B).

### 4.3.3 Generation and characterisation of transgene-independent BNC2-iPSCs

As the second step of the validation process, CKS-B was introduced into Oct4::IRES-PuroEGFP MEFs using doxycycline inducible constructs and maintained in doxycycline for 21 days to obtain iPSCs. Oct4::IRES-PuroEGFP MEFs consist of genetic modifications at the endogenous Oct4 locus, where a fusion construct combining genes encoding for puromycin resistance and EGFP was inserted within the 3' UTR of Oct4. Surprisingly, reprogramming efficiencies using the doxycycline inducible system was significantly lower than using the CAG-promoter mediated system. Transfection of 1 x $10^6$ MEFs gave rise to two iPSC colonies, hereby designated as BNC2-iPSCs. A control experiment where only PB-TRE-CKS was introduced did not generate iPSC colonies. Low reprogramming efficiencies elicited by CKS-B could be attributed to the obligatory requirement for large doses of BNC2 to trigger the dedifferentiation process, and could be overcome by the amendments to the transfection cocktail. Examination of transposon insertion sites within successfully reprogrammed cells reveal a large number of integrated transposon cassettes (**Figure 4.8**), reinforcing the notion that high amounts of reprogramming factors are required to initiate reprogramming events. The use of splinkerette PCR to determine the number of integration sites is cursory and does not allow different transposon cassettes to be distinguished. As reprogramming cocktail CKS-B comprises of three different transposon cassettes (PB-TRE-CKS, PB-TRE-BNC2, PB-CAG-rtTA), it is not clear if multiple transposon integration sites in BNC2-iPSCs were attributable to high amounts of BNC2. On the other hand, additional experiments have demonstrated that high amounts of the conventional reprogramming cocktail are not obligatory for the formation of iPSCs (data not shown), speculating that a high proportion of insertion sites belongs to transposons encoding for BNC2. Alternatively, deep sequencing of genomic DNA from BNC2-iPSCs would provide an indication to the proportion of insertion sites attributed to PB-TRE-BNC2.

To ensure the purity of iPSCs generated using CKS-B, it is critical to verify the absence of contaminating plasmids consisting of exogenous Oct4. Detection of transgenes using genomic DNA reveals that only transposons which correspond to PB-TRE-CKS were observed (**Figure 4.9**), and transposons consisting of ectopic Oct4 were absent.

Reprogrammed colonies were picked and expanded. Upon successful establishment of the colonies in culture, doxycycline was withdrawn after 21 days. The surviving population of cells represent fully reprogrammed cells which retain their pluripotent nature independent of transgene expression. As an additional step to isolate a homogenous subset of naïve pluripotent cells, growth media was switched from serum-LIF to 2i-LIF conditions (Ying et al., 2008). As previously described, the addition of 2i-LIF gives rise to a homogenous population of iPSCs which exhibit "ground state" pluripotency and eliminates reprogramming intermediates which may confound the analysis of subsequent data (Ying et al., 2008). These cells were designated as BNC2-iPSCs and subsequently characterised.

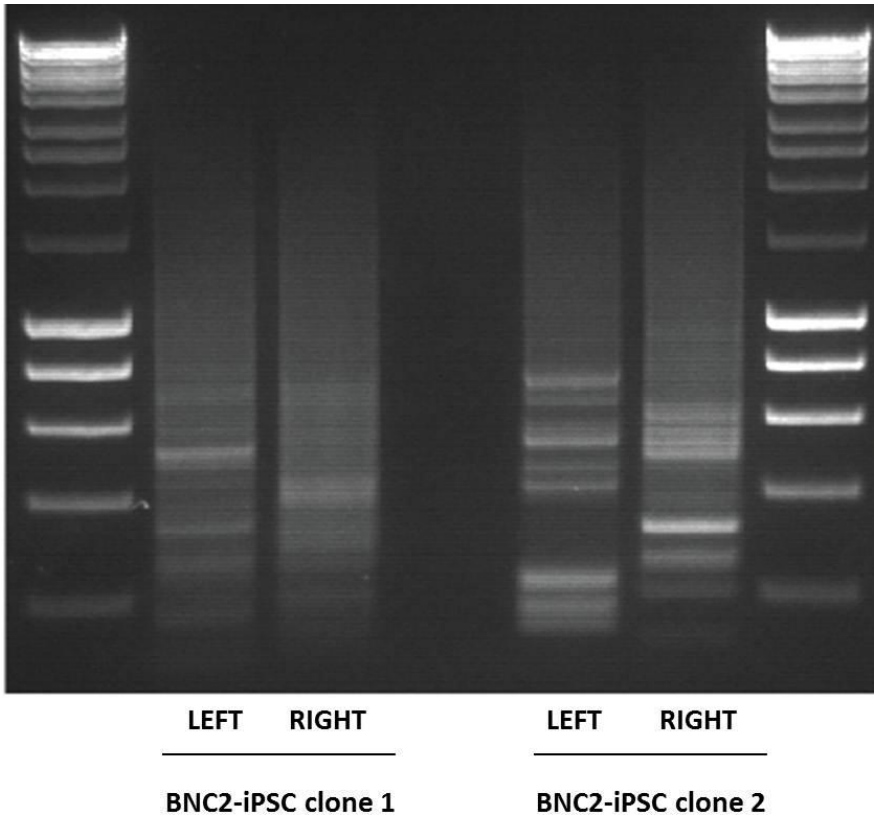



**Figure 4.8 Splinkerette Analysis of integration sites in BNC2-iPSCs** BNC2-iPSCs were generated using PB-TRE-CKS, PB-TRE-BNC2 and PB-CAG-rtTA. DNA from BNC2-iPSCs was harvested and splinkerette PCR was performed. Left and Right indicate the primers against the respective left or right transposon arm that were used during splinkerette PCR. Amplified products were visualised using an agarose gel and each band corresponds to an integration site.

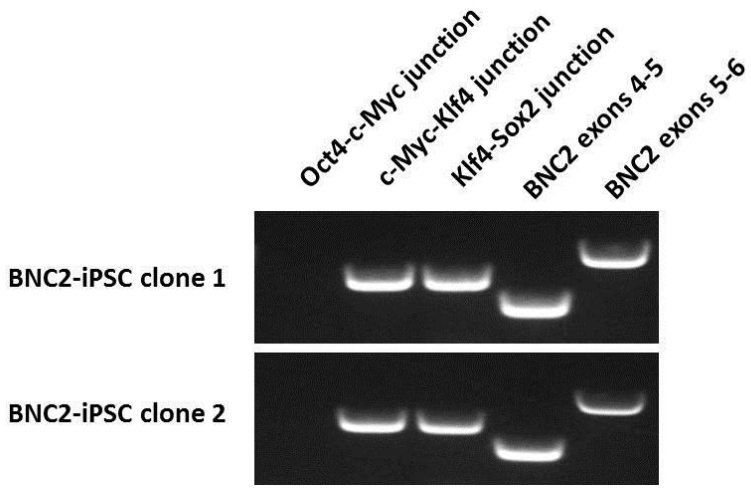



**Figure 4.9 Validation of the expression of transposon cassettes** Two BNC2-iPSC colonies generated from the introduction of PB-TRE-CKS (1μg), PB-TRE-BNC2 (1μg), CAG-PBase (2μg) and PB-CAG-rtTA (1μg) were expanded and maintained in the absence of doxycycline. RNA was extracted from each clone and converted to cDNA. Amplification of the junctions within the transposon cassettes was performed to determine the expression of the transgenes. Oct4-c-Myc junction was included to exclude the possibility of contamination from a separate vector.

#### 4.3.3.1 Characterisation of BNC2-iPSCs

In order to be certain that BNC2 can stimulate reprogramming events in the absence of Oct4, it is necessary to demonstrate that BNC2-iPSCs resemble ESCs. A battery of tests, as described in **Figure 4.1**, highlights the steps that must be overcome to verify the pluripotent nature of iPSCs.

As a rudimentary approach, pluripotent cells should stain positive for the expression of AP, a stem cell marker. **Figure 4.10** shows images of doxycycline independent BNC2-iPSCs which express AP. To obtain a pure population of iPSCs devoid of contaminating fibroblasts, BNC2-iPSCs were maintained in a feeder-free environment.

As ESCs have unique gene signatures that encompass the expression of the core pluripotency apparatus, BNC2-iPSCs were assessed if they exhibited a similar transcription profile. Qualitative examination of transcript levels using RT-PCR reveals that BNC2-iPSCs express a list of pluripotency markers at similar levels to ESCs and iPSCs derived by delivering the conventional reprogramming cocktail using PB-TRE-CKS and PB-TRE-Oct4 (Oct4-iPSCs) (**Figure 4.11**). Quantification of transcript levels using real-time PCR displayed an analogous pattern (**Figure 4.12**), reinforcing the similarities between BNC2-iPSCs, ESCs and Oct4-iPSCs. To ensure that these transcripts were efficiently processed and translated, protein levels of Oct4, Nanog and SSEA-1 were determined through immunofluorescence (**Figure 4.13**). BNC2-iPSCs displayed high levels of the examined markers, reaffirming the presence of the ES cell coupled transcriptional machinery.

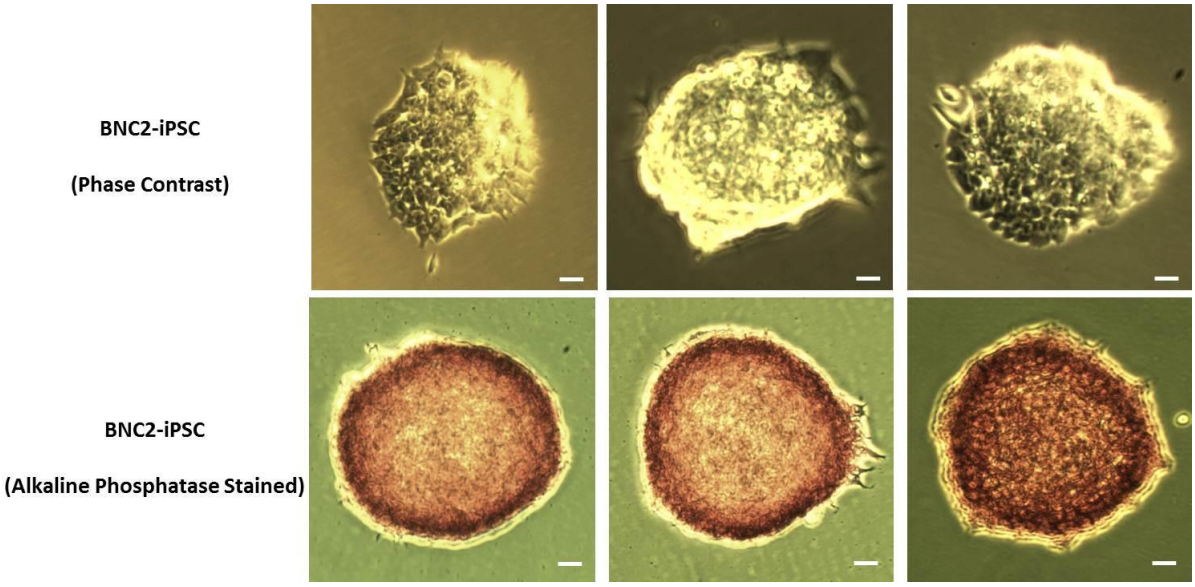



**Figure 4.10 Images of doxycycline independent BNC2-iPSC colonies** BNC2-iPSC colonies generated from the introduction of PB-TRE-CKS (1μg), PB-TRE-BNC2 (1μg), CAG-PBase (2μg) and PB-CAG-rtTA (1μg) were expanded and maintained in 2i+LIF, in the absence of doxycycline, and on gelatin coated plates. Three images of unmodified colonies were captured in brightfield (Top). The colonies were also assessed for the presence of AP, as depicted by the red colour (Bottom). Scale bar: 50μm.

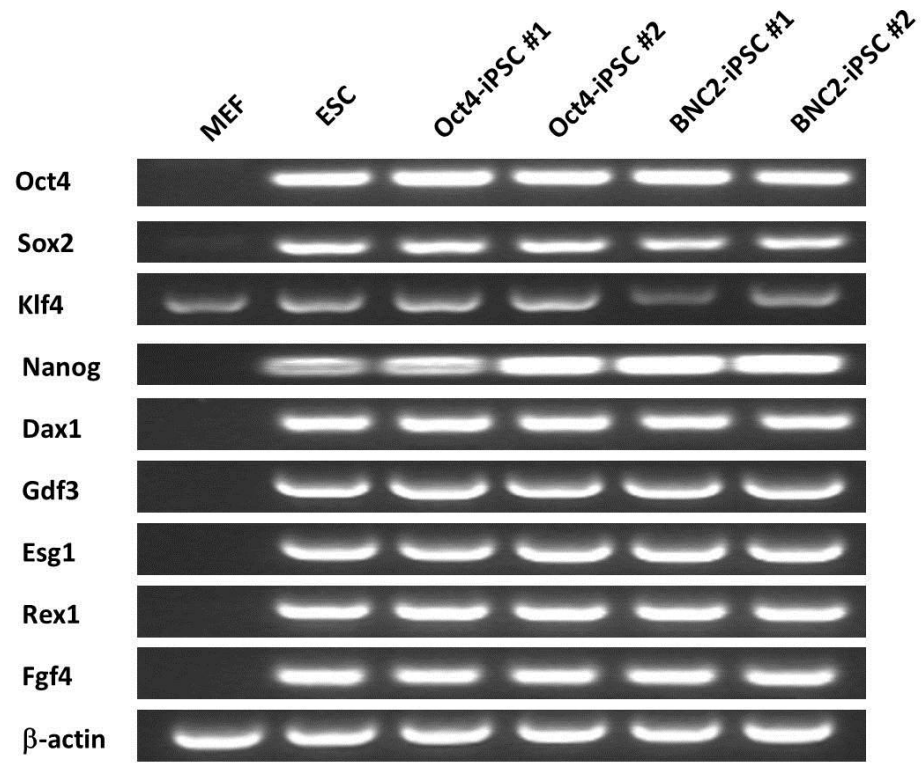



**Figure 4.11 Expression levels of pluripotency markers in BNC2-iPSCs** RT-PCR of the various pluripotency markers was performed in MEF (lane1), ESCs (lane2), iPSCs generated using c-Myc, Klf4, Sox2 and Oct4 (lanes 3 and 4), and iPSCs generated using c-Myc, Klf4, Sox2 and BNC2 (lanes 5 and 6). Expression levels are described qualitatively by the intensity of the amplified products viewed using an agarose gel. The array of pluripotency markers tested is labelled to the left of the image.

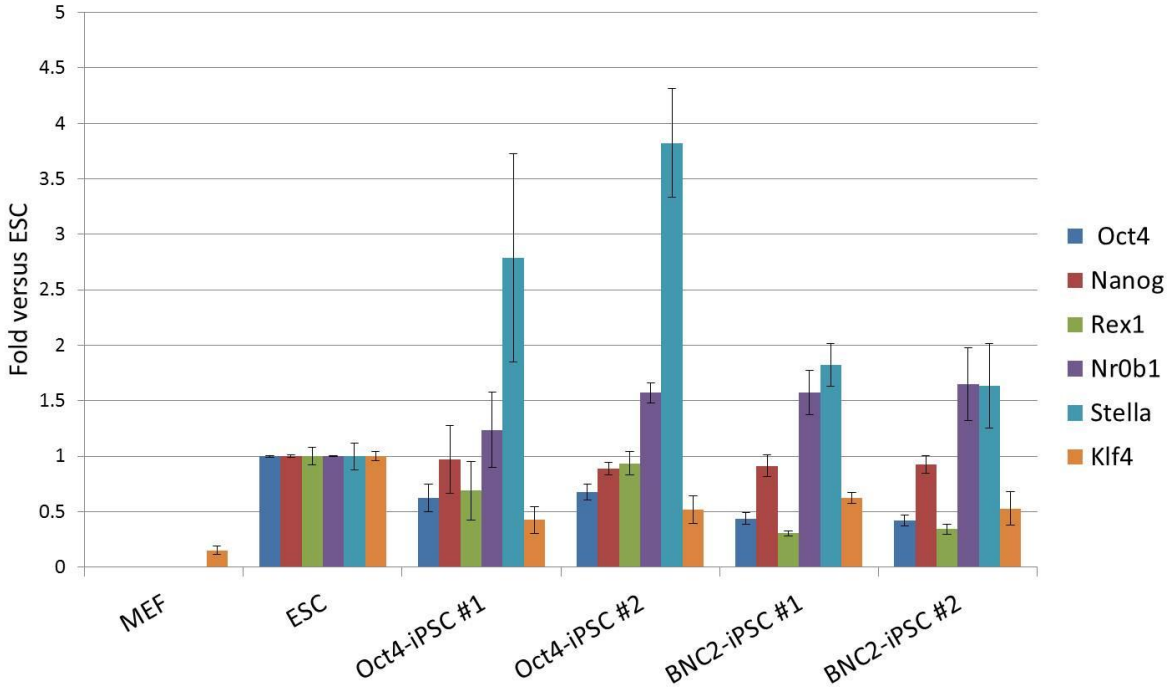



**Figure 4.12 Quantitative expression levels of pluripotency markers in BNC2-iPSCs** q-PCR of the various pluripotency markers was performed in MEF (lane1), ESC (lane2), Oct4-iPSCs (lanes 3 and 4), and BNC2-iPSCs (lanes 5 and 6). Expression levels of pluripotency markers are normalised to GAPDH amounts and ESC expression levels

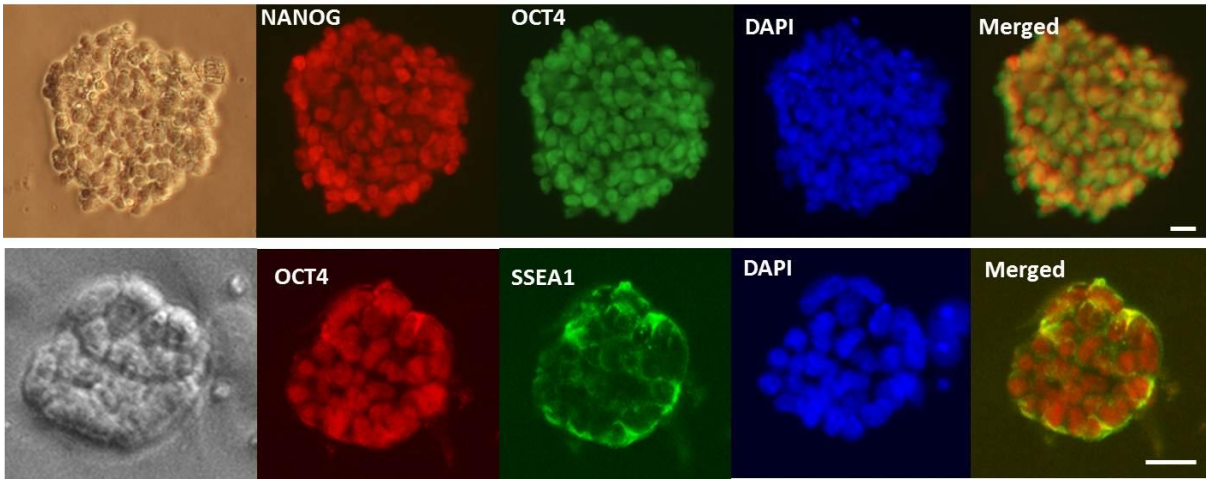



**Figure 4.13 Immunostaining of pluripotency markers in BNC2-iPSCs** Doxycycline independent BNC2-iPSCs were fixed and stained with antibodies against, (Top) Nanog (red) and Oct4 (green), and (Bottom) Oct4 (red) and SSEA-1 (green). Images of representative colonies are shown, displaying the respective fluorescence, phase contrast and DAPI stain. Scale bar: 50μm.

As reprogramming is accompanied by remodelling of the epigenetic landscape, it is crucial to determine that promoter loci of genetic elements associated to pluripotency have been modified to an ES cell-like state. Bisulfite sequencing of promoter loci corresponding to Rex1, Nanog and Oct4 in BNC2-iPSCs indicated demethylation on several sites, disclosing distinct methylation patterns from MEFs but a similar relationship to ESCs (**Figure 4.14**).

Although it is evident that genes associated to pluripotency are activated in BNC2-iPSCs, as determined by transcript, protein and epigenetic analysis, and BNC2-iPSCs bear resemblance to ESCs and Oct4-iPSCs, it is pivotal to assess the function of BNC2-iPSCs. Pluripotent stem cells are depicted by the ability to derive three germ layers upon differentiation. To this end, BNC2-iPSCs were grown onto low attachment dishes in the absence of LIF to form embryoid bodies. Seeding of the embryoid bodies led to the development of mesodermal, ectodermal and endodermal derivatives (**Figure 4.15**), reinforcing the pluripotent nature of BNC2-iPSCs.

Having demonstrated that BNC2-iPSCs are able to generate derivatives of the three germ layers on the petri dish, it is critical to determine the pluripotent capacity of BNC2-iPSCs *in vivo*. This can be done by embedding the cells in a physiologically relevant environment and determining their differentiation potential. To this end, BNC2-iPSCs were introduced into mouse blastocysts and immune-compromised mice to determine their contribution to the development of somatic tissues and teratomas respectively.

As the constant expression of the reprogramming factors may possess deleterious effects when implanted into the mouse (Okita et al., 2007), it is important to attest to the silencing of the transgenes. cDNA obtained from BNC2-iPSCs maintained in the absence or presence of doxycycline was used as a template to amplify junctions between transgenes. Clearly illustrated in **Figure 4.16**, transgenes were minimally expressed. This was observed in one of two BNC2-iPSC lines that were established and the cell line which

displayed minimal transgene expression was employed for further tests, affirming that transgene expression should not confound the ability of BNC2-iPSCs to contribute to somatic tissues and the germline in mice.

Another aspect which may affect the implantation of iPSCs into the mouse blastocyst or subcutaneous flank is genomic integrity. Gross chromosomal anomalies can result in detrimental effects in the live animal. As a mode of assessment, fluorescence *in situ* hybridisation was performed to analyse the chromosomal entirety. Inspection of twenty metaphase-spreads revealed genomic abnormalities in all examined cells. Centromeric fusion at chromosome 11 was the most frequently observed abnormality (**Figure 4.17)**. Chromosome 11 has been reported to play important developmental roles where the inclusion of a balancer chromosome results in prenatal and postnatal death, small size, developmental delay, craniofacial or neurological abnormalities (Kile et al., 2003). As such, it is possible that BNC2-iPSCs may lead to abnormal phenotypes when introduced into mice.

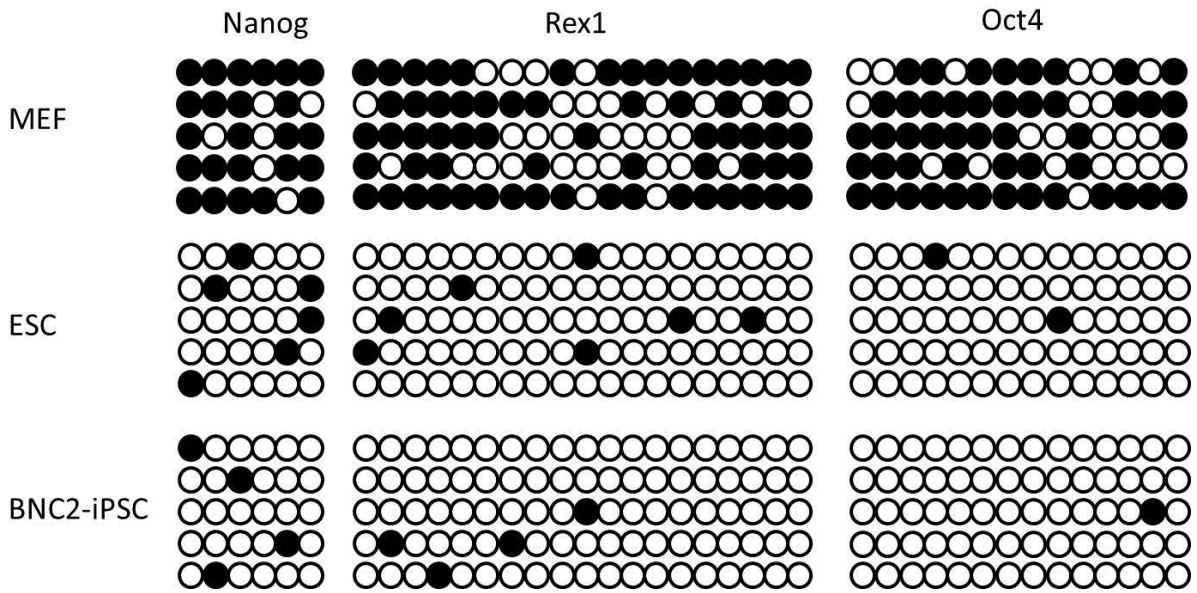



**Figure 4.14 Analysis of methylation at promoter loci** DNA was obtained from MEF (Top), ESCs (Middle) and doxycycline independent BNC2-iPSCs (Bottom) and treated with sodium bisulphite. PCR directed at promoter loci of Nanog (Left), Rex1 (Centre) and Oct4 (Right) were performed. Amplified products were analysed and summarised in the diagram. Each circle represents a CpG dinucleotide and five methylation patterns are displayed for each promoter loci, as indicated by the five rows. 6, 19 and 14 CpG dinucleotides were examined at the Nanog, Rex1 and Oct4 promoters respectively. A shaded circle indicates a methylated site, whereas an unshaded circle indicates an unmethylated site.

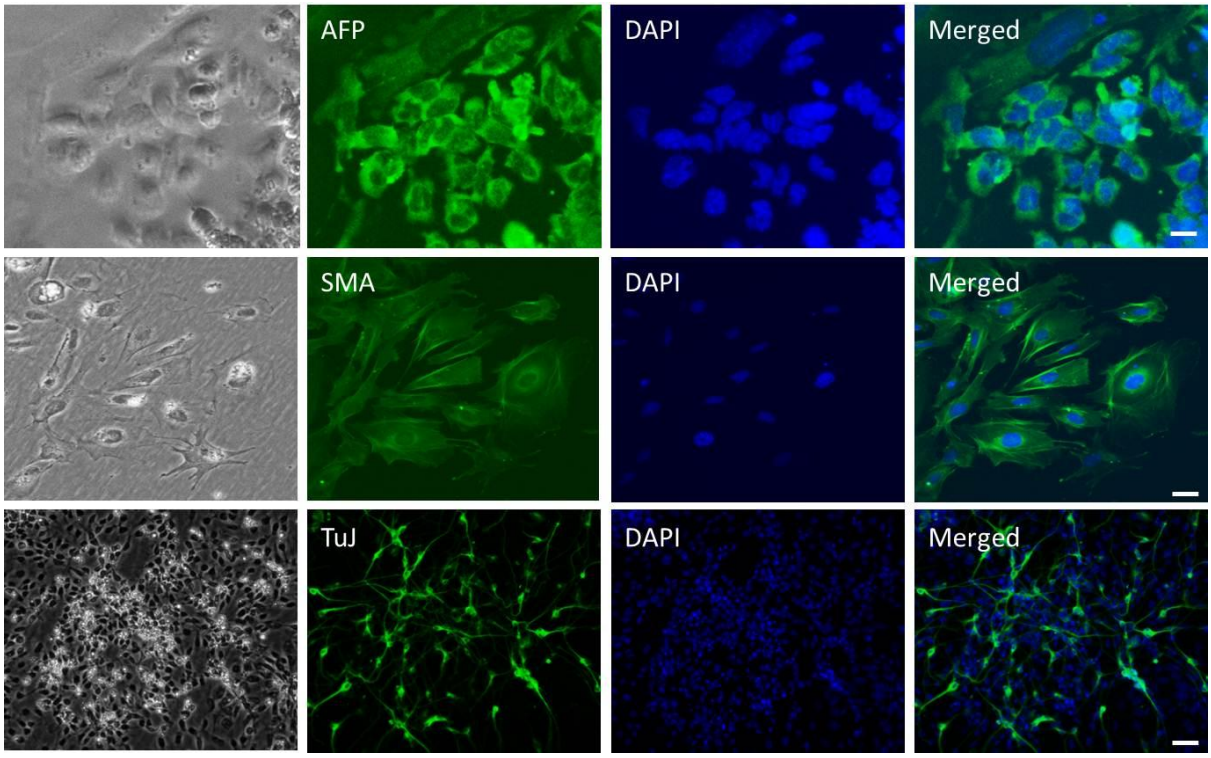



**Figure 4.15 BNC2-iPSCs can differentiate into three germ layers *in vitro*** BNC2-iPSCs were differentiated into the three germ layers *in vitro* using various culture conditions. Differentiated cells representing the endodermal, mesodermal and ectodermal lineages were immunostained against alpha feto-protein (Top), smooth muscle actin (Middle), and β-tubulin III (Bottom) respectively. The markers are labelled in green, DAPI which stains the cell nuclei is labelled in blue. Scale bar: 50μm.

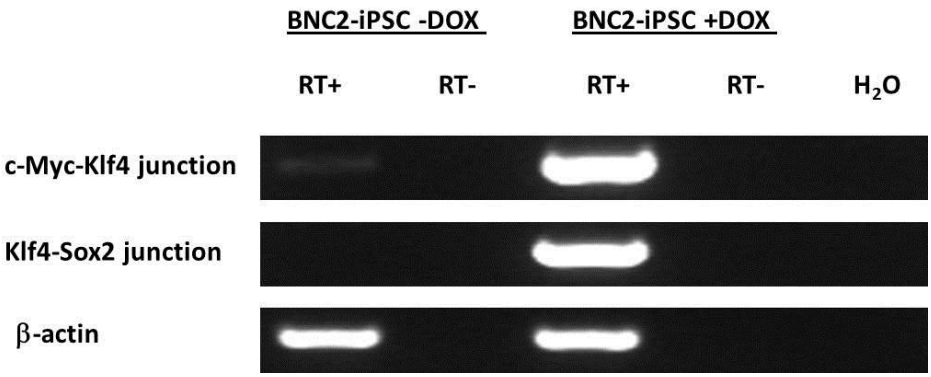



**Figure 4.16 Transgenes are silent in the absence of doxycycline** BNC2-iPSCs were cultivated in the presence and absence of doxycycline (DOXYCYCLINE). RNA was extracted from these cells and (i) converted to cDNA using reverse transcriptase (RT+) or (ii) tested for contamination with genomic DNA in the absence of RT (RT-). Amplification of the junctions between c-Myc and Klf4, and Klf4 and Sox2 was performed to determine the expression of the transgenes. Water only control (last lane) was included as an additional layer of control.

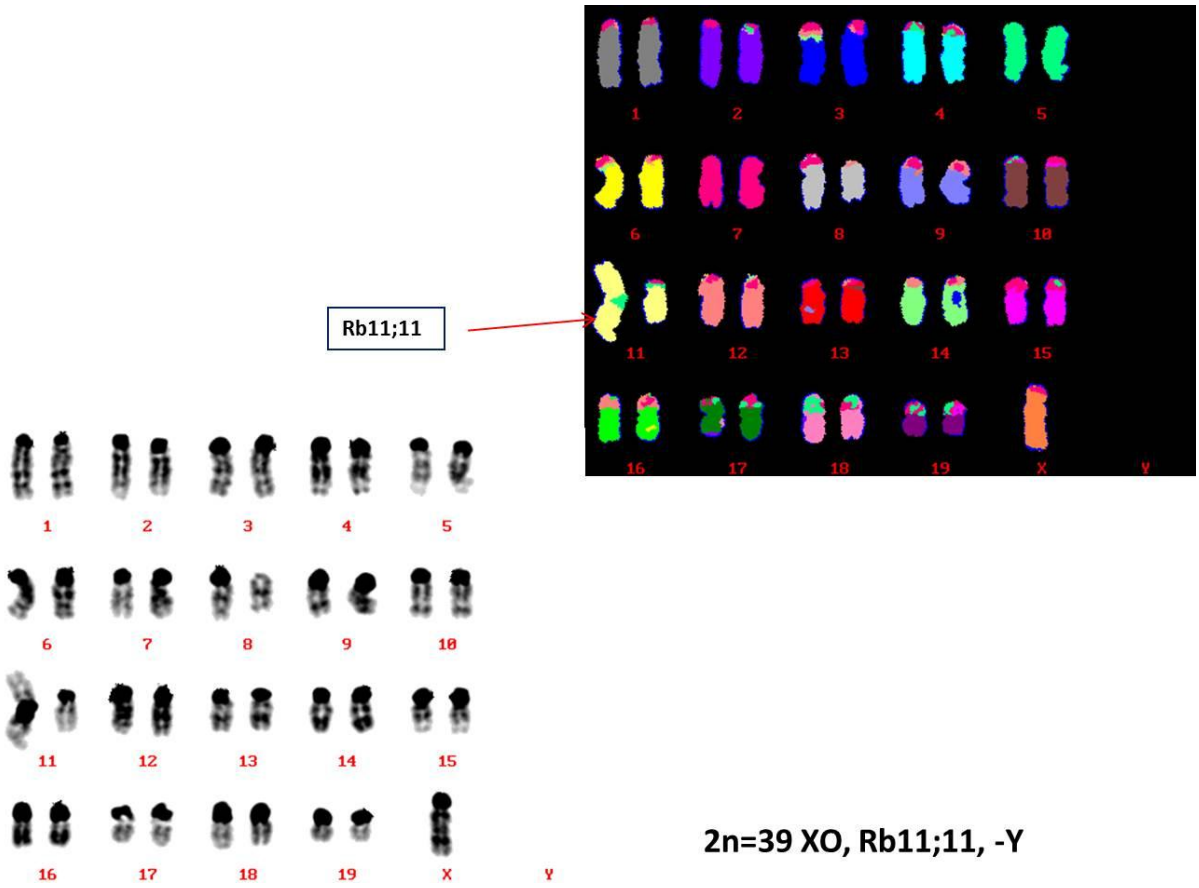



**Figure 4.17 Karyotype analysis of BNC2-iPSCs** BNC2-iPSCs were mitotically arrested and metaphase spreads were achieved. Using Fluorescence *in situ* hybridisation (FISH), each chromosome was accounted and analysed. The array of colours depicts different chromosomes and the arrow indicates a centromeric fusion at chromosome 11. The diploid chromosome count and abnormalities encountered are displayed at the bottom.

However, subcutaneous injection of 1 x $10^6$ BNC2-iPSCs into each flank of immune-compromised mice led to the development of hard masses of one centimetre in diameter two weeks after injection. The mice were sacrificed and tumours were extracted. **Figure 4.18** illustrates representative histological sections of the tumours, portraying contribution to the three germ layers.

Surprisingly, micro-injection of BNC2-iPSCs into the inner cell mass of mouse blastocysts generated live chimeric pups. As the starting material used to generate BNC2-iPSCs was derived from crosses between C57BL/6J and 1295S, contribution by BNC2-iPSCs can be assessed by black or agouti coats of fur. In contrast, blastocysts that received injections possessed a recessive mutation at the Tyrosinase locus which is required for synthesis of melanin. As a result, the extent of embryonic development dictated by BNC2-iPSCs can be estimated by the proportion of black or agouti contribution to the fur coats of pups. **Figure 4.19** depicts images of chimeras obtained from the micro-injection of BNC2-iPSCs. Thirteen mice exhibiting 5-80% chimerism based on coat-colour were generated. When crossed with albino mice, one male chimera ($F_0$) produced offspring with black coats ($F_1$), signifying contribution to the germline. Germline competence of BNC2-iPSCs was unexpected as the injected BNC2-iPSC line exhibited an abnormal karyotype. To verify the karyotype of the chimeric mouse which demonstrated germline competence of BNC2-iPSCs, splenocytes were extracted and analysed. Interestingly, the splenocytes displayed normal karyotypes (**Figure 4.20**), surmising that a subset of BNC2-iPSCs exhibited normal karyotypes. As only twenty cells were initially analysed (**Figure 4.17**), it is possible that karyotypically normal cells existed in low proportions and were undetected. As a result, injection of the population of cells into the mouse blastocyst acted as a selection process to only allow karyotypically normal cells to contribute to chimerism and display germline competence.

A mouse colony was established through crossing between $F_1$ mice. To determine that offspring from the colony retained transposon cassettes similar to BNC2-iPSCs, ear biopsies from $F_2$ mice were genotyped. **Figure 4.21** illustrates that $F_2$ mice exhibited

combinations of transposons present in BNC2-iPSCs, reaffirming that progeny were initially derived from BNC2-iPSCs. These pieces of evidence portray the capacity of BNC2 to competently reprogram MEFs in the absence of exogenous Oct4, resulting in BNC2-iPSCs which exhibit naïve pluripotency.

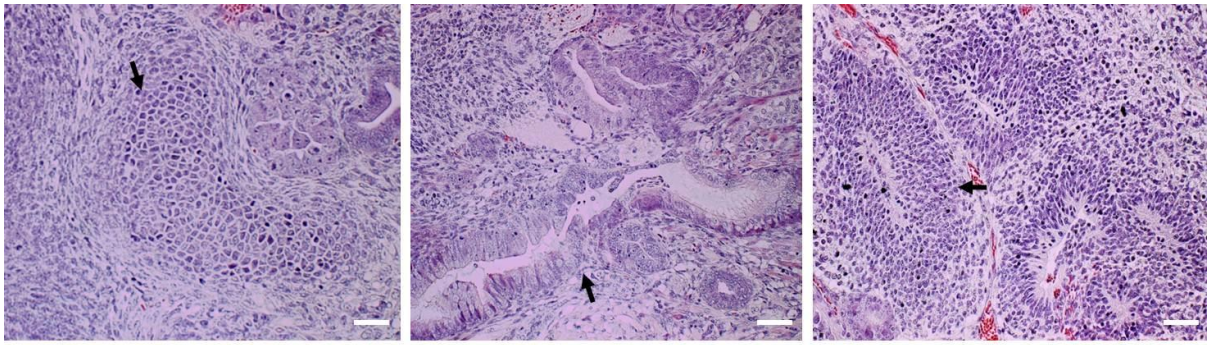



**Figure 4.18 BNC2-iPSCs can differentiate into three germ layers *in vivo*** BNC2-iPSCs were injected subcutaneously into both flanks of immunocompromised NOD.Cg-Prkdc$^{scid}$ Il2rg$^{tm1Wjl}$/SzJ (NSG) mice. After 2 weeks, tumours were observed. Histological analyses of sections of these tumours are shown. Differentiation into endodermal, mesodermal and ectodermal lineages are displayed and labelled. The arrows indicate the position of the differentiated structure. Scale bar: 50μm.

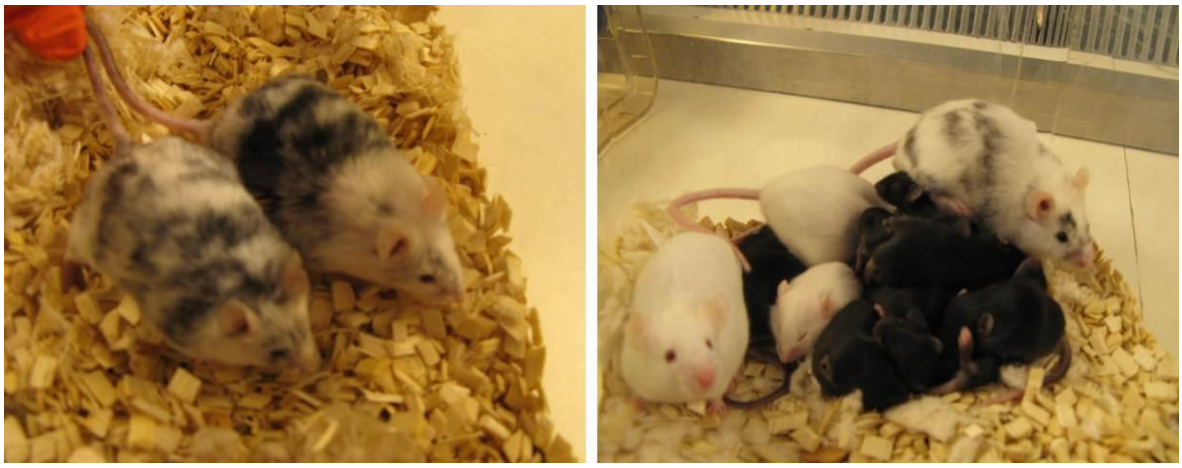


**Figure 4.19 BNC2-iPSCs are capable of generating chimeras and contributing to the germline** (LEFT) Chimeras (50% and 70%) were generated through the injection of BNC2-iPSCs into mouse blastocysts. (RIGHT) Crossing between a 30% chimera and albino CALB mice resulted in the generation of pups carrying a black fur coat.

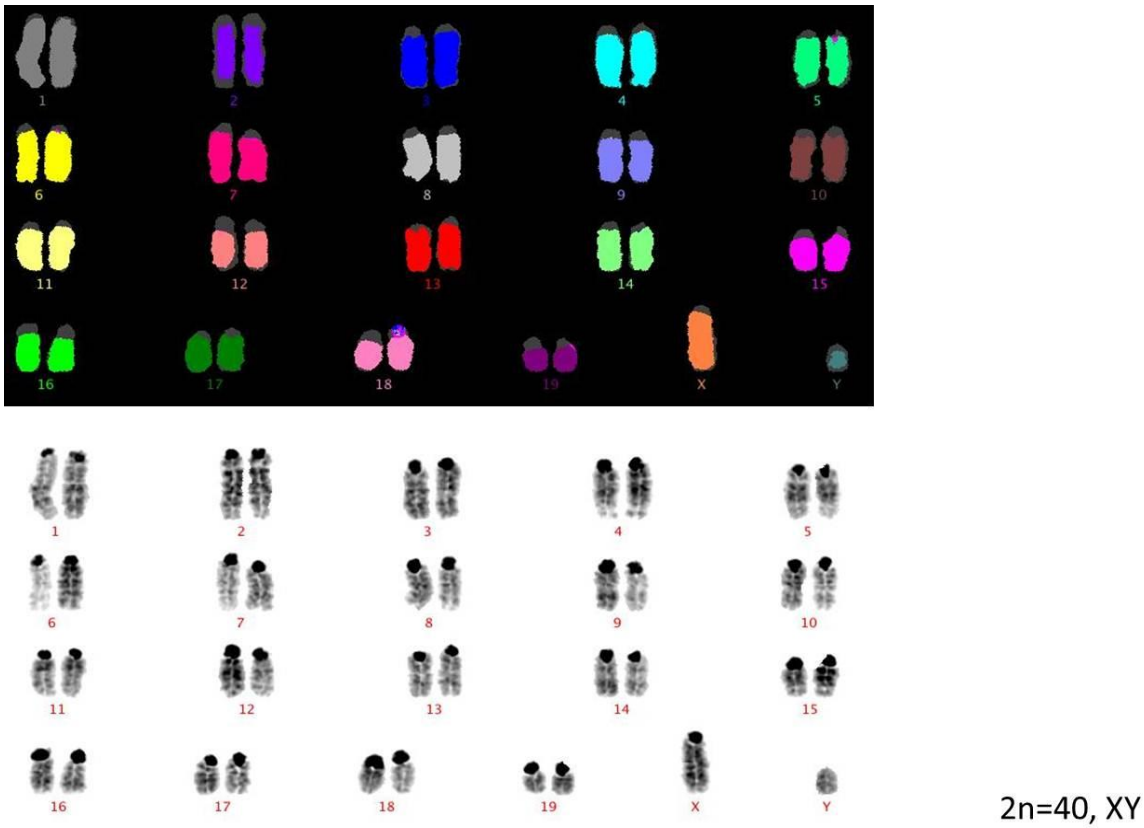



**Figure 4.20 Karyotype analysis of splenocytes from a mouse chimera derived using BNC2-iPSCs** Splenocytes were extracted from the spleen of a chimeric mouse. Splenocytes were mitotically arrested and metaphase spreads were achieved. Using Fluorescence *in situ* hybridisation (FISH), each chromosome was accounted and analysed. The array of colours depicts different chromosomes.

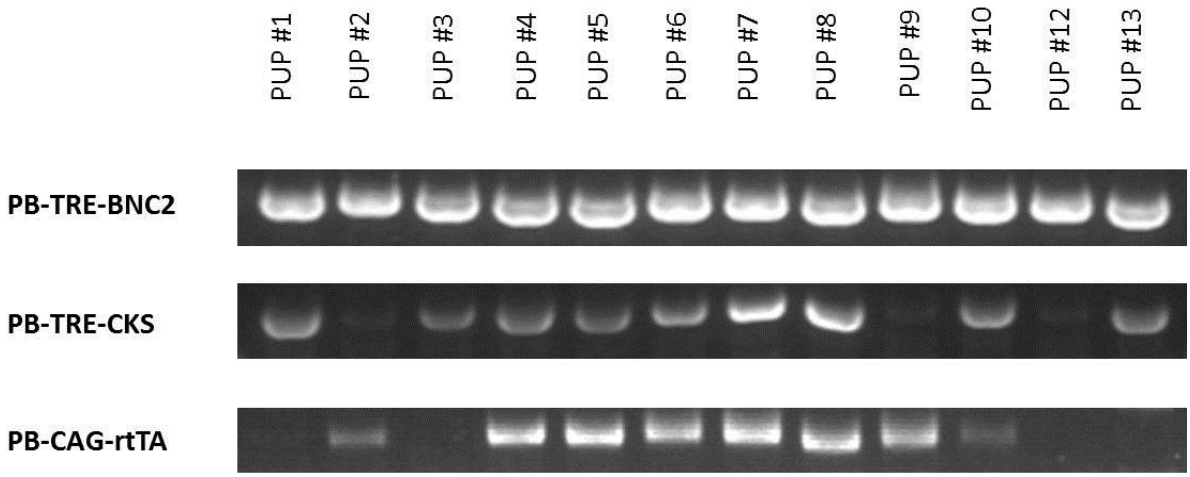



**Figure 4.21 Genotyping of mouse colony established from BNC2-iPSCs** Pups which displayed germline contribution of BNC2-iPSCs were crossed among themselves to establish a mouse colony. 13 pups belonging to F2 generation were genotyped for the presence of transposon cassettes. DNA was retrieved from ear biopsies and PCR was performed to amplify regions specific to PB-TRE-BNC2, PB-TRE-CKS and PB-CAG-rtTA. The amplified products were visualised using agarose gel and the presence of a band signifies the presence of the respective transgene.

### 4.3.4 Generation and characterisation of transgene-independent RARβ iPSCs

Employing a similar strategy to BNC2, the ability of RARβ to negotiate the requirement for ectopic Oct4 during the initiation of reprogramming events was reaffirmed through the transfection of CKS-Rβ into MEFs in a doxycycline inducible approach. Rex1::EGFP-IRES-Puro MEFs were utilised as a platform to assess reprogramming. As described earlier, Rex1::EGFP-IRES-Puro MEFs consist of genetic modifications at the endogenous Rex1 locus where an expression cassette for EGFP and puromycin resistance was inserted adjacent to exon 1. As a result, successful reprogramming of MEFs can be examined by transcriptional activation of Rex1 and EGFP expression.

Three weeks following the introduction and expression of CKS-Rβ, 12 ES cell-like colonies were picked and expanded. Unlike BNC2, although the use of the drug inducible approach significantly reduced the number of reprogrammed colonies when compared against the constitutively active system, the difference was not as stark as BNC2. Overall, reprogramming efficiencies incited by the introduction of CKS-Rβ and CKS-B in a doxycycline inducible manner was reduced by 10- and 1400-fold respectively. This could be attributable to the inherent properties of BNC2 and RARβ. As BNC2 participates in mRNA processing and may function as an Oct4 substitute indirectly, large volumes of BNC2 may be required. In contrast, RARβ has been described to affect Oct4 expression directly (Ben-Shushan et al., 1995), potentially reducing the amount of RARβ necessary to trigger reprogramming events in the absence of Oct4 and lowering its susceptibility to gene dosage effects.

Upon the establishment of 12 iPSC clones, doxycycline was withdrawn to obtain subsets of cells which were not reliant on transgene expression for the maintenance of pluripotency. An additional selection pressure was inflicted on surviving cells through a switch in growth media, from serum-LIF to 2i-LIF. This allows the isolation of a homogenous population of cells which exhibit naïve pluripotency. Resultant clones are henceforth labelled as RARβ-iPSCs and were characterised.

### 4.3.4 Characterisation of RARβ-iPSCs

To ascertain the pluripotent potential of RARβ-iPSCs and verify the competence of RARβ as an effective substitute for the ectopic requirement of Oct4 during reprogramming, a battery of tests described in **Figure 4.1** was performed. As a primary assessment of the pluripotent nature of RARβ-iPSCs, the cells were inspected under the microscope to detect EGFP fluorescence. As the MEFs which underwent reprogramming harboured a genetic modification at the endogenous Rex1 locus, activation of Rex1 was coupled to EGFP expression and puromycin resistance. **Figure 4.22** displays homogenous EGFP expression in RARβ-iPSCs, suggestive of the activation of Rex1, a pluripotency marker.

To gain insight into the molecular circuitry within RARβ-iPSCs, transcription of pluripotency markers were assessed and compared to ESCs and Oct4-iPSCs. To this end, a list of pluripotency-associated factors was examined qualitatively and quantitatively. **Figure 4.23** illustrates the qualitative assessment of a string of pluripotency markers in RARβ-iPSCs. Comparison of transcript patterns to ESCs and Oct4-iPSCs reveal a close relationship. Quantitative analysis of pluripotency gene expression levels is depicted in **Figure 4.24**. As a general trend, it is evident that RARβ-iPSCs express pluripotency markers at levels comparable to ESCs and Oct4-iPSCs.

Having demonstrated that transcripts of pluripotency markers exist at similar levels between RARβ-iPSCs and other pluripotent derivatives, it is crucial to determine that transcript levels can be extrapolated to the amounts of their translated products. Using immunofluorescence, proteins coding for Oct4, Nanog and SSEA-1 were examined. **Figure 4.25** illustrates images which exhibit clear and strong delineation of Oct4 and Nanog within the nucleus and SSEA-1 at the cell surface. As an additional layer to investigate the re-activation of pluripotency associated genes, promoter loci of pluripotency markers were examined for epigenetic modifications. **Figure 4.26** depicts methylation patterns of CpG dinucleotides interspersed along the promoter regions of

Nanog, Rex1 and Oct4. Comparisons of methylation patterns disclose a close relationship between RARβ-iPSCs and ESCs, but large disparities between RARβ-iPSCs and MEFs. These findings highlight restructuring of transcription machinery during RARβ-assisted reprogramming, where pluripotency associated factors were modified at transcript, protein and epigenetic levels.

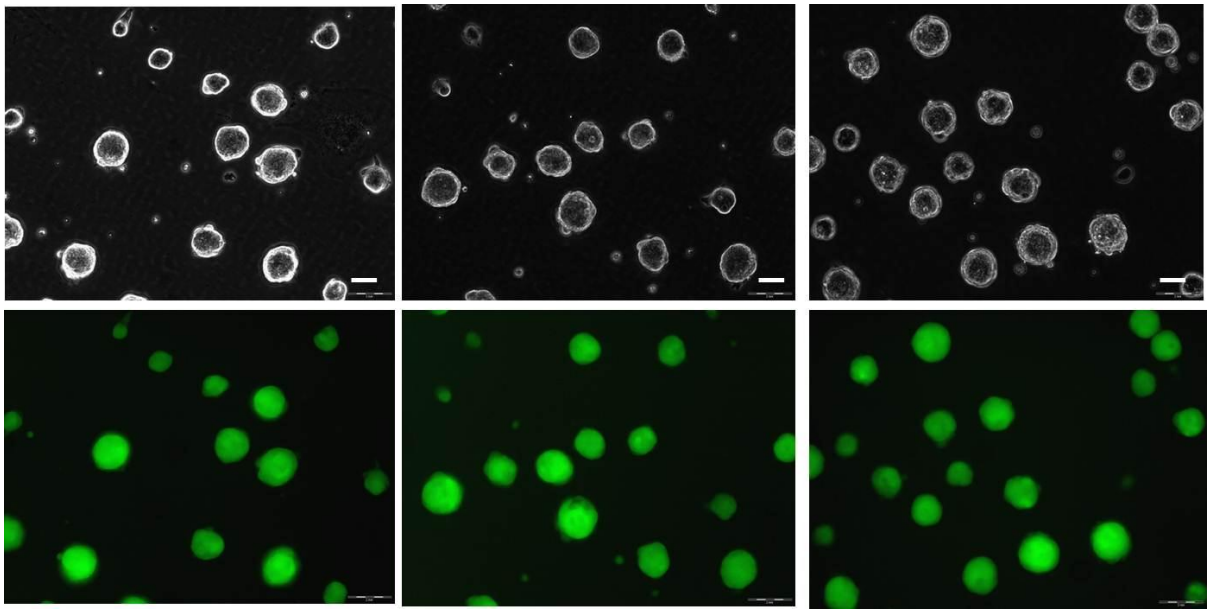


**Figure 4.22 Rex1 expressing RARβ-iPSCs** RARβ-iPSC colonies generated from the introduction of PB-TRE-CKS (1μg), RARβ (1μg), CAG-PBase (2μg) and PB-CAG-rtTA (1μg) were expanded and maintained in the absence of doxycycline, in 2i+LIF, on gelatin coated plates. As the starting MEFs contain a reporter cassette at the endogenous Rex1 locus, the images depict the activation of Rex1, as observed by the presence of GFP signals. Scale bar: 100μm.

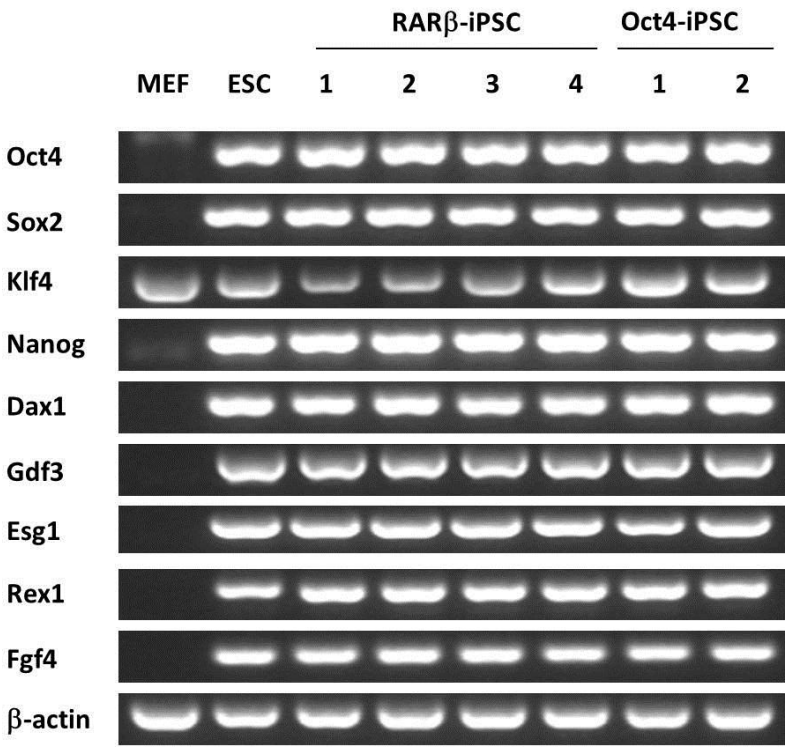



**Figure 4.23 Expression levels of pluripotency markers in RARβ-iPSCs** RT-PCR of the various pluripotency markers was performed in MEF (lane1), ESCs (lane2), iPSCs generated using c-Myc, Klf4, Sox2 and RARβ (lanes 3-6), and iPSCs generated using c-Myc, Klf4, Sox2 and Oct4 (lanes 7 and 8). Expression levels were described qualitatively by the intensity of the amplified products viewed using an agarose gel. The array of pluripotency markers tested is labelled to the left of the image.

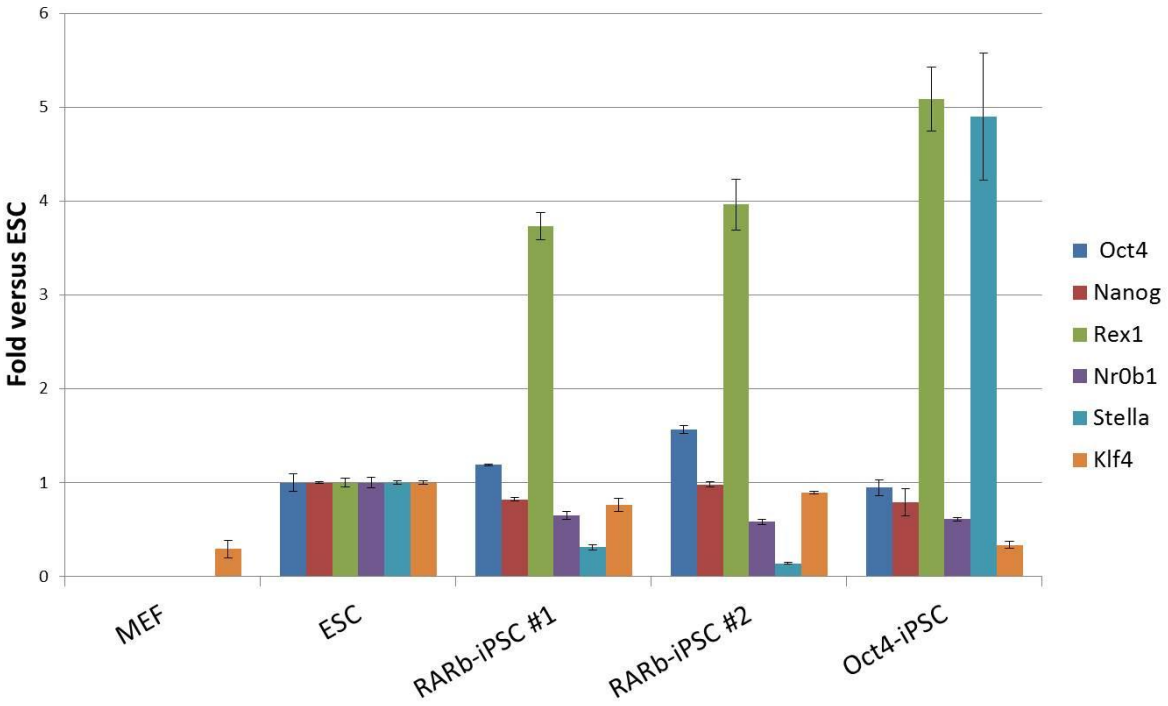



**Figure 4.24 Quantitative expression levels of pluripotency markers in RARβ-iPSCs** q-PCR of the various pluripotency markers was performed in MEF (lane1), ESCs (lane2), RARβ-iPSCs (lanes 3 and 4), and Oct4-iPSCs (lanes 5). Expression level of each pluripotency marker was normalised to GAPDH amounts and ESC expression levels.

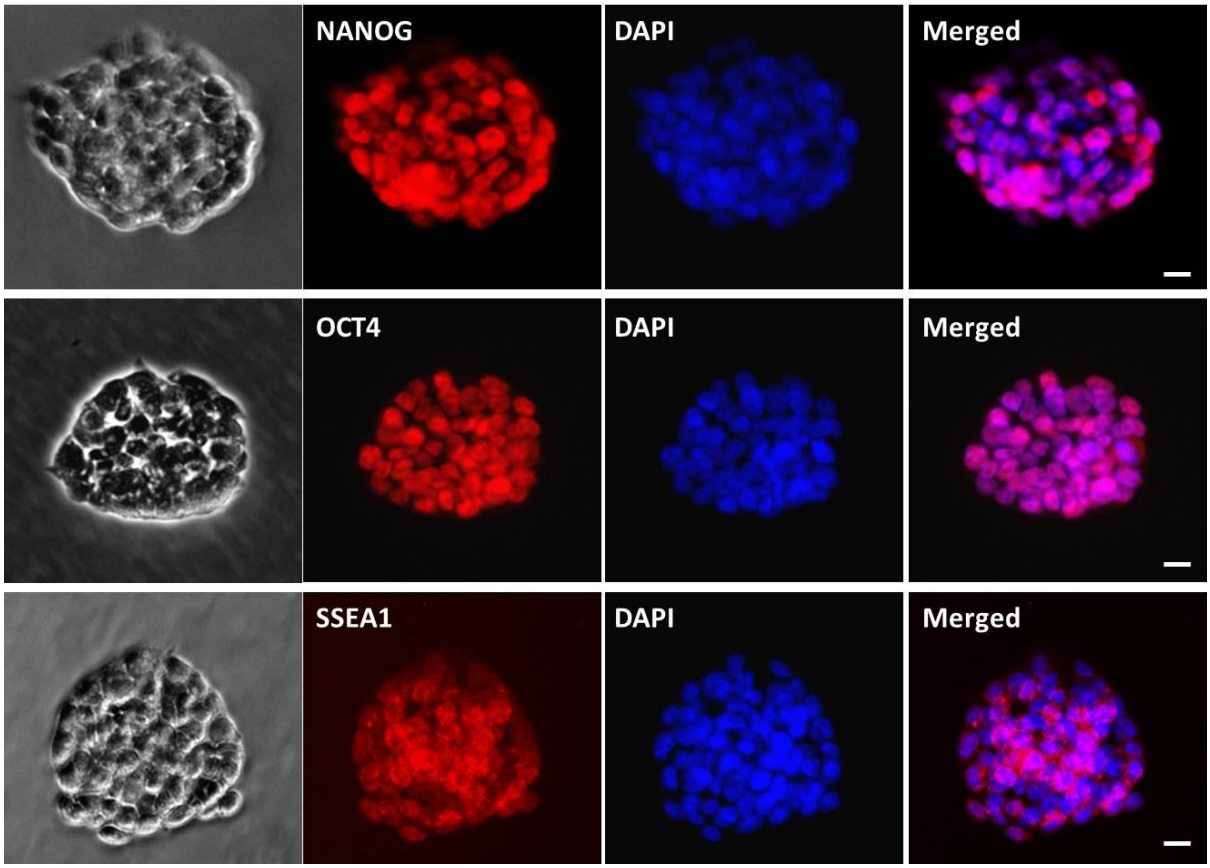



**Figure 4.25 Immunostaining of pluripotency markers in RARβ-iPSCs** Doxycycline independent RARβ-iPSCs were fixed and stained with antibodies against, Nanog (Top), Oct4 (Middle) and SSEA-1 (Bottom). Images of representative colonies are shown, displaying the respective fluorescence, brightfield and DAPI stain. Scale bar: 50μm.

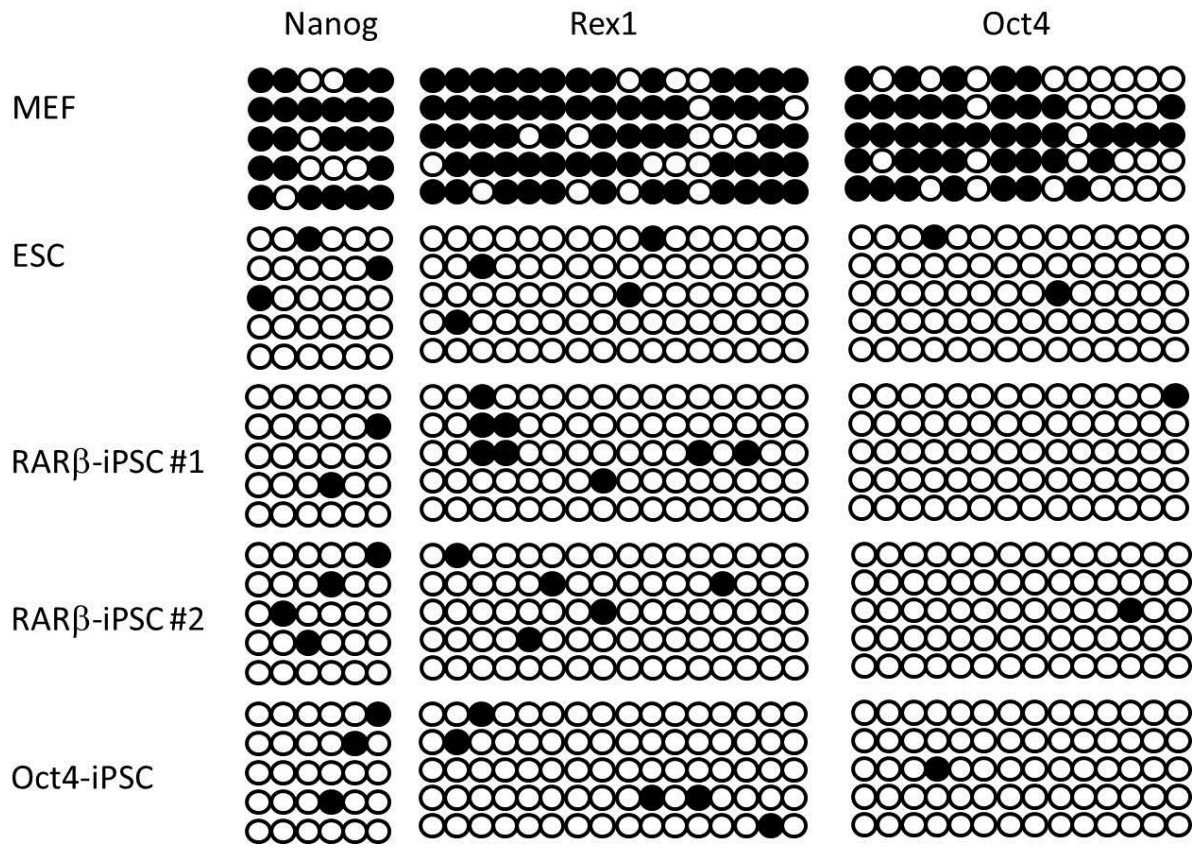



**Figure 4.26 Analysis of methylation at promoter loci in RARβ-iPSCs** DNA was obtained from MEF, ESCs, two RARβ-iPSC clones and Oct4-iPSCs and treated with sodium bisulphite. PCR directed at promoter loci of Nanog (LEFT), Rex1 (CENTRE) and Oct4 (RIGHT) were performed. Amplified products were analysed and summarised in the diagram. Each circle represents a CpG dinucleotide and five methylation patterns are displayed for each promoter loci, as indicated by the five rows. 6, 19 and 14 CpG dinucleotides were examined at the Nanog, Rex1 and Oct4 promoters respectively. A shaded circle indicates a methylated site, whereas an unshaded circle indicates an unmethylated site.

Having determined that pluripotency markers were re-activated in RARβ-iPSCs, these iPSCs were further examined for their ability to contribute to the three germ layers. This acted as a functional test, assessing the pluripotent potential of iPSCs from a separate perspective. To achieve this, LIF was withdrawn from the growth media to form embryoid bodies. By providing a platform for adherence, embryoid bodies competently differentiated into the three germ layers and were visualised through immunofluorescence methods using antibodies that recognise lineage-specific markers (**Figure 4.27**).

Having explored the differentiation capacity of RARβ-iPSCs in an artificial setting, it is pivotal to recapitulate these findings *in vivo*. This can be determined through the implantation of iPSCs into mouse blastocysts or subcutaneous flanks of immune-compromised mice to derive chimeras or teratomas respectively. As described earlier, it is crucial to ensure that the transgenes are switched off and not expressed. Attempts at detecting the transgene expression levels in two RARβ iPSC lines revealed that the transgenes were largely silent in the absence of doxycycline (**Figure 4.28**).

It was essential that no gross chromosomal abnormalities were present before RARβ-iPSCs could be innoculated into mice for determination of their pluripotent nature. Inspection of the genomic integrity in four RARβ-iPSC lines was performed using fluorescence *in situ* hybridisation. **Figure 4.29** demonstrates that RARβ-iPSCs harbour a normal set of chromosomes and do not exhibit detectable anomalies. This was reproduced in all four cell lines analysed.

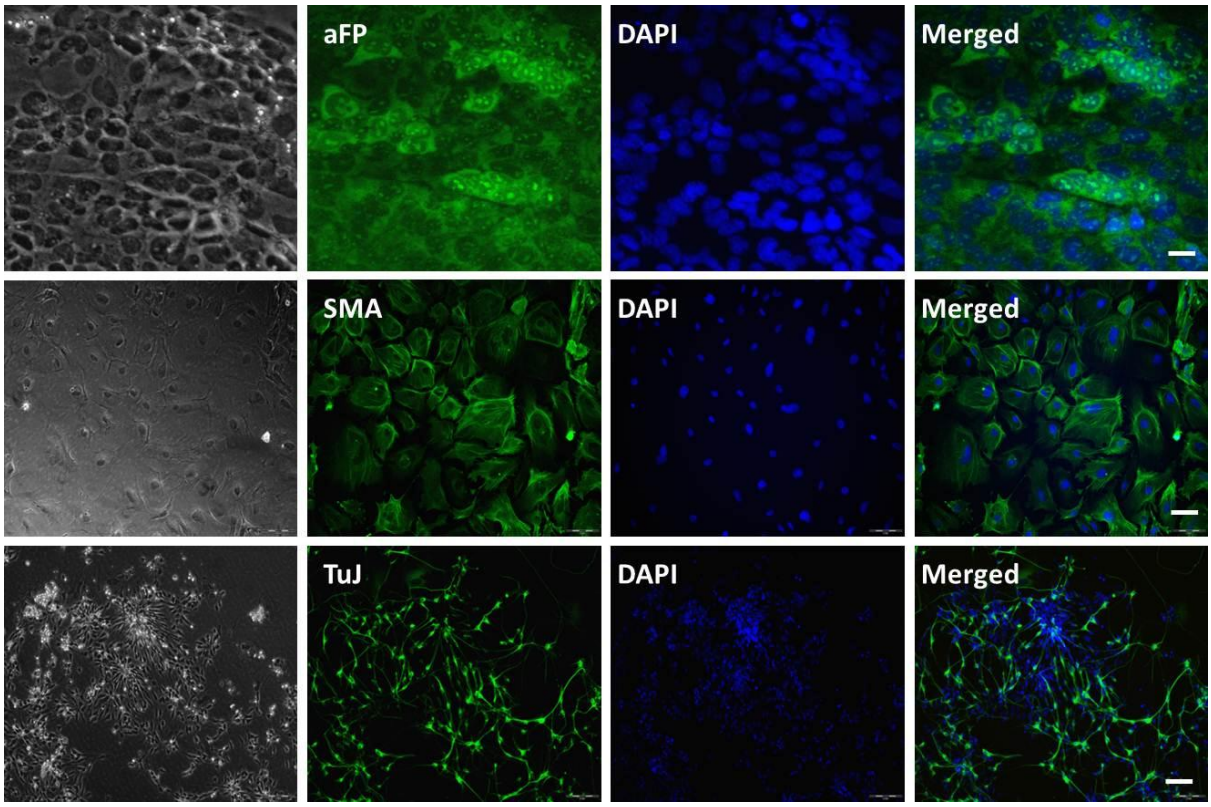



**Figure 4.27 RARβ iPSCs can differentiate into three germ layers *in vitro*** RARβ-iPSCs were differentiated into the three germ layers in vitro using various culture conditions. Differentiated cells representing the endodermal, mesodermal and ectodermal lineages are immunostained against alpha feto-protein (Top), smooth muscle actin (Middle), and tubulin (Bottom) respectively. The markers are labelled in green, DAPI which stains the cell nuclei is labelled in blue. Scale bar: 50μm.

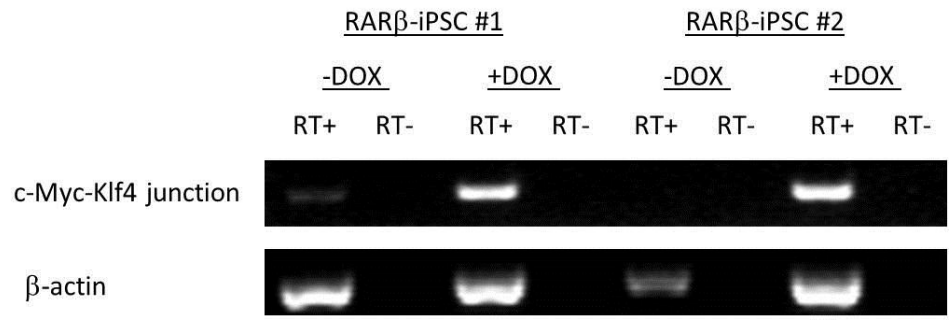



**Figure 4.28 Transgenes are silent in the absence of doxycycline** Two RARβ-iPSC clones were cultivated in the presence and absence of doxycycline. RNA was extracted from these cells and (i) converted to cDNA using reverse transcriptase (RT+) or (ii) tested for contamination with genomic DNA in the absence of RT (RT-). Amplification of the junction between c-Myc and Klf4 was performed to determine the expression of the transgenes.

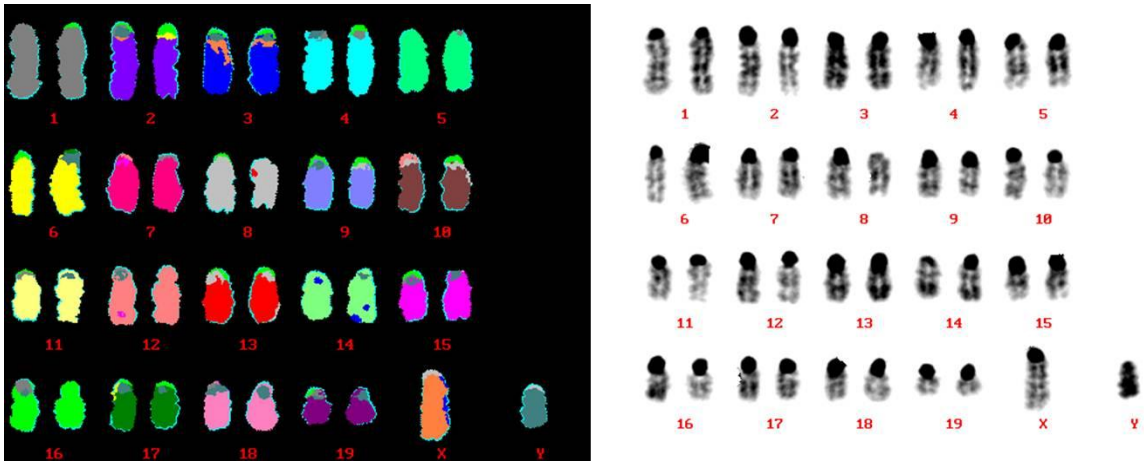



**Figure 4.29 Karyotype analysis of RARβ-iPSCs** RARβ-iPSCs were mitotically arrested and metaphase spreads were achieved. Using Fluorescence *in situ* hybridisation (FISH), each chromosome was accounted and analysed. The array of colours depicts different chromosomes.

1 x $10^6$ RARβ-iPSCs were subcutaneously injected into the left and right flanks of immune-compromised mice to determine their capability to contribute to the somatic tissues *in vivo*. After two weeks, hard masses were observed and extracted. Histological examination of the tumour masses was performed and **Figure 4.30** provides a representation of the presence of mesodermal, endodermal and ectodermal lineages.

Serving as the benchmark to test for naïve pluripotency, RARβ-iPSCs were injected into mouse blastocysts to observe for chimera formation. Chimeras of various percentages (10-50%) were derived, indicating the ability of RARβ-iPSCs to participate in normal murine development. Crosses between chimeras and wildtype mice displaying white coats of fur led to the generation of pups carrying black coats of fur (**Figure 4.31**), signifying the ability of RARβ-iPSCs to contribute to an entire organism. Similar to BNC2, RARβ has been proven to be competent in generating iPSCs without ectopic requirement of Oct4, and resultant iPSCs were capable of contributing to the mouse germline and somatic tissues, exemplifying naïve pluripotency.

Amid the attempts to validate BNC2, RARβ and RORα as candidate genes which are capable of replacing ectopic Oct4 during reprogramming, two factors remain unvalidated: Cald1 and DCAF5. Although these genes are not obvious candidates, they may unwittingly provide insightful findings in the mechanisms behind reprogramming. Therefore, it will be intriguing to investigate the ability of Cald1 and DCAF5 in acting as ectopic Oct4 substitutes in the acquisition of pluripotency.

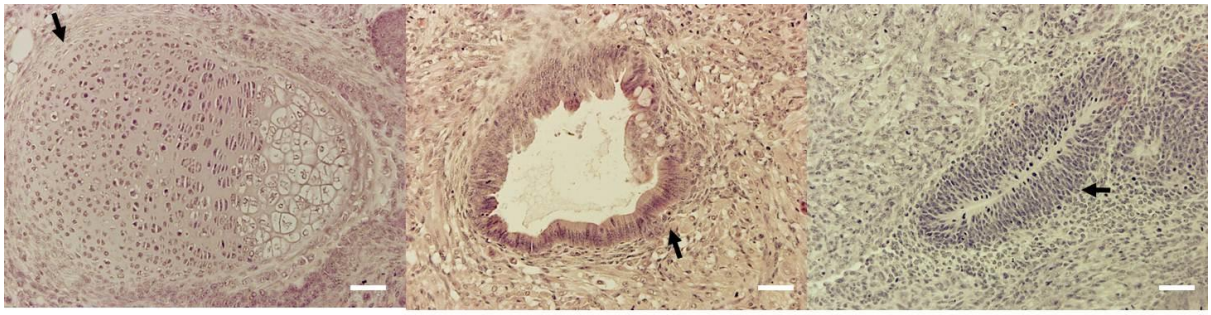



**Figure 4.30 RARβ-iPSCs can differentiate into three germ layers *in vivo*** RARβ-iPSCs were injected subcutaneously into both flanks of immune-compromised NOD.Cg-Prkdc$^{scid}$ Il2rg$^{tm1Wjl}$/SzJ (NSG) mice. After 2 weeks, tumours were observed. Histological analyses of sections of these tumours are shown. Differentiation into endodermal, mesodermal and ectodermal lineages are displayed and labelled. The arrows indicate the position of the differentiated structure. Scale bar: 50μm.

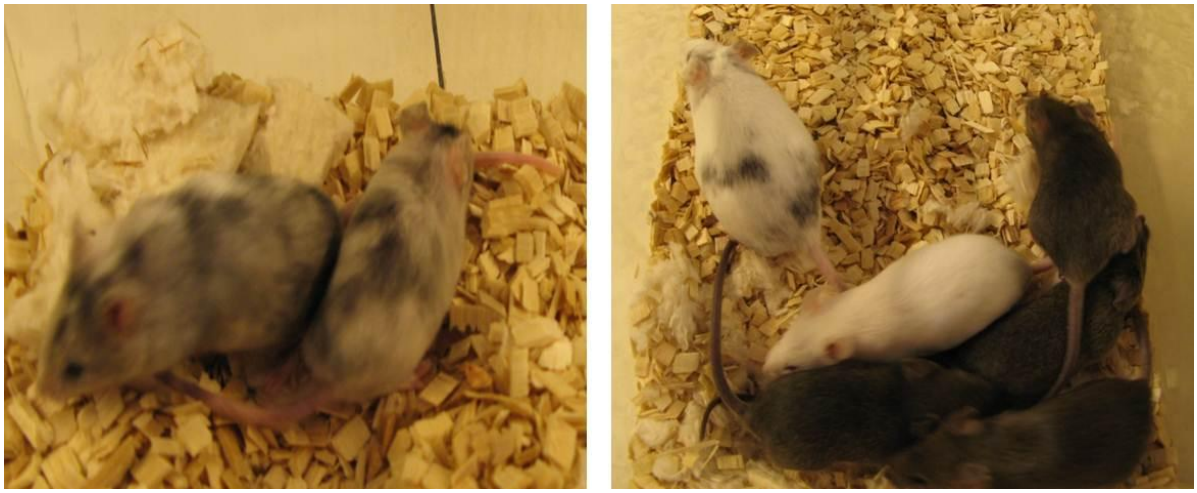


**Figure 4.31 RARβ-iPSCs are capable of generating chimeras and contributing to the germline** (LEFT) Chimeras (40% and 50%) were generated through the injection of RARβ-iPSCs into mouse blastocysts. (RIGHT) Crossing between a 10% chimera and albino CALB mouse results in the generation of pups carrying a black fur coat.

## 4.4 Discussion

### 4.4.1 Variable reprogramming efficiencies

This chapter discusses the ability of BNC2 and RARβ in triggering dedifferentiation events in the absence of exogenous Oct4. From validation experiments, it is evident that gene dosage plays a role in determining reprogramming efficiencies.

Disparate promoter systems were demonstrated to trigger varying reprogramming efficiencies, suggesting that the generation of iPSCs is sensitive to discrepancies in reprogramming factor dosage. BNC2 was described to generate iPSCs efficiently (0.28%) via the use of a CAG-promoter driven system, and low reprogramming efficiencies (0.0002%) when a doxycycline-inducible system was employed. It was subsequently observed that iPSC colonies developed using a doxycycline-dependent approach exhibited large numbers of integration sites. This is a stark contrast to the generation of iPSCs through the conventional set of reprogramming factors CKS-O where CAG-promoter and doxycycline-inducible delivery methods gave rise to reprogramming efficiencies of 0.3% and 0.15% respectively. Similarly, RARβ exhibited mild differences in reprogramming efficiencies when distinct promoter systems were employed, where comparisons between CAG-promoter and doxycycline-inducible delivery methods resulted in reprogramming efficiencies of 0.12% and 0.02% respectively.

The stoichiometric ratio of reprogramming factors and the resultant number of iPSC colonies have been described to be closely related (Carey et al., 2011; Tiemann et al., 2011). High ectopic expression of Oct4 has been associated with higher reprogramming efficiency and naïve pluripotency. As such, it is within anticipation that modifying expression levels of Oct4 and its substitutes through the application of different promoters would generate varying number of iPSC colonies.

As RARβ and BNC2 are candidate genes that were identified through piggyBac-assisted mutagenesis, it is plausible that truncated gene products that were generated during the screen had a functional advantage during the acquisition of pluripotency and improve reprogramming efficiencies. As such, close examination of insertion sites identified during the screen may illuminate the mechanisms employed by the candidate genes to initiate reprogramming events in the absence of ectopic Oct4.

Analysis of the insertion sites found within RARβ revealed that one integration site led to the disruption of the ligand independent transactivation domain, whereas the other was observed in the opposite direction to RARβ transcription. As the CAG promoter exhibits enhancer properties, it is conceivable that the latter integration posed as an enhancer to RARβ transcription. RARβ mutants have been described in humans and *Xenopus Laevis*, conceived through the mutation of the ligand binding domain, towards the C-terminal end of the protein (Shen et al., 1993; van der Wees et al., 1998). However, as both the ligand binding domain and DNA binding domain are encoded by the same exon, it is not possible to affect the ligand binding domain without affecting the DNA binding domain with the design of the screen. Nevertheless, analysis of information retrieved from the screen may unearth a relationship between the nature and function of the insertion. Alternatively, it will be interesting to compare the potency of mutant and wildtype RARβ in the generation of iPSCs in the absence of exogenous Oct4.

On a different note, the genomic locus encoding BNC2 has been described to generate 90,000 mRNA transcripts using bioinformatics prediction tools (Vanhoutteghem et al., 2007). This was attributed to permutations between multiple promoters driving combinations of minor and/or major exons. Although an equivalent phenomenon has not been described in the mouse isoform, it may prove difficult to envisage the significance behind the insertion sites obtained from the screen. However, a representative model has been delineated, where the fifth and sixth major exons encode for most of the critical domains, including three pairs of zinc fingers, nuclear localisation signal and serine stripe.

As all three insertions are within the first and second exons, it is unlikely that the resultant truncation product abolishes the domains necessary for BNC2 function.

### 4.4.2 RORα and its inability to trigger pluripotency

RORα was a candidate gene sifted from the large number of insertion sites by virtue of possessing more than one independent transposon integration. However, introduction of RORα, c-Myc, Klf4 and Sox2, driven by a CAG-promoter did not give rise to iPSC colonies. There are several reasons which could explain the poor validation of RORα. First, RORα alone may not be sufficient to act as an Oct4 substitute, and co-operating partners of RORα may be required to elicit reprogramming events. Closer inspection of additional integration sites in clones where RORα was identified could shed light on synergistic factors or pathways that facilitate the reprogramming process. Second, the genetic screen which identified RORα incorporates the use of a strong constitutive CAG promoter. As CAG is a hybrid consisting of a cytomegalovirus (CMV) early enhancer element and chicken beta-actin promoter, it can function as both a promoter and enhancer. As a result, it can affect neighbouring genes (Hasegawa et al., 2002). To this effect, it is plausible that genes adjacent to RORα were critical in generation of iPSCs during the performance of the screen. RORα lies in a gene-rich region and is in close proximity to additional genes involved in developmental pathways. Third, the genomic locus encoding for RORα spans 0.7Mb, considerably larger than an average gene locus of 0.1Mb, as predicted by assuming equal spatial distribution of 30,000 genes over 3Gb of the genome. The large size implicates high probabilistic occurrences of transposon integrations. With this in mind, although two independent insertions were identified in a span of 0.7Mb, this could be due to chance and exist as a passenger mutation, instead of a driver mutation.

### 4.4.3 Genomic instability in BNC2-iPSCs

In this thesis, two BNC2-iPSC lines were characterised and genomic aberrations were observed in both lines. One of the two lines portrayed karyotypic abormalities and exhibited centromeric fusion at chromosome 11. Surprisingly, these cells displayed

pluripotent properties and contributed to the mouse germline. Interestingly, karyotype analysis of splenocytes obtained from chimeric mice derived from BNC2-iPSCs showed a normal phenotype. Moreover, genomic material of mice arising from $F_1$ generation was demonstrated to possess the reprogramming transgenes corresponding to PB-TRE-CKS and PB-TRE-BNC2, reinforcing the notion that BNC2-iPSCs contributed to the mouse germline. This phenomenon can be attributable to heterogeneity within the BNC2-iPSC clone where a small subset of BNC2-iPSCs harbours a normal set of chromosomes. Over the course of fluorescence *in situ* hybridisation, twenty metaphase spreads were scrutinised for gross aberrations. In the event that less than 5% of the cell population display normal karyotypic properties, it is unlikely to be captured during the analysis. It is possible that a karyotypically normal iPSC was selected during the injection process and incorporated into the mouse blastocyst. This postulation is reinforced by a low proportion of chimeric mice that exhibited germline competence (**Figure 4.32**), In contrast, the extent of which chimeric mice generated from RARβ-iPSCs exhibited germline competence was four-fold higher (**Figure 4.32**) and resembled efficiencies reported in the literature (Hansen et al., 2008).

In addition, chromosome 11 has been described to play pivotal roles in haematopoiesis, craniofacial and cardiovascular development and fertility, where mutations incurred by N-ethyl-N-nitrosourea (ENU) treatment resulted in developmental roadblocks (Kile et al., 2003). Rearrangements of regions spanning from 3-4cM to 100kb in chromosome 11 has also been depicted to result in developmental abnormalities or lethality (Liu et al., 1998). These pieces of evidence reveal the importance of genomic integrity at chromosome 11, reinforcing the notion that a subset of the BNC2-iPSCs do not display chromosomal abnormalities and contributed to the mouse germline.

| iPSC | Number of embryos injected | Number of pups born | Number of Chimeras | Number of Chimeras which possessed germline competence |
|---|---|---|---|---|
| BNC2-iPSCs | 351 | 71 | 13 | 1 (7.7%) |
| RARβ-iPSCs | 85 | 17 | 3 | 1 (33.3%) |

**Figure 4.32 Table indicating the number of mice which exhibited germline transmission of BNC2-iPSCs and RARβ-iPSCs** BNC2-iPSCs and RARβ-iPSCs were injected into blastocysts of albino wildtype mice to test for chimera contribution. The table reflects the number of embryos that were injected with the respective iPSCs, resultant number of live pups born, number of chimeras as indicated by coat colour, and number of chimeras demonstrating germline transmission (percentage in brackets indicate the proportion of chimeras born that displayed germline competence).

### 4.4.4 Conclusion

This chapter serves as a validation process to determine competent substitutes of exogenous Oct4 during the initiation of reprogramming events. Five candidate genes were identified from a transposon-mediated genetic screen. By testing a subset of three genes, RARβ and BNC2 were demonstrated to instigate the acquisition of pluripotency without the ectopic introduction of Oct4. Resultant cells were able to fulfil requirements of pluripotency, establishing remodelling and expression levels of pluripotency markers and displaying the potential to contribute to somatic tissues and the mouse germline. These findings validate the dependability of the genetic screen described in **Chapter 3**. Armed with the knowledge of novel genes involved in the reprogramming process, molecular mechanisms behind RARβ will be dissected in **Chapter 5**.

# CHAPTER 5 : RETINOIC ACID RECEPTORS and the REGULATION OF OCT4 EXPRESSION

## 5.1 Introduction

### 5.1.1 Retinoic Acid signalling during Development

In **Chapter 4**, I described that BNC2 and RARβ can effectively replace ectopic requirement of Oct4 during reprogramming. Although the reprogramming efficiency of BNC2 was superior to RARβ, co-operation between RARβ, c-Myc, Klf4 and Sox2 in the reprogramming cocktail led to the generation of iPSC colonies expressing endogenous Rex1 in a homogenous manner. Together with evidence describing that RARs can associate with Oct4 regulatory elements and affect Oct4 expression (Ben-Shushan et al., 1995; Wang et al., 2011c), it is tempting to speculate that Retinoic Acid (RA) signalling may play a role in the reprogramming process. With this in mind, **Chapter 5** aims to unravel the mechanisms behind RARβ and its family members in their abilities to influence the acquisition of pluripotency.

Retinoic Acid (RA) signalling has been illustrated in developmental processes in a multi-dimensional manner. Identified as the first candidate diffusible morphogen in vertebrates, attempts to dissect the role of RA during development have been inundating (Thaller and Eichele, 1987). Initial exemplification of the obligatory need of Vitamin A, a precursor of RA, during embryonic development (Hale, 1933; Warkany and Schraffenberger, 1946; Wilson and Warkany, 1948; Wilson et al., 1953) led to the discovery that specific elimination of retinoic acid during development leads to malformation of the neural crest, ocular and nervous system (Dickman et al., 1997; White et al., 1998). Conversely, excessive amounts of retinoic acid have been demonstrated to initiate aberrations in embryonic development across species, from *Xenopus laevis* to zebrafish, mice and humans (Sive et al., 1990; Holder and Hill, 1991; Papalopulu et al., 1991; Morriss-Kay et al., 1991; Conlon and Rossant, 1992; Lammer et al., 1985).

From a canonical perspective, RA elicits a cascade of events through its association to Retinoic Acid Receptors (RARs). RARs are molecular components which recognise and bind to RA (Giguere et al., 1987; Petkovich et al., 1987; Benbrook et al., 1988; Brand et al., 1988; Krust et al., 1989; Zelent et al., 1989). There are three members in the RAR family, namely RAR-alpha (RARα), RAR-beta (RARβ) and RAR-gamma (RARγ). All three family members share features such as a ligand binding domain (LBD) that associates with RA and a DNA binding domain (DBD) which tethers to Retinoic Acid Response Elements (RAREs) delineated by the motif (A/G)G(G/T)TC(A/G) and its variants (reviewed by Chambon, 1996). Analagous to RA, RARs are essential during embryonic development. Due to close similarities between RARs, functional redundancy is apparent where mutations in single RARs do not instigate developmental defects. In contrast, simultaneous ablation of two RARs results in overt malformation and death *in utero* (Ghyselinck et al., 1997; Lohnes et al., 1994; Kastner et al., 1994; Mendelsohn et al., 1994).

With an affinity for RAREs, RARs reside in the nucleus and bind to cognate recognition motifs. In the absence of its ligand, RARs are postulated to associate with co-repressor complexes, thus inhibiting the initiation of transcription. Conversely, in the presence of its ligand, RA binds to the LBD of RARs and results in the dissociation of co-repressor complexes. This is ensued by the recruitment of co-activator modules and transcriptional machinery to prompt the transcriptional process (reviewed by Rochette-Egly and Germain, 2009).

Apart from the canonical signalling pathway, non-genomic responses of RARs have been delineated. Acting through kinase cascades, RARs have been demonstrated to activate p38 MAPK, MSK1 and PI3K (Alsayed et al., 2001; Gianni et al., 2002; Gianni et al., 2006; Bruck et al., 2009; Masia et al., 2007). These signalling events are independent of the transcriptional capabilities of RARs and have been attributed to two serine phosphorylation sites within the LBD and N-terminal domain (Bruck et al., 2009; Rochette-Egly et al., 1997; Bastien et al., 2000). In addition to non-genomic modes of

action, RARs can also achieve ligand independence. These incidences include co-operation with other receptors and epigenetic modifications. To this end, RARα and Estrogen Receptor (ER) have been demonstrated to co-occupy regulatory elements within the chromatin in an ER dependent manner (Ross-Innes et al., 2010). In addition, RARα has been described to influence transcriptionally permissive histone modifications at the promoter locus of Mest in the absence of RA (Laursen et al., 2012).

### 5.1.2 Retinoic Acid Receptors and Oct4 regulation

Amid the plethora of genomic interactions exhibited by RARs at its cognate binding sequences, RAREs have been elucidated within the cis-regulatory elements of Oct4 (Schoorlemmer et al., 1994; Pikarsky et al., 1994; Okazawa et al., 1991). The discovery of RAREs within the Oct4 regulatory elements was incited by the amalgamation of various pieces of information, namely (i) the importance of Oct4 during early development (Nichols et al., 1998) and (ii) the loss of pluripotency upon high doses of retinoic acid (Okamoto et al., 1990). With this in mind, our current understanding of the regulation of Oct4 expression, with primary focus on RA signalling, will be illustrated below.

#### 5.1.2.1 Regulation of Oct4 expression levels

The fleeting and obligatory expression of Oct4 during early stages of embryonic development has motivated the elucidation of its functional network. With the advent of sophisticated analysis methods such as chromatin immunoprecipitation coupled to sequencing, the transcriptional landscape during early stages of embryonic development has been feasible. Using mouse and human ESCs as a surrogate to study early development, transcriptional complexes associated with pluripotency have been unearthed (Boyer et al., 2005; Loh et al., 2006; Kim et al., 2008; Wang et al., 2006; Marson et al., 2008b; Chen et al., 2008). These analyses revealed the presence of a core transcriptional apparatus consisting of Oct4, Sox2 and Nanog which actively participates in the maintenance of pluripotency. Examination of the genomic regions bound by pluripotency-associated transcription factors at high resolution revealed that the distal enhancer of Oct4

depicts the densest binding locus. Transcriptional regulators which tether to this region include Stat3, Klf4, Smad1, Esrrb, Nanog, Sox2, Tcf3 and c-Myc (Chen et al., 2008; Marson et al., 2008b). In addition, p300 binding and activating methylation marks at lysines 4 and 36 of histone 3 are heavily represented in the enhancer, promoter and transcribed regions of Oct4 within ESCs (Marson et al., Chen et al., 2008; Schnetz et al., 2010). This strengthens the notion that Oct4 regulatory elements are heavily occupied in pluripotent cells, hence unravelling of the regulatory mechanisms at these elements will illuminate the molecular switches involved in the acquisition of induced pluripotency and differentiation.

#### 5.1.2.2 Retinoic Acid Receptors mediate Oct4 expression

To initiate the understanding of Oct4 transcriptional regulation, phenomenon described to accompany changes in Oct4 expression levels was employed as a learning platform. To this end, RA-induced differentiation which leads to a decrease in Oct4 expression levels pioneered initial studies and sought to determine the effect of RA signalling on Oct4 regulation (Schoorlemmer et al., 1994; Pikarsky et al., 1994; Okazawa et al., 1991).

Using a combination of DNase footprinting, chloroamphenicol acetyltransferase (CAT) assays and computational prediction to study the effects of RA-induced differentiation, three regions that encompass RAREs were disclosed. They were subsequently labelled as the distal enhancer, proximal enhancer and proximal promoter, positioned approximately 2kb, 1kb and 100bp upstream of the Oct4 transcriptional start site respectively. These elements were portrayed to be active in EC cells and murine embryos, and inactive in the presence of high doses of retinoic acid (Schoorlemmer et al., 1994; Pikarsky et al., 1994; Okazawa et al., 1991; Yeom et al., 1996).

The three regulatory elements control the expression of Oct4 at different stages of development. The distal enhancer drives Oct4 expression within the pre-implantation epiblast, whereas the proximal enhancer is responsible for inciting Oct4 expression in the

post-implantation epiblast (Yeom et al., 1996). EpiSCs reflect a sub-population from the late epiblast and rely on the proximal enhancer for expression of Oct4, whereas derivatives from the early epiblast depend on the distal enhancer for Oct4 expression (Tesar et al., 2007; Guo et al., 2010; Han et al., 2010a). Despite dissimilar regulatory circuits, the expression levels of Oct4 are similar between EpiSCs and ESCs (Tesar et al., 2007). Although the control of Oct4 expression has been delineated in various pluripotent states, regulatory elements which govern Oct4 expression during the acquisition of pluripotency have not been explored.

The RARE situated at the proximal promoter (RAREoct) has been described to be pivotal in the regulation of Oct4 expression through competitive binding between Coup-TFI, Coup-TFII and the RARs (Ben-Shushan et al., 1995). Coup-TFI and Coup-TFII have the capacity to tether to RAREoct and repress Oct4 expression, whereas increasing amounts of RARs counteract this phenomenon by displacing negative regulators and inciting Oct4 expression. Additional regulatory components have since been elucidated to adhere to RAREoct (Fuhrmann et al., 1999; Fuhrmann et al., 2001; Barnea and Bergman, 2000). For example, an SF1 (Nr5a1) recognition site was identified adjacent to RAREoct and the application of CAT assays reveals that a combination of SF1 and RARα or RARγ results in enhanced transcriptional activity at RAREoct (Barnea and Bergman 2000). As SF1 shares similar recognition sites to LRH1 (Nr5a2), it was further delineated that LRH1 competently occupies RAREoct (Gu et al., 2005) and co-operation between LRH1 and RARγ instigates rapid and efficient generation of iPSCs (Wang et al., 2011c). As exploration of the role of RARs in the acquisition of induced pluripotency is in its infancy, it will be intriguing to extend these findings to RARβ and determine its ability to function as a surrogate for ectopic Oct4 during reprogramming.

### 5.1.3 Chapter Aim

This chapter aims to delineate the mechanisms behind RARβ in its ability to replace ectopic Oct4 during the initiation of reprogramming events. As RARs have been

implicated in the regulation of Oct4 through their association with Oct4 cis-regulatory elements, these findings serve as a basis to understand the role of RARβ in negotiating the requirement for ectopic Oct4 during the acquisition of pluripotency and may offer a distinct perspective of the reprogramming process.

## 5.2 Results

### 5.2.1 Retinoic Acid enhances reprogramming efficiency

As an initial step to dissect the role of RARβ in the acquisition of pluripotency, its ligand, all trans-RA (ATRA), was assessed for its ability to imitate the effects of RARβ and elicit an effect on the reprogramming process. To this end, ATRA was introduced to the conventional reprogramming cocktail of Oct4, c-Myc, Klf4 and Sox2. High concentrations of RA have been well-established to incite differentiation into neural lineages (Okada et al., 2004; Schuldiner et al., 2001). As such, low concentrations of ATRA were employed. 0.1nM and 1nM of ATRA were introduced at the initial stages of reprogramming, and number of colonies generated acted as a surrogate to determine reprogramming efficiencies. To heighten the sensitivity of the screen, high passage MEFs were utilised and few colonies were observed when the conventional reprogramming cocktail was introduced on its own (**Figure 5.1**). This allows clear delineation of any effect on reprogramming efficiencies. Interestingly, addition of 1nM ATRA escalated the number of iPSC colonies generated, whereas 0.1nM of ATRA did not yield a similar effect. A similar phenomenon was observed when a separate form of RA, 9-cis RA, was used. 9-cis RA has been described to be reciprocal ligand recognised by RXRs (Heyman et al., 1992; Levin et al., 1992), suggesting that RXRs may also be involved in reprogramming events. The experiment was replicated three times and a similar effect was observed. **Figure 5.2** illustrates a quantitative view of the colony numbers generated by the addition of 1nM RA. These results ascertain the role of RA signalling in the reprogramming and provide supporting evidence for the ability of RARβ to competently act as a substitute for exogenous Oct4 during reprogramming.

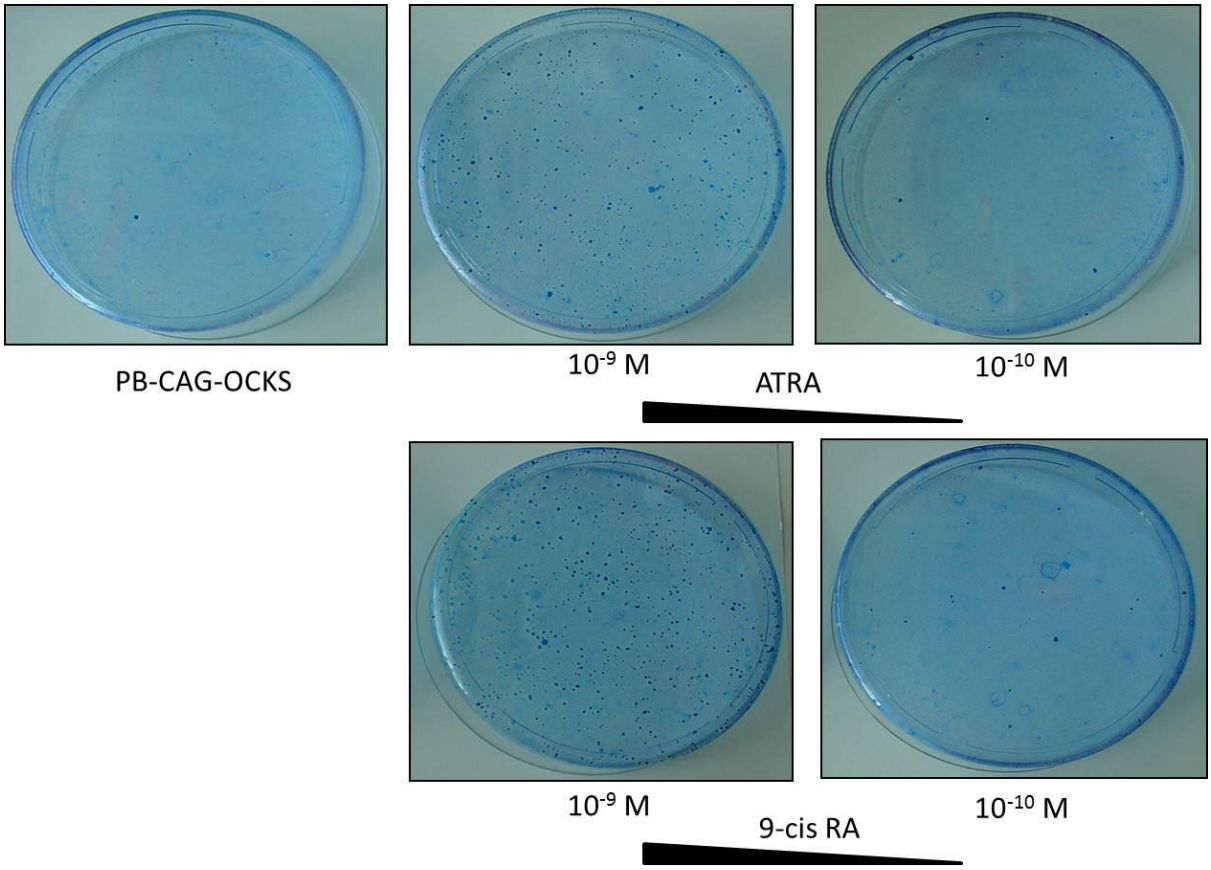



**Figure 5.1 Addition of all-trans retinoic acid affects reprogramming efficiencies** To assess the influence of all-trans retinoic acid (ATRA) on reprogramming, Oct4, c-Myc, Klf4 and Sox2 were transfected into MEFs and electroporated MEFs were seeded onto a mitotically inactive feeder layer in the presence of ATRA . Cells were exposed to the indicated ATRA concentrations for 5 days. 21 days after reprogramming, the cells were stained with bromophenol blue for visualisation.

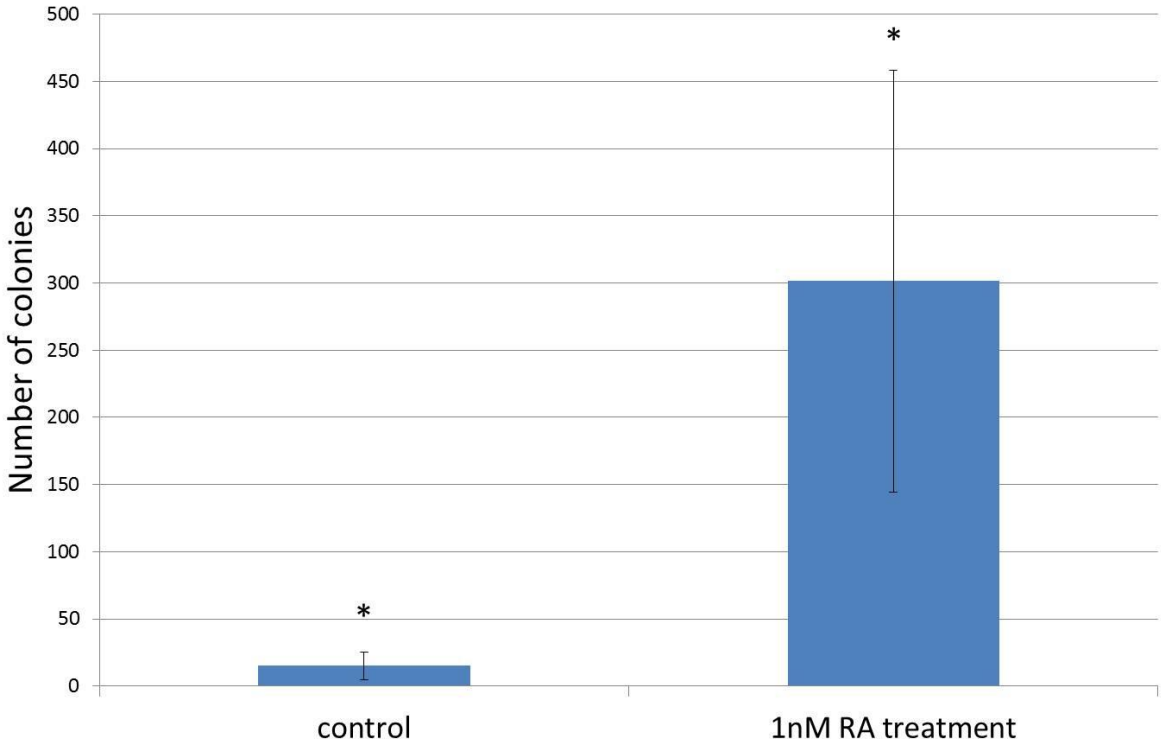



**Figure 5.2 Addition of 1nM all-trans retinoic acid improves reprogramming efficiencies** Tripicates were conducted to study the effect of 1nM ATRA on reprogramming efficiencies. 1nM ATRA was introduced for 5 days to MEFs which had been transfected with Oct4, c-Myc, Klf4 and Sox2. 21 days after transfection, cells were stained with bromophenol blue and counted. The numbers of observed colonies are reflected in the bar chart above where the error bars signify the standard deviation calculated from the values obtained from 3 separate experiments. Statistical analysis was performed using a student's T-test. *$p<0.05$

### 5.2.1 RARβ binds to the enhancer and promoter regions of Oct4

Having delineated that the natural ligand of RARβ can enhance reprogramming efficiencies, it was pertinent to understand the mechanisms behind the ability of RARβ in acting as a substitute for exogenous Oct4 in the acquisition of induced pluripotency. Recent findings have highlighted the association of RARα and RARγ to the distal enhancer element of Oct4 (Delacroix et al., 2010). As RARβ was omitted from the study, it is intriguing to determine if RARβ possesses similar qualities. To this end, chromatin immunoprecipitation (ChIP) was performed using an antibody specific to RARβ. As RARβ is not highly expressed in ESCs, MEFs were used as a starting material. Fragmented genomic regions which were tethered to RARβ were eluted and analysed. Using GenePalette to identify consensus motifs recognised by the RAR family, eleven potential RAREs were annotated 10kb upstream of the Oct4 transcriptional start site (TSS) (**Figure 5.3**). Primers against each of the eleven sites were designed and introduced to the ChIP eluate for amplification. **Figure 5.4** illustrates quantification of the amplified products corresponding to eleven putative RAREs preceding Oct4 TSS. RAREs between 100bp and 2.17kb upstream of Oct4 TSS displayed positive association to RARβ. This region comprises of the sites which have been previously demonstrated to show susceptibility to control by RARs, namely the distal enhancer, proximal enhancer and proximal promoter (RAREoct) (Yeom et al., 1996). To inject an additional level of validation, DNA immunoprecipitated using the abovementioned RARβ antibody was sequenced using primers corresponding to 2.17kb upstream of Oct4 TSS, exhibiting identity to the distal enhancer of Oct4. (**Figure 5.5**).

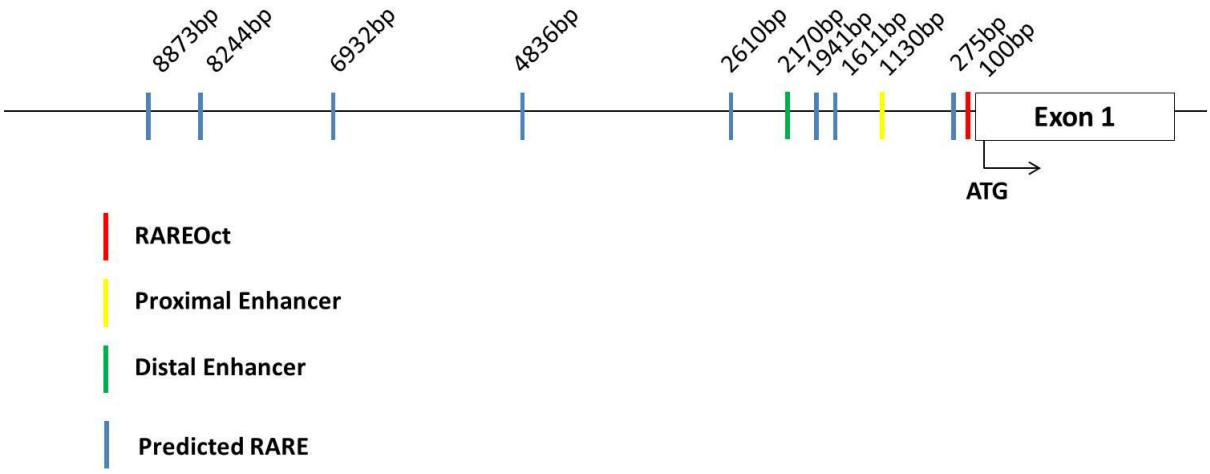



**Figure 5.3 Predicted Retinoic Acid Response Elements (RAREs) upstream of Oct4** RAREs which fulfil the consensus motif (A/G)G(G/T)TC(A/G) were identified using GenePalette. Predicted RAREs are displayed in blue boxes and their distances from the transcriptional start site are indicated. The boxes in red, yellow and green depict RAREs representative of the proximal promoter (RAREoct), proximal enhancer and distal enhancer respectively.

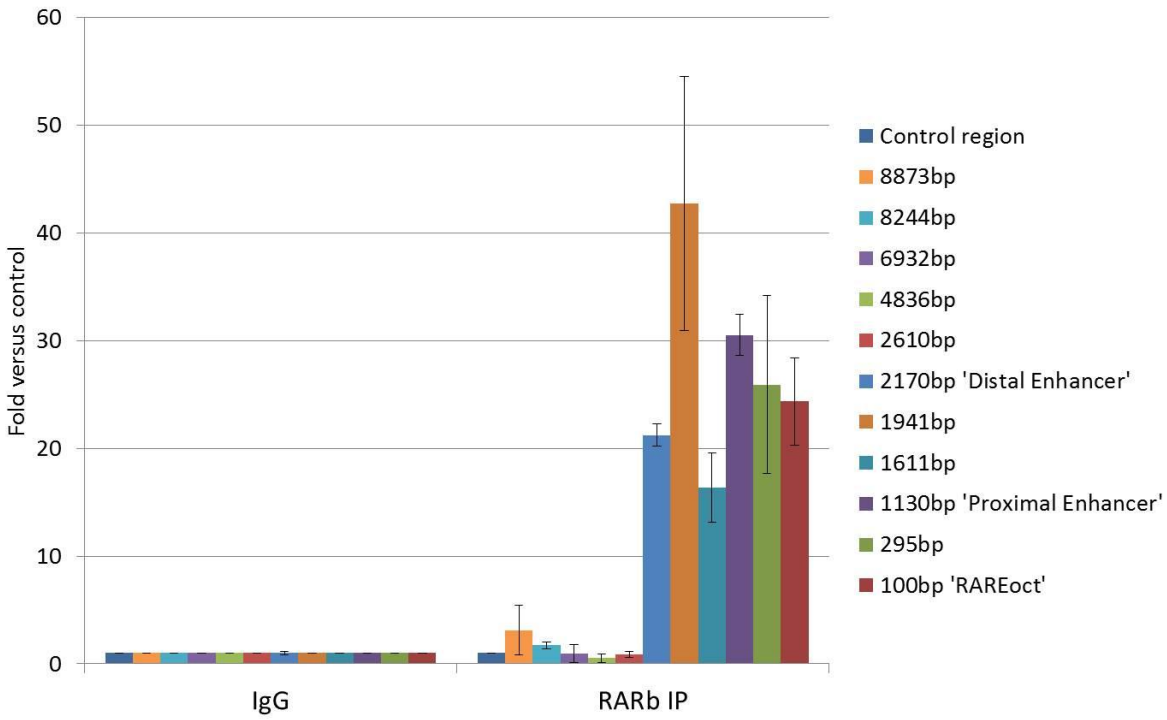



**Figure 5.4 RARβ binds to RAREs up to 2kb upstream of Oct4** Chromatin Immunoprecipitation (ChIP) was performed using antibodies against RARβ in MEFs. Using eluate obtained from the pull-down, qPCR was conducted using primers designed against the 11 predicted RAREs. The values were normalised against a control region 24kb upstream of Oct4, and IgG.

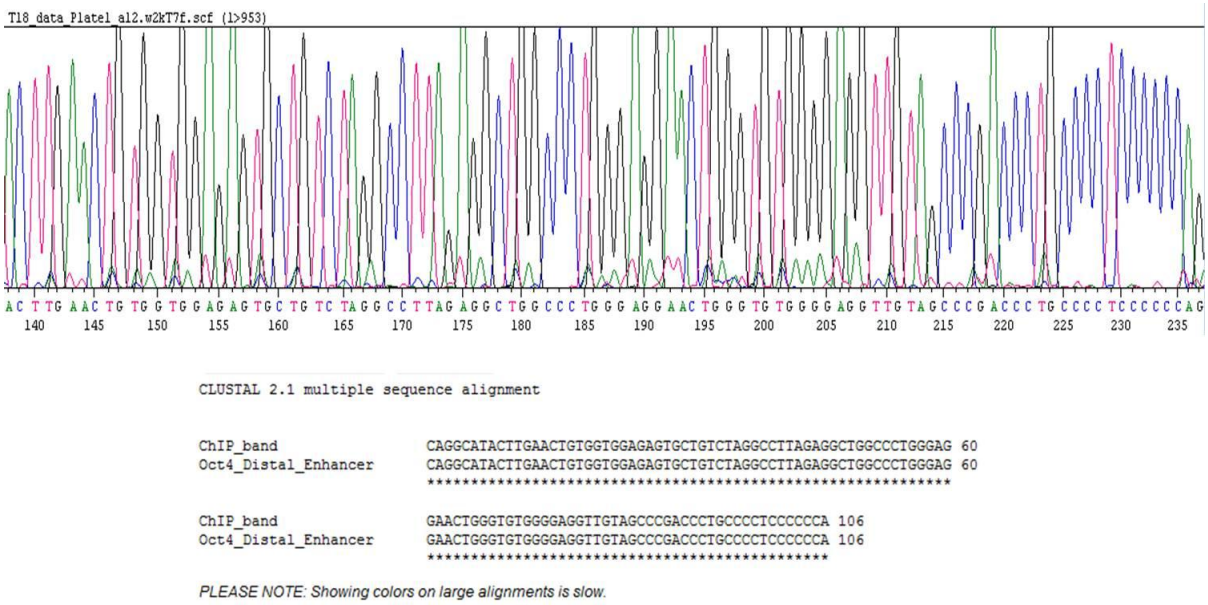



**Figure 5.5 Sequence of amplified band using primers against Oct4 Distal Enhancer** Chromatin Immunoprecipitation (ChIP) was performed using antibodies against RARβ in MEFs. Using eluate obtained from the pull-down, PCR was conducted using primers designed against the distal enhancer of Oct4. Amplified products were visualised on an agarose gel before extraction and purification. The purified amplicon was sequenced (Top) and aligned to the genomic sequence of Oct4 using ClustalW2 (Bottom).

### 5.2.2 RARβ elicits transcriptional activation through interaction with the distal enhancer of Oct4

To determine if binding of RARβ to the regulatory elements of Oct4 elicits transcriptional activation, various DNA fragment sizes of the Oct4 genomic locus were cloned into luciferase reporter constructs (**Figure 5.6**). Having delineated that binding of RARβ was restricted to 2.17kb upstream of the Oct4 TSS, I chose to study DNA spanning 2.4kb which encompassed the distal and proximal enhancers of Oct4 and RAREoct (DE-PE-RAREoct). In addition, a 1.3kb fragment comprising of both Oct4 proximal enhancer and RAREoct (PE-RAREoct), and a 466bp fragment solely consisting of RAREoct (RAREoct), were inserted into reporter constructs preceding cDNA encoding for firefly luciferase. In order to study the effect of RARβ on the enhancer elements alone, additional constructs were generated. First, a 450bp genomic fragment containing merely the distal enhancer was inserted adjacent to the luciferase reporter construct (DE-minP). A basic promoter consisting of a TATA box (minP) was included preceding the luciferase reporter gene to allow for efficient promotion of gene expression in the presence of enhancers. Extending this construct, genomic fragments of increasing length, namely a 2kb fragment consisting of both the distal and proximal enhancer (DE-PE-minP), and a 2.4kb fragment encompassing the distal and proximal enhancer and RAREoct (DE-PE-RAREoct-minP), were inserted in front of the minimal promoter and cDNA encoding for firefly luciferase.

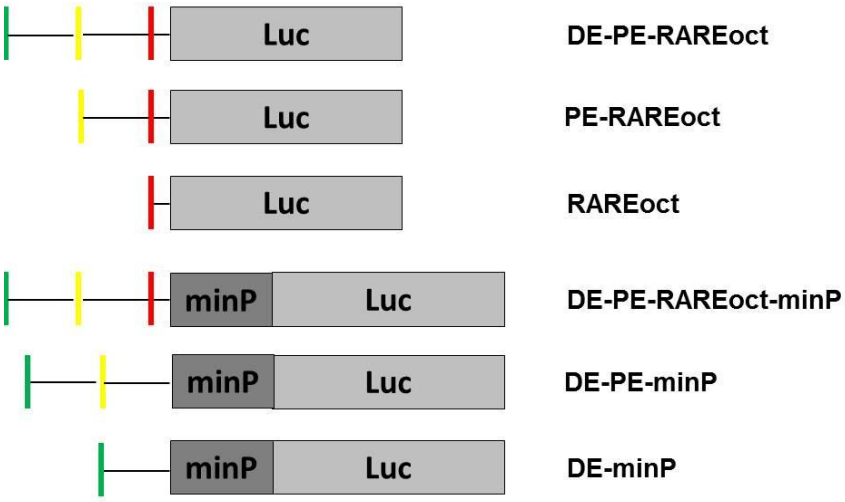



**Figure 5.6 Schematic Diagram of luciferase constructs generated** Luciferase constructs comprising of a combination of elements upstream of Oct4 were designed. The red, yellow and green boxes depict RAREs representative of the proximal promoter (RAREoct), proximal enhancer and distal enhancer of Oct4 respectively. The top three constructs involve the modification of pGL3-basic, which allows the insertion of a promoter of interest adjacent to a luciferase reporter gene. The bottom three constructs involve the modification of pGL4.23, which allows the insertion of desired enhancer regions adjacent to a minimal promoter (minP) preceding a luciferase reporter gene.

It was previously described that a genomic fragment of length 18kb which contained the regulatory elements of Oct4 was able to recapitulate the expression of Oct4 in a developing mouse embryo (Yeom et al., 1996). To verify the functionality of the luciferase reporter constructs, each luciferase vector was introduced into MEFs or ESCs, representing somatic or pluripotent cell types. **Figure 5.7** demonstrates that DE-PE-RAREoct is highly activated in ESCs, as reflected by high luciferase activity. Conversely, the same construct is repressed in MEFs. Similarly, PE-RAREoct is twice as active in ESCs than MEFs. On the contrary, RAREoct was observed to be repressed in ESCs but active in MEFs. This could be attributed to the presence of repressors such as Coup-TFI or Coup-TFII in ESCs which have been described to bind to RAREoct (Ben-Shushan et al., 1995). **Figure 5.8** fortifies these observations where DE-PE-RAREoct-minP is 8-fold more active in ESCs than MEFs. In parallel, DE-minP and DE-PE-minP exhibited 2-fold and 3-fold increased activity in ESCs compared to MEFs respectively, suggesting synergy between the distal enhancer and RAREoct for the manifestation of maximum Oct4 expression.

Having demonstrated the functional integrity of the luciferase constructs, these reporter vectors were introduced to MEFs in combination with reprogramming cocktails. **Figure 5.9** offers a graphical representation of the luciferase readings triggered by the presence of the reprogramming factors. Given that the distal enhancer encompasses several recognition sites for a spectrum of transcription factors, including c-Myc, Klf4 and Sox2 (CKS), it was expected that the presence of CKS activated DE-PE-RAREoct. The observed activation of DE-PE-RAREoct was augmented when RARβ was co-introduced with CKS. Intriguingly, the proximal enhancer and proximal promoter regions of Oct4 were not affected in the presence of RARβ. On the contrary, RARγ has recently been implicated in the regulation of Oct4 expression levels via interaction at the proximal promoter (Wang et al., 2011c). In a bid to explore if RARγ could function in a similar manner as RARβ, RARγ was introduced into the experimental study. Interestingly, when RARγ was introduced in place of RARβ, a similar phenomenon was observed where CKS and RARγ triggered activation of DE-PE-RAREoct, reinforcing the observation that RARs elicit a transcriptional effect at the Oct4 distal enhancer.

The co-operative role between LRH1 and RARγ in the acceleration and enhancement of reprogramming events has recently been established (Wang et al., 2011c). To determine if LRH1 has a similar additive effect in the presence of RARβ and CKS, five factors (CKS, RARβ and LRH1) were introduced into MEFs and observed for transcriptional activation as indicated by luciferase activity. Interestingly, the addition of LRH1 to the reprogramming cocktail containing CKS and RARβ enhanced luciferase activity. However, comparable to previous findings, the increased luciferase activity was only observed in the luciferase construct that encompasses the distal enhancer of Oct4 (DE-PE-RAREoct) (**Figure 5.9**). An analogous result was observed when RARγ was applied instead of RARβ, where the presence of LRH1, RARγ and CKS displayed elevated levels of luciferase activity.

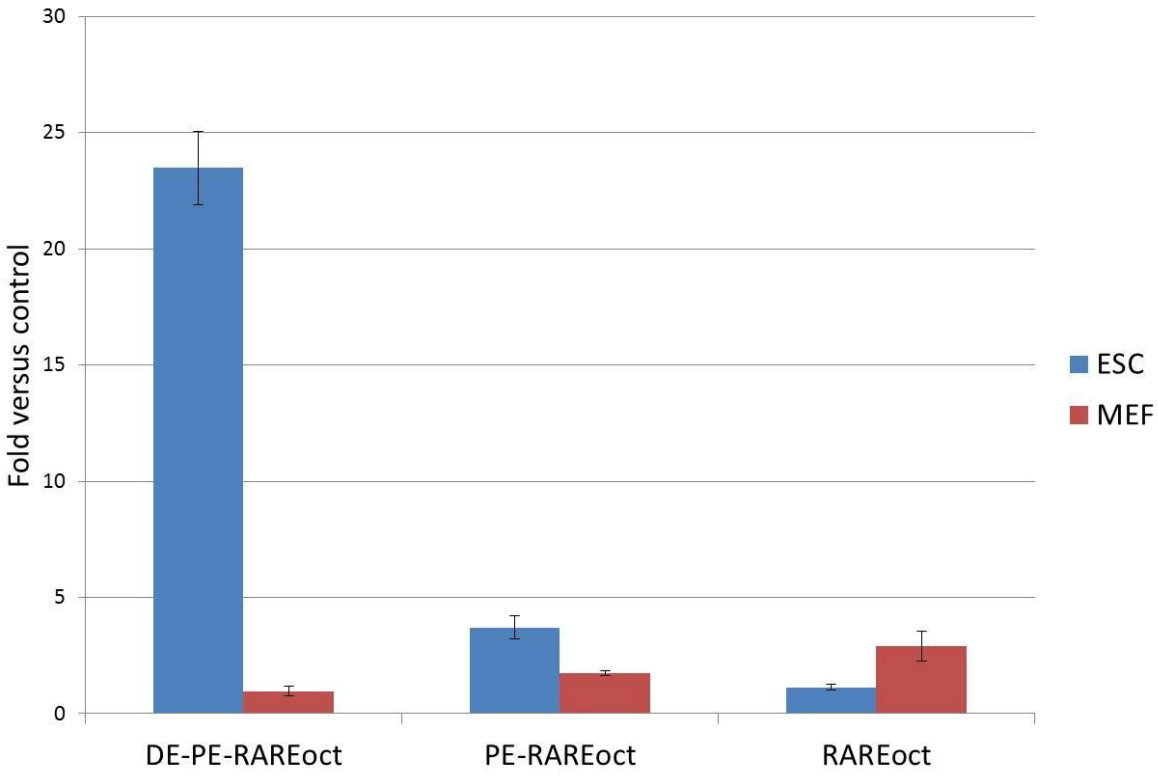



**Figure 5.7 Activity of luciferase constructs in MEFs and ESCs** To determine the functionality of each luciferase construct, the reporter vectors were separately transfected into MEFs or ESCs. After 48 hours, transfected cells were lysed and analysed for luciferase actvity through the addition of luciferin. Renilla was co-introduced with each luciferase reporter construct to account for transfection variability. The values plotted on the graph were normalised to an empty pGL3-basic vector and their corresponding Renilla control readings.

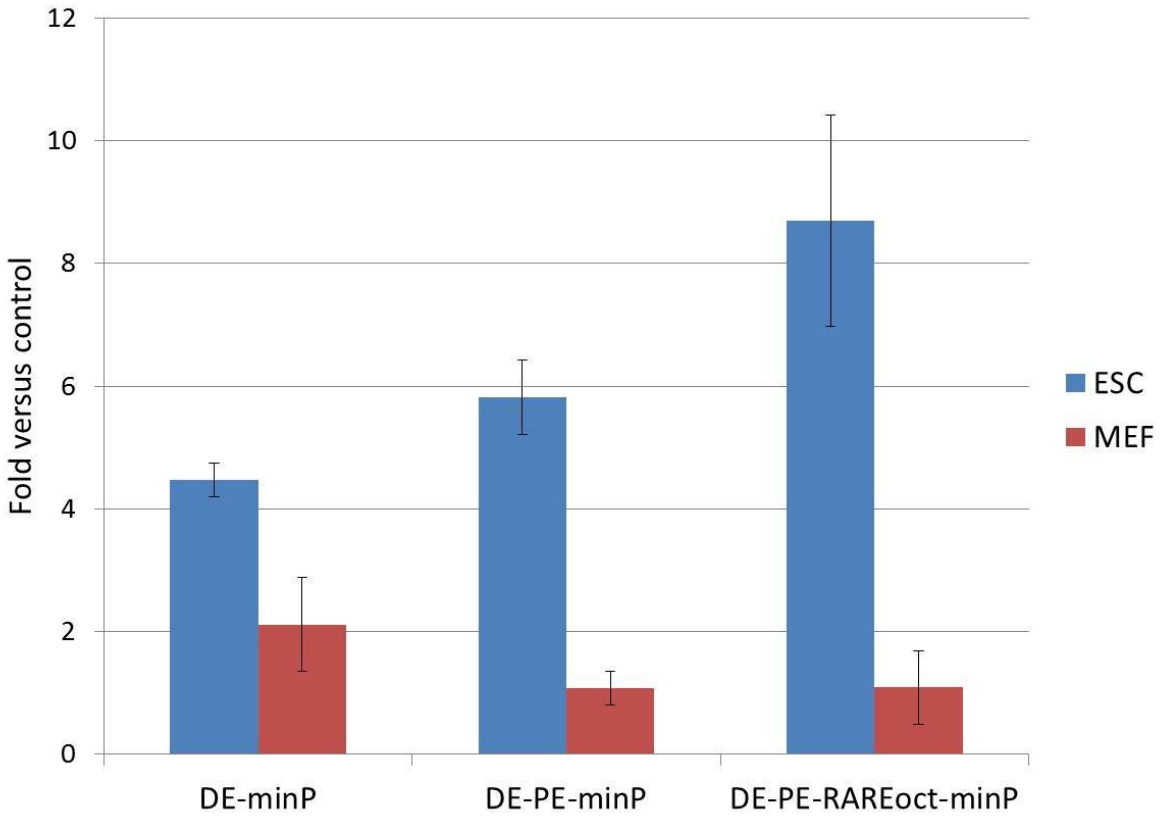



**Figure 5.8 Activity of luciferase constructs in MEFs and ESCs** To determine the functionality of each luciferase construct, the reporter vectors were separately transfected into MEFs or ESCs. After 48 hours, transfected cells were lysed and analysed for luciferase actvity through the addition of luciferin. Renilla was co-introduced with each luciferase reporter construct to account for transfection variability. The values plotted on the graph were normalised to pGL4.23 and their corresponding Renilla control readings.

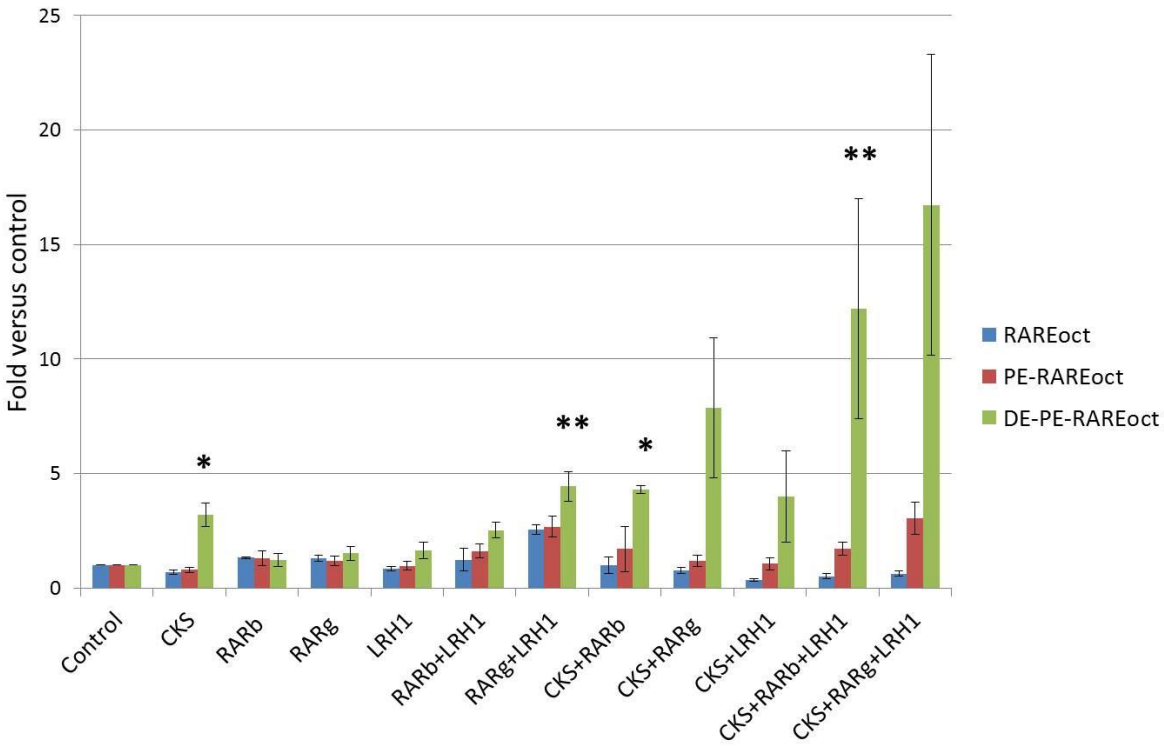



**Figure 5.9 RARβ and RARγ act synergistcally with LRH1 to transcriptionally activate the distal enhancer of Oct4** Combinations reflective of the reprogramming cocktails were transfected into MEFs. The combinations of transgenes introduced are labelled along the horizontal axis. Each of the three luciferase constructs (RAREoct, PE-RAREoct, and DE-PE-RAREoct) was co-introduced with each transgenic combination. After 48 hours, the cells were lysed and assessed for luciferase activity. The values are displayed in the bar chart, where the three colours depict the different luciferase construct analysed. The readings are normalised against Renilla and pGL3-basic. Statistical analysis was performed using a student's T-test. *p=0.05; **p<0.05

### 5.2.3 RARs can function as replacements to exogenous Oct4 during the generation of iPSCs

Having described that both RARγ and RARβ positively regulate the distal enhancer of Oct4, I next investigated if RARγ and RARα could similarly trigger reprogramming events in the absence of ectopic Oct4 expression. Using a doxycycline inducible system, individual RAR family members were introduced into MEFs, in the absence or presence of LRH1. LRH1 was included on the premise it can act synergistically with RARγ and RARβ to activate Oct4 expression levels, as suggested by **Figure 5.9**. To distinguish reprogrammed cells, AP which acts as a general stem cell marker was employed. **Figure 5.10** illustrates the number of AP expressing colonies obtained from the various reprogramming mixes. The presence of RARs or CKS individually did not generate any iPSC colonies. In contrast, combination of CKS and each RAR family member incited the formation of iPSC colonies (RARγ-iPSCs and RARβ-iPSCs). The presence of LRH1 and CKS was observed to trigger reprogramming events (LRH1-iPSCs), reproducing observations which were previously established (Heng et al., 2010). Analagous to findings obtained from **Figure 5.9**, addition of LRH1 to reprogramming cocktails consisting of CKS and each RAR family member promoted the number of reprogrammed colonies (RARγ+LRH1-iPSCs and RARβ+LRH1-iPSCs), reiterating the co-operation between LRH1 and RARs in the activation of the Oct4 regulatory elements.

These observations were reaffirmed in triplicates using larger volumes of starting material. **Figure 5.11** provides a numerical representation on the various reprogramming efficiencies. Mirroring previous observations in **Figure 5.10**, RARβ and RARγ were competent at replacing exogenous Oct4 to generate iPSCs and this ability was augmented with the inclusion of LRH1. However, the conventional reprogramming cocktail (Oct4 and CKS) was the most efficient at producing iPSC colonies, surmising that the stoichiometric ratios of the replacement factors were sub-optimal.

To ensure that iPSCs generated in the presence of LRH1 and/or RARγ were sustainable and exhibited pluripotency, transcript levels of pluripotency markers in these cells were determined. **Figure 5.12** illustrates that iPSCs generated in the presence of LRH1 and RARγ, where exogenous Oct4 was absent, exhibited comparable levels of pluripotency markers to ESCs and Oct4-iPSCs, reinforcing that RARγ and LRH1 can derive pluripotent derivatives in the absence of ectopic Oct4.

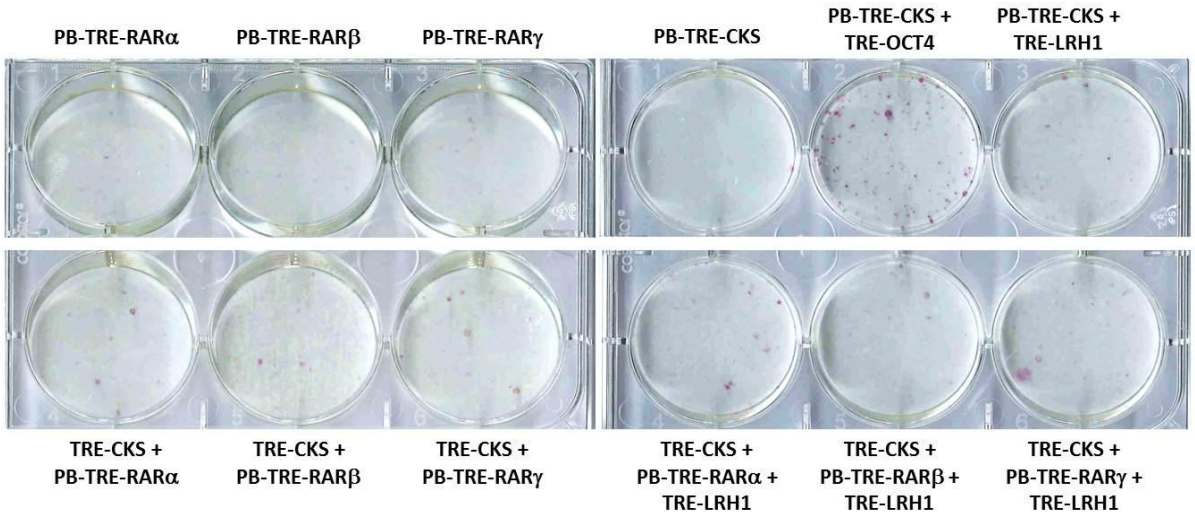



**Figure 5.10 RAR family members act synergistcally with LRH1 to increase reprogramming efficiencies in the presence of PB-TRE-CKS** To determine the reprogramming capabilities of RARβ, RARγ and RARα, each factor was introduced into MEFs, in co-operation with PB-TRE-CKS and PB-CAG-rtTA . Transfected cells were maintained over three weeks on a feeder layer to observe for colonies. iPSC colonies were stained for the expression of AP, as depicted in red.

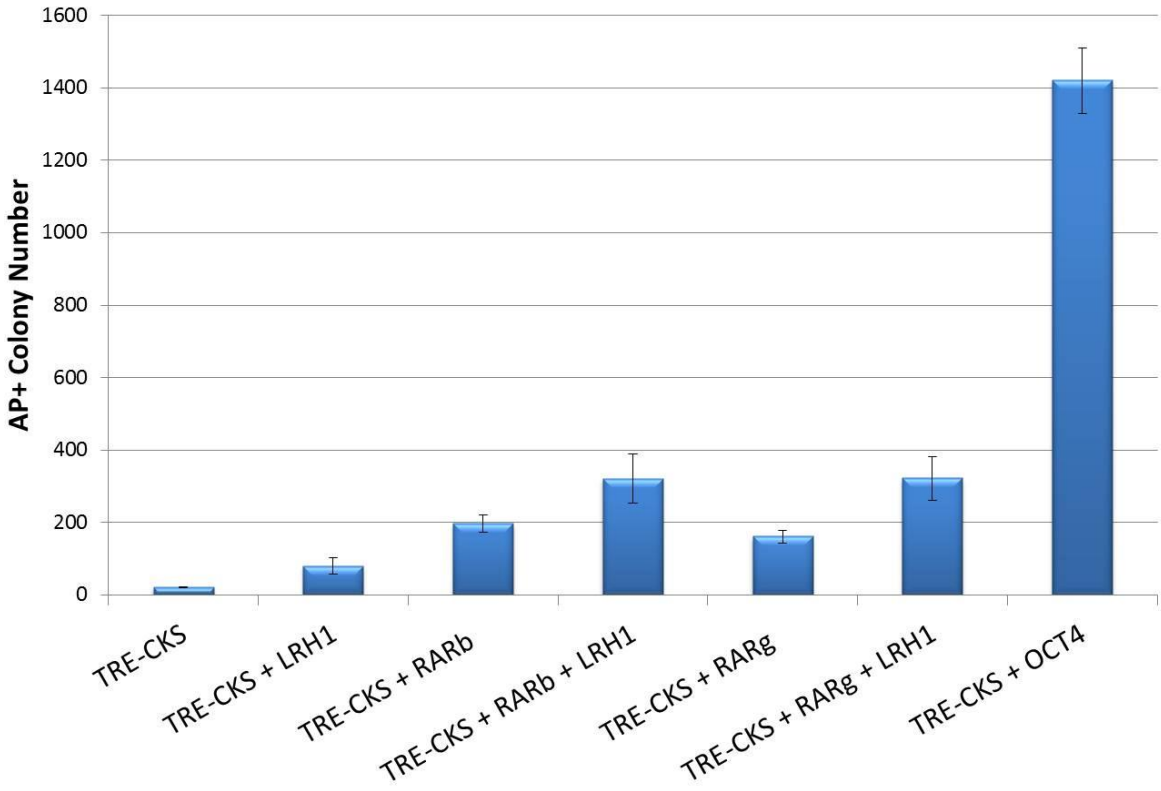



**Figure 5.11 Comparison of reprogramming efficiencies triggered by RARβ and RARγ, in synergy with LRH1** To compare the reprogramming capabilities of RARβ and RARγ, both factors were introduced into MEFs, in the presence or absence of LRH1, and the addition of PB-TRE-CKS and PB-CAG-rtTA . Transfected cells were maintained over three weeks on a feeder layer to observe for colonies. Number of colonies generated in each transgenic combination were calculated and indicated in the bar chart. Triplicates for each reprogramming combination was performed and the error bars indicate the standard deviation derived from the three independent experiments.

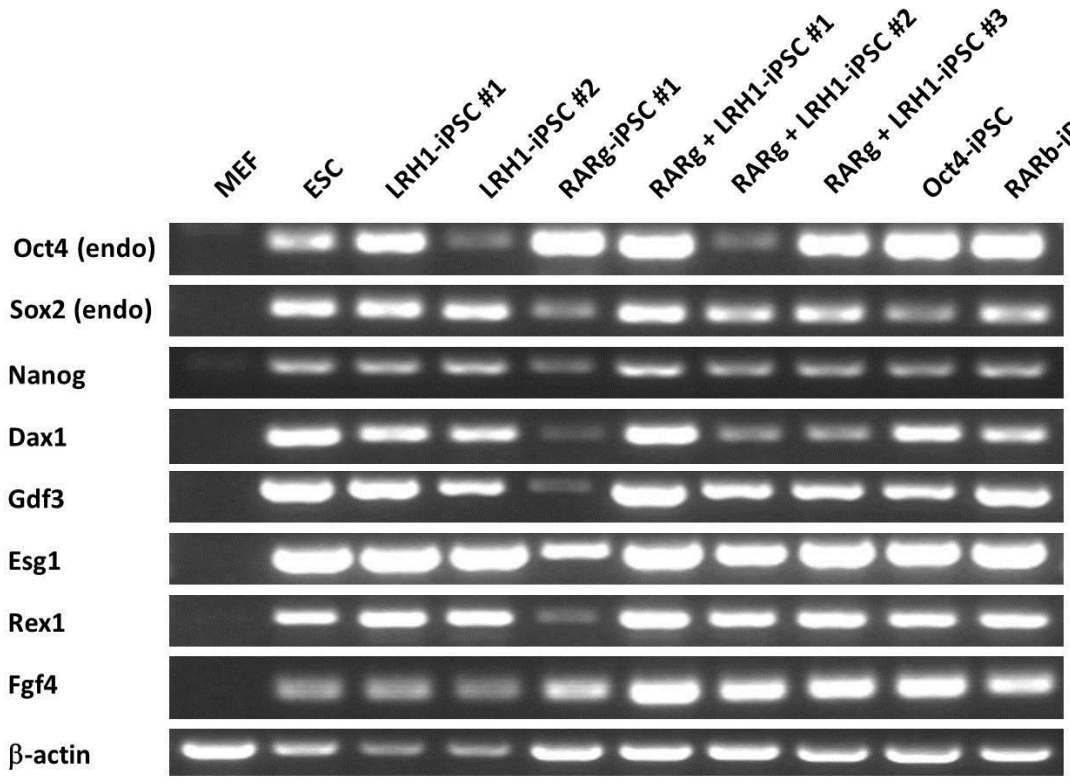



**Figure 5.12 Expression levels of pluripotency markers in iPSCs generated by RARγ and/or LRH1** iPSCs were generated using combinations of RARβ, RARγ and LRH1, in the presence of PB-TRE-CKS. iPSC colonies were picked and expanded in doxycycline free-growth medium. cDNA was obtained from established iPSC clones and expression of pluripotency markers were assessed by PCR and visualised on an agarose gel. The intensities of the bands are indicative of the amount of transcript present.

### 5.2.4 Oct4, c-Myc, Klf4, Sox2, RARγ and LRH1 (6F) are the most potent activators of the Oct4 distal enhancer

This chapter has described that RARs and LRH1 play a positive transcriptional role by tethering to the Oct4 distal enhancer in the absence of ectopic Oct4. Recent work has described the remarkable ability of Oct4, c-Myc, Klf4, Sox2, RARγ and LRH1 (6F) in improving reprogramming speed and efficiencies (Wang et al., 2011c). As ectopic Oct4 has been omitted from the experiments thus far, it was interesting to determine if the reprogramming advantage instigated by 6F was also attributable to transcriptional activation at the distal enhancer. To this end, reprogramming cocktails inclusive of Oct4 were co-introduced with luciferase constructs described in **Figure 5.6** into MEFs. Luciferase assays were performed at 48 hours post transfection and **Figure 5.13** illustrates the results of the luciferase assays. Evident from the graphical representation of luciferase readings, the presence of 6F initiated luciferase activity of DE-PE-RAREoct up to 20-fold, when compared to activity incited by the presence of Oct4, c-Myc, Klf4 and Sox2 (4F). Interestingly, substitution of RARγ with RARβ also led to a 10-fold increment in luciferase activity compared to 4F. These results reinforce the notion that RARs can regulate Oct4 expression, especially when facilitated by the introduction of ectopic Oct4 and LRH1.

To align the luciferase results to a functional test, an identical spectrum of reprogramming combinations in **Figure 5.13** was introduced into MEFs and observed for the formation of iPSC colonies. **Figure 5.14** illustrates a numerical comparison of iPSC colonies obtained from the respective reprogramming conditions. Analagous to data retrieved from **Figure 5.13**, the addition of RARβ and RARγ to the conventional reprogramming mix (4F) improved reprogramming efficiencies. The efficiency was substantially further increased in co-operation with LRH1. These results complement previous findings and suggest that RAR family members elicit similar effects in assisting reprogramming events, but the 6F combination is most competent at inciting Oct4 expression and reprogramming efficiencies.

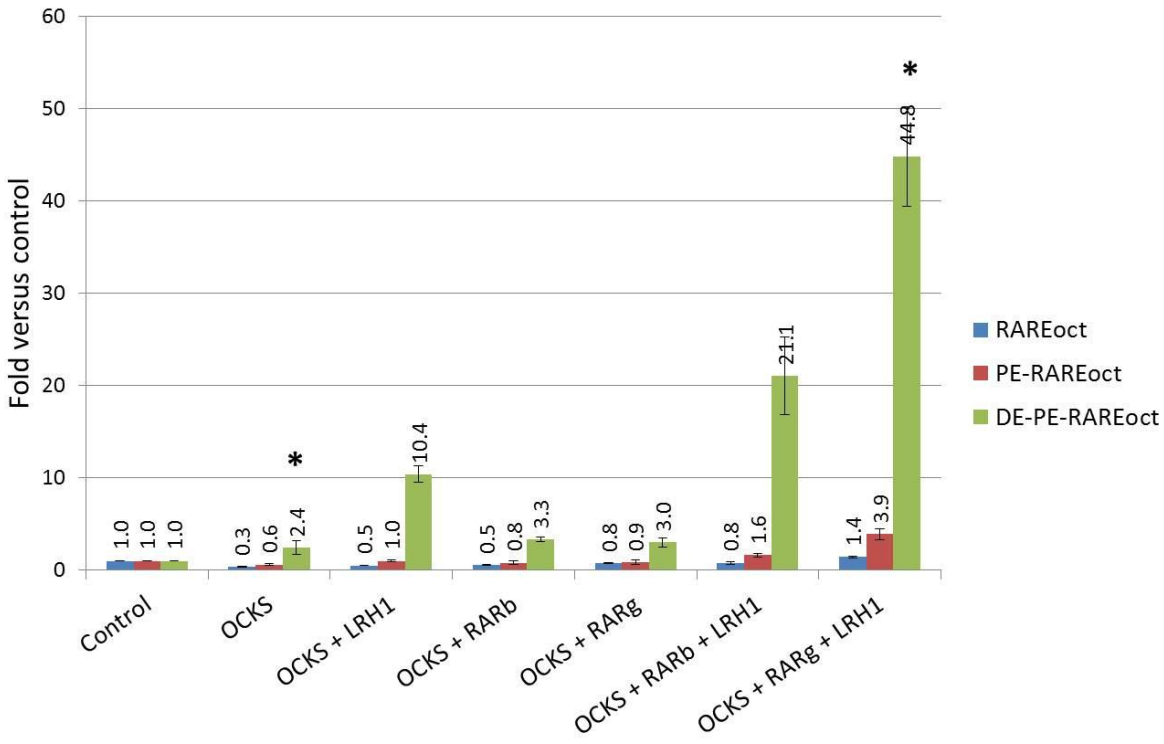



**Figure 5.13 RARβ and RARγ act synergistcally with LRH1 to transcriptionally activate the distal enhancer of Oct4 in the presence of conventional four reprogramming factors**
Combinations reflective of the reprogramming cocktails were transfected into MEFs. The combinations of transgenes introduced are labelled along the horizontal axis. Each of the three luciferase constructs (RAREoct, PE-RAREoct, and DE-PE-RAREoct) was co-introduced with each transgenic combination. After 48 hours, the cells were lysed and assessed for luciferase activity. The values are displayed in the bar chart, where the three colours depict the different luciferase construct analysed. The readings are normalised against Renilla and pGL3-basic. Statistical analysis was performed using student's T-test. *$p<0.05$

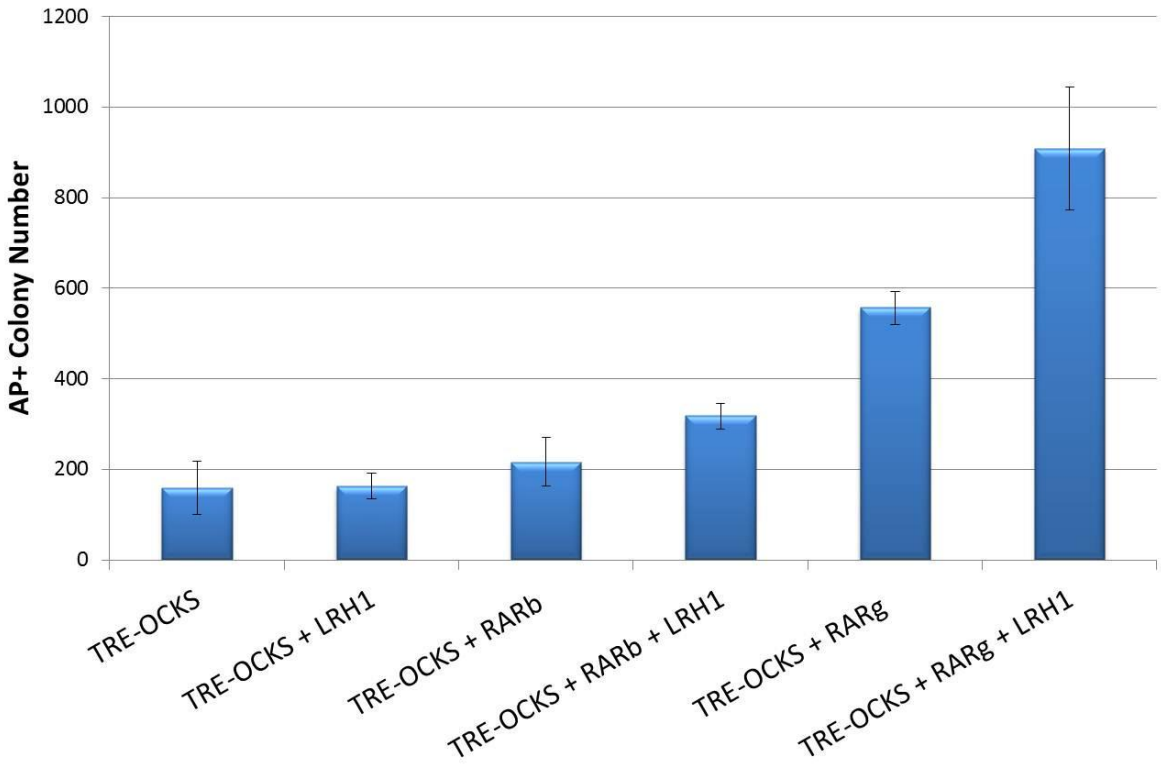



**Figure 5.14 Comparison of reprogramming efficiencies triggered by RARβ and RARγ, in synergy with LRH1 and the four conventional reprogramming factors** To compare the reprogramming capabilities of RARβ and RARγ, both factors were introduced into MEFs, in the presence or absence of LRH1, and the addition of PB-TRE-OCKS and PB-CAG-rtTA. Transfected cells were maintained over three weeks on a feeder layer to observe for colonies. Number of colonies generated in each transgenic combination were calculated and indicated in the bar chart. Triplicates for each reprogramming combination was performed and the error bars indicate the standard deviation derived from the three independent experiments.

### 5.2.5 LRH1 binds to the regulatory elements of Oct4

The synergy between RARγ and LRH1 in activating Oct4 expression at RAREoct has been attributed to co-operative binding (Ben-Shushan et al., 1995; Wang et al., 2011c). With the discovery that RARβ and RARγ work in unison with LRH1 to promote Oct4 expression at the distal enhancer, it is imperative to determine if LRH1 is able to tether to similar upstream regions of Oct4. To pursue this question, ChIP was performed in ESCs using an antibody against LRH1. Using the same set of primers described earlier (**Figure 5.3**), 11 putative RAREs within 10kb upstream of the Oct4 TSS were amplified. **Figure 5.15** illustrates quantification of the amplified products at the 11 potential sites. Remarkably, there exists a close resemblance between the patterns observed for LRH1 and RARβ. Motifs situated beyond 2.17kb of the Oct4 TSS were poorly represented, whereas regions closer to the Oct4 TSS were heavily represented. This supports the notion that LRH1 competently binds to the distal and proximal enhancers of Oct4 and RAREoct, explicating its ability to co-regulate the Oct4 distal enhancer with RAR family members.

Results from the luciferase assay suggest that RARγ interacts more effectively with LRH1 than RARβ to activate transcription at the Oct4 distal enhancer (**Figure 5.9** and **Figure 5.13**). This could be attributable to mutual exclusivity in their inherent expression profiles. Examination of transcript levels of the RAR family members and LRH1 in ESCs and MEFs illustrate that RARβ and LRH1 are unlikely to co-express in a natural setting (**Figure 5.16**). In contrast, LRH1, RARα and RARγ are expressed in ESCs, surmising the innate ability of RARγ and LRH1 to co-operate and trigger transcription at the Oct4 distal enhancer (**Figure 5.9** and **Figure 5.13**). These discrepancies in expression levels reflect the effective pairing between RARγ and LRH1, but not RARβ and LRH1. Subsequent co-immunoprecipitation experiments using endogenous proteins could allow better understanding of the natural synergy between RARs and LRH1. As RARα and RARγ are expressed in similar patterns, it will be interesting to determine if RARα and LRH1 display synergistic abilities to promote the transcriptional activation of Oct4.

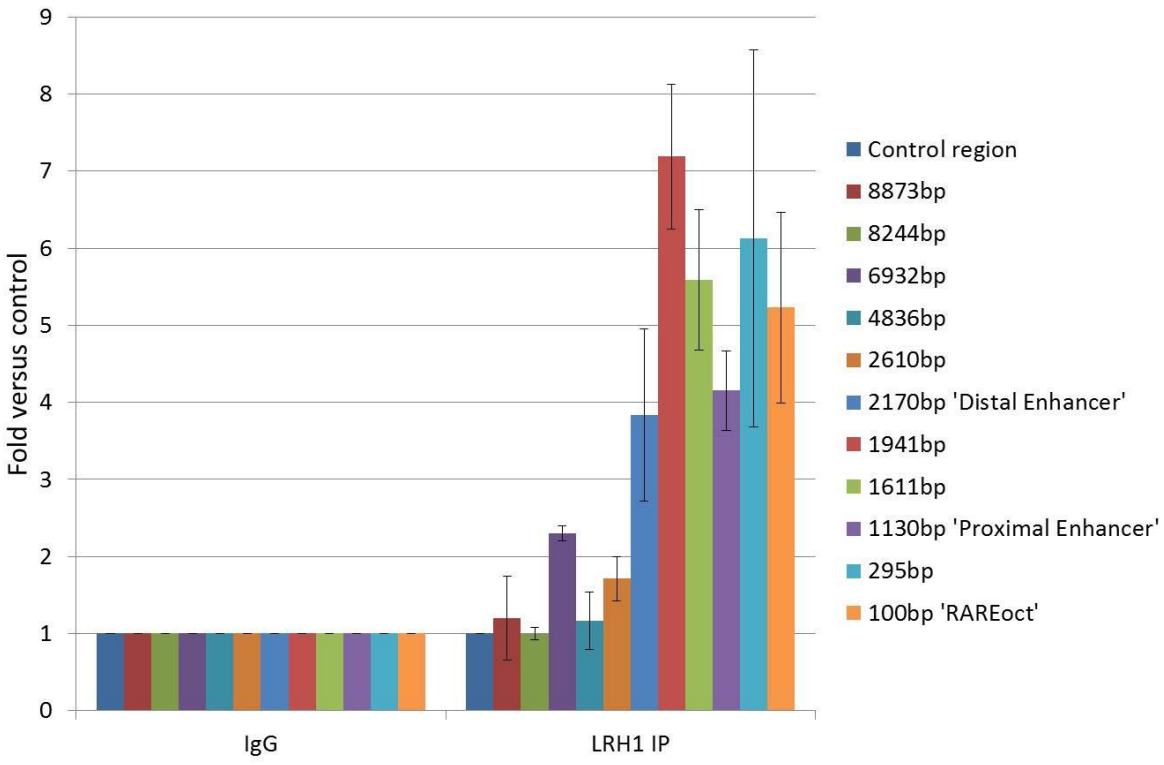



**Figure 5.15 LRH1 binds to RAREs up to 2kb upstream of Oct4** Chromatin Immunoprecipitation (ChIP) was performed using antibodies against LRH1 in ESCs. Using eluate obtained from the pull-down, qPCR was conducted using primers designed against the 11 predicted RAREs. The values were normalised against a control region 24kb upstream of Oct4, and IgG.

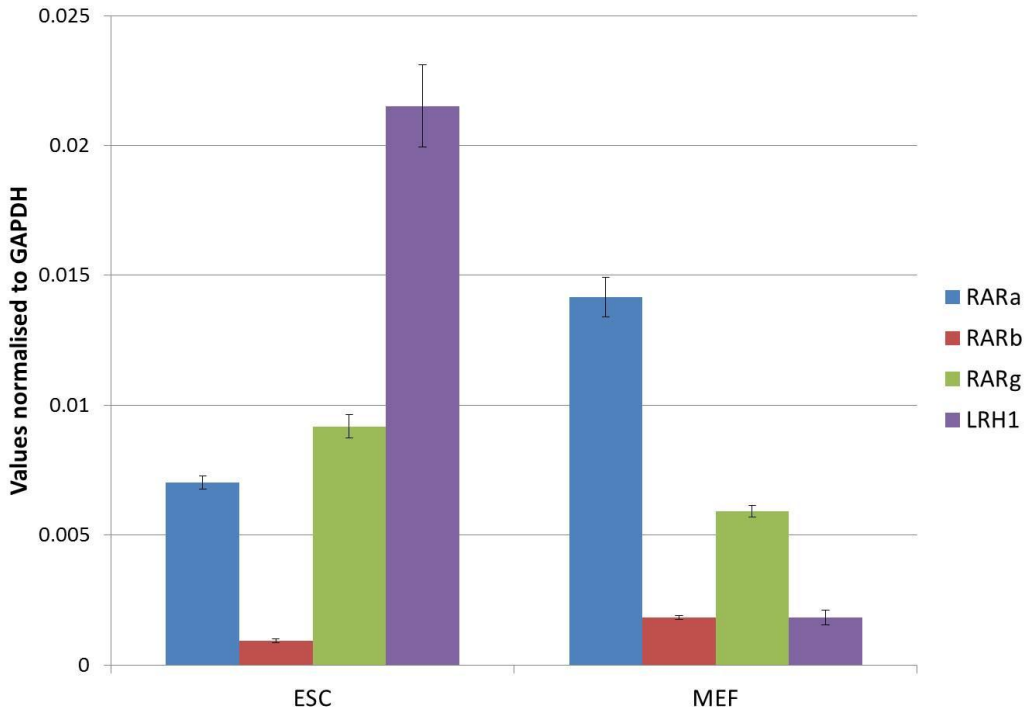



**Figure 5.16 Expression profile of RAR family members and LRH1 in MEFs and ESCs** Primers were designed against exon junctions of RARα, RARβ, RARγ and LRH1. Using cDNA from ESCs and MEFs, the transcript amounts of RARα, RARβ, RARγ and LRH1 were quantified through qPCR using corresponding primers and SYBR-Green and normalised against GAPDH expression levels.

### 5.2.6 Oct4 expression is obligatory during reprogramming

Although RAR family members have been portrayed to induce nuclear reprogramming of mouse fibroblasts to a primitive pluripotent state and eradicate the requirement for exogenous Oct4 expression, it is not clear if endogenous Oct4 expression is obligatory during the reprogramming process. To address this concern, temporal control of Oct4 in ZHBTc4 ESCs was employed (Niwa et al., 2000). **Figure 5.17** delineates the genetic modifications received by ZHBTc4 ESCs. In essence, both endogenous copies of Oct4 had been replaced by antibiotic resistance through homologous recombination. To sustain pluripotency, exogenous Oct4 cDNA controlled by a tetracycline repressible system was introduced into ZHBTc4 ESCs. These genetic manipulations allow self-renewal and the maintenance of pluripotency. In addition, Oct4 levels can be altered according to the addition of doxycycline, a potent derivative of tetracycline, where escalating amounts of doxycycline represses Oct4 expression to an increasing extent.

To determine if endogenous Oct4 is required during reprogramming, ZHBTc4 ESCs were differentiated for 6 days in the presence of 0.1μM ATRA and 1μg/mL doxycycline before reprogramming factors were introduced to determine competence in the acquisition of pluripotency (**Figure 5.18**). A combination of high doses of retinoic acid and doxycycline instigated differentiation of compact ESC colonies into flat extended cells. It has been previously established that exposure to doxycycline triggers trophoblast differentiation (Niwa et al., 2000), whereas high amounts of ATRA have been described to instigate differentiation cues (Strickland and Mahdavi, 1978; Schuldiner et al., 2001; Okada et al., 2004). A combination of two differentiation stimuli results in a population of differentiated cells that are distinct from ESCs and exhibit a complete loss of Oct4 (**Figure 5.19**). With the ablation of Oct4 expression, reprogramming cocktails were introduced to assess the reprogramming potential of the differentiated ESCs. Upon the introduction of reprogramming factors, the transfected cells were maintained over two weeks in the absence or presence of doxycycline and scored for the number of iPSC colonies generated.

As a control experiment, I made an initial attempt to determine the pliability of Oct4 expression after the withdrawal of doxycycline. ZHBTc4 ESCs which had been differentiated using ATRA and doxycycline over 6 days were seeded in the absence of doxycycline. Cells were collected at 1, 2, 3 and 7 days after doxycycline withdrawal and assessed for Oct4 expression (**Figure 5.19**). Upon 24 hours of doxycycline treatment at 0.1μg/mL, Oct4 expression was markedly reduced to levels resembling MEFs. Interestingly, after 7 days of doxycycline withdrawal, Oct4 transcript amounts were only regained to 20% of the initial intensity, and reversion to ESC-like morphology was not observed. This could be attributed to epigenetic modifications of the Oct4 transgene in ZHBTc4 cells during the differentiation process.

To assess the acquisition of pluripotency in the absence of Oct4, differentiated ZHBTc4 cells were transfected with various combinations of reprogramming factors (**Figure 5.20**). Continuous exposure to doxycycline and repression of Oct4 ablated the reprogramming potential of RARs and LRH1, implying the obligatory requirement for Oct4 during the initiation of reprogramming events. The absence of reprogramming events was rectified through the introduction of ectopic Oct4, reinforcing the notion that Oct4 is essential during reprogramming. However, the observation that ectopic expression of Oct4 was sufficient to initiate reprogramming events suggests that endogenous expression of Oct4 is unnecessary. Apart from ectopic introduction of Oct4, doxycycline which repressed Oct4 expression was withdrawn to stimulate Oct4 expression and generate iPSC colonies in the presence of CKS and RARs (**Figure 5.20**). This observation was compounded with the co-introduction of LRH1. Surprisingly, withdrawal of doxycycline in cells which only encompassed ectopic CKS could not initiate the formation of iPSCs. This could be attributable to an incomplete re-expression of Oct4 following doxycycline withdrawal as described in **Figure 5.19**. To offer a graphical representation of colony numbers, **Figure 5.21** depicts a cumulative bar chart reflecting the number of iPSC colonies obtained from four separate reprogramming experiments.

Having determined that Oct4 expression is essential to trigger the generation of iPSCs, it was next explored if increasing Oct4 expression levels could affect reprogramming efficiency. To this end, compounded amounts of Oct4 facilitated by introduction of ectopic Oct4 and re-expression of Oct4 transgene in the absence of doxycycline, were demonstrated to escalate reprogramming efficiencies (**Figure 5.20**). This proposes that amount of Oct4 is a limiting factor for the initiation of reprogramming. Additionally, inclusion of RARs and LRH1 into the reprogramming mix further enhanced reprogramming efficiencies (**Figure 5.20**), once again emphasising the capacity of RARs and LRH1 to promote the generation of iPSCs  As Oct4 expression in ZHBTc4 cells is not controlled by its genomic regulatory elements, it suggests that RARs have alternative roles in the reprogramming process. To this end, RARs have been described to function through non-canonical pathways and may account for the discrepancies observed  (Alsayed et al., 2001; Gianni et al., 2002; Gianni et al., 2006; Bruck et al., 2009; Masia et al., 2007; Ross-Innes et al., 2010; Laursen et al., 2012).

Overall, these results surmise the unequivocal need for the presence of Oct4 during reprogramming. Ectopic expression of Oct4 through the delivery of a foreign vector is sufficient to trigger reprogramming events but heightened amounts of Oct4 can enhance the generation of iPSCs. Furthermore, inclusion of RARs into the reprogramming cocktail can augment reprogramming efficiencies, substantiating previous experimental observations and proposing a new dimension to the role of RARs in the reprogramming process.

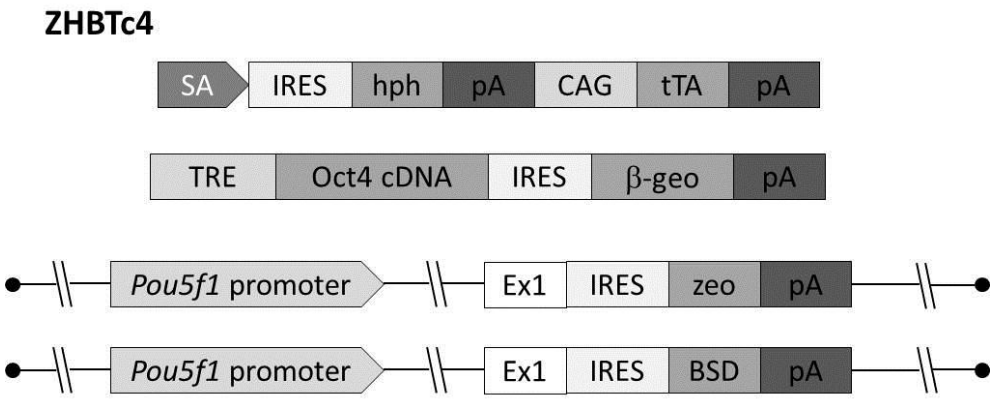



**Figure 5.17 Genomic modifications of ZHBTc4 ESCs** ZHBTc4 ESCs comprise of transgenes encoding (i) tetracycline transactivator and (ii) Oct4 cDNA driven by a tetracycline response element. The endogenous Oct4 loci are disrupted through homologous recombination and exon 1 is spliced with zeocin and blasticidin resistance genes.

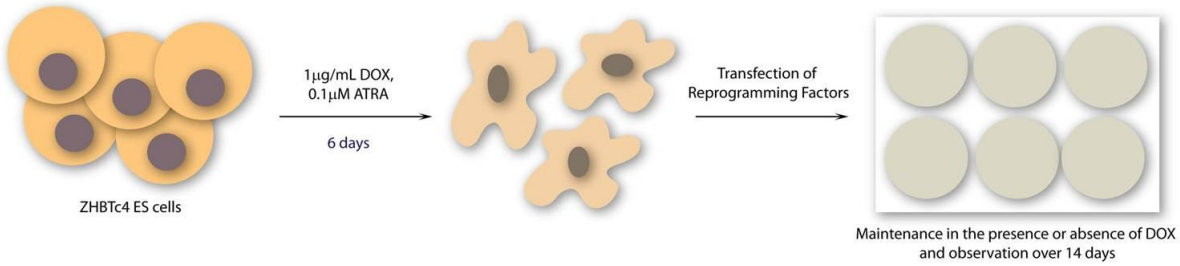



**Figure 5.18 Experimental Strategy to determine involvement of Oct4 in reprogramming** To determine the role of Oct4 in reprogramming, Oct4 expression levels were manipulated using ZHBTc4 ESCs. 1μg/mL DOXYCYCLINE and 0.1μM ATRA were introduced to ZHBTc4 ESCs for 6 days to trigger differentiation. Combinations of reprogramming factors were transfected into the differentiated cells and maintained in ESC growth media for 14 days. Colonies obtained were stained for the presence of AP and quantified.

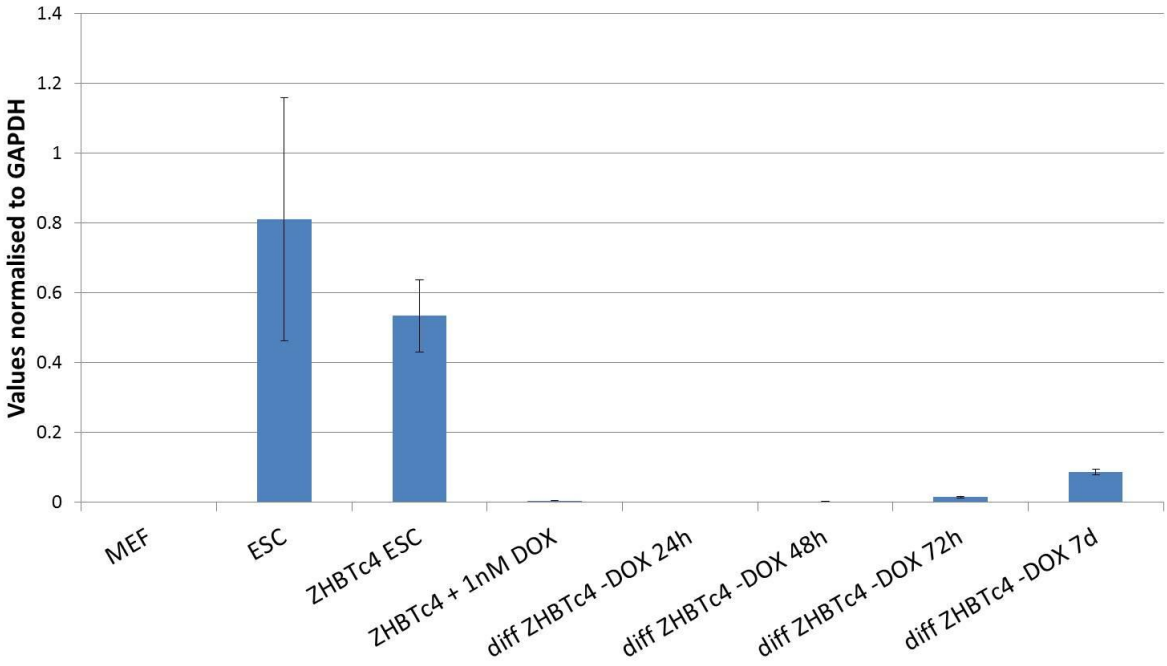



**Figure 5.19 Manipulation of Oct4 expression levels in ZHBTc4 ESCs** To verify the changes in Oct4 expression levels in ZHBTc4 ESCs, Oct4 transcript amounts were quantified using qPCR. Lane 1 shows negligible Oct4 levels in MEFs. Unmodified ZHBTc4 ESCs (Lane 3) exhibit Oct4 expression comparable to ESCs that carry both endogenous copies of Oct4 (Lane 2). Introduction of 1nM DOXYCYCLINE to ZHBTc4 ESCs over a period of 24 hours (Lane 4) drastically reduces Oct4 expression levels. Lanes 5-8 correspond to the re-expression of Oct4 in differentiated ZHBTc4 cells upon the withdrawal of DOXYCYCLINE over a 7 day period.

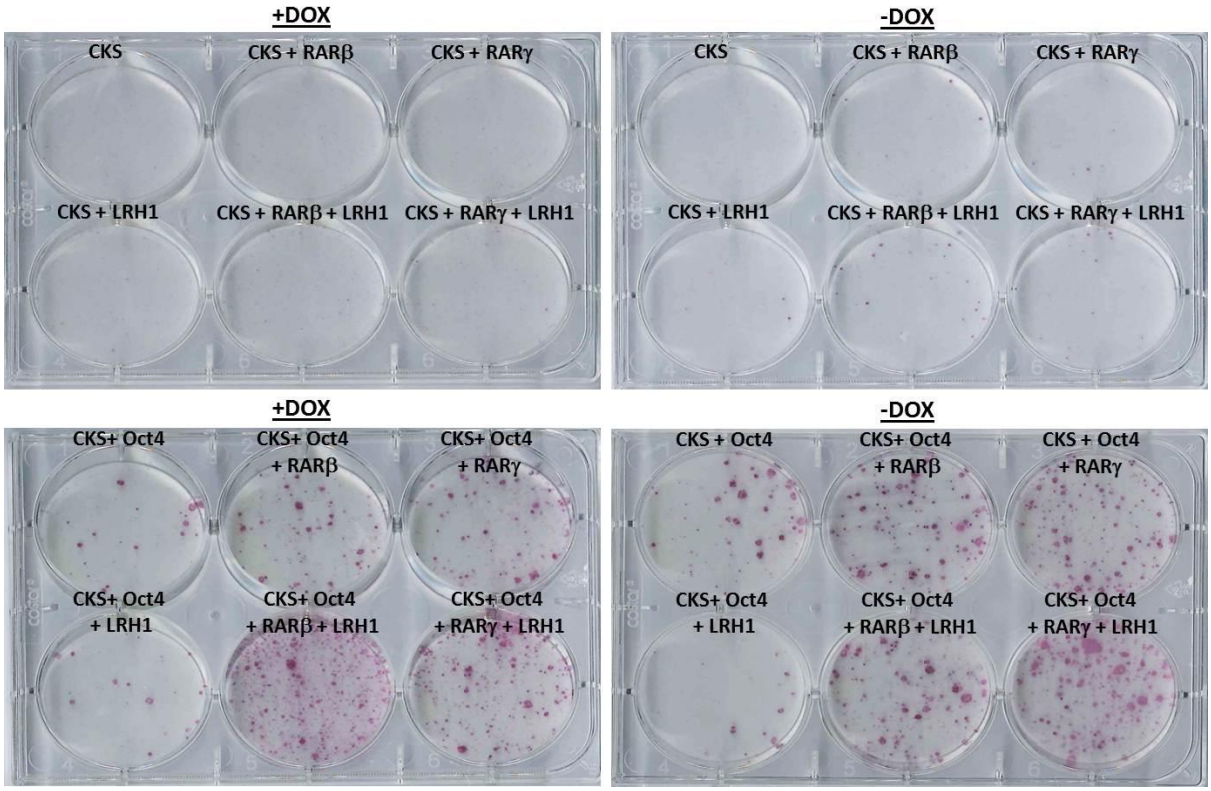



**Figure 5.20 Reprogramming of differentiated ZHBTc4 ESCs in the presence or absence of Oct4** ZHBTc4 ESCs were differentiated in 0.1μM ATRA and 1μg/mL DOXYCYCLINE for 6 days before combinations of CAG-promoter driven reprogramming factors were introduced by electroporation. Transfected cells were maintained in ESC growth media for 14 days, before reprogrammed colonies were stained for the presence of AP. AP activity, as indicated by red colour is a marker for pluripotency.

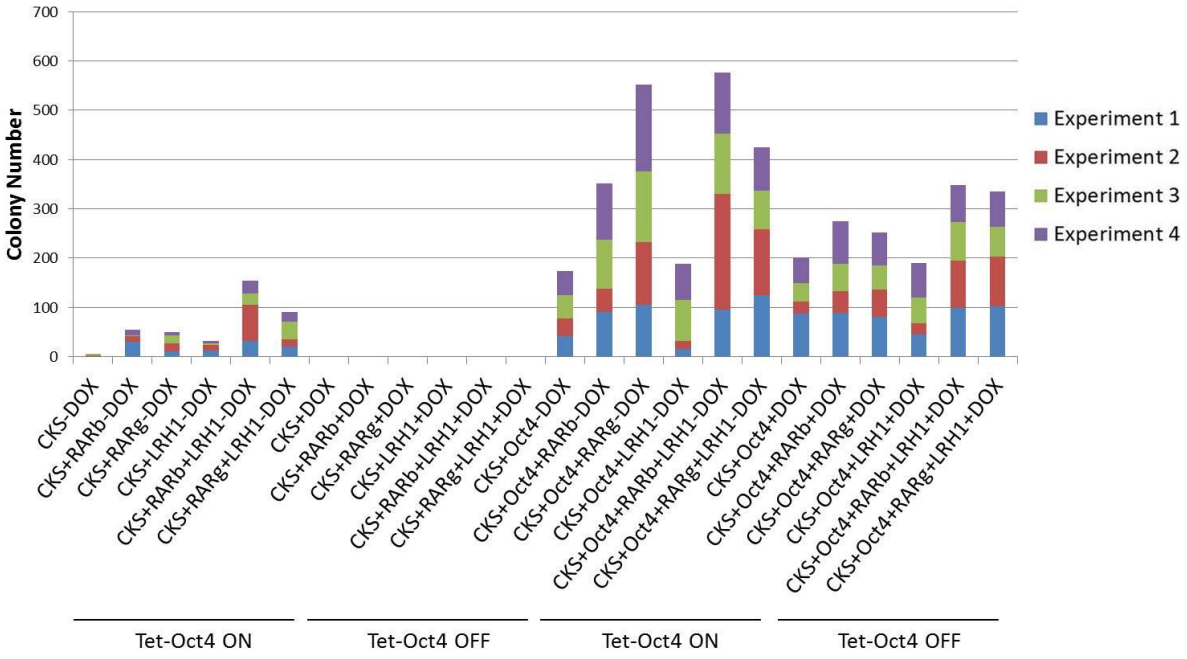



**Figure 5.21 Bar chart reflecting reprogrammed colonies from differentiated ZHBTc4 ESCs** ZHBTc4 ESCs were differentiated in 0.1μM ATRA and 1μg/mL DOXYCYCLINE for 6 days before combinations of reprogramming factors were introduced by electroporation. Transfected cells were maintained in ESC growth media in the absence (Tet-Oct4 ON) or presence (Tet-Oct4 OFF) of doxycycline for 14 days, before reprogrammed colonies were stained for the presence of AP. Number of colonies expressing AP is reflected in the bar chart. Each colour represents an independent experiment. Cumulative values for four separate experiments are illustrated.

### 5.2.7 Understanding the kinetics of activating Oct4 regulatory elements during reprogramming

RARs have been described as potent triggers of the reprogramming process, with RARγ and LRH1 exhibiting impactful capacities to positively regulate the Oct4 distal enhancer and RAREoct. With this in mind, it is pertinent to dissect the molecular kinetics of Oct4 activation assisted by RARγ and LRH1 during reprogramming. In order to assess the kinetics behind reprogramming, recordings of a homogenous population of cells undergoing reprogramming would be ideal. To this end, iPSCs were generated through the introduction of 4F or 6F in a doxycycline inducible approach. Both populations of iPSCs were differentiated through the addition of 0.1μM ATRA over 6 days (**Figure 5.22**). This creates a consistent platform to examine snapshots of the initial stages of reprogramming. Differentiated iPSCs were then exposed to doxycycline to elicit the re-expression of reprogramming factors. Comparison of 4F and 6F reprogramming kinetics highlights differences attributed to the presence of RARγ and LRH1. Cells were collected at various time points after doxycycline treatment to inspect epigenetic alterations at the Oct4 regulatory elements. Extending this, luciferase assays were performed in these cells to determine the activity of reporter constructs described in **Figure 5.6** at 24, 48 and 72 hours after doxycycline treatment. This method permits the dissection of changes akin to epigenetic modifications and activating stimuli at the regulatory components of the Oct4 locus during the first 72 hours of reprogramming and illuminates the participation of RARγ and LRH1 during the acquisition of pluripotency.

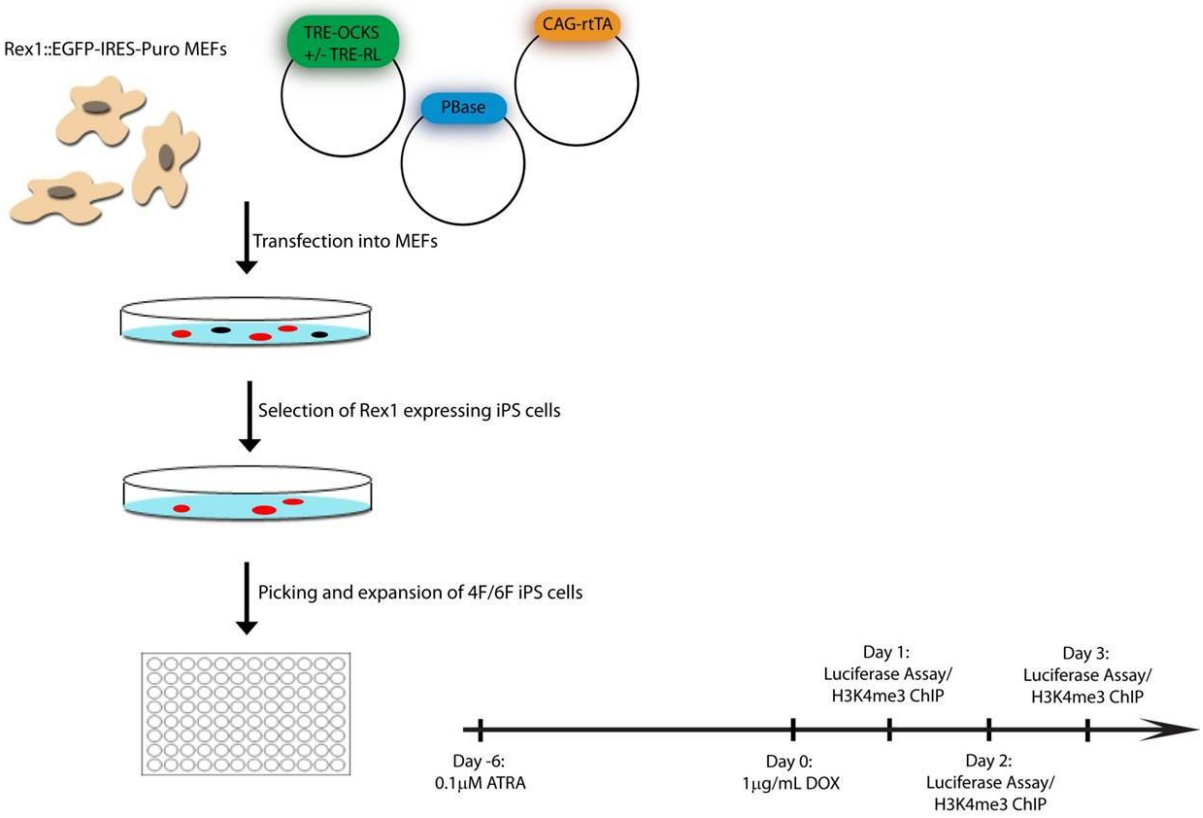



**Figure 5.22 Experimental Strategy to study the kinetics of Oct4 activation during reprogramming** Using a doxycycline (DOXYCYCLINE) inducible platform, 4F (Oct4, c-Myc, Klf4 and Sox2) or 6F (Oct4, c-Myc, Klf4, Sox2, RARγ and LRH1) were transfected into Rex1::EGFP-IRES-Puro MEFs. iPSCs expressing Rex1 were selected for using Puromycin and surviving colonies were picked and expanded. Upon expansion, DOXYCYCLINE was withdrawn to establish iPSC lines independent of the reprogramming transgenes. The cells were differentiated using 0.1μM ATRA for 6 days, before DOXYCYCLINE was re-introduced to trigger the re-expression of transgenes. At 24, 48, and 72 hours after transgene expression, cells were analysed for the activation of Oct4 enhancer/promoter regions using luciferase assays and chromatin immunoprecipitation of H3K4me3.

To determine if six days of ATRA treatment was sufficient to elicit differentiation of iPSCs and silence Oct4 transcription to levels that resemble MEFs, luciferase constructs described in **Figure 5.6** were introduced into differentiated 4F and 6F iPSCs. In unmodified 4F and 6F iPSCs, transcriptional activities of luciferase constructs were similar to ESCs (**Figure 5.23**). On the other hand, differentiation of iPSCs led to silencing of luciferase constructs to levels similar to MEFs (**Figure 5.24**). These results reaffirm the dependability of differentiated iPSCs as an assessment of regulatory elements upstream of Oct4 using luciferase assays.

Having determined that six days of ATRA treatment was sufficient to reduce luciferase activity in differentiated iPSCs to levels similar in MEFs, the experiment was repeated and doxycycline was introduced for various lengths of time, immediately after six days of ATRA treatment. The addition of doxycycline acted as a stimulus for the re-expression of reprogramming factors and initiated the acquisition of pluripotency in differentiated 4F and 6F iPSCs. **Figure 5.25** and **Figure 5.26** offer a graphical representation of luciferase readings measured using luciferase reporter constructs described in **Figure 5.6**, at specified durations after the addition of doxycycline in differentiated 4F and 6F iPSCs.

At 24 hours, luciferase constructs that comprised of either the distal enhancer (DE-minP) or proximal promoter (RAREoct) alone triggered transcriptional activation in differentiated 6F iPSCs, suggesting that expression of 6F could activate both the distal enhancer and RAREoct independent of each other. However, after 48 hours, these levels plateau. In contrast, a marked increase in luciferase activities was observed in luciferase constructs driven by the presence of both the distal enhancer and RAREoct (DE-PE-RAREoct and DE-PE-RAREoct-minP), 48 hours after re-expression of reprogramming transgenes in differentiated 6F iPSCs. Luciferase activity continued to escalate at 72 hours, proposing synergistic co-operation between the distal enhancer of Oct4 and RAREoct. The presence of the proximal enhancer did not significantly affect transcriptional activity, reinforcing previous observations that RARs and LRH1 did not significantly regulate Oct4 expression through the proximal enhancer (**Figure 5.9** and **Figure 5.13**). Interestingly,

differentiated 4F iPSCs did not result in the activation of luciferase reporter constructs between 24 and 72 hours, suggesting that observed differences in the luciferase activities in differentiated 6F iPSCs were attributable to the presence of the extra two factors, RARγ and LRH1.

From the luciferase assay, it is evident that the Oct4 distal enhancer and RAREoct are rapidly activated upon reprogramming, attributable to the expression of RARγ and LRH1. To offer a distinct perspective to the kinetics of reprogramming elicited by the presence of RARγ and LRH1, transcriptionally permissive histone modifications at the distal enhancer of Oct4 and RAREoct were examined in differentiated 4F and 6F iPSCs at 24, 48 and 72 hours after transgene re-activation. To this end, immunoprecipitation of trimethylated lysine 4 of histone 3 (H3K4me3) would allow the co-precipitation of actively transcribed genomic regions (**Figure 5.22**). Detection of Oct4 distal enhancer elements and RAREoct would suggest that these regions have undergone epigenetic modifications to permit transcription.

To allow a fair comparison to findings from the luciferase assay (**Figure 5.25** and **Figure 5.26**), 4F and 6F iPSCs were similarly differentiated in 0.1μM ATRA before doxycycline was introduced to trigger the re-expression of reprogramming factors. Analogous to the results obtained from the luciferase assay, both the Oct4 distal enhancer and RAREoct were associated with H3K4me3 in differentiated 6F iPSCs after 24 hours of doxycycline treatment (**Figure 5.27**). However, between 24 and 48 hours, pervasive epigenomic modifications ensued and both the Oct4 distal enhancer and RAREoct in 6F cells were heavily tethered to H3K4me3 activating marks. This association exhibited up to an eight-fold increment when compared to unmodified 6F iPSCs. However, between 48 and 72 hours, H3K4me3 levels were reduced to resemble unmodified 6F iPSCs. Reminiscent of the luciferase assay, the Oct4 regulatory elements in differentiated 4F iPSCs were not associated to H3K4me3 up to 72 hours of transgene re-activation, again demonstrating the notion that rapid activation of the distal enhancer and proximal promoter of Oct4 was triggered by the presence of RARγ and LRH1.

Overall, two distinct techniques were employed to study the kinetics behind the initiation of reprogramming events and findings from both experiments concur, implicating RARγ and LRH1 in the prompt activation of Oct4 within 24 hours of reprogramming. To associate the rapid activation of Oct4 with a functional outcome, differentiated 4F or 6F iPSCs were seeded onto a feeder layer in the presence of doxycycline and assessed for the formation of iPSCs. At the end of 7 or 14 days, AP expressing colonies were counted and compared to determine differences in reprogramming efficiencies (**Figure 5.28**). In the absence of doxycycline, both differentiated 4F iPSCs and 6F iPSCs gave rise to iPSC colonies, suggesting aberrant silencing of transgenes. Overall, differentiated 6F iPSCs generated 12.5 and 8 fold more iPSC colonies than differentiated 4F iPSCs at 7 days and 14 days respectively, reiterating the capacity of RARγ and LRH1 in the prompt activation of Oct4 regulatory elements and initiation of reprogramming events.

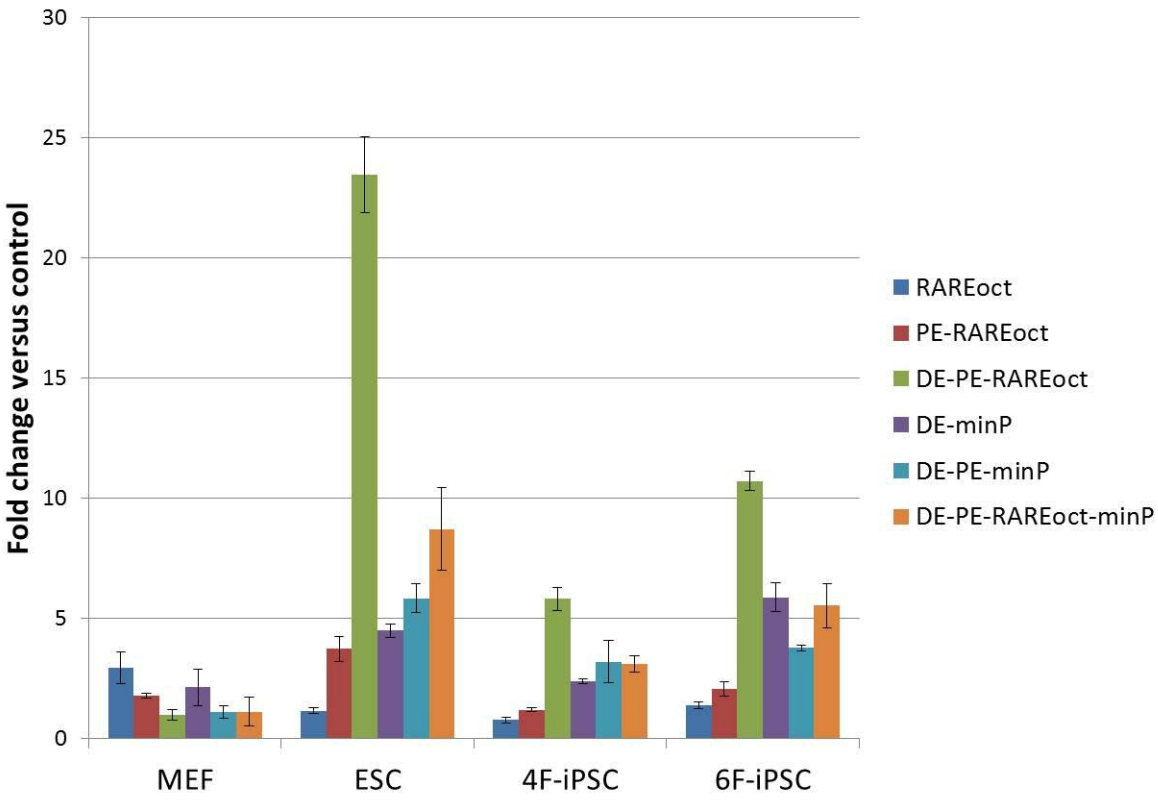



**Figure 5.23 Activity of luciferase constructs in unmodified 4F and 6F iPSCs** To determine the activity of luciferase constructs depicted in **Figure 5.6**, 10μg of each construct was co-electroporated with 1μg of Renilla luciferase driven by a constitutive TK promoter. Each coloured bar denotes a luciferase construct. The constructs were introduced into MEFs, ESCs and 4F and 6F iPSCs, which were obtained as described in **Figure 5.19**. The firefly luciferase values were normalised to their corresponding Renilla readings and respective empty vector. Normalised values are indicated on the bar chart.

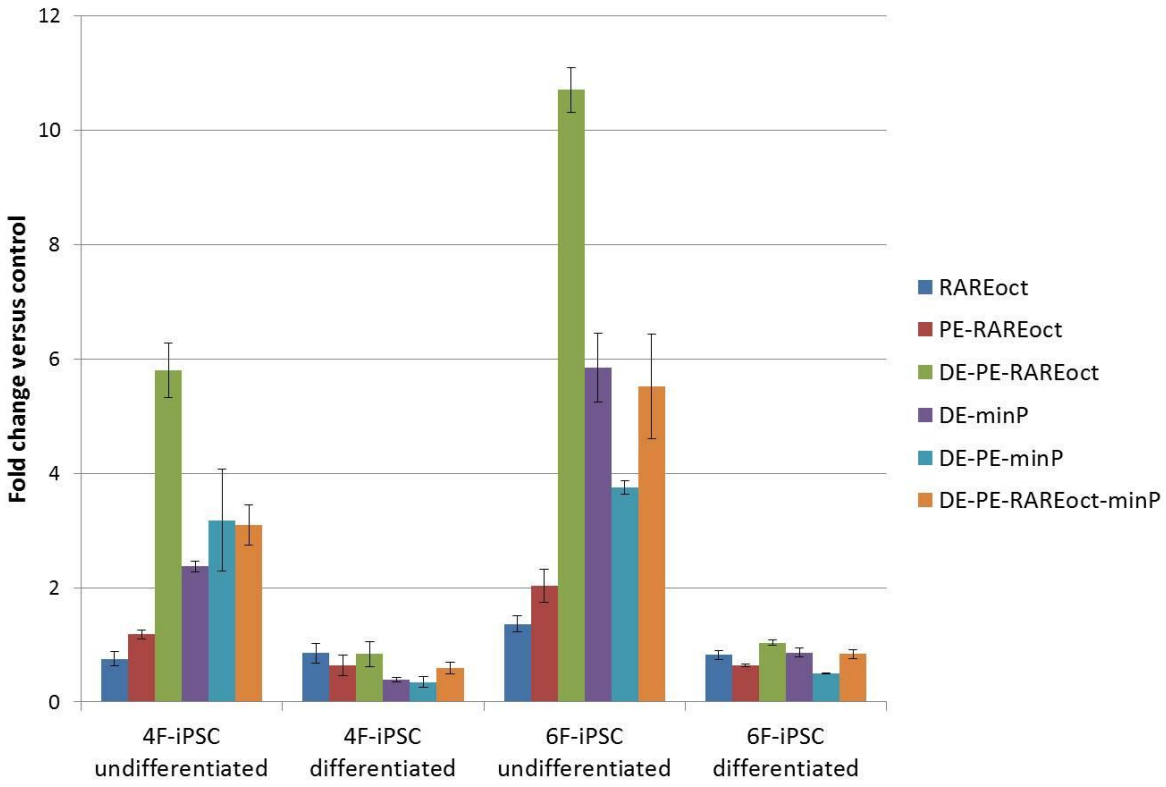



**Figure 5.24 Differentiation of 4F and 6F iPSCs reduces luciferase activity** To ensure that 6 days of 0.1μM ATRA treatment was sufficient to induce differentiation, luciferase constructs listed in **Figure 5.6** were introduced into both undifferentiated and differentiated iPSCs. 10μg of each luciferase construct was co-electroporated with 1μg of Renilla contstruct into undifferentiated and differentiated derivatives of 4F and 6F iPSCs. Firefly luciferase readings were normalised to Renilla luciferase values and the respective empty vector. Normalised values are shown on the bar chart.

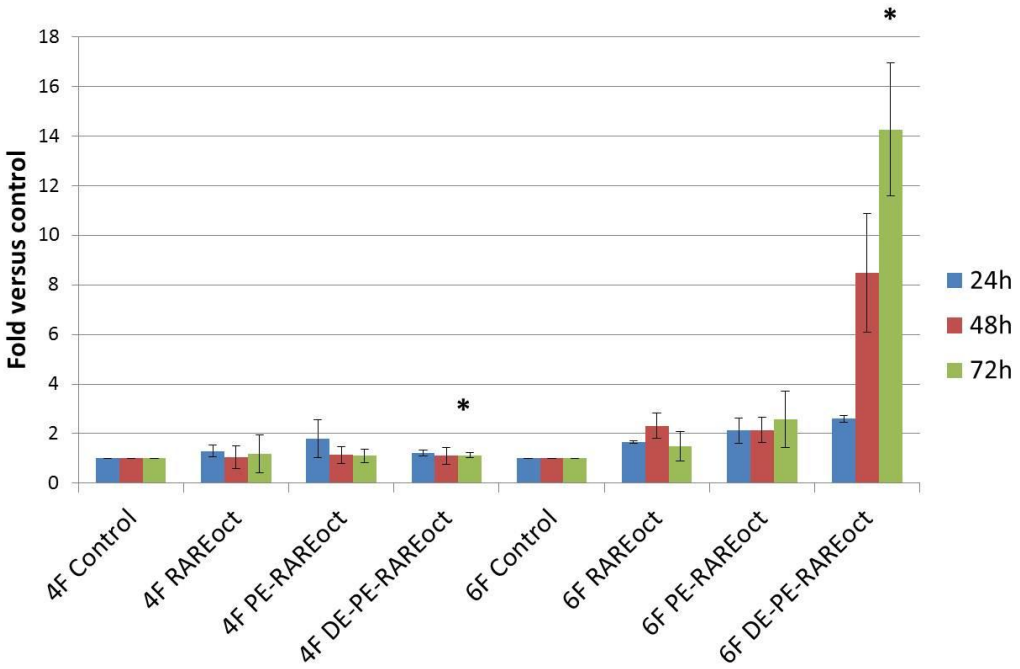



**Figure 5.25 Oct4 Distal Enhancer and Proximal Promoter are activated within 24h of ectopic transgene expression** To determine the onset of activation at the distal enhancer, proximal enhancer and proximal promoter of Oct4, luciferase constructs as labelled on the horizontal axis were introduced into differentiated 4F and 6F iPSCs. Transfected cells were exposed to DOXYCYCLINE for 24, 48 and 72hours until luciferase activity was assayed. Renilla luciferase constructs were co-introduced with the luciferase reporter vectors for normalisation. Values observed in the bar chart reflect normalisation to Renilla luciferase readings, empty pGL3 vector and activity in the absence of DOXYCYCLINE. Statistical analysis was performed using a student's T-test. *$p<0.05$

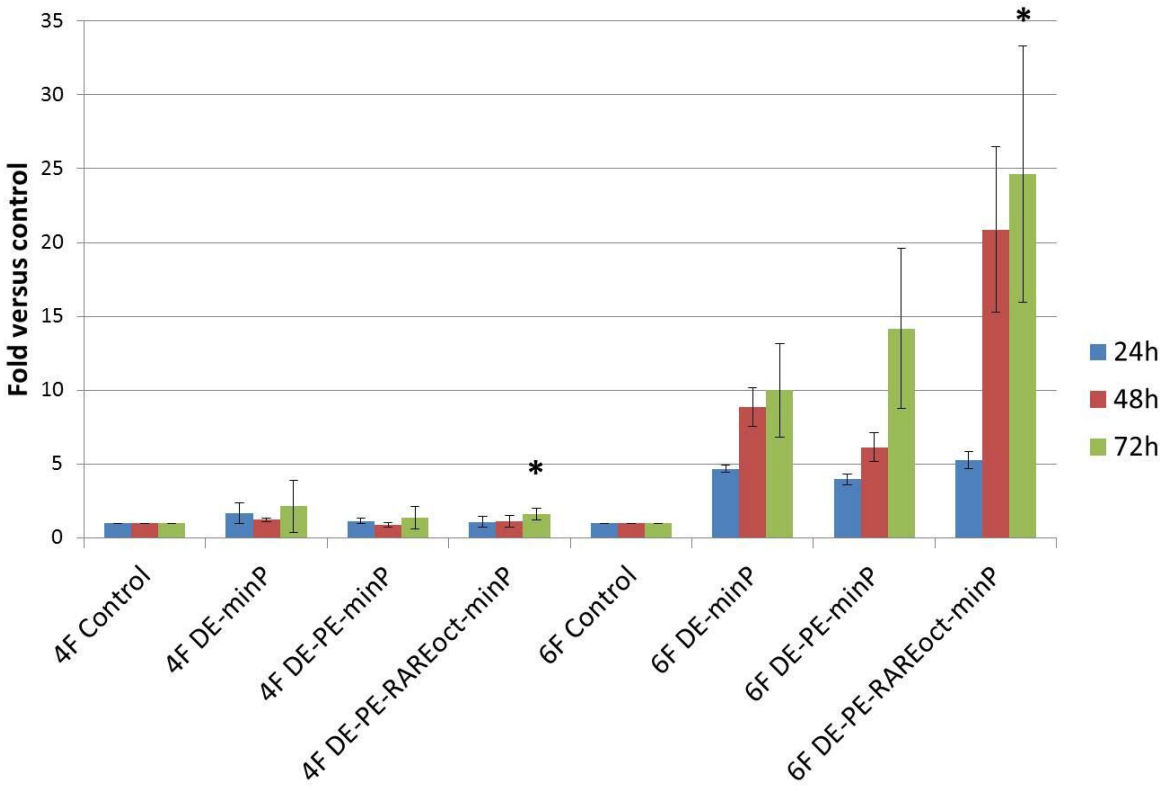



**Figure 5.26 Oct4 Distal Enhancer and Proximal Promoter are activated within 24h of ectopic transgene expression** To determine the onset of activation at the distal enhancer, proximal enhancer and proximal promoter of Oct4, luciferase constructs as labelled on the horizontal axis were introduced into differentiated 4F and 6F iPSCs. Transfected cells were exposed to doxycycline for 24, 28 and 72hours until luciferase activity was assayed. Renilla luciferase constructs were co-introduced with the luciferase reporter vectors for normalisation. Values observed in the bar chart reflect normalisation to Renilla luciferase readings, empty pGL4.23 vectors and activity in the absence of doxycycline. Statistical analysis was performed using a student's T-test. *$p<0.05$

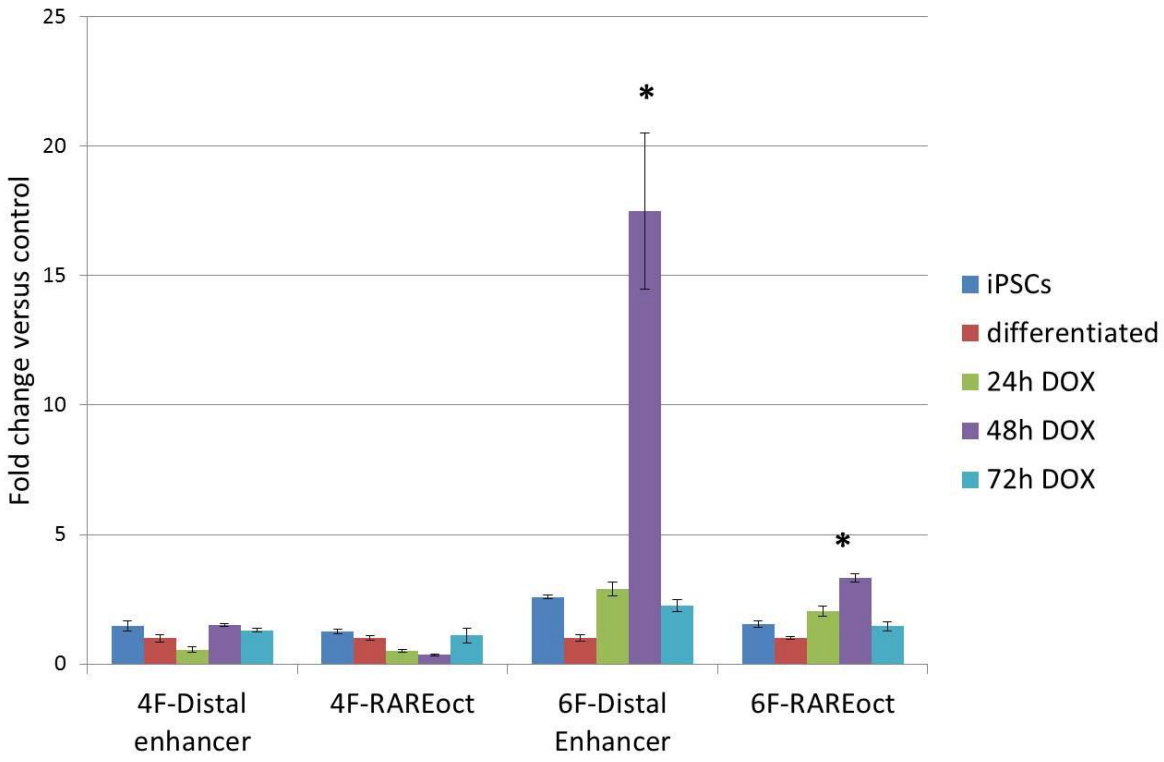



**Figure 5.27 Oct4 Distal Enhancer and Proximal Promoter are associated to activating histone marks within 24h of ectopic transgene expression** To determine the epigenomic landscape of the Oct4 distal enhancer and proximal promoter during reprogramming, differentiated 4F and 6F iPSCs were exposed to doxycycline for 24, 48 and 72 hours. After incubation for the respective lengths of time, chromatin immunoprecipitation of H3K4me3 activating marks was performed. Using the eluate, qPCR was performed using primers against the distal enhancer or RAREoct, as indicated on the bar chart. Readings were normalised against IgG and plotted on the bar chart. Statistical analysis were performed using a student's T-test. *$p<0.05$

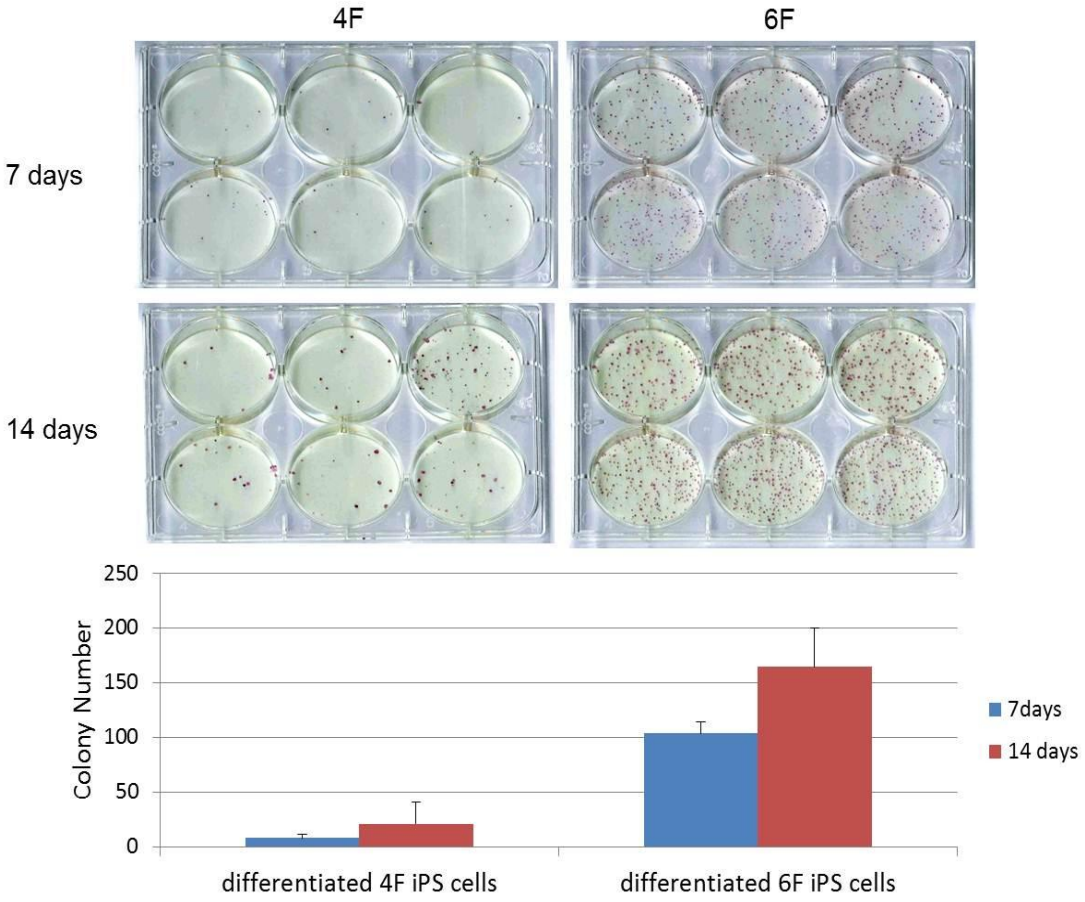



**Figure 5.28 Comparison of reprogramming efficiencies in differentiated 4F and 6F iPSCs** iPSCs generated through the introduction of 4F and 6F in a doxycycline dependent manner were differentiated for 6 days in ATRA. Once differentiated, 50 x $10^3$ cells were seeded into each well of a 6 well plate. The cells were exposed to 1μg/ml doxycycline to re-express 4F or 6F reprogramming factors. (Top) After 7 or 14 days, the reprogrammed cells were stained for the expression of AP. The expression of AP is indicated by the colour red. (Bottom) Quantification of colony numbers.

### 5.2.8 6F triggers rapid expression of GFP controlled by endogenous Oct4 regulatory elements

Previous findings in this chapter delineated the role of RARγ and LRH1 in the activation of the Oct4 distal enhancer and RAREoct within 24 hours of reprogramming. To this end, a significant proportion of supporting evidence originated from the employment of luciferase assays. However, analysis of transcriptional activity using luciferase constructs which are ectopically introduced poorly reflects the dynamics that pertain to chromatin architecture at the endogenous locus. As such, it is imperative to draw parallels to transcriptional activity at the endogenous locus. To address this, MEFs that encompass a transgene consisting of Oct4 regulatory elements spanning 18kb controlling EGFP expression (Oct4-EGFP MEFs) were employed to generate a new population of 4F and 6F iPSCs (Ying et al., 2002). As such, EGFP expression acts as a surrogate to assess transcriptional activation of endogenous Oct4. 4F and 6F were introduced into Oct4-EGFP MEFs in a doxycycline inducible manner and reprogrammed colonies were expanded in the absence of doxycycline. Resultant iPSCs were designated as 4F-O-iPSCs and 6F-O-iPSCs respectively (**Figure 5.29**).

To determine if expression of RARγ and LRH1 leads to rapid activation of the endogenous Oct4 locus, a similar strategy to **Chapter 5.2.7** was employed to enable fair comparisons. 4F-O-iPSCs and 6F-O-iPSCs were differentiated using ATRA over 6 days before doxycycline was employed to trigger re-expression of the reprogramming factors. As EGFP expression acts as an indicator for Oct4 activity, cells which were exposed to doxycycline were observed under the microscope for EGFP expression.

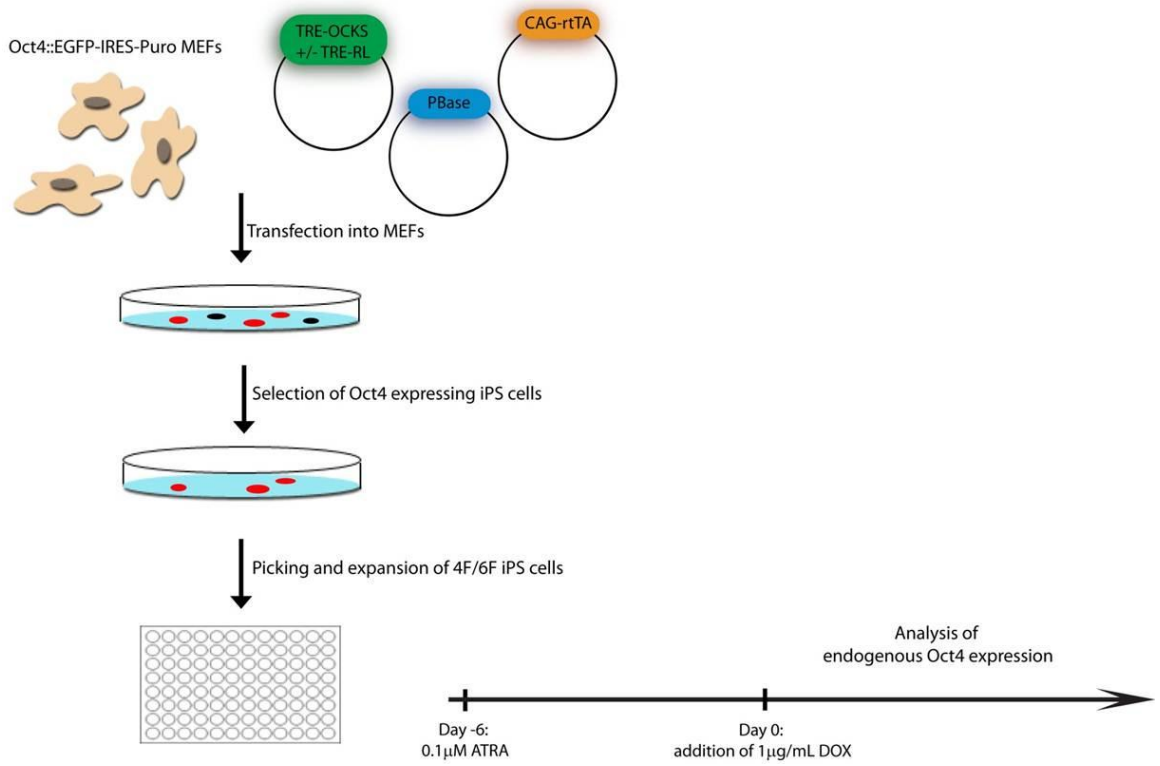



**Figure 5.29 Experimental Strategy to observe expression of endogenous Oct4 during reprogramming** MEFs which carry genes encoding EGFP expression and Puromycin resistance under the endogenous Oct4 promoter were innoculated with either 4F or 6F in a doxycycline inducible manner. Resultant iPSCs (4F-O-iPSCs and 6F-O-iPSCs) which were transcriptionally active at their endogenous Oct4 loci were selected for using puromycin. Surviving cells were picked and expanded. Differentiation was induced using 0.1μM ATRA for 6 days before doxycycline was introduced. Re-expression of Oct4 was tracked under the microscope over the course of doxycycline treatment.

As a control, it was first determined if 4F-O-iPSCs and 6F-O-iPSCs exhibited faithful recapitulation of endogenous Oct4 expression as denoted by EGFP expression. **Figure 5.30** illustrates homogenous iPSC colonies that express GFP, reiterating high levels of Oct4 expression. In addition, 4F-O-iPSCs and 6F-O-iPSCs express pluripotency markers in a similar pattern to ESCs (**Figure 5.31**), indicating the attainment of pluripotency. To omit the possibility of contamination between 4F-O-iPSCs and 6F-O-iPSCs, genomic material from both populations was inspected for the presence of incongruent transgenes. **Figure 5.32** illustrates that 4F-O-iPSCs and 6F-O-iPSCs consist of the right combination of transgenes, reaffirming the integrity of the iPSCs.

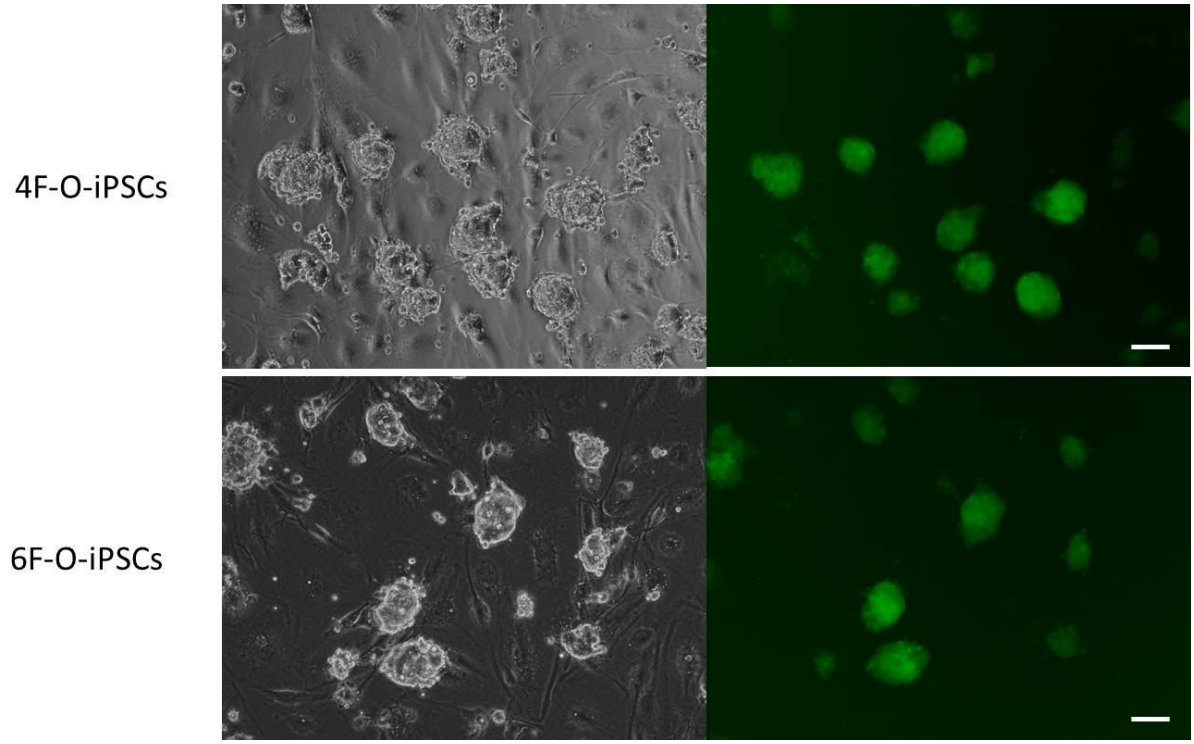



**Figure 5.30 Microscope images of 4F-O-iPSCs and 6F-O-iPSCs** iPSCs were obtained as described in **Figure 5.25.** iPSCs generated from the introduction of 4F and 6F into Oct4::EGFP-IRES-Puro MEFs were designated as 4F-O-iPSCs and 6F-O-iPSCs respectively. The cells were maintained in 2i-LIF growth media, in the absence of doxycycline. The images on the left reflect their morphology in brightfield, whereas the images on the right demonstrate EGFP expression from the endogenous Oct4 locus. Scale Bar: 50μm.

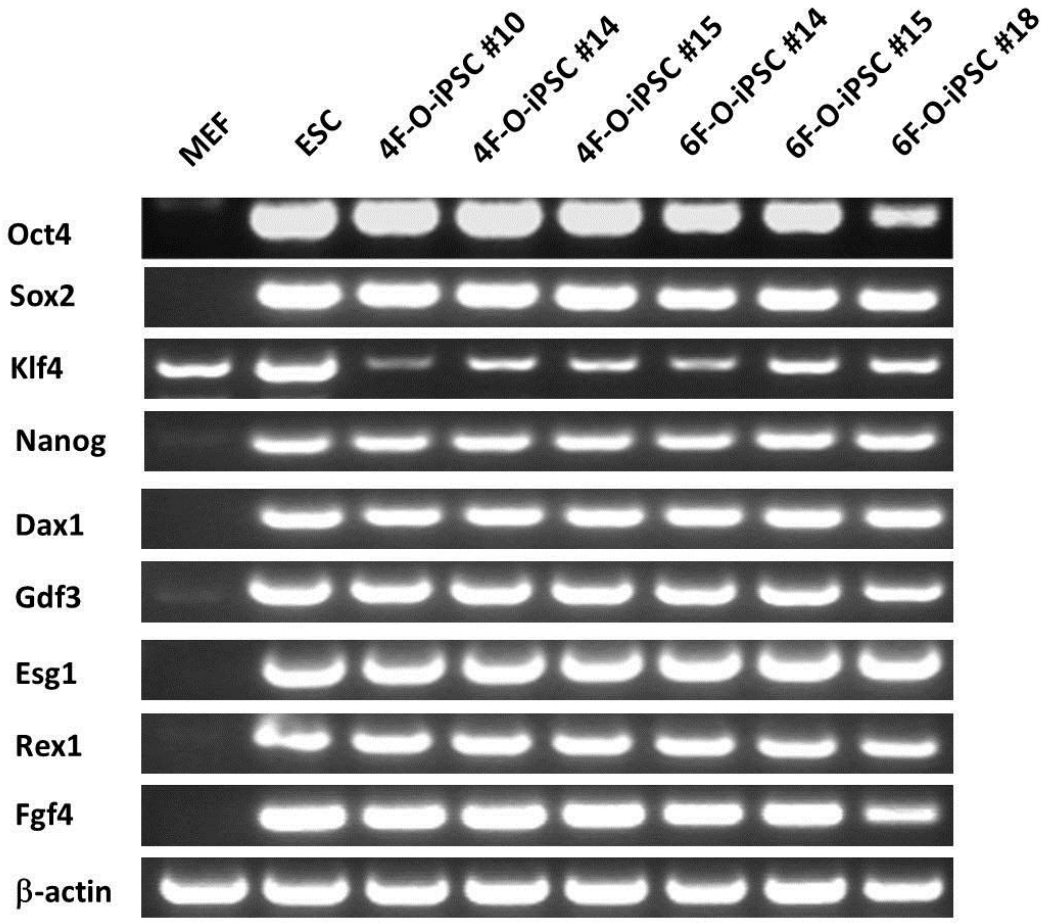



**Figure 5.31 Expression levels of Pluripotency Markers in 4F-O-iPSCs and 6F-O-iPSCs** RT-PCR of the various pluripotency markers was performed in MEF (lane1), ESCs (lane2), 4F-O-iPSCs (lanes 3-5), and 6F-O-iPSCs (lanes 6-8). Expression levels are described qualitatively by the intensity of the amplified products viewed using an agarose gel. The array of pluripotency markers tested is labelled to the left of the image.

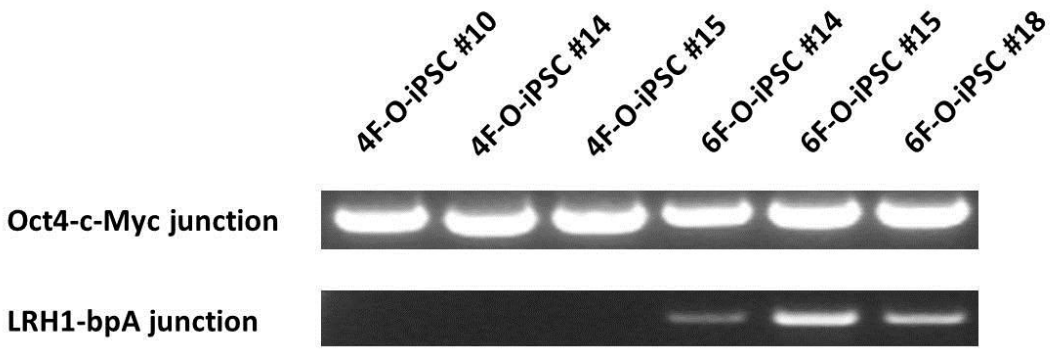



**Figure 5.32 Presence of transgenes in 4F-O-iPSCs and 6F-O-iPSCs** To ensure that the right combination of transgenes was present in 4F-O-iPSCs and 6F-O-iPSCs, genomic DNA was retrieved from 3 clones of 4F-O-iPSCs and 3 clones of 6F-O-iPSCs. Primers specific to the transgenes were designed. To detect insertions containing the vector encompassing 4F, primers against the junction of Oct4 and c-Myc were desgined. To detect insertions containing the vector containing RARγ and LRH1, primers were designed against the junction corresponding to LRH1 and bpA. PCR was performed using the primers indicated to the left and the amplified products were visualised using an agarose gel.

Having ascertained that 4F-O-iPSCs and 6F-O-iPSCs exhibited pluripotency, it was essential to ensure that differentiation of both iPSC populations would result in ablation of EGFP expression. As depicted by **Figure 5.34** and **Figure 5.35**, Oct4 expression levels were drastically reduced in the presence of ATRA, exemplifying the dependability of the system to study the effect of RARγ and LRH1 on the re-emergence of Oct4 expression.

As part of the experimental design, differentiated 4F-O-iPSCs and 6F-O-iPSCs were exposed to doxycycline to trigger re-expression of the reprogramming factors. However, as observed in **Figure 5.19**, differentiation of pluripotent cells may induce modifications which compromise transgene re-expression. To address this possibility, transgene re-expression in differentiated 4F-O-iPSCs and 6F-O-iPSCs was assessed. **Figure 5.33** demonstrates efficient re-expression of the reprogramming transgenes within 24 hours of doxycycline treatment, strengthening the reliability of the strategy chosen for the study of reprogramming kinetics in the presence of RARγ and LRH1.

Having obtained 4F-O-iPSCs and 6F-O-iPSCs which differentiate readily and exhibit the capacity for transgene re-activation, both iPSC lines were compared to determine if RARγ and LRH1 exhibited an effect on the expression of endogenous Oct4 during the initiation of reprogramming. To this end, 4F-O-iPSCs and 6F-O-iPSCs were differentiated in the presence of 0.1μM ATRA over 6 days before being exposed to doxycycline to re-establish transgene expression. Cells incubated in doxycycline were observed under a microscope at 24 hour intervals over a span of three days to capture transcriptional activation of the Oct4 locus, as indicated by green fluorescence. **Figure 5.34** and **Figure 5.35** depict representative images of the reprogramming events that occurred. In both populations of cells that harbour 4F and 6F, cells were amenable to morphological transformations within 72 hours, where an epithelial-like conformation was adopted. In contrast, EGFP expression was observed in reprogramming intermediates of differentiated 6F-O-iPSCs at 48 hours but not in equivalent 4F counterparts, suggesting that endogenous Oct4 transcription was activated at an earlier time due to the presence of RARγ and LRH1, thus conforming to previous findings.

As microscope images reflect the re-activation of Oct4 transcription at single cell resolution, it is pivotal to determine if a similar phenomenon is observed in a larger population of cells. To address this concern, differentiated 4F-O-iPSCs and 6F-O-iPSCs that had been exposed to doxycycline over a period of 1, 3 and 6 days were collected and analysed. **Figure 5.36** depicts a qualitative perspective of endogenous Oct4 transcript levels during the reprogramming of differentiated 4F-O-iPSCs and 6F-O-iPSCs. Concurring with previous observations, Oct4 transcripts were detected at 24 hours after doxycycline treatment of differentiated 6F-O-iPSCs, whereas the presence of Oct4 transcripts was detected at 72 hours after doxycycline induction in differentiated 4F-O-iPSCs. Overall, these findings reaffirm that co-operation between RARγ and LRH1 results in rapid activation of the distal enhancer of Oct4 and RAREoct, prompting transcriptional activity.

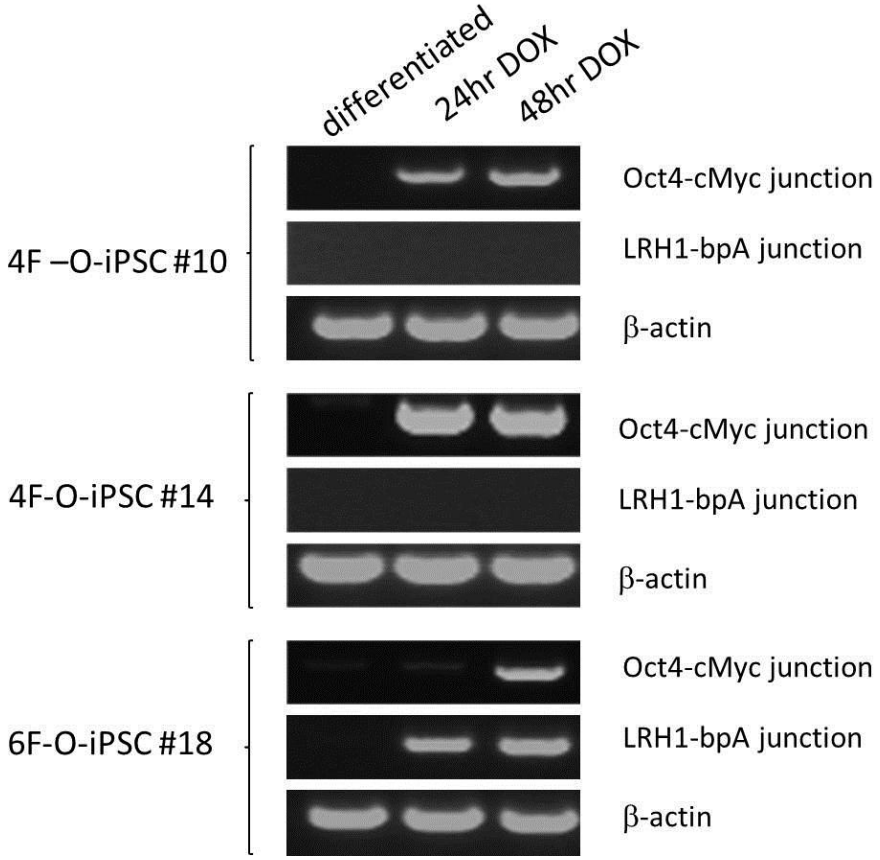



**Figure 5.33 Re-activation of transgenes in differentiated 4F-O-iPSCs and 6F-O-iPSCs** 4F-O-iPSCs and 6F-O-iPSCs were differentiated in the presence of 0.1μM ATRA over 6 days. To ensure that transgenes are readily re-expressed upon the addition of doxycycline, differentiated cells were seeded onto 6 well plates and 1μg/mL doxycycline was introduced for 24 and 48 hours. Cells were collected and RNA was extracted and converted to cDNA. PCR was performed using cDNA as a template to determine transcript amounts as recognised by primers specific to the junctions between cDNAs present only in the transgene.

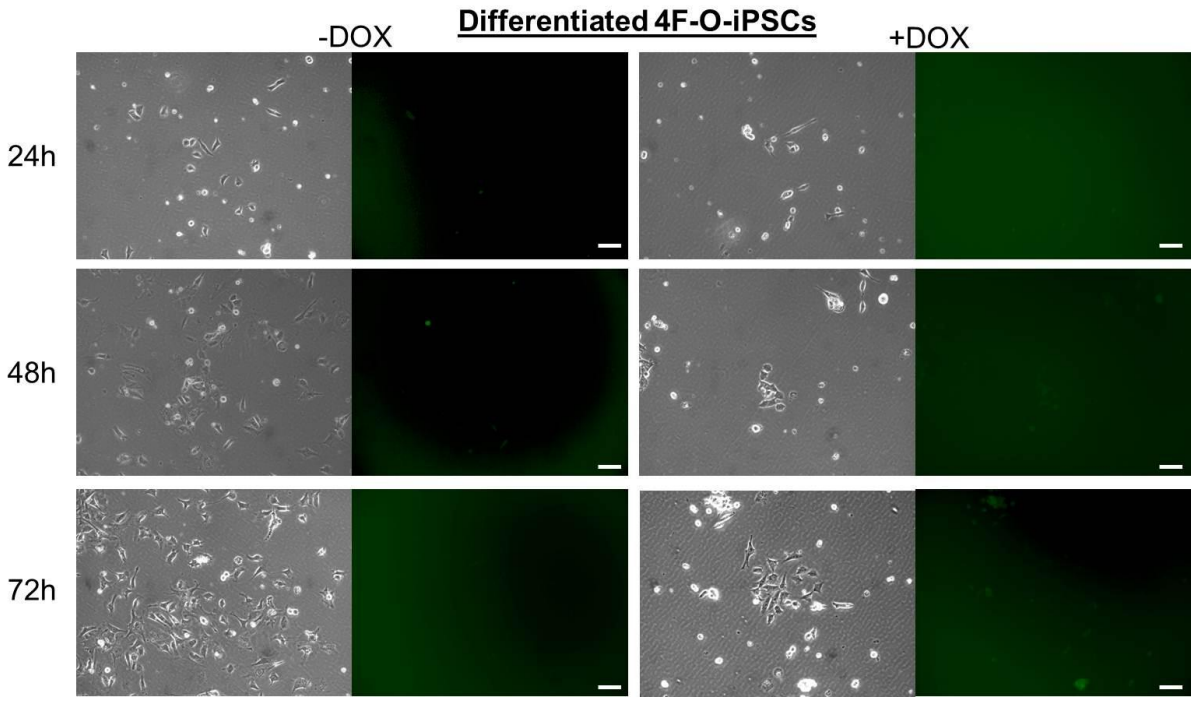



**Figure 5.34 Re-establishment of Oct4 expression through the addition of doxycycline to differentiated 4F-O-iPSCs** 4F-O-iPSCs were differentiated using 0.1μM ATRA for 6 days. 1μg/ml doxycycline (DOX) was added to the differentiated cells over a period of 72 hours. Changes in morphology and the re-expression of endogenous Oct4, as depicted by green fluorescence (200x), were traced every 24 hours. As a control, cells were grown in ESC media, in the absence of doxycycline. Scale bar: 50μm.

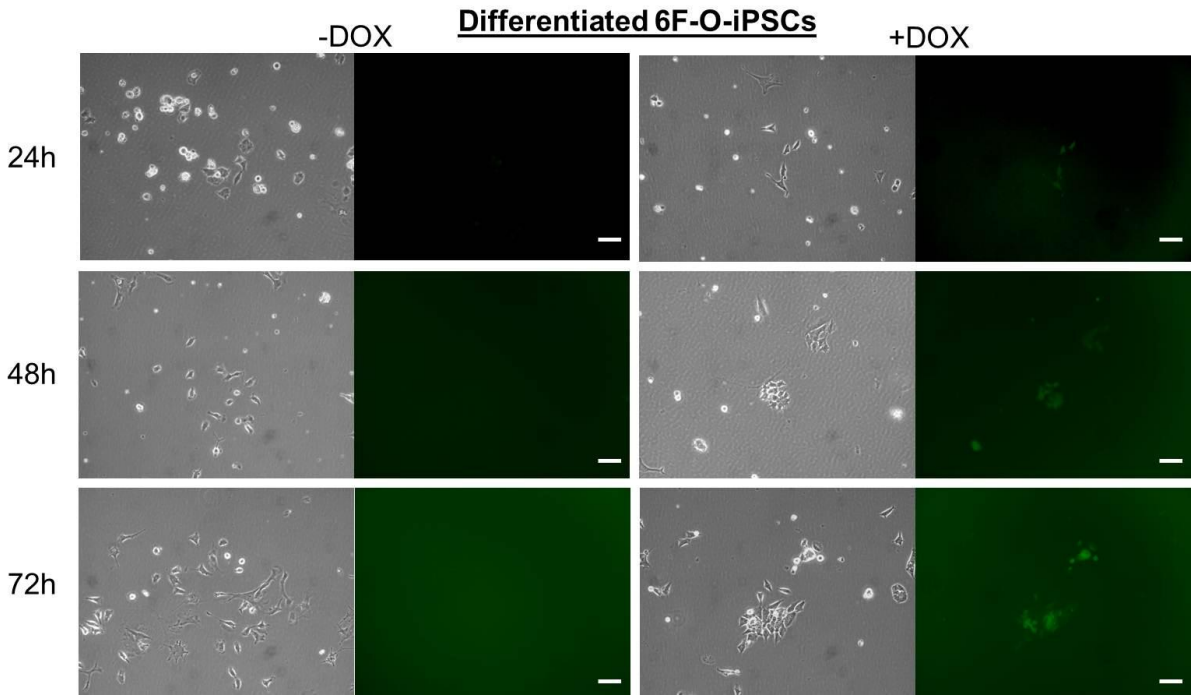



**Figure 5.35 Re-establishment of Oct4 expression through the addition of doxycycline to differentiated 6F-O-iPSCs** 6F-O-iPSCs were differentiated using 0.1μM ATRA for 6 days. 1μg/ml doxycycline (DOX) was added to the differentiated cells over a period of 72 hours. Changes in morphology and the re-expression of endogenous Oct4, as depicted by green fluorescence (200x), were traced every 24 hours. As a control, cells were grown in ESC media, in the absence of doxycycline. Scale bar: 50μm.

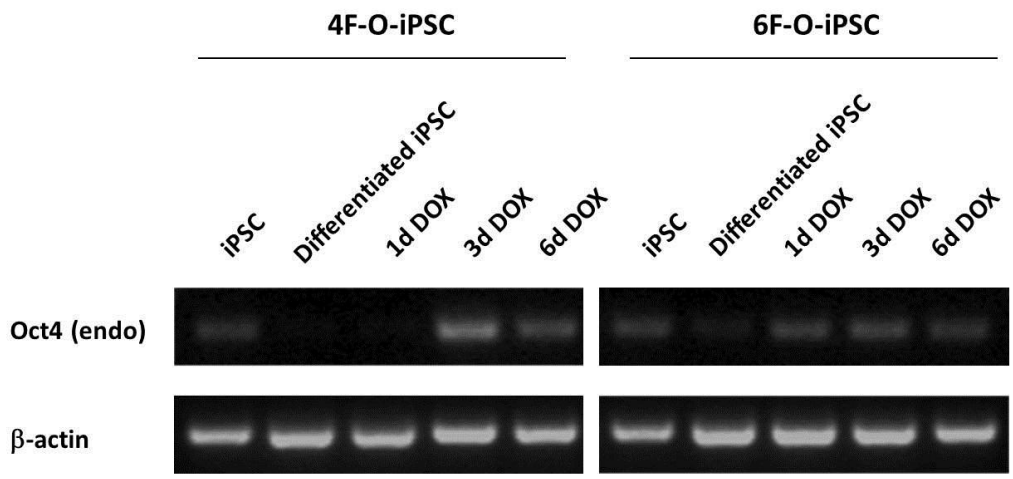



**Figure 5.36 Analysis of Oct4 transcript levels after the addition of doxycycline to differentiated 4F-O-iPSCs and 6F-O-iPSCs** 4F-O-iPSCs and 6F-O-iPSCs were differentiated using 0.1μM ATRA for 6 days. 1μg/ml doxycycline (DOX) was added to the differentiated cells over a period of 6 days. RNA was collected after 1, 3 and 6 days after doxycycline induction and the presence of Oct4 transcripts was assessed through PCR amplification using primers specific to endogenous Oct4. β-actin levels were also assessed as an internal control.

## 5.3 Discussion

### 5.3.1 Epigenetic changes in Oct4 regulatory elements

Chromatin remodelling has been portrayed to play a significant role in the initiation of the reprogramming process, as delineated by the profound effect elicited by chemicals which modify the epigenomic landscape (Mikkelsen et al., 2008; Huangfu et al., 2008a). Furthermore, components that participate in Brahma-related gene 1-Associated Factors (BAF) chromatin remodelling processes and Activation-Induced cytidine Deaminase (AID) dependent demethylation have been described to enhance reprogramming efficiencies (Singhal et al., 2010; Bhutani et al., 2010). In addition, impairment of transcriptionally repressive histone modifications by ablation of Utx1 has been described to affect pluripotency, cell fate decisions and reprogramming (Welstead et al., 2012; Mansour et al., 2012). Several studies have also depicted discrepancies in epigenetic profiles exhibited by iPSCs and ESCs, surmising a relationship between epigenetic disparities and variations in the pluripotent nature of reprogrammed cells (Lister et al., 2011; Stadtfeld et al., 2010).

Although it is evident that genome wide epigenomic alterations are pivotal during the acquisition of pluripotency, precise epigenetic changes have not been dissected. This chapter proposes that an initial amendment to the epigenomic landscape occurs at the distal enhancer of Oct4 and RAREoct. Activation of both regulatory elements was demonstrated to be facilitated by RARγ and LRH1, leading to the enhancement of reprogramming speed and efficiencies. Remarkably, the influence of RARγ and LRH1 ensues within 24 hours of reprogramming. This was examined from distinct perspectives, employing chromatin immunoprecipitation experiments which inspect transcriptionally permissive histone modifications at endogenous gene loci, to luciferase assays which assess the functional aspects of the regulatory elements. Ultimately, endogenous Oct4 transcriptional activity was assessed through the measurement of endogenous Oct4 transcripts and visualisation of EGFP as a marker. Results from these experiments concur and suggest that the presence of RARγ and LRH1 leads to rapid activation of the distal enhancer of Oct4 and RAREoct, resulting in efficient transcription of Oct4.

The distal enhancer of Oct4 and RAREoct were originally discovered due to their sensitivity to RA treatment (Schoorlemmer et al., 1994; Pikarsky et al., 1994; Okazawa et al., 1991; Sylvester and Schöler, 1994; Barnea and Bergman, 2000; Minucci et al., 1996). Exploration of the regulatory elements revealed that the distal enhancer influences Oct4 transcription in naïve pluripotent cells such as ESCs, but not EC cells or EpiSCs (Yeom et al., 1996; Tesar et al., 2007; Bao et al., 2009; Han et al., 2010a). Conversely, the proximal promoter plays a role in driving Oct4 expression in primed pluripotent cells such as EpiSCs (Tesar et al., 2007; Guo et al., 2010; Han et al., 2010a). As the generation of iPSCs entails the conversion of somatic cells to ESC-like derivatives, this chapter proposes that activation of the distal enhancer of Oct4 is critical in the initial stages of reprogramming. The addition of RARγ and LRH1 to the reprogramming mix was demonstrated to elicit a selective effect on the distal enhancer but not proximal enhancer of Oct4, thereby improving reprogramming conditions and reinforcing the significance of distal enhancer during the acquisition of naïve pluripotency.

On a separate note, studies on RAREoct have delineated positive and negative regulatory components which exert effects on Oct4 expression levels (Ben-Shushan et al., 1995; Fuhrmann et al., 1999; Fuhrmann et al., 2001). The RAR family has been portrayed to positively and negatively regulate the proximal promoter of Oct4 (Ben-Shushan et al., 1995; Barnea and Bergman, 2000; Sylvester and Schöler, 1994). As the locus is heavily occupied by regulators, the multifunctional role of RARs could be narrowed down to co-operating partners. This notion is reinforced by observations that the inclusion of LRH1 into various reprogramming mixes containing either RARγ or RARβ enhanced Oct4 transcriptional activity and reprogramming efficiencies. In addition, as ligand independent effects of RARs have been described (Ross-Innes et al., 2010; Laursen et al., 2012), it is plausible that the reprogramming capacities of RARs may be partly attributable to ligand independence. As RARα has been described to co-occupy regulatory regions with ER (Ross-Innes et al., 2010), it will be interesting to investigate the plausible role of ER signalling during somatic cell reprogramming. It has been established that Estrogen

Receptor Related Receptor-beta (Esrrb) can mediate reprogramming events in the mere presence of Oct4 and Sox2 (Feng et al., 2009). Although Esrrb remains to be an orphan receptor, these findings implicate the possibility of an estrogen related mechanism during the acquisition of pluripotency.

This chapter identified epigenetic modifications at Oct4 regulatory elements which accompany the initiation of reprogramming events. This was determined through the employment of chromatin immunoprecipitation of H3K4me3, followed by site specific quantification. Histone modifications allow an epigenetic snapshot of a genomic locus and have been demonstrated to embody dynamic changes during retinoic acid-mediated differentiation (You et al., 2011a). Trimethylated forms of histone H3 at lysine 4 (H3K4me3) have been identified at transcriptional start site of active promoters (Kim et al., 2005; Heintzman et al., 2007; Roh et al., 2005). By examining genomic regions which associate with H3K4me3, genes which are transcriptionally active can be isolated. To exploit this phenomenon, H3K4me3 was employed to examine molecular changes at the regulatory elements of Oct4. However, mono- and di-methylation at lysine 4 of histone 3 (H3K4me1 and H3K4me2) have been described to possess predictive power for the activation of enhancer regions (Kim et al., 2005; Heintzman et al., 2007). Enhancers in Oct4 regulatory regions are no exception and have been described to associate with H3K4me1 in ESCs (Whyte et al., 2012). As a result, examination of genomic regions encompassing H3K4me1/2 marks during the reprogramming process triggered by 4F and 6F may provide better insight to the activation status of the Oct4 enhancers. Nevertheless, the versatility of chromatin often entails looping between enhancer and promoter regions (Kagey et al., 2010) and analysis of the genomic locus corresponding to Oct4 in ESCs describes common components observed in promoter and enhancer regions, illustrating a native looped conformation in ESCs. With this in mind, H3K4me3 activating histone modifications are indicative of transcriptionally active gene loci, regardless of enhancer or promoter elements, and were employed in this study.

### 5.3.2 Identification of RAR-associated genomic regions

RARs were originally identified through homology to glucocorticoid and thyroid hormone receptors (Dejean et al., 1986; Giguere et al., 1987). Steroid hormone receptors possess DNA binding domains which recognise the canonical sequence (A/G)G(G/T)TC(A/G) and its variants. Traditional methods to identify RAR-regulated genes include analysis of transcriptional profiles stimulated by retinoic acid treatment, deletion mapping to identify the binding region, and predictions using computational techniques. With the advent of chromatin immunoprecipitation techniques coupled to sequencing (ChIP-seq), genome wide exploration of transcription factor binding events can be performed (Nielsen et al., 2008). To this end, genomic regions bound to RARα and RARγ have been elucidated using ChIP-seq (Delacroix et al., 2010; Ross-Innes et al., 2010). In contrast, there has been no attempt at delineating genome-wide RARβ binding regions. This study demonstrates an execution of ChIP using antibodies that recognise RARβ. The cis-regulatory elements of Oct4 were identified and verified using 3730 Sanger sequencing of amplified products corresponding to the regions-of-interest. With broader interest in genome-wide targets of RARβ, it will be intriguing to perform deep sequencing of eluates obtained from immunoprecipitation of RARβ. These findings may illuminate additional roles of RARβ in its ability to replace the requirement of exogenous Oct4 during the initiation of reprogramming events. Moreover, RARβ but not RARγ was identified from the genome wide screen performed in **Chapter 3**. Although this observation could be attributable to technical limitations, it is plausible that RARβ may participate in alternative pathways, or possess distinct interacting partners that facilitated its identification during the screen. To reinforce this notion, it has been established that RARβ and RARγ display specific expression patterns during development (reviewed by Dolle, 2009), suggesting that both RARs exhibit distinct cellular functions.

### 5.3.3 Implications of delineating transcriptional regulation of Oct4

Oct4 exists in a fleeting manner during development, where zygotic expression begins before 3.5dpc and abrogates between 4.5dpc and 10.5 dpc in somatic cells or in the germline respectively. Repression of Oct4 expression marks the first cell fate decision

where differentiation occurs to adopt a trophectoderm fate (Palmieri et al., 1994; Nichols et al., 1998). Deregulation of Oct4 expression levels leads to corrupted developmental cues, resulting in a spectrum of phenotypes, ranging from embryonic lethality to inappropriate differentiation accompanied by the loss of pluripotency (Nichols et al., 1998; Niwa et al., 2000). During the acquisition of pluripotency, the significance of ectopic expression of Oct4 is unequivocal (Nakagawa et al., 2008; Kim et al., 2009b; Kim et al., 2009c). The importance of Oct4 has been suggested to be attributable to its ability to elicit transcriptional activation (Hammachi et al., 2012). Findings from this chapter reiterate the importance of Oct4 two-fold. First, c-Myc, Klf4 and Sox2 alone were unable to elicit reprogramming events and second, silencing of Oct4 in ZHBTc4 cells abrogates the ability to acquire pluripotency. In addition to the generation of iPSCs, Oct4 has also been described to play roles in lineage switching (Efe et al., 2011; Szabo et al., 2010). As nuclear reprogramming entails immense amounts of chromatin remodelling, it will be interesting to inspect molecular changes at the regulatory elements of endogenous Oct4 during alternative methods of nuclear reprogramming and draw parallels to findings from this study.

### 5.3.4 Alternative platforms to study Oct4 activation during reprogramming events

The experimental strategy employed in this chapter focuses on secondary reprogramming events originating from differentiated iPSCs. A drawback of this method is exemplified by remnants of pluripotent cells which are recalcitrant to retinoic acid driven differentiation. The presence of these cells may mask the evaluation of obtained results. To circumvent this, various platforms can be utilised.

In order to obtain a homogenous platform to study the kinetics of reprogramming, secondary MEFs can be derived from chimeras derived from the injection of iPSCs in mouse blastocysts. Cells generated through this method would have been exposed to various developmental cues and no residual pluripotent cells should persist. To this end, several studies have employed the use of secondary MEFs to study the reprogramming process in detail (Hanna et al., 2009; Samavarchi-Tehrani et al., 2010). On the other hand,

secondary MEFs have participated in embryonic development and possess inherent barriers to reprogramming cues. This is demonstrated by the forced re-expression of reprogramming factors in secondary fibroblasts, where 20% of cells displayed early pluripotency markers and a mere 1.2% of these cells re-activated their endogenous Nanog locus after 16 days (Mikkelsen et al., 2008). As this chapter describes rapid epigenetic changes at the regulatory elements of Oct4, analysis of transcriptionally permissive histone modifications, H3K4me3, may not be reliably recapitulated using a population of secondary MEFs which exhibits resistance to reprogramming.

From another perspective, it will be interesting to explore if rapid activation of the Oct4 distal enhancer is observed during reprogramming of primary fibroblasts. However, heterogeneity in starting material may impede the ability to assess epigenetic changes. It has been suggested that reprogramming occurs in a stochastic fashion (Hanna et al., 2009) and various sources of starting material possess inherently distinct reprogramming efficiencies (reviewed by Hochedlinger and Plath, 2009). These confounding factors will pose difficulties in the examination of Oct4 epigenetic changes in a sub-population of cells undergoing nuclear reprogramming.

With these difficulties in mind, experiments conducted in this chapter serve as a building block to decipher epigenetic changes that accompany the acquisition of pluripotency. Differentiation of iPSCs using ATRA allows a consistent treatment to a population of cells, producing a powerful platform to study initial events during the course of reprogramming.

### 5.4 Conclusion

This chapter attempts to decipher the molecular mechanisms that explicate the ability of RARβ to replace exogenous Oct4 during the acquisition of pluripotency. Bioinformatics and experimental data have described several Retinoic Acid Response Elements (RAREs) in the enhancer and promoter regions of Oct4. Given that RAREs are motifs recognisable

by RARs, ChIP and luciferase assays were performed to demonstrate that RARβ binds to the RARE within the distal enhancer of Oct4 and activates Oct4 expression. Similar attributes were mirrored in RARγ, another RAR family member, which has been described to promote reprogramming speed and efficiencies in synergy with LRH1. RARγ also had the capacity to positively regulate Oct4 levels through the Oct4 distal enhancer. Moreover, a combination of RARγ, LRH1, Oct4, c-Myc, Klf4 and Sox2 (6F) represents the most efficient way to activate Oct4 expression in the luciferase assay.

As the 6F reprogramming cocktail (RARγ, LRH1, Oct4, c-Myc, Klf4, Sox2) improves the speed, quality, and efficiency of reprogramming, both the distal enhancer and RAREoct were examined during the initial three days of reprogramming to determine rapid epigenetic changes. Using ChIP and luciferase assays, it was demonstrated that both regions were associated to activating histone marks within 24 hours of the ectopic expression of 6F, drawing a parallel to rapid reprogramming. In contrast, the Oct4 distal enhancer remained silent for the first three days when only four reprogramming factors (Oct4, c-Myc, Klf4, Sox2) were used. These findings dissect the kinetics of reprogramming through the observation of epigenomic changes at the Oct4 locus, culminating in the better understanding of molecular events during nuclear reprogramming.

# CHAPTER 6 : GENERAL DISCUSSION

Findings from this thesis can be divided into three components. First, a transposon-mediated genetic screen was performed to identify candidate genes which display the capacity to replace the ectopic requirement of Oct4 during the acquisition of pluripotency. Second, two genetic factors, BNC2 and RARβ, were identified from the genetic screen and exhibited competence in acting as exogenous Oct4 substitutes. Third, molecular dissection of the role played by RARβ to negotiate the need for ectopic Oct4 during the attainment of pluripotent derivatives inadvertently shed light on the ability of RARs to transcriptionally activate the Oct4 distal enhancer and in certain conditions, RAREoct. This activation triggers rapid and efficient reprogramming, providing insight to the epigenetic changes that accompany the acquisition of pluripotency. In this chapter, the three areas will be discussed further, culminating in the future prospects of this thesis.

## 6.1 Transposon assisted mutagenesis serves as a dependable tool to identify genetic candidates

The origin of iPSC technology was attributed to a candidate gene-based screen comprising of 24 ESC specific transcription factors. Although this screen has led to the inception of a landmark discovery in nuclear reprogramming, succeeding studies aimed at expanding the list of factors involved in the maintenance of pluripotency have employed alternative approaches. Diverse methods include chemical screens, siRNA library screens and transcriptional profiling of the reprogramming process (Ichida et al., 2009; Chia et al., 2010; Samavarchi-Tehrani et al., 2010). Transposon mediated mutagenesis is an emerging gene discovery tool (Guo et al., 2010). Initially portrayed as a tool for cancer gene discovery, its potential has diversified to cell-based assays in a spectrum of phenotypic assays (Guo et al., 2010; Guo et al., 2011; Wang et al., 2011b) transcending across species (Li et al., 2011a; You et al., 2011b).

**Chapter 3** reiterates the ability of transposon mutagenesis to isolate genes capable of substituting exogenous Oct4 during the reprogramming process. Unbiased assessment of

genes is exemplified in the observation that most genes incurring transposon integrations are poorly characterised. To embody the reliability of the screen, a significant proportion of genes which received insertion sites represented pathways critical during development. As an evaluation of the dependability of the screen, candidate genes identified were validated through independent experiments. The abilities of BNC2 and RARβ in generating iPSC colonies in the absence of ectopic Oct4 reaffirmed the validity of the genetic screen. On the other hand, of a total of 177 genes identified to harbour integration sites only three genes were validated and two exhibited potency as exogenous Oct4 substitutes. 98.3% of candidate genes were ignored as they did not fulfil the selection criteria of possessing more than one independent insertion site. Keeping in mind that the genetic screen attained genomic coverage of an average of one integration per 10kb of genomic material, a similar genetic screen can be conducted in a larger scale to increase the prevalence of integrations and common insertion sites, ultimately improving the probability of uncovering causal mutations. Alternatively, introduction of a variety of transposons with distinct properties may facilitate saturation of the host genome (Izsvak and Ivics, 2005). However, the detection of integration sites in a transposon-mediated screen relies on several layers of experimental procedures. These include the amplification of transposon arms, insertion of the amplicons into an expression vector and capillary sequencing of the inserted fragments. Technical drawbacks at each stage results in a diminished breadth of integration sites. To circumvent this, next generation sequencing techniques can be employed (reviewed by Metzker, 2010). The low cost and large sequence data volume allows sequencing of genomic DNA and identification of every transposon integration site. This will result in an exhaustive list of candidate genes and permit improved analysis to identify prominent candidate genes.

Analysis of integration sites to ascertain candidate genes which drive the phenotype-of-interest requires intensive statistical computation (Brett et al., 2011) and persists as a bottleneck in the employment of transposon mediated mutagenesis in genetic screens. In contrast, directed genetic screens examine a pre-determined set of candidate genes (Takahashi and Yamanaka, 2006; Chia et al., 2010), dispensing the requirement of rigorous analysis. To circumvent this caveat, recent advances in piggyBac mediated

transposition have led to the generation of a library containing 14,000 individual gene-trap ESC clones (Wang et al., 2009). In co-operation with Bloom deficiency, homozygosity at mutated loci can be achieved to facilitate the performance of loss of function genetic screens. As each ESC clone carries a single genomic integration, laborious analysis of insertion site will be avoided. Alternatively, the use haploid cells which encompass a single set of chromosomes overcomes the need for bi-allelic mutants to observe a phenotype-of-interest and eliminates the requirement for extensive computational analysis (Leeb and Wutz, 2011; Yang et al., 2012; Carette et al., 2009).

Overall, the immense potential of transposon-mediated mutagenesis can be clearly illustrated from the discovery of novel genes involved in the pathogenesis in an array of cancer types (Rad et al., 2010; Dupuy et al., 2005). In addition to its inherent ability to integrate into the host genome in a random fashion, there has been heightened interest in its ability to deliver genes efficiently temporally with the option of excision (Balciunas et al., 2006; VandenDriessche et al., 2009; Kang et al., 2009). The pleiotropic properties of transposons highlight its prime candidacy to serve as gene delivery and discovery tools in a clinical setting. As a non-viral delivery method, transposon mediated gene delivery has been appealing in its ability to perform stable, ectopic and reversible manipulations an array of human cell types (Chen et al., 2010; Belay et al., 2011). Uses of transposon mediated gene delivery include *ex vivo* gene therapy (reviewed by Di Matteo et al., 2012), circumventing the occurrence of detrimental effects from the use of retroviruses (Blaese et al., 1995). Although there have been some success in performing gene therapy using retroviruses (Aiuti et al., 2009), a non-integrative and amenable approach would be an attractive alternative avenue.

### 6.2 Replacements of ectopic Oct4 during the initiation of reprogramming events

A number of studies have investigated the possibility to substitute exogenous requirements of Oct4 with alternative factors (Heng et al., 2010; Redmer et al., 2011). Findings from this thesis reiterate the observations that ectopic Oct4 can be substituted and two effective genetic replacements were identified, BNC2 and RARβ.

The identification of RARβ as a potent substitute for the ectopic requirement of Oct4 during the generation of iPSCs is worth deliberating over. Although a number of studies have attributed RA signalling to the loss of pluripotency, there are several observations which suggest otherwise (Ben-Shushan et al., 1995; Wang et al., 2011c). RA has been described to possess anti-proliferative and differentiation inducing properties, representing an effective deterrent during tumour formation (Schuldiner et al., 2001; Jiang et al., 2007; Yamashita et al., 2005; Metallo et al., 2008; Strickland and Mahdavi, 1978; Flynn et al., 1983). However, opposing cellular responses have been witnessed where RA signalling results in proliferation and the initation of tumorigenesis (Ledda-Columbano et al., 2004; Liu et al., 2011). Furthermore, exposure to RA in human and mouse subjects reveals confounding results and promotes tumorigenesis (Mikkelsen et al., 1998; Albright et al., 2004; Mollersen et al., 2004; Ommen et al., 1996). These pieces of evidence highlight our poor understanding of the RA signalling pathway. The discovery that RARs are potent instigators of the reprogramming process may provide a distinct dimension to their participation in developmental pathways.

Unravelling of factors which play co-operative roles in the activation of endogenous Oct4 during the acquisition of pluripotency offers insight to the roles performed by Oct4 during the reprogramming process. Epitomising the first transcription factor that participated in the maintenance of pluripotency, Oct4 has received voluminous amounts of attention in deciphering the molecular pathways that accompany it. Large scale genomic and proteomic studies have elucidated its role as the core transcriptional machinery in co-operation with Nanog and Sox2, and an exhaustive list of binding regions have been delineated (Boyer et al., 2005; Loh et al., 2006; Wang et al., 2006). However, post translational modifications of Oct4 have been side-lined. A number of studies have determined that phosphorylation, SUMO-ylation, ubiquitination, and glycosylation of Oct4 affect its transcriptional activity and stability (Brehm et al., 1997; Saxe et al., 2009; Brumbaugh et al., 2012; Wei et al., 2007; Zhang et al., 2007; Xu et al., 2009; Jang et al., 2012). To this end, it will be intriguing to determine if factors in the reprogramming milieu

exhibit affinities toward sites which are susceptible to post translational modifications. This knowledge will further enhance our knowledge on the mechanisms behind Oct4 and reprogramming.

Interestingly, porcine fibroblasts have been described to defy the notion that exogenous Oct4 or a suitable replacement is quintessential during the reprogramming process (Montserrat et al., 2011). c-Myc, Klf4 and Sox2 were described to be sufficient in generating porcine iPSCs. Reprogrammed cells exhibited Oct4 expression from its endogenous loci and were able to contribute to the formation of three germ layers when injected into immune-compromised mice. Conversely, the addition of ectopic Oct4 to the reprogramming mix deteriorated the reprogramming process and increased the propensity of reprogrammed cells to differentiate. However, Oct4 has been portrayed to be expressed at low levels in porcine fibroblasts and trophoectoderm, indicating a separate role of Oct4 in distinct vertebrate developmental processes. Elucidation of the roles played by Oct4 during reprogramming leaves much to be addressed and experiments performed in this thesis offer some insight to unravel the unknowns. Overall, this study reinforces the notion that ectopic Oct4 can be replaced or dispensed, but endogenous Oct4 remains crucial for the maintenance of pluripotency and the generation of iPSCs.

### 6.3 Epigenetic Regulation of Oct4

The regulatory elements of Oct4 are conserved across species (Nordhoff et al., 2001; van Eijk et al., 1999), reflecting their importance from an evolutionary perspective. Classified as Conserved Regions 1-4 (CR1-4), murine, human and bovine genomes exhibit these features within major histocompatibility complexes (MHC) on syntenic chromosomes (Abdelrahman et al., 1995; Takeda et al., 1992; van Eijk et al., 1999). Methylation profiles at these conserved regions have been described to indicate the transcriptional activity of Oct4. Differentiation of human pluripotent cells has been depicted to accompany methylation at the Oct4 distal enhancer and diminuition in Oct4 expression levels (Deb-Rinker et al., 2005; Yeo et al., 2007; Cheong et al., 2010), whereas DNA demethylation is synonymous with the activation of the Oct4 locus (Mikkelsen et al., 2008).

Hypermethylation can be observed as early as the first cell fate decision where the genomic locus of Oct4 is inflicted by increased methylation marks, resulting in the repression of Oct4 expression levels and differentiation into the trophoblastic lineage (Hattori et al., 2004). On the contrary, the presence of DNA demethylating agents or ablation of Dnmt1 resurrects Oct4 expression, highlighting the importance of epigenetic remodelling on the expression of Oct4.

This thesis proposes that RAR family members facilitate the activation of Oct4 cis-regulatory elements by imposing H3K4me3 activating histone marks. Extrication of histone modifications has been employed to study alternative cellular processes such as the silencing of p16(INK4a) during tumorigenesis and Jarid1a/b mediated H3K4 demethylation during senescence (Yao et al., 2010; Chicas et al., 2012). As specific epigenetic alterations associated to the acquisition of pluripotency remain concealed, an unbiased large scale analysis of promoter and enhancer sequences can be performed to elucidate novel molecular changes that complement the conversion of somatic cells to pluripotent derivatives (Patwardhan et al., 2012; Melnikov et al., 2012; Sharon et al., 2012).

Another layer of expression regulation is through alternative splicing to produce transcript variants and four protein isoforms. The human Oct4 locus is reported to generate three transcript variants and four protein isoforms. Clear delineation of the functions exhibited by each variant remains to be uncovered but it has been suggested that they exhibit distinct functions in the maintenance of pluripotency and cellular stress response (Cauffman et al., 2006; Lee et al., 2006; Papamichos et al., 2009). As an additional layer of complexity, Oct4 has been described to possess multiple transcription start sites and pseudogenes (van Eijk et al., 1999; Lin et al., 2007; Panagopoulos et al., 2008; Singh et al., 2012). Having elucidated the importance of the cis-regulatory elements of Oct4, it will be intriguing to determine their influence on the expression of the transcript variants and pseudogenes. Moreover, the participation and effect of these variants during the reprogramming process may also contribute to our understanding of Oct4 and its role in pluripotency.

## 6.4 Kinetics of Reprogramming

Since the inception of nuclear reprogramming, molecular kinetics that dictate the revival of transcriptional machinery underpinning the conversion of cellular identity have been of immense interest. This thesis attempts to dissect molecular events that influence the kinetics behind the dedifferentiation process. By retrieving epigenetic snapshots of the Oct4 regulatory elements over the first 72 hours of reprogramming, it was illustrated that activation of the distal enhancer of Oct4 and RAREoct is a key impetus to stimulate the dedifferentiation process. Alternative approaches to determine molecular changes during the generation of iPSCs include the examination of transcriptomes over the course of nuclear reprogramming (Foshay et al., 2012; Samavarchi-Tehrani et al., 2010; Mikkelsen et al., 2008). Despite the attainment and analysis of transcription profiles that accompany reprogramming intermediates, epigenetic changes reflect subtle molecular changes that precede fluctuations in transcript levels and possess better predictive value when delineating molecular events during reprogramming. To address this, methylomes have been illustrated in somatic cells and pluripotent cells (Meissner et al., 2008; Hawkins et al., 2010; Marks et al., 2012). However, an equivalent study to explore genome-wide histone modifications during the conversion of somatic cells to iPSCs has not been performed. To this end, this thesis presents as an initial effort to explore the tip of the iceberg, and inspects microscopic changes in histone conformation at precise regulatory elements of Oct4 over the course of reprogramming.

Exploration of kinetics that influence nuclear reprogramming events provides a perspective on cascades of molecular events that take place. This is exemplified through the examination of molecular changes during cell fusion which suggests biphasic reprogramming. The initial phase composes of trans-acting events that are independent of DNA replication and involve the cellular environment, whereas the subsequent phase embodies DNA replication dependent cis-acting chromatin modifiers that reawaken silent genes (Foshay et al., 2012). In contrast, the generation of iPSCs has been proposed to be divided into three phases: initiation, maturation and stabilisation (Samavarchi-Tehrani et

al., 2010). The initial phase involves mesenchymal-to-epithelial transition, whereas the maturation and stablisation phase are delineated by the up-regulation of a subset of pluripotency associated genes.

Findings from this study strengthen our understanding of acquired pluripotency, hence extrapolation of these findings and examination of histone modifications on a genome-wide scale will yield copious amounts of information about the clockwork behind the generation of iPSCs.

### 6.5 Future Direction

The objective of this thesis was to identify genetic replacements of exogenous Oct4 during the initiation of reprogramming events, thereby illuminating the role of Oct4 during the acquisition of pluripotency. To this end, two factors were identified and verified. Through the elucidation of the mechanism behind RARβ, it was delineated that activation of the distal enhancer and proximal promoter of Oct4 is integral during the reprogramming process. The presence of RARγ and LRH1 facilitates this process and accelerates iPSC colony formation and improves reprogramming efficiencies.

iPSC technology holds immense potential in a clinical setting. Derivation of patient specific iPSCs enables the execution of disease modelling, drug screening and cell therapy. As such, it will be interesting to determine if RARβ and BNC2 are able to replace the exogenous requirement of Oct4 in reprogramming human somatic cells. Interestingly, RARβ has been described to possess tumour suppressor properties (Widschwendter et al., 1997; Xu et al., 1997), whereas the conventional set of reprogramming factors have been established to promote tumorigenesis (Hochedlinger et al., 2005; Okita et al., 2007). Hence, application of RARβ in place of Oct4 may be desirable. However, reprogramming of mouse and human fibroblasts are distinct, where introduction of an identical reprogramming cocktail consisting of Oct4, c-Myc, Klf4 and Sox2 leads to iPSCs that

exhibit distinct properties and morphologies (reviewed by De Los Angeles et al., 2012). Attempts at replacing reprogramming factors have only been performed in limited cell types or with the use of chemicals (Giorgetti et al., 2009; Kim et al., 2009b; Kim et al., 2009c; Huangfu et al., 2008b), and have to be pursued with caution.

Although conventional reprogramming methods convert human somatic cells to iPSCs that resemble human ESCs, addition of RARγ and LRH1 to the traditional reprogramming mix has been described to generate mouse ESC-like derivatives (SH-iPSCs) which are disparate from human ESCs (Wang et al., 2011c). Unlike human ESCs, SH-iPSCs are easily amenable and able to survive in a single cell suspension. As such, it will be fascinating to determine if RARγ and LRH1 elicit a similar effect on the distal enhancer and proximal promoter of human Oct4, explaining the peculiar cellular phenotype of SH-iPSCs. Retinoic acid differentiation of human ESCs incites hypermethylation of the Oct4 distal enhancer (You et al., 2011a; Yeo et al., 2007). As such, it is tempting to speculate that a similar phenomenon takes place during the acquisition of pluripotency in human cells and activation of both the distal enhancer and proximal promoter of human Oct4 is imperative. It has been demonstrated that cell fusion results in epigenetic modifications of the distal enhancer and proximal promoter of Oct4 (Freberg et al., 2007). Keeping this in mind, it will be interesting to learn if RARγ and LRH1 also have the capacity to enhance reprogramming efficiencies through cell fusion.

Findings from this thesis set a foundation for deeper comprehension of the reprogramming process in human and murine cells. Elucidation of the activation of endogenous Oct4 during the early stages reprogramming may allow us to improve the reprogramming protocol, where the endogenous genes are targeted, dispensing the need for ectopic introduction of reprogramming factors. To this end, transcription activator-like effectors (TALEs) have been designed to recognise the Oct4 proximal promoter. Introduction of TALEs into ESCs stimulates Oct4 transcriptional activity, whereas epigenetic modifiers are required for TALEs to elicit an effect on neural stem cells (Bultmann et al., 2012). This surmises the possibility that activation of the Oct4 distal enhancer could overcome the

epigenetic barrier faced in neural stem cells. Overall, this thesis proposes that epigenetic changes at the endogenous Oct4 locus act as reprogramming checkpoints and can be surmountable by the presence of RARγ and LRH1. Accordingly, unravelling of microscopic alterations during the generation of iPSCs may resolve uncertainties in the molecular processes that occur during nuclear reprogramming.

## REFERENCES


Aasen T, Raya A, Barrero MJ, Garreta E, Consiglio A, Gonzalez F, Vassena R, Bilic J, Pekarik V, Tiscornia G et al: Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat Biotechnol 2008, 26(11):1276-1284.

Abdelrahman B, Fiddler M, Rappolee D, Pergament E: Expression of Transcription Regulating Genes in Human Preimplantation Embryos. Hum Reprod 1995, 10(10):2787-2792.

Agarwal S, Loh YH, McLoughlin EM, Huang J, Park IH, Miller JD, Huo H, Okuka M, Dos Reis RM, Loewer S et al: Telomere elongation in induced pluripotent stem cells from dyskeratosis congenita patients. Nature 2010, 464(7286):292-296.

Aiuti A, Cattaneo F, Galimberti S, Benninghoff U, Cassani B, Callegaro L, Scaramuzza S, Andolfi G, Mirolo M, Brigida I et al: Gene Therapy for Immunodeficiency Due to Adenosine Deaminase Deficiency. New Engl J Med 2009, 360(5):447-458.

Akagi T, Ito T, Kato M, Jin Z, Cheng Y, Kan T, Yamamoto G, Olaru A, Kawamata N, Boult J et al: Chromosomal abnormalities and novel disease-related regions in progression from Barrett's esophagus to esophageal adenocarcinoma. International journal of cancer Journal international du cancer 2009, 125(10):2349-2359.

Albright CD, Salganik RI, Van Dyke T: Dietary depletion of vitamin E and vitamin A inhibits mammary tumor growth and metastasis in transgenic mice. J Nutr 2004, 134(5):1139-1144.

Alsayed Y, Uddin S, Mahmud N, Lekmine F, Kalvakolanu DV, Minucci S, Bokoch G, Platanias LC: Activation of Rac1 and the p38 mitogen-activated protein kinase pathway in response to all-trans-retinoic acid. The Journal of biological chemistry 2001, 276(6):4012-4019.

Amps K, Andrews PW, Anyfantis G, Armstrong L, Avery S, Baharvand H, Baker J, Baker D, Munoz MB, Beil S et al: Screening ethnically diverse human embryonic stem

cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nat Biotechnol 2011, 29(12):1132-1144.

Amsterdam A, Nissen RM, Sun Z, Swindell EC, Farrington S, Hopkins N: Identification of 315 genes essential for early zebrafish development. P Natl Acad Sci USA 2004, 101(35):12792-12797.

An MC, Zhang NZ, Scott G, Montoro D, Wittkop T, Mooney S, Melov S, Ellerby LM: Genetic Correction of Huntington's Disease Phenotypes in Induced Pluripotent Stem Cells. Cell stem cell 2012, 11(2):253-263.

Andre E, Gawlas K, Becker-Andre M: A novel isoform of the orphan nuclear receptor RORbeta is specifically expressed in pineal gland and retina. Gene 1998, 216(2):277-283.

Ang HL, Duester G: Initiation of retinoid signaling in primitive streak mouse embryos: spatiotemporal expression patterns of receptors and metabolic enzymes for ligand synthesis. Dev Dyn 1997, 208(4):536-543.

Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, Zhang Y, Yang W, Gruber PJ, Epstein JA et al: Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell stem cell 2011, 8(4):376-388.

Aoi T, Yae K, Nakagawa M, Ichisaka T, Okita K, Takahashi K, Chiba T, Yamanaka S: Generation of pluripotent stem cells from adult mouse liver and stomach cells. Science 2008, 321(5889):699-702.

Aristotle: Generation of Animals. Cambridge, MA: Harvard University Press; 1979.

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R: Multipotent cell lineages in early mouse development depend on SOX2 function. Gene Dev 2003, 17(1):126-140.

Balciunas D, Wangensteen KJ, Wilber A, Bell J, Geurts A, Sivasubbu S, Wang X, Hackett PB, Largaespada DA, McIvor RS et al: Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. PLoS genetics 2006, 2(11):e169.

Balmer JE, Blomhoff R: Gene expression regulation by retinoic acid. Journal of lipid research 2002, 43(11):1773-1808.

Bao S, Tang F, Li X, Hayashi K, Gillich A, Lao K, Surani MA: Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature 2009, 461(7268):1292-1295.

Bard-Chapeau EA, Jeyakani J, Kok CH, Muller J, Chua BQ, Gunaratne J, Batagov A, Jenjaroenpun P, Kuznetsov VA, Wei CL et al: Ecotopic viral integration site 1 (EVI1) regulates multiple cellular processes important for cancer and is a synergistic partner for FOS protein in invasive tumors. P Natl Acad Sci USA 2012, 109(6):2168-2173.

Barnea E, Bergman Y: Synergy of SF1 and RAR in activation of Oct-3/4 promoter. The Journal of biological chemistry 2000, 275(9):6608-6619.

Bastien J, Adam-Stitah S, Riedl T, Egly JM, Chambon P, Rochette-Egly C: TFIIH interacts with the retinoic acid receptor gamma and phosphorylates its AF-1-activating domain through cdk7. The Journal of biological chemistry 2000, 275(29):21896-21904.

Beattie GM, Lopez AD, Bucay N, Hinton A, Firpo MT, King CC, Hayek A: Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. Stem Cells 2005, 23(4):489-495.

Becker M, Becker A, Miyara F, Han Z, Kihara M, Brown DT, Hager GL, Latham K, Adashi EY, Misteli T: Differential in vivo binding dynamics of somatic and oocyte-specific linker histones in oocytes and during ES cell nuclear transfer. Molecular biology of the cell 2005, 16(8):3887-3895.

Becker-Andre M, Andre E, DeLamarter JF: Identification of nuclear receptor mRNAs by RT-PCR amplification of conserved zinc-finger motif sequences. Biochem Bioph Res Co 1993, 194(3):1371-1379.

Belay E, Dastidar S, VandenDriessche T, Chuah MKL: Transposon-Mediated Gene Transfer into Adult and Induced Pluripotent Stem Cells. Curr Gene Ther 2011, 11(5):406-413.

Benbrook D, Lernhardt E, Pfahl M: A new retinoic acid receptor identified from a hepatocellular carcinoma. Nature 1988, 333(6174):669-672.

Bendall AJ, Sturm RA, Danoy PA, Molloy PL: Broad binding-site specificity and affinity properties of octamer 1 and brain octamer-binding proteins. European journal of biochemistry / FEBS 1993, 217(3):799-811.

Ben-Shushan E, Sharir H, Pikarsky E, Bergman Y: A dynamic balance between ARP-1/COUP-TFII, EAR-3/COUP-TFI, and retinoic acid receptor:retinoid X receptor heterodimers regulates Oct-3/4 expression in embryonal carcinoma cells. Mol Cell Biol 1995, 15(2):1034-1048.

Bessereau JL, Wright A, Williams DC, Schuske K, Davis MW, Jorgensen EM: Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line. Nature 2001, 413(6851):70-74.

Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, Blau HM: Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature 2010, 463(7284):1042-1047.

Bingham PM, Kidwell MG, Rubin GM: The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell 1982, 29(3):995-1004.

Blaese RM, Culver KW, Miller AD, Carter CS, Fleisher T, Clerici M, Shearer G, Chang L, Chiang YW, Tolstoshev P et al: T-Lymphocyte-Directed Gene-Therapy for Ada(-) Scid - Initial Trial Results after 4 Years. Science 1995, 270(5235):475-480.

Blau HM, Chiu CP, Webster C: Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell 1983, 32(4):1171-1180.

Bouchard C, Dittrich O, Kiermaier A, Dohmann K, Menkel A, Eilers M, Luscher B: Regulation of cyclin D2 gene expression by the Myc/Max/Mad network: Myc-dependent TRRAP recruitment and histone acetylation at the cyclin D2 promoter. Gene Dev 2001, 15(16):2042-2047.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG et al: Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 2005, 122(6):947-956.

Bradley A, Evans M, Kaufman MH, Robertson E: Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature 1984, 309(5965):255-256.

Brand N, Petkovich M, Krust A, Chambon P, de The H, Marchio A, Tiollais P, Dejean A: Identification of a second human retinoic acid receptor. Nature 1988, 332(6167):850-853.

Brehm A, Ohbo K, Scholer H: The carboxy-terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain. Mol Cell Biol 1997, 17(1):154-162.

Bretscher A, Lynch W: Identification and Localization of Immunoreactive Forms of Caldesmon in Smooth and Nonmuscle Cells - a Comparison with the Distributions of Tropomyosin and Alpha-Actinin. J Cell Biol 1985, 100(5):1656-1663.

Brett BT, Berquam-Vrieze KE, Nannapaneni K, Huang J, Scheetz TE, Dupuy AJ: Novel molecular and computational methods improve the accuracy of insertion site analysis in Sleeping Beauty-induced tumors. PloS one 2011, 6(9):e24668.

Briggs R, King TJ: Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. P Natl Acad Sci USA 1952, 38(5):455-463.

Brimble SN, Zeng X, Weiler DA, Luo Y, Liu Y, Lyons IG, Freed WJ, Robins AJ, Rao MS, Schulz TC: Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001. Stem cells and development 2004, 13(6):585-597.

Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA et al: Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature 2007, 448(7150):191-195.

Bruck N, Vitoux D, Ferry C, Duong V, Bauer A, de The H, Rochette-Egly C: A coordinated phosphorylation cascade initiated by p38MAPK/MSK1 directs RARalpha to target promoters. The EMBO journal 2009, 28(1):34-47.

Brumbaugh J, Hou ZG, Russell JD, Howden SE, Yu PZ, Ledvina AR, Coon JJ, Thomson JA: Phosphorylation regulates human OCT4. P Natl Acad Sci USA 2012, 109(19):7162-7168.

Buecker C, Chen HH, Polo JM, Daheron L, Bu L, Barakat TS, Okwieka P, Porter A, Gribnau J, Hochedlinger K et al: A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. Cell stem cell 2010, 6(6):535-546.

Buehr M, Meek S, Blair K, Yang J, Ure J, Silva J, McLay R, Hall J, Ying QL, Smith A: Capture of authentic embryonic stem cells from rat blastocysts. Cell 2008, 135(7):1287-1298.

Bultmann S, Morbitzer R, Schmidt CS, Thanisch K, Spada F, Elsaesser J, Lahaye T, Leonhardt H: Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. Nucleic Acids Res 2012.

Burdon T, Stracey C, Chambers I, Nichols J, Smith A: Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Developmental biology 1999, 210(1):30-43.

Buzzard JJ, Gough NM, Crook JM, Colman A: Karyotype of human ES cells during extended culture. Nat Biotechnol 2004, 22(4):381-382; author reply 382.

Byrne JA, Simonsson S, Western PS, Gurdon JB: Nuclei of adult mammalian somatic cells are directly reprogrammed to oct-4 stem cell gene expression by amphibian oocytes. Curr Biol 2003, 13(14):1206-1213.

Cadinanos J, Bradley A: Generation of an inducible and optimized piggyBac transposon system. Nucleic Acids Res 2007, 35(12):e87.

Carette JE, Guimaraes CP, Varadarajan M, Park AS, Wuethrich I, Godarova A, Kotecki M, Cochran BH, Spooner E, Ploegh HL et al: Haploid Genetic Screens in Human Cells Identify Host Factors Used by Pathogens. Science 2009, 326(5957):1231-1235.

Carey BW, Markoulaki S, Hanna J, Saha K, Gao Q, Mitalipova M, Jaenisch R: Reprogramming of murine and human somatic cells using a single polycistronic vector. P Natl Acad Sci USA 2009, 106(1):157-162.

Carey BW, Markoulaki S, Hanna JH, Faddah DA, Buganim Y, Kim J, Ganz K, Steine EJ, Cassady JP, Creyghton MP et al: Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. Cell stem cell 2011, 9(6):588-598.

Carlberg C, Hooft van Huijsduijnen R, Staple JK, DeLamarter JF, Becker-Andre M: RZRs, a new family of retinoid-related orphan receptors that function as both monomers and homodimers. Mol Endocrinol 1994, 8(6):757-770.

Cartwright P, McLean C, Sheppard A, Rivett D, Jones K, Dalton S: LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development 2005, 132(5):885-896.

Casci T: Developmental biology - Retinoic acid passes the morphogen test. Nat Rev Genet 2008, 9(1):7-7.

Cauffman G, Liebaers I, Van Steirteghem A, de Velde HV: POU5F1 isoforms show different expression patterns in human embryonic stem cells and preimplantation embryos. Stem Cells 2006, 24(12):2685-2691.

Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A: Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 2003, 113(5):643-655.

Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A: Nanog safeguards pluripotency and mediates germline development. Nature 2007, 450(7173):1230-1234.

Chambon P: A decade of molecular biology of retinoic acid receptors. Faseb J 1996, 10(9):940-954.

Chen J, Liu J, Yang J, Chen Y, Chen J, Ni S, Song H, Zeng L, Ding K, Pei D: BMPs functionally replace Klf4 and support efficient reprogramming of mouse fibroblasts by Oct4 alone. Cell research 2011, 21(1):205-212.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J et al: Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 2008, 133(6):1106-1117.

Chen Y, Yee D, Dains K, Chatterjee A, Cavalcoli J, Schneider E, Om J, Woychik RP, Magnuson T: Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells. Nat Genet 2000, 24(3):314-317.

Chen YT, Furushima K, Hou PS, Ku AT, Deng JM, Jang CW, Fang H, Adams HP, Kuo ML, Ho HN et al: PiggyBac transposon-mediated, reversible gene transfer in human embryonic stem cells. Stem cells and development 2010, 19(6):763-771.

Cheong HS, Lee HC, Park BL, Kim H, Jang MJ, Han YM, Kim SY, Kim YS, Shin HD: Epigenetic modification of retinoic acid-treated human embryonic stem cells. Bmb Rep 2010, 43(12):830-835.

Chew JL, Loh YH, Zhang W, Chen X, Tam WL, Yeap LS, Li P, Ang YS, Lim B, Robson P et al: Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol Cell Biol 2005, 25(14):6031-6046.

Chia NY, Chan YS, Feng B, Lu X, Orlov YL, Moreau D, Kumar P, Yang L, Jiang J, Lau MS et al: A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature 2010, 468(7321):316-320.

Chicas A, Kapoor A, Wang X, Aksoy O, Evertts AG, Zhang MQ, Garcia BA, Bernstein E, Lowe SW: H3K4 demethylation by Jarid1a and Jarid1b contributes to retinoblastoma-mediated gene silencing during cellular senescence. P Natl Acad Sci USA 2012, 109(23):8971-8976.

Chickarmane V, Enver T, Peterson C: Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. PLoS computational biology 2009, 5(1):e1000268.

Chiu CP, Blau HM: 5-Azacytidine permits gene activation in a previously noninducible cell type. Cell 1985, 40(2):417-424.

Claassen GF, Hann SR: A role for transcriptional repression of p21CIP1 by c-Myc in overcoming transforming growth factor beta -induced cell-cycle arrest. P Natl Acad Sci USA 2000, 97(17):9498-9503.

Conlon RA, Rossant J: Exogenous retinoic acid rapidly induces anterior ectopic expression of murine Hox-2 genes in vivo. Development 1992, 116(2):357-368.

Consortium Hi: Induced Pluripotent Stem Cells from Patients with Huntington's Disease Show CAG-Repeat-Expansion-Associated Phenotypes. Cell stem cell 2012, 11(2):264-278.

Cooley L, Kelley R, Spradling A: Insertional mutagenesis of the Drosophila genome with single P elements. Science 1988, 239(4844):1121-1128.

Cowan CA, Atienza J, Melton DA, Eggan K: Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science 2005, 309(5739):1369-1373.

Daheron L, Opitz SL, Zaehres H, Lensch MW, Andrews PW, Itskovitz-Eldor J, Daley GQ: LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. Stem Cells 2004, 22(5):770-778.

De Los Angeles A, Loh YH, Tesar PJ, Daley GQ: Accessing naive human pluripotency. Current opinion in genetics & development 2012.

Deb-Rinker P, Ly D, Jezierski A, Sikorska M, Walker PR: Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. J Biol Chem 2005, 280(8):6257-6260.

Dejean A, Bougueleret L, Grzeschik KH, Tiollais P: Hepatitis B virus DNA integration in a sequence homologous to v-erb-A and steroid receptor genes in a hepatocellular carcinoma. Nature 1986, 322(6074):70-72.

Dejosez M, Krumenacker JS, Zitur LJ, Passeri M, Chu LF, Songyang Z, Thomson JA, Zwaka TP: Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. Cell 2008, 133(7):1162-1174.

Delacroix L, Moutier E, Altobelli G, Legras S, Poch O, Choukrallah MA, Bertin I, Jost B, Davidson I: Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells. Mol Cell Biol 2010, 30(1):231-244.

Di Matteo M, Belay E, Chuah MK, Vandendriessche T: Recent developments in transposon-mediated gene therapy. Expert opinion on biological therapy 2012, 12(7):841-858.

DiBerardino MA, Hoffner NJ: Gene reactivation in erythrocytes: nuclear transplantation in oocytes and eggs of Rana. Science 1983, 219(4586):862-864.

Dickman ED, Thaller C, Smith SM: Temporally-regulated retinoic acid depletion produces specific neural crest, ocular and nervous system defects. Development 1997, 124(16):3111-3121.

Dimos JT, Rodolfa KT, Niakan KK, Weisenthal LM, Mitsumoto H, Chung W, Croft GF, Saphier G, Leibel R, Goland R et al: Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science 2008, 321(5893):1218-1221.

Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T: Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell 2005, 122(3):473-483.

Do JT, Scholer HR: Cell fusion-induced reprogramming. Methods in molecular biology (Clifton, NJ 2010, 636:179-190.

Doetschman T, Williams P, Maeda N: Establishment of hamster blastocyst-derived embryonic stem (ES) cells. Developmental biology 1988, 127(1):224-227.

Dolle P: Developmental expression of retinoic acid receptors (RARs). Nuclear receptor signaling 2009, 7:e006.

Doulazmi M, Frederic F, Capone F, Becker-Andre M, Delhaye-Bouchaud N, Mariani J: A comparative study of Purkinje cells in two RORalpha gene mutant mice: staggerer and RORalpha(-/-). Brain research Developmental brain research 2001, 127(2):165-174.

Doulazmi M, Frederic F, Lemaigre-Dubreuil Y, Hadj-Sahraoui N, Delhaye-Bouchaud N, Mariani J: Cerebellar Purkinje cell loss during life span of the heterozygous staggerer mouse (Rora(+)/Rora(sg)) is gender-related. The Journal of comparative neurology 1999, 411(2):267-273.

Driesch H: Analytische theorie der organischen entwicklung: Leipzig, W. Engelmann; 1894.

Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA: Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature 2005, 436(7048):221-226.

Dupuy AJ, Rogers LM, Kim J, Nannapaneni K, Starr TK, Liu P, Largaespada DA, Scheetz TE, Jenkins NA, Copeland NG: A modified sleeping beauty transposon system that can be used to model a wide variety of human cancers in mice. Cancer research 2009, 69(20):8150-8156.

Dussault I, Fawcett D, Matthyssen A, Bader JA, Giguere V: Orphan nuclear receptor ROR alpha-deficient mice display the cerebellar defects of staggerer. Mechanisms of development 1998, 70(1-2):147-153.

Ebert AD, Yu J, Rose FF, Jr., Mattis VB, Lorson CL, Thomson JA, Svendsen CN: Induced pluripotent stem cells from a spinal muscular atrophy patient. Nature 2009, 457(7227):277-280.

Efe JA, Hilcove S, Kim J, Zhou H, Ouyang K, Wang G, Chen J, Ding S: Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. Nat Cell Biol 2011, 13(3):215-U261.

Eilken HM, Nishikawa S, Schroeder T: Continuous single-cell imaging of blood generation from haemogenic endothelium. Nature 2009, 457(7231):896-900.

Ellinger-Ziegelbauer H, Dreyer C: A retinoic acid receptor expressed in the early development of Xenopus laevis. Gene Dev 1991, 5(1):94-104.

Esteban MA, Wang T, Qin B, Yang J, Qin D, Cai J, Li W, Weng Z, Chen J, Ni S et al: Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. Cell stem cell 2010, 6(1):71-79.

Esteban MA, Xu J, Yang J, Peng M, Qin D, Li W, Jiang Z, Chen J, Deng K, Zhong M et al: Generation of induced pluripotent stem cell lines from Tibetan miniature pig. The Journal of biological chemistry 2009, 284(26):17634-17640.

Evans MJ: The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. Journal of embryology and experimental morphology 1972, 28(1):163-176.

Evans MJ, Kaufman MH: Establishment in culture of pluripotential cells from mouse embryos. Nature 1981, 292(5819):154-156.

Ezashi T, Das P, Roberts RM: Low O2 tensions and the prevention of differentiation of hES cells. P Natl Acad Sci USA 2005, 102(13):4783-4788.

Felgner PL, Gadek TR, Holm M, Roman R, Chan HW, Wenz M, Northrop JP, Ringold GM, Danielsen M: Lipofection - a Highly Efficient, Lipid-Mediated DNA-Transfection Procedure. P Natl Acad Sci USA 1987, 84(21):7413-7417.

Feng B, Jiang J, Kraus P, Ng JH, Heng JC, Chan YS, Yaw LP, Zhang W, Loh YH, Han J et al: Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol 2009, 11(2):197-203.

Fibrici G: De Formatione Ovi, et Pulli Tractatus Accuratissimus: Patavii [Padua]: ex Officina Aloysii Bencii Bibliopolae; 1621.

Flasza M, Shering AF, Smith K, Andrews PW, Talley P, Johnson PA: Reprogramming in inter-species embryonal carcinoma-somatic cell hybrids induces expression of pluripotency and differentiation markers. Cloning and stem cells 2003, 5(4):339-354.

Flynn PJ, Miller WJ, Weisdorf DJ, Arthur DC, Brunning R, Branda RF: Retinoic acid treatment of acute promyelocytic leukemia: in vitro and in vivo observations. Blood 1983, 62(6):1211-1217.

Forsyth NR, Musio A, Vezzoni P, Simpson AH, Noble BS, McWhir J: Physiologic oxygen enhances human embryonic stem cell clonal recovery and reduces chromosomal abnormalities. Cloning and stem cells 2006, 8(1):16-23.

Foshay KM, Looney TJ, Chari S, Mao FF, Lee JH, Zhang L, Fernandes CJ, Baker SW, Clift KL, Gaetz J et al: Embryonic Stem Cells Induce Pluripotency in Somatic Cell Fusion through Biphasic Reprogramming. Mol Cell 2012, 46(2):159-170.

Frank SR, Schroeder M, Fernandez P, Taubert S, Amati B: Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. Gene Dev 2001, 15(16):2069-2082.

Fraser MJ, Ciszczon T, Elick T, Bauser C: Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. Insect molecular biology 1996, 5(2):141-151.

Freberg CT, Dahl JA, Timoskainen S, Collas P: Epigenetic reprogramming of OCT4 and NANOG regulatory regions by embryonal carcinoma cell extract. Molecular biology of the cell 2007, 18(5):1543-1553.

French AJ, Adams CA, Anderson LS, Kitchen JR, Hughes MR, Wood SH: Development of human cloned blastocysts following somatic cell nuclear transfer with adult fibroblasts. Stem Cells 2008, 26(2):485-493.

Fuhrmann G, Chung AC, Jackson KJ, Hummelke G, Baniahmad A, Sutter J, Sylvester I, Scholer HR, Cooney AJ: Mouse germline restriction of Oct4 expression by germ cell nuclear factor. Developmental cell 2001, 1(3):377-387.

Fuhrmann G, Sylvester I, Scholer HR: Repression of Oct-4 during embryonic cell differentiation correlates with the appearance of TRIF, a transiently induced DNA-binding factor. Cell Mol Biol (Noisy-le-grand) 1999, 45(5):717-724.

Furst DO, Cross RA, Demey J, Small JV: Caldesmon Is an Elongated, Flexible Molecule Localized in the Actomyosin Domains of Smooth-Muscle. Embo J 1986, 5(2):251-257.

Fusaki N, Ban H, Nishiyama A, Saeki K, Hasegawa M: 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. Proceedings of the Japan Academy 2009, 85(8):348-362.

Gao S, Chung YG, Parseghian MH, King GJ, Adashi EY, Latham KE: Rapid H1 linker histone transitions following fertilization or somatic cell nuclear transfer: evidence for a uniform developmental program in mice. Developmental biology 2004, 266(1):62-75.

Gerami-Naini B, Dovzhenko OV, Durning M, Wegner FH, Thomson JA, Golos TG: Trophoblast differentiation in embryoid bodies derived from human embryonic stem cells. Endocrinology 2004, 145(4):1517-1524.

Geurts AM, Collier LS, Geurts JL, Oseth LL, Bell ML, Mu D, Lucito R, Godbout SA, Green LE, Lowe SW et al: Gene mutations and genomic rearrangements in the mouse as a result of transposon mobilization from chromosomal concatemers. PLoS genetics 2006, 2(9):1413-1423.

Ghyselinck NB, Dupe V, Dierich A, Messaddeq N, Garnier JM, Rochette-Egly C, Chambon P, Mark M: Role of the retinoic acid receptor beta (RARbeta) during mouse development. The International journal of developmental biology 1997, 41(3):425-447.

Gianni M, Bauer A, Garattini E, Chambon P, Rochette-Egly C: Phosphorylation by p38MAPK and recruitment of SUG-1 are required for RA-induced RAR gamma degradation and transactivation. The EMBO journal 2002, 21(14):3760-3769.

Gianni M, Parrella E, Raska I, Jr., Gaillard E, Nigro EA, Gaudon C, Garattini E, Rochette-Egly C: P38MAPK-dependent phosphorylation and degradation of SRC-3/AIB1 and RARalpha-mediated transcription. The EMBO journal 2006, 25(4):739-751.

Giguere V, Beatty B, Squire J, Copeland NG, Jenkins NA: The orphan nuclear receptor ROR alpha (RORA) maps to a conserved region of homology on human chromosome 15q21-q22 and mouse chromosome 9. Genomics 1995, 28(3):596-598.

Giguere V, Ong ES, Segui P, Evans RM: Identification of a receptor for the morphogen retinoic acid. Nature 1987, 330(6149):624-629.

Giguere V, Tini M, Flock G, Ong E, Evans RM, Otulakowski G: Isoform-specific amino-terminal domains dictate DNA-binding properties of ROR alpha, a novel family of orphan hormone nuclear receptors. Gene Dev 1994, 8(5):538-553.

Gillespie RF, Gudas LJ: Retinoic acid receptor isotype specificity in F9 teratocarcinoma stem cells results from the differential recruitment of coregulators to retinoic response elements. The Journal of biological chemistry 2007, 282(46):33421-33434.

Gillespie RF, Gudas LJ: Retinoid regulated association of transcriptional co-regulators and the polycomb group protein SUZ12 with the retinoic acid response elements of Hoxa1, RARbeta(2), and Cyp26A1 in F9 embryonal carcinoma cells. Journal of molecular biology 2007, 372(2):298-316.

Giorgetti A, Montserrat N, Aasen T, Gonzalez F, Rodriguez-Piza I, Vassena R, Raya A, Boue S, Barrero MJ, Corbella BA et al: Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2. Cell stem cell 2009, 5(4):353-357.

Gloor GB, Nassif NA, Johnson-Schlitz DM, Preston CR, Engels WR: Targeted gene replacement in Drosophila via P element-induced gap repair. Science 1991, 253(5024):1110-1117.

Godin I, Wylie C, Heasman J: Genital Ridges Exert Long-Range Effects on Mouse Primordial Germ-Cell Numbers and Direction of Migration in Culture. Development 1990, 108(2):357-363.

Goff SP, Witte ON, Gilboa E, Rosenberg N, Baltimore D: Genome structure of Abelson murine leukemia virus variants: proviruses in fibroblasts and lymphoid cells. Journal of virology 1981, 38(2):460-468.

Golling G, Amsterdam A, Sun Z, Antonelli M, Maldonado E, Chen W, Burgess S, Haldi M, Artzt K, Farrington S et al: Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. Nat Genet 2002, 31(2):135-140.

Gonzalez F, Boue S, Izpisua Belmonte JC: Methods for making induced pluripotent stem cells: reprogramming a la carte. Nat Rev Genet 2011, 12(4):231-242.

Gore A, Li Z, Fung HL, Young JE, Agarwal S, Antosiewicz-Bourget J, Canto I, Giorgetti A, Israel MA, Kiskinis E et al: Somatic coding mutations in human induced pluripotent stem cells. Nature 2011, 471(7336):63-67.

Gossen M, Bujard H: Tight Control of Gene-Expression in Mammalian-Cells by Tetracycline-Responsive Promoters. P Natl Acad Sci USA 1992, 89(12):5547-5551.

Gossen M, Freundlieb S, Bender G, Muller G, Hillen W, Bujard H: Transcriptional Activation by Tetracyclines in Mammalian-Cells. Science 1995, 268(5218):1766-1769.

Graves KH, Moreadith RW: Derivation and characterization of putative pluripotential embryonic stem cells from preimplantation rabbit embryos. Molecular reproduction and development 1993, 36(4):424-433.

Gresch O, Engel FB, Nesic D, Tran TT, England HM, Hickman ES, Korner I, Gan L, Chen S, Castro-Obregon S et al: New non-viral method for gene transfer into primary cells. Methods 2004, 33(2):151-163.

Gu P, Goodwin B, Chung AC, Xu X, Wheeler DA, Price RR, Galardi C, Peng L, Latour AM, Koller BH et al: Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. Mol Cell Biol 2005, 25(9):3492-3505.

Guo G, Huang Y, Humphreys P, Wang X, Smith A: A PiggyBac-based recessive screening method to identify pluripotency regulators. PloS one 2011, 6(4):e18189.

Guo G, Smith A: A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. Development 2010, 137(19):3185-3192.

Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, Smith A: Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development 2009, 136(7):1063-1069.

Guo HQ, Wang CLA: Specific disruption of smooth muscle caldesmon expression in mice. Biochem Bioph Res Co 2005, 330(4):1132-1137.

Gupta P, Ho PC, Huq MM, Ha SG, Park SW, Khan AA, Tsai NP, Wei LN: Retinoic acid-stimulated sequential phosphorylation, PML recruitment, and SUMOylation of nuclear receptor TR2 to suppress Oct4 expression. P Natl Acad Sci USA 2008, 105(32):11424-11429.

Gurdon JB: The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. Journal of embryology and experimental morphology 1962, 10:622-640.

Gurdon JB, Brennan S, Fairman S, Mohun TJ: Transcription of muscle-specific actin genes in early Xenopus development: nuclear transplantation and cell dissociation. Cell 1984, 38(3):691-700.

Gurdon JB, Elsdale TR, Fischberg M: Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature 1958, 182(4627):64-65.

Gurdon JB, Laskey RA, Reeves OR: The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs. Journal of embryology and experimental morphology 1975, 34(1):93-112.

Hale F: Pigs born without eyeballs. Journal of Heredity 1933, 24:105-106.

Hall J, Guo G, Wray J, Eyres I, Nichols J, Grotewold L, Morfopoulou S, Humphreys P, Mansfield W, Walker R et al: Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. Cell stem cell 2009, 5(6):597-609.

Hall VJ, Compton D, Stojkovic P, Nesbitt M, Herbert M, Murdoch A, Stojkovic M: Developmental competence of human in vitro aged oocytes as host cells for nuclear transfer. Human reproduction (Oxford, England) 2007, 22(1):52-62.

Hamanaka RB, Chandel NS: Targeting glucose metabolism for cancer therapy. The Journal of experimental medicine 2012, 209(2):211-215.

Hamilton BA, Frankel WN, Kerrebrock AW, Hawkins TL, FitzHugh W, Kusumi K, Russell LB, Mueller KL, van Berkel V, Birren BW et al: Disruption of the nuclear hormone receptor RORalpha in staggerer mice. Nature 1996, 379(6567):736-739.

Hammachi F MAM, Sharov A.A., Livigni A., Narayan S., Papapetrou E.P., O'Malley J., Kaji K., Ko M.S.H., Ptashne M., Brickman J.M.: Transcriptional Activation by Oct4 Is Sufficient for the Maintenance and Induction of Pluripotency. Cell Reports 2012, 1(2):99-109.

Han DW, Tapia N, Joo JY, Greber B, Arauzo-Bravo MJ, Bernemann C, Ko K, Wu G, Stehling M, Do JT et al: Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. Cell 2010, 143(4):617-627.

Han J, Yuan P, Yang H, Zhang J, Soh BS, Li P, Lim SL, Cao S, Tay J, Orlov YL et al: Tbx3 improves the germ-line competency of induced pluripotent stem cells. Nature 2010, 463(7284):1096-1100.

Hanna J, Cheng AW, Saha K, Kim J, Lengner CJ, Soldner F, Cassady JP, Muffat J, Carey BW, Jaenisch R: Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. P Natl Acad Sci USA 2010, 107(20):9222-9227.

Hanna J, Markoulaki S, Schorderet P, Carey BW, Beard C, Wernig M, Creyghton MP, Steine EJ, Cassady JP, Foreman R et al: Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell 2008, 133(2):250-264.

Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R: Direct cell reprogramming is a stochastic process amenable to acceleration. Nature 2009, 462(7273):595-601.

Hanna J, Wernig M, Markoulaki S, Sun CW, Meissner A, Cassady JP, Beard C, Brambrink T, Wu LC, Townes TM et al: Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science 2007, 318(5858):1920-1923.

Hansen GM, Markesich DC, Burnett MB, Zhu QH, Dionne KM, Richter LJ, Finnell RH, Sands AT, Zambrowicz BP, Abuin A: Large-scale gene trapping in C57BL/6N mouse embryonic stem cells. Genome research 2008, 18(10):1670-1679.

Hansis C, Barreto G, Maltry N, Niehrs C: Nuclear reprogramming of human somatic cells by xenopus egg extract requires BRG1. Curr Biol 2004, 14(16):1475-1480.

Hansis C, Grifo JA, Krey LC: Oct-4 expression in inner cell mass and trophectoderm of human blastocysts. Mol Hum Reprod 2000, 6(11):999-1004.

Hansis C, Tang YX, Grifo JA, Krey LC: Analysis of Oct-4 expression and ploidy in individual human blastomeres. Mol Hum Reprod 2001, 7(2):155-161.

Harris H: Behaviour of differentiated nuclei in heterokaryons of animal cells from different species. Nature 1965, 206(984):583-588.

Harris H, Watkins JF, Ford CE, Schoefl GI: Artificial heterokaryons of animal cells from different species. J Cell Sci 1966, 1(1):1-30.

Harvey W: Exercitationes de generatione animalium: Hagae Comitis, apud A. Leers, 1680.; 1651.

Hasegawa K, Nakatsuji N: Insulators prevent transcriptional interference between two promoters in a double gene construct for transgenesis. FEBS letters 2002, 520(1-3):47-52.

Hattori N, Nishino K, Ko YG, Hattori N, Ohgane J, Tanaka S, Shiota K: Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. J Biol Chem 2004, 279(17):17063-17069.

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S et al: Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell stem cell 2010, 6(5):479-491.

He YW, Deftos ML, Ojala EW, Bevan MJ: RORgamma t, a novel isoform of an orphan receptor, negatively regulates Fas ligand expression and IL-2 production in T cells. Immunity 1998, 9(6):797-806.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA et al: Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 2007, 39(3):311-318.

Heng JC, Feng B, Han J, Jiang J, Kraus P, Ng JH, Orlov YL, Huss M, Yang L, Lufkin T et al: The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell stem cell 2010, 6(2):167-174.

Heyman RA, Mangelsdorf DJ, Dyck JA, Stein RB, Eichele G, Evans RM, Thaller C: 9-Cis Retinoic Acid Is a High-Affinity Ligand for the Retinoid-X Receptor. Cell 1992, 68(2):397-406.

Hirose T, Smith RJ, Jetten AM: ROR gamma: the third member of ROR/RZR orphan receptor subfamily that is highly expressed in skeletal muscle. Biochem Bioph Res Co 1994, 205(3):1976-1983.

Hochedlinger K, Plath K: Epigenetic reprogramming and induced pluripotency. Development 2009, 136(4):509-523.

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

Holder N, Hill J: Retinoic acid modifies development of the midbrain-hindbrain border and affects cranial ganglion formation in zebrafish embryos. Development 1991, 113(4):1159-1170.

Hong H, Takahashi K, Ichisaka T, Aoi T, Kanagawa O, Nakagawa M, Okita K, Yamanaka S: Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature 2009, 460(7259):1132-1135.

Hu X, Lazar MA: The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. Nature 1999, 402(6757):93-96.

Huang X, Guo HF, Tammana S, Jung YC, Mellgren E, Bassi P, Cao Q, Tu ZJ, Kim YC, Ekker SC et al: Gene Transfer Efficiency and Genome-Wide Integration Profiling of Sleeping Beauty, Tol2, and PiggyBac Transposons in Human Primary T Cells (vol 18, pg 1803, 2010). Mol Ther 2010, 18(11):2038-2038.

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

Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA: Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nat Biotechnol 2008, 26(11):1269-1275.

Huggenvik JI, Collard MW, Kim YW, Sharma RP: Modification of the retinoic acid signaling pathway by the catalytic subunit of protein kinase-A. Mol Endocrinol 1993, 7(4):543-550.

Humphrey RK, Beattie GM, Lopez AD, Bucay N, King CC, Firpo MT, Rose-John S, Hayek A: Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. Stem Cells 2004, 22(4):522-530.

Hussein SM, Batada NN, Vuoristo S, Ching RW, Autio R, Narva E, Ng S, Sourour M, Hamalainen R, Olsson C et al: Copy number variation and selection during reprogramming to pluripotency. Nature 2011, 471(7336):58-62.

Ichida JK, Blanchard J, Lam K, Son EY, Chung JE, Egli D, Loh KM, Carter AC, Di Giorgio FP, Koszka K et al: A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell stem cell 2009, 5(5):491-503.

Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, Srivastava D: Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell 2010, 142(3):375-386.

Inoue A, Zhang Y: Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science 2011, 334(6053):194.

Itzhaki I, Maizels L, Huber I, Zwi-Dantsis L, Caspi O, Winterstern A, Feldman O, Gepstein A, Arbel G, Hammerman H et al: Modelling the long QT syndrome with induced pluripotent stem cells. Nature 2011, 471(7337):225-229.

Iuchi S, Easley K, Matsuzaki K, Weiner L, O'Connor N, Green H: Alternative subcellular locations of keratinocyte basonuclin. Experimental dermatology 2000, 9(3):178-184.

Iuchi S, Green H: Nuclear localization of basonuclin in human keratinocytes and the role of phosphorylation. P Natl Acad Sci USA 1997, 94(15):7948-7953.

Ivics Z, Hackett PB, Plasterk RH, Izsvak Z: Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 1997, 91(4):501-510.

Iwasaki H, Mizuno S, Arinobu Y, Ozawa H, Mori Y, Shigematsu H, Takatsu K, Tenen DG, Akashi K: The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. Gene Dev 2006, 20(21):3010-3021.

Izsvak Z, Ivics Z: Sleeping Beauty hits them all: transposon-mediated saturation mutagenesis in the mouse germline. Nature methods 2005, 2(10):735-736.

Jaenisch R: Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. P Natl Acad Sci USA 1976, 73(4):1260-1264.

James D, Noggle SA, Swigut T, Brivanlou AH: Contribution of human embryonic stem cells to mouse blastocysts. Developmental biology 2006, 295(1):90-102.

Jang H, Kim TW, Yoon S, Choi SY, Kang TW, Kim SY, Kwon YW, Cho EJ, Youn HD: O-GlcNAc Regulates Pluripotency and Reprogramming by Directly Acting on Core Components of the Pluripotency Network. Cell stem cell 2012, 11(1):62-74.

Jang J, Kang HC, Kim HS, Kim JY, Huh YJ, Kim DS, Yoo JE, Lee JA, Lim B, Lee J et al: Induced pluripotent stem cell models from X-linked adrenoleukodystrophy patients. Annals of neurology 2011, 70(3):402-409.

Jetten AM: Retinoid-related orphan receptors (RORs): critical roles in development, immunity, circadian rhythm, and cellular metabolism. Nuclear receptor signaling 2009, 7:e003.

Ji Z, Tian B: Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. PloS one 2009, 4(12):e8419.

Jiang J, Chan YS, Loh YH, Cai J, Tong GQ, Lim CA, Robson P, Zhong S, Ng HH: A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol 2008, 10(3):353-360.

Jiang W, Shi Y, Zhao D, Chen S, Yong J, Zhang J, Qing T, Sun X, Zhang P, Ding M et al: In vitro derivation of functional insulin-producing cells from human embryonic stem cells. Cell research 2007, 17(4):333-344.

Jin JP, Arias EE, Chen J, Harper JW, Walter JC: A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1. Mol Cell 2006, 23(5):709-721.

Judson RL, Babiarz JE, Venere M, Blelloch R: Embryonic stem cell-specific microRNAs promote induced pluripotency. Nat Biotechnol 2009, 27(5):459-461.

Jullien J, Astrand C, Halley-Stott RP, Garrett N, Gurdon JB: Characterization of somatic cell nuclear reprogramming by oocytes in which a linker histone is required for pluripotency gene reactivation. P Natl Acad Sci USA 2010, 107(12):5483-5488.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS et al: Mediator and cohesin connect gene expression and chromatin architecture. Nature 2010, 467(7314):430-435.

Kahan BW, Ephrussi B: Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. Journal of the National Cancer Institute 1970, 44(5):1015-1036.

Kaji K, Norrby K, Paca A, Mileikovsky M, Mohseni P, Woltjen K: Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature 2009, 458(7239):771-U112.

Kajimura S, Seale P, Kubota K, Lunsford E, Frangioni JV, Gygi SP, Spiegelman BM: Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-beta transcriptional complex. Nature 2009, 460(7259):1154-1158.

Kallen J, Schlaeppi JM, Bitsch F, Delhon I, Fournier B: Crystal structure of the human RORalpha Ligand binding domain in complex with cholesterol sulfate at 2.2 A. The Journal of biological chemistry 2004, 279(14):14033-14038.

Kang Y, Zhang XY, Jiang W, Wu CQ, Chen CM, Zheng YF, Gu JR, Xu CJ: Tumor-directed gene therapy in mice using a composite nonviral gene delivery system consisting of the piggyBac transposon and polyethylenimine. Bmc Cancer 2009, 9.

Kastner P, Grondona JM, Mark M, Gansmuller A, LeMeur M, Decimo D, Vonesch JL, Dolle P, Chambon P: Genetic analysis of RXR alpha developmental function: convergence of RXR and RAR signaling pathways in heart and eye morphogenesis. Cell 1994, 78(6):987-1003.

Kawakami K, Shima A, Kawakami N: Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. P Natl Acad Sci USA 2000, 97(21):11403-11408.

Kawamura T, Suzuki J, Wang YV, Menendez S, Morera LB, Raya A, Wahl GM, Izpisua Belmonte JC: Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature 2009, 460(7259):1140-1144.

Kelley MR, Kidd S, Berg RL, Young MW: Restriction of P-Element Insertions at the Notch Locus of Drosophila-Melanogaster. Mol Cell Biol 1987, 7(4):1545-1548.

Kemler I, Schreiber E, Muller MM, Matthias P, Schaffner W: Octamer transcription factors bind to two different sequence motifs of the immunoglobulin heavy chain promoter. The EMBO journal 1989, 8(7):2001-2008.

Keng VW, Villanueva A, Chiang DY, Dupuy AJ, Ryan BJ, Matise I, Silverstein KA, Sarver A, Starr TK, Akagi K et al: A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. Nat Biotechnol 2009, 27(3):264-274.

Kile BT, Hentges KE, Clark AT, Nakamura H, Salinger AP, Liu B, Box N, Stockton DW, Johnson RL, Behringer RR et al: Functional genetic analysis of mouse chromosome 11. Nature 2003, 425(6953):81-86.

Kim D, Kim CH, Moon JI, Chung YG, Chang MY, Han BS, Ko S, Yang E, Cha KY, Lanza R et al: Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell stem cell 2009, 4(6):472-476.

Kim J, Chu J, Shen X, Wang J, Orkin SH: An extended transcriptional network for pluripotency of embryonic stem cells. Cell 2008, 132(6):1049-1061.

Kim JB, Greber B, Arauzo-Bravo MJ, Meyer J, Park KI, Zaehres H, Scholer HR: Direct reprogramming of human neural stem cells by OCT4. Nature 2009, 461(7264):649-643.

Kim JB, Sebastiano V, Wu G, Arauzo-Bravo MJ, Sasse P, Gentile L, Ko K, Ruau D, Ehrich M, van den Boom D et al: Oct4-induced pluripotency in adult neural stem cells. Cell 2009, 136(3):411-419.

Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LI et al: Epigenetic memory in induced pluripotent stem cells. Nature 2010, 467(7313):285-290.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B: A high-resolution map of active promoters in the human genome. Nature 2005, 436(7052):876-880.

King TJ, Briggs R: Changes in the Nuclei of Differentiating Gastrula Cells, as Demonstrated by Nuclear Transplantation. P Natl Acad Sci USA 1955, 41(5):321-325.

Klein G: Specific chromosomal translocations and the genesis of B-cell-derived tumors in mice and men. Cell 1983, 32(2):311-315.

Kleinsmith LJ, Pierce GB, Jr.: Multipotentiality of Single Embryonal Carcinoma Cells. Cancer research 1964, 24:1544-1551.

Klemm JD, Rould MA, Aurora R, Herr W, Pabo CO: Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell 1994, 77(1):21-32.

Klempnauer KH, Gonda TJ, Bishop JM: Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene. Cell 1982, 31(2 Pt 1):453-463.

Kliewer SA, Umesono K, Mangelsdorf DJ, Evans RM: Retinoid X receptor interacts with nuclear receptors in retinoic acid, thyroid hormone and vitamin D3 signalling. Nature 1992, 355(6359):446-449.

Kobayashi T, Yamaguchi T, Hamanaka S, Kato-Itoh M, Yamazaki Y, Ibata M, Sato H, Lee YS, Usui J, Knisely AS et al: Generation of rat pancreas in mouse by interspecific blastocyst injection of pluripotent stem cells. Cell 2010, 142(5):787-799.

Kopp JL, Ormsbee BD, Desler M, Rizzino A: Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells 2008, 26(4):903-911.

Krust A, Kastner P, Petkovich M, Zelent A, Chambon P: A third human retinoic acid receptor, hRAR-gamma. P Natl Acad Sci USA 1989, 86(14):5310-5314.

Kuehn MR, Bradley A, Robertson EJ, Evans MJ: A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature 1987, 326(6110):295-298.

Kuijk EW, van Tol LT, Van de Velde H, Wubbolts R, Welling M, Geijsen N, Roelen BA: The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. Development 2012, 139(5):871-882.

Kulessa H, Frampton J, Graf T: GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. Gene Dev 1995, 9(10):1250-1262.

Kunath T, Saba-El-Leil MK, Almousailleakh M, Wray J, Meloche S, Smith A: FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development 2007, 134(16):2895-2902.

Laiosa CV, Stadtfeld M, Xie H, de Andres-Aguayo L, Graf T: Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity 2006, 25(5):731-744.

Lammer EJ, Chen DT, Hoar RM, Agnish ND, Benke PJ, Braun JT, Curry CJ, Fernhoff PM, Grix AW, Jr., Lott IT et al: Retinoic acid embryopathy. The New England journal of medicine 1985, 313(14):837-841.

Lang MR, Patterson LB, Gordon TN, Johnson SL, Parichy DM: Basonuclin-2 Requirements for Zebrafish Adult Pigment Pattern Development and Female Fertility. PLoS genetics 2009, 5(11).

Laskey RA, Gurdon JB: Genetic content of adult somatic cells tested by nuclear transplantation from cultured cells. Nature 1970, 228(5278):1332-1334.

Laursen KB, Wong PM, Gudas LJ: Epigenetic regulation by RARalpha maintains ligand-independent transcriptional activity. Nucleic Acids Res 2012, 40(1):102-115.

Ledda-Columbano GM, Pibiri M, Molotzu F, Cossu C, Sanna L, Simbula G, Perra A, Columbano A: Induction of hepatocyte proliferation by retinoic acid. Carcinogenesis 2004, 25(11):2061-2066.

Lee G, Papapetrou EP, Kim H, Chambers SM, Tomishima MJ, Fasano CA, Ganat YM, Menon J, Shimizu F, Viale A et al: Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. Nature 2009, 461(7262):402-406.

Lee J, Kim HK, Rho JY, Han YM, Kim J: The human OCT-4 isoforms differ in their ability to confer self-renewal. J Biol Chem 2006, 281(44):33554-33565.

Lee JM, Kim IS, Kim H, Lee JS, Kim K, Yim HY, Jeong J, Kim JH, Kim JY, Lee H et al: RORalpha attenuates Wnt/beta-catenin signaling by PKCalpha-dependent phosphorylation in colon cancer. Mol Cell 2010, 37(2):183-195.

Leeb M, Wutz A: Derivation of haploid embryonic stem cells from mouse embryos. Nature 2011, 479(7371):131-134.

Leeb M, Wutz A: Derivation of haploid embryonic stem cells from mouse embryos. Nature 2011, 479(7371):131-134.

Lefort N, Feyeux M, Bas C, Feraud O, Bennaceur-Griscelli A, Tachdjian G, Peschanski M, Perrier AL: Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. Nat Biotechnol 2008, 26(12):1364-1366.

Leid M, Kastner P, Lyons R, Nakshatri H, Saunders M, Zacharewski T, Chen JY, Staub A, Garnier JM, Mader S et al: Purification, cloning, and RXR identity of the

HeLa cell factor with which RAR or TR heterodimerizes to bind target sequences efficiently. Cell 1992, 68(2):377-395.

Lengner CJ, Gimelbrant AA, Erwin JA, Cheng AW, Guenther MG, Welstead GG, Alagappan R, Frampton GM, Xu P, Muffat J et al: Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell 2010, 141(5):872-883.

Lenhoff SG, Lenhoff HM: Hydra and the birth of experimental biology-1744: Abraham Trembley's memoirs concerning the natural history of a type of freshwater polyp with arms shaped like horns. . Pacific Grove, CA: Boxwood Press; 1986.

Levin AA, Sturzenbecker LJ, Kazmer S, Bosakowski T, Huselton C, Allenby G, Speck J, Kratzeisen C, Rosenberger M, Lovey A et al: 9-cis retinoic acid stereoisomer binds and activates the nuclear receptor RXR alpha. Nature 1992, 355(6358):359-361.

Li H, Collado M, Villasante A, Strati K, Ortega S, Canamero M, Blasco MA, Serrano M: The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature 2009, 460(7259):1136-1139.

Li J, Shen H, Himmel KL, Dupuy AJ, Largaespada DA, Nakamura T, Shaughnessy JD, Jr., Jenkins NA, Copeland NG: Leukaemia disease genes: large-scale cloning and pathway predictions. Nat Genet 1999, 23(3):348-353.

Li J, Zhang JM, Li X, Suo F, Zhang MJ, Hou W, Han J, Du LL: A piggyBac transposon-based mutagenesis system for the fission yeast Schizosaccharomyces pombe. Nucleic Acids Res 2011, 39(6):e40.

Li MA, Turner DJ, Ning Z, Yusa K, Liang Q, Eckert S, Rad L, Fitzgerald TW, Craig NL, Bradley A: Mobilization of giant piggyBac transposons in the mouse genome. Nucleic Acids Res 2011, 39(22):e148.

Li R, Liang J, Ni S, Zhou T, Qing X, Li H, He W, Chen J, Li F, Zhuang Q et al: A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell stem cell 2010, 7(1):51-63.

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

Liang Q, Kong J, Stalker J, Bradley A: Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. Genesis 2009, 47(6):404-408.

Liao B, Bao X, Liu L, Feng S, Zovoilis A, Liu W, Xue Y, Cai J, Guo X, Qin B et al: MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. The Journal of biological chemistry 2011, 286(19):17359-17364.

Liao J, Cui C, Chen S, Ren J, Chen J, Gao Y, Li H, Jia N, Cheng L, Xiao H et al: Generation of induced pluripotent stem cell lines from adult rat cells. Cell stem cell 2009, 4(1):11-15.

Lin H, Shabbir A, Molnar M, Lee T: Stem cell regulatory function mediated by expression of a novel mouse Oct4 pseudogene. Biochem Bioph Res Co 2007, 355(1):111-116.

Lin JJ, Li Y, Eppinga RD, Wang Q, Jin JP: Chapter 1: roles of caldesmon in cell motility and actin cytoskeleton remodeling. International review of cell and molecular biology 2009, 274:1-68.

Lin S, Gaiano N, Culp P, Burns JC, Friedmann T, Yee JK, Hopkins N: Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. Science 1994, 265(5172):666-669.

Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S et al: Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 2011, 471(7336):68-73.

Liu H, Ye Z, Kim Y, Sharkis S, Jang YY: Generation of endoderm-derived human induced pluripotent stem cells from primary hepatocytes. Hepatology (Baltimore, Md 2010, 51(5):1810-1819.

Liu H, Zhu F, Yong J, Zhang P, Hou P, Li H, Jiang W, Cai J, Liu M, Cui K et al: Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts. Cell stem cell 2008, 3(6):587-590.

Liu XX, Nugoli M, Laferrierea J, Saleh SM, Rodrigue-Gervais IG, Saleh M, Park M, Hallett MT, Muller WJ, Giguere V: Stromal retinoic acid receptor beta promotes mammary gland tumorigenesis. P Natl Acad Sci USA 2011, 108(2):774-779.

Lluis F, Ombrato L, Pedone E, Pepe S, Merrill BJ, Cosma MP: T-cell factor 3 (Tcf3) deletion increases somatic cell reprogramming by inducing epigenome modifications. P Natl Acad Sci USA 2011, 108(29):11912-11917.

Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S et al: Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet 2010, 42(12):1113-1117.

Loh YH, Agarwal S, Park IH, Urbach A, Huo H, Heffner GC, Kim K, Miller JD, Ng K, Daley GQ: Generation of induced pluripotent stem cells from human blood. Blood 2009, 113(22):5476-5479.

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J et al: The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 2006, 38(4):431-440.

Lohle M, Hermann A, Glass H, Kempe A, Schwarz SC, Kim JB, Poulet C, Ravens U, Schwarz J, Scholer HR et al: Differentiation efficiency of induced pluripotent stem cells depends on the number of reprogramming factors. Stem Cells 2012, 30(3):570-579.

Lohnes D, Kastner P, Dierich A, Mark M, LeMeur M, Chambon P: Function of retinoic acid receptor gamma in the mouse. Cell 1993, 73(4):643-658.

Lohnes D, Mark M, Mendelsohn C, Dolle P, Dierich A, Gorry P, Gansmuller A, Chambon P: Function of the retinoic acid receptors (RARs) during development (I).

Craniofacial and skeletal abnormalities in RAR double mutants. Development 1994, 120(10):2723-2748.

Lufkin T, Lohnes D, Mark M, Dierich A, Gorry P, Gaub MP, LeMeur M, Chambon P: High postnatal lethality and testis degeneration in retinoic acid receptor alpha mutant mice. P Natl Acad Sci USA 1993, 90(15):7225-7229.

Luo GB, Ivics Z, Izsvak Z, Bradley A: Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. P Natl Acad Sci USA 1998, 95(18):10769-10773.

Luo J, Sucov HM, Bader JA, Evans RM, Giguere V: Compound mutants for retinoic acid receptor (RAR) beta and RAR alpha 1 reveal developmental functions for multiple RAR beta isoforms. Mechanisms of development 1996, 55(1):33-44.

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL: Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 2012.

Maekawa M, Yamaguchi K, Nakamura T, Shibukawa R, Kodanaka I, Ichisaka T, Kawamura Y, Mochizuki H, Goshima N, Yamanaka S: Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. Nature 2011, 474(7350):225-229.

Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R et al: Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell stem cell 2007, 1(1):55-70.

Maki N, Suetsugu-Maki R, Sano S, Nakamura K, Nishimura O, Tarui H, Del Rio-Tsonis K, Ohsumi K, Agata K, Tsonis PA: Oocyte-type linker histone B4 is required for transdifferentiation of somatic cells in vivo. Faseb J 2010, 24(9):3462-3467.

Mann KM, Ward JM, Yew CC, Kovochich A, Dawson DW, Black MA, Brett BT, Sheetz TE, Dupuy AJ, Chang DK et al: Sleeping Beauty mutagenesis reveals

cooperating mutations and pathways in pancreatic adenocarcinoma. P Natl Acad Sci USA 2012.

Mansour AA, Gafni O, Weinberger L, Zviran A, Ayyash M, Rais Y, Krupalnik V, Zerbib M, Amann-Zalcenstein D, Maza I et al: The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. Nature 2012, 488(7411):409-413.

Mansour AA, Gafni O, Weinberger L, Zviran A, Ayyash M, Rais Y, Krupalnik V, Zerbib M, Amann-Zalcenstein D, Maza I et al: The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. Nature 2012, 488(7411):409-413.

Marion RM, Strati K, Li H, Murga M, Blanco R, Ortega S, Fernandez-Capetillo O, Serrano M, Blasco MA: A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature 2009, 460(7259):1149-1153.

Mark M, Ghyselinck NB, Chambon P: Function of retinoic acid receptors during embryonic development. Nuclear receptor signaling 2009, 7:e002.

Marks H, Kalkan T, Menafra R, Denissov S, Jones K, Hofemeister H, Nichols J, Kranz A, Stewart AF, Smith A et al: The transcriptional and epigenomic foundations of ground state pluripotency. Cell 2012, 149(3):590-604.

Marson A, Foreman R, Chevalier B, Bilodeau S, Kahn M, Young RA, Jaenisch R: Wnt signaling promotes reprogramming of somatic cells to pluripotency. Cell stem cell 2008, 3(2):132-135.

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J et al: Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell 2008, 134(3):521-533.

Marston SB, Lehman W: Caldesmon Is a Ca-2+-Regulatory Component of Native Smooth-Muscle Thin-Filaments. Biochem J 1985, 231(3):517-522.

Martin GR: Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. P Natl Acad Sci USA 1981, 78(12):7634-7638.

Martin GR, Evans MJ: The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. Cell 1974, 2(3):163-172.

Masaki H, Ishikawa T, Takahashi S, Okumura M, Sakai N, Haga M, Kominami K, Migita H, McDonald F, Shimada F et al: Heterogeneity of pluripotent marker gene expression in colonies generated in human iPS cell induction culture. Stem Cell Res 2008, 1(2):105-115.

Masia S, Alvarez S, de Lera AR, Barettino D: Rapid, nongenomic actions of retinoic acid on phosphatidylinositol-3-kinase signaling pathway mediated by the retinoic acid receptor. Mol Endocrinol 2007, 21(10):2391-2402.

Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA et al: Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol 2007, 9(6):625-635.

Matkovits T, Christakos S: Ligand occupancy is not required for vitamin D receptor and retinoid receptor-mediated transcriptional activation. Mol Endocrinol 1995, 9(2):232-242.

Matsui Y, Toksoz D, Nishikawa S, Nishikawa SI, Williams D, Zsebo K, Hogan BLM: Effect of Steel Factor and Leukemia Inhibitory Factor on Murine Primordial Germ-Cells in Culture. Nature 1991, 353(6346):750-752.

Matsui Y, Zsebo K, Hogan BLM: Derivation of Pluripotential Embryonic Stem-Cells from Murine Primordial Germ-Cells in Culture. Cell 1992, 70(5):841-847.

Matsui Y, Zsebo K, Hogan BLM: Derivation of Pluripotential Embryonic Stem-Cells from Murine Primordial Germ-Cells in Culture. Cell 1992, 70(5):841-847.

Mayshar Y, Ben-David U, Lavon N, Biancotti JC, Yakir B, Clark AT, Plath K, Lowry WE, Benvenisty N: Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell stem cell 2010, 7(4):521-531.

McClintock B: The origin and behavior of mutable loci in maize. P Natl Acad Sci USA 1950, 36(6):344-355.

McGrath J, Solter D: Nuclear transplantation in the mouse embryo by microsurgery and cell fusion. Science 1983, 220(4603):1300-1302.

McMahon AP, Bradley A: The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. Cell 1990, 62(6):1073-1085.

McNamara S, Wang H, Hanna N, Miller WH, Jr.: Topoisomerase IIbeta negatively modulates retinoic acid receptor alpha function: a novel mechanism of retinoic acid resistance. Mol Cell Biol 2008, 28(6):2066-2077.

Medvedev A, Yan ZH, Hirose T, Giguere V, Jetten AM: Cloning of a cDNA encoding the murine orphan receptor RZR/ROR gamma and characterization of its response element. Gene 1996, 181(1-2):199-206.

Meeker TC, Nagarajan L, ar-Rushdi A, Croce CM: Cloning and characterization of the human PIM-1 gene: a putative oncogene related to the protein kinases. Journal of cellular biochemistry 1987, 35(2):105-112.

Meir YJJ, Weirauch MT, Yang HS, Chung PC, Yu RK, Wu SCY: Genome-wide target profiling of piggyBac and Tol2 in HEK 293: pros and cons for gene discovery and gene therapy. Bmc Biotechnol 2011, 11.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB et al: Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 2008, 454(7205):766-770.

Meissner A, Wernig M, Jaenisch R: Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. Nat Biotechnol 2007, 25(10):1177-1181.

Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG, Jr., Kinney JB et al: Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol 2012, 30(3):271-277.

Mendelsohn C, Lohnes D, Decimo D, Lufkin T, LeMeur M, Chambon P, Mark M: Function of the retinoic acid receptors (RARs) during development (II). Multiple abnormalities at various stages of organogenesis in RAR double mutants. Development 1994, 120(10):2749-2771.

Mermod N, O'Neill EA, Kelly TJ, Tjian R: The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. Cell 1989, 58(4):741-753.

Metallo CM, Ji L, de Pablo JJ, Palecek SP: Retinoic acid and bone morphogenetic protein signaling synergize to efficiently direct epithelial differentiation of human embryonic stem cells. Stem Cells 2008, 26(2):372-380.

Metzker ML: Sequencing technologies - the next generation. Nat Rev Genet 2010, 11(1):31-46.

Meyer T, Kneissel M, Mariani J, Fournier B: In vitro and in vivo evidence for orphan nuclear receptor RORalpha function in bone metabolism. P Natl Acad Sci USA 2000, 97(16):9197-9202.

Mikkelsen S, Berne B, Staberg B, Vahlquist A: Potentiating effect of dietary vitamin A on photocarcinogenesis in hairless mice. Carcinogenesis 1998, 19(4):663-666.

Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A: Dissecting direct reprogramming through integrative genomic analysis. Nature 2008, 454(7200):49-55.

Minucci S, Botquin V, Yeom YI, Dey A, Sylvester I, Zand DJ, Ohbo K, Ozato K, Scholer HR: Retinoic acid-mediated down-regulation of Oct3/4 coincides with the loss of promoter occupancy in vivo. The EMBO journal 1996, 15(4):888-899.

Miyoshi N, Ishii H, Nagano H, Haraguchi N, Dewi DL, Kano Y, Nishikawa S, Tanemura M, Mimori K, Tanaka F et al: Reprogramming of mouse and human cells to pluripotency using mature microRNAs. Cell stem cell 2011, 8(6):633-638.

Mohan M, Malayer JR, Geisert RD, Morgan GL: Expression of retinol-binding protein messenger RNA and retinoic acid receptors in preattachment bovine embryos. Molecular reproduction and development 2001, 60(3):289-296.

Mollersen L, Paulsen JE, Olstorn HB, Knutsen HK, Alexander J: Dietary retinoic acid supplementation stimulates intestinal tumour formation and growth in multiple intestinal neoplasia (Min)/+ mice. Carcinogenesis 2004, 25(1):149-153.

Montserrat N, de Onate L, Garreta E, Gonzalez F, Adamo A, Eguizabal C, Hafner S, Vassena R, Belmonte JC: Generation of feeder free pig induced pluripotent stem cells without Pou5f1. Cell transplantation 2011.

Moretti A, Bellin M, Welling A, Jung CB, Lam JT, Bott-Flugel L, Dorn T, Goedel A, Hohnke C, Hofmann F et al: Patient-specific induced pluripotent stem-cell models for long-QT syndrome. The New England journal of medicine 2011, 363(15):1397-1409.

Morishita K, Parker DS, Mucenski ML, Jenkins NA, Copeland NG, Ihle JN: Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. Cell 1988, 54(6):831-840.

Morriss-Kay GM, Murphy P, Hill RE, Davidson DR: Effects of retinoic acid excess on expression of Hox-2.9 and Krox-20 and on morphological segmentation in the hindbrain of mouse embryos. The EMBO journal 1991, 10(10):2985-2995.

Mucenski ML, Taylor BA, Ihle JN, Hartley JW, Morse HC, 3rd, Jenkins NA, Copeland NG: Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors. Mol Cell Biol 1988, 8(1):301-308.

Nagamatsu G, Kosaka T, Kawasumi M, Kinoshita T, Takubo K, Akiyama H, Sudo T, Kobayashi T, Oya M, Suda T: A germ cell-specific gene, Prmt5, works in somatic cell reprogramming. The Journal of biological chemistry 2011, 286(12):10641-10648.

Nagy A, Gocza E, Diaz EM, Prideaux VR, Ivanyi E, Markkula M, Rossant J: Embryonic stem cells alone are able to support fetal development in the mouse. Development 1990, 110(3):815-821.

Nagy A, Rossant J, Nagy R, Abramow-Newerly W, Roder JC: Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. P Natl Acad Sci USA 1993, 90(18):8424-8428.

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S: Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol 2008, 26(1):101-106.

Nakagawa S, Watanabe M, Inoue Y: Prominent expression of nuclear hormone receptor ROR alpha in Purkinje cells from early development. Neuroscience research 1997, 28(2):177-184.

Nakagawa Y, O'Leary DD: Dynamic patterned expression of orphan nuclear receptor genes RORalpha and RORbeta in developing mouse forebrain. Developmental neuroscience 2003, 25(2-4):234-244.

Narlikar GJ, Fan HY, Kingston RE: Cooperation between complexes that regulate chromatin structure and transcription. Cell 2002, 108(4):475-487.

Narsinh KH, Sun N, Sanchez-Freire V, Lee AS, Almeida P, Hu SJ, Jan T, Wilson KD, Leong D, Rosenberg J et al: Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells. J Clin Invest 2011, 121(3):1217-1221.

Neel BG, Hayward WS, Robinson HL, Fang J, Astrin SM: Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. Cell 1981, 23(2):323-334.

Nichols J, Chambers I, Taga T, Smith A: Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development 2001, 128(12):2333-2339.

Nichols J, Jones K, Phillips JM, Newland SA, Roode M, Mansfield W, Smith A, Cooke A: Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. Nature medicine 2009, 15(7):814-818.

Nichols J, Silva J, Roode M, Smith A: Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development 2009, 136(19):3215-3222.

Nichols J, Smith A: Naive and primed pluripotent states. Cell stem cell 2009, 4(6):487-492.

Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A: Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 1998, 95(3):379-391.

Nielsen R, Pedersen TA, Hagenbeek D, Moulos P, Siersbaek R, Megens E, Denissov S, Borgesen M, Francoijs KJ, Mandrup S et al: Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. Gene Dev 2008, 22(21):2953-2967.

Niwa H, Burdon T, Chambers I, Smith A: Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Gene Dev 1998, 12(13):2048-2060.

Niwa H, Miyazaki J, Smith AG: Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 2000, 24(4):372-376.

Niwa H, Yamamura K, Miyazaki J: Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene 1991, 108(2):193-199.

Noggle S, Fung HL, Gore A, Martinez H, Satriani KC, Prosser R, Oum K, Paull D, Druckenmiller S, Freeby M et al: Human oocytes reprogram somatic cells to a pluripotent state. Nature 2011, 478(7367):70-75.

Nordhoff V, Hubner K, Bauer A, Orlova I, Malapetsa A, Scholer HR: Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences. Mamm Genome 2001, 12(4):309-317.

Nori S, Okada Y, Yasuda A, Tsuji O, Takahashi Y, Kobayashi Y, Fujiyoshi K, Koike M, Uchiyama Y, Ikeda E et al: Grafted human-induced pluripotent stem-cell-derived neurospheres promote motor functional recovery after spinal cord injury in mice. P Natl Acad Sci USA 2011, 108(40):16825-16830.

Notarianni E, Galli C, Laurie S, Moor RM, Evans MJ: Derivation of pluripotent, embryonic cell lines from the pig and sheep. Journal of reproduction and fertility 1991, 43:255-260.

Notarianni E, Laurie S, Moor RM, Evans MJ: Maintenance and differentiation in culture of pluripotential embryonic cell lines from pig blastocysts. Journal of reproduction and fertility 1990, 41:51-56.

Novy RE, Lin JLC, Lin JJC: Characterization of Cdna Clones Encoding a Human Fibroblast Caldesmon Isoform and Analysis of Caldesmon Expression in Normal and Transformed-Cells. J Biol Chem 1991, 266(25):16917-16924.

O'Brien RN, Shen Z, Tachikawa K, Lee PA, Briggs SP: Quantitative proteome analysis of pluripotent cells by iTRAQ mass tagging reveals post-transcriptional regulation of proteins required for ES cell self-renewal. Molecular & cellular proteomics : MCP 2010, 9(10):2238-2251.

Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ et al: Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat Cell Biol 2011, 13(5):541-549.

Okada Y, Shimazaki T, Sobue G, Okano H: Retinoic-acid-concentration-dependent acquisition of neural cell identity during in vitro differentiation of mouse embryonic stem cells. Developmental biology 2004, 275(1):124-142.

Okamoto K, Okazawa H, Okuda A, Sakai M, Muramatsu M, Hamada H: A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell 1990, 60(3):461-472.

Okazawa H, Okamoto K, Ishino F, Ishino-Kaneko T, Takeda S, Toyoda Y, Muramatsu M, Hamada H: The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. The EMBO journal 1991, 10(10):2997-3005.

Okita K, Ichisaka T, Yamanaka S: Generation of germline-competent induced pluripotent stem cells. Nature 2007, 448(7151):313-317.

Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S: Generation of mouse induced pluripotent stem cells without viral vectors. Science 2008, 322(5903):949-953.

O'Leary T, Heindryckx B, Lierman S, van Bruggen D, Goeman JJ, Vandewoestyne M, Deforce D, Lopes SMCD, De Sutter P: Tracking the progression of the human inner cell mass during embryonic stem cell derivation. Nat Biotechnol 2012, 30(3):278-+.

Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL, Valanis B, Williams JH et al: Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. New Engl J Med 1996, 334(18):1150-1155.

Pain B, Clark ME, Shen M, Nakazawa H, Sakurai M, Samarut J, Etches RJ: Long-term in vitro culture and characterisation of avian embryonic stem cells with multiple morphogenetic potentialities. Development 1996, 122(8):2339-2348.

Palmieri SL, Peter W, Hess H, Scholer HR: Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. Developmental biology 1994, 166(1):259-267.

Panagopoulos I, Moller E, Collin A, Mertens F: The POU5F1P1 pseudogene encodes a putative protein similar to POU5F1 isoform 1. Oncol Rep 2008, 20(5):1029-1033.

Pander CH: Beiträge zur Entwickelungsgeschichte des Hühnchens im Eye: Wurzburg; 1817.

Papalopulu N, Clarke JD, Bradley L, Wilkinson D, Krumlauf R, Holder N: Retinoic acid causes abnormal development and segmental patterning of the anterior hindbrain in Xenopus embryos. Development 1991, 113(4):1145-1158.

Papamichos SI, Kotoula V, Tarlatzis BC, Agorastos T, Papazisis K, Lambropoulos AF: OCT4B1 isoform: the novel OCT4 alternative spliced variant as a putative marker of stemness. Mol Hum Reprod 2009, 15(5):269-270.

Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM, Choudhary J: An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell stem cell 2010, 6(4):382-395.

Park IH, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, Lensch MW, Cowan C, Hochedlinger K, Daley GQ: Disease-specific induced pluripotent stem cells. Cell 2008, 134(5):877-886.

Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM et al: Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol 2012, 30(3):265-270.

Pavlath GK, Blau HM: Expression of muscle genes in heterokaryons depends on gene dosage. The Journal of cell biology 1986, 102(1):124-130.

Pease S, Braghetta P, Gearing D, Grail D, Williams RL: Isolation of embryonic stem (ES) cells in media supplemented with recombinant leukemia inhibitory factor (LIF). Developmental biology 1990, 141(2):344-352.

Pereira CF, Terranova R, Ryan NK, Santos J, Morris KJ, Cui W, Merkenschlager M, Fisher AG: Heterokaryon-based reprogramming of human B lymphocytes for pluripotency requires Oct4 but not Sox2. PLoS genetics 2008, 4(9):e1000170.

Pesce M, Wang X, Wolgemuth DJ, Scholer H: Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. Mechanisms of development 1998, 71(1-2):89-98.

Petkovich M, Brand NJ, Krust A, Chambon P: A human retinoic acid receptor which belongs to the family of nuclear receptors. Nature 1987, 330(6147):444-450.

Petroski MD, Deshaies RJ: Function and regulation of Cullin-RING ubiquitin ligases. Nat Rev Mol Cell Bio 2005, 6(1):9-20.

Phanstiel DH, Brumbaugh J, Wenger CD, Tian S, Probasco MD, Bailey DJ, Swaney DL, Tervo MA, Bolin JM, Ruotti V et al: Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nature methods 2011, 8(10):821-827.

Picard E, Seguin C, Monhoven N, Rochette-Egly C, Siat J, Borrelly J, Martinet Y, Martinet N, Vignaud JM: Expression of retinoid receptor genes and proteins in non-small-cell lung cancer. Journal of the National Cancer Institute 1999, 91(12):1059-1066.

Pikarsky E, Sharir H, Ben-Shushan E, Bergman Y: Retinoic acid represses Oct-3/4 gene expression through several retinoic acid-responsive elements located in the promoter-enhancer region. Mol Cell Biol 1994, 14(2):1026-1038.

Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T et al: Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol 2010, 28(8):848-855.

Prasad SM, Czepiel M, Cetinkaya C, Smigielska K, Weli SC, Lysdahl H, Gabrielsen A, Petersen K, Ehlers N, Fink T et al: Continuous hypoxic culturing maintains activation of Notch and allows long-term propagation of human embryonic stem cells without spontaneous differentiation. Cell proliferation 2009, 42(1):63-74.

Prigione A, Fauler B, Lurz R, Lehrach H, Adjaye J: The Senescence-Related Mitochondrial/Oxidative Stress Pathway is Repressed in Human Induced Pluripotent Stem Cells. Stem Cells 2010, 28(4):721-733.

Rad R, Rad L, Wang W, Cadinanos J, Vassiliou G, Rice S, Campos LS, Yusa K, Banerjee R, Li MA et al: PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. Science 2010, 330(6007):1104-1107.

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA: c-Myc regulates transcriptional pause release. Cell 2010, 141(3):432-445.

Ramirez-Solis R, Davis AC, Bradley A: Gene targeting in embryonic stem cells. Methods in enzymology 1993, 225:855-878.

Rashid ST, Corbineau S, Hannan N, Marciniak SJ, Miranda E, Alexander G, Huang-Doran I, Griffin J, Ahrlund-Richter L, Skepper J et al: Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. The Journal of clinical investigation 2010, 120(9):3127-3136.

Raya A, Rodriguez-Piza I, Guenechea G, Vassena R, Navarro S, Barrero MJ, Consiglio A, Castella M, Rio P, Sleep E et al: Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. Nature 2009, 460(7251):53-59.

Redmer T, Diecke S, Grigoryan T, Quiroga-Negreira A, Birchmeier W, Besser D: E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. EMBO reports 2011, 12(7):720-726.

Rhinn M, Dolle P: Retinoic acid signalling during development. Development 2012, 139(5):843-858.

Rochette-Egly C, Adam S, Rossignol M, Egly JM, Chambon P: Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIIH and phosphorylation by CDK7. Cell 1997, 90(1):97-107.

Rochette-Egly C, Germain P: Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). Nuclear receptor signaling 2009, 7:e005.

Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P: Transcriptional regulation of nanog by OCT4 and SOX2. The Journal of biological chemistry 2005, 280(26):24731-24737.

Roh TY, Cuddapah S, Zhao K: Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Gene Dev 2005, 19(5):542-552.

Romano RA, Li H, Tummala R, Maul R, Sinha S: Identification of Basonuclin2, a DNA-binding zinc-finger protein expressed in germ tissues and skin keratinocytes. Genomics 2004, 83(5):821-833.

Roode M, Blair K, Snell P, Elder K, Marchant S, Smith A, Nichols J: Human hypoblast formation is not dependent on FGF signalling. Developmental biology 2012, 361(2):358-363.

Rosenthal MD, Wishnow RM, Sato GH: In vitro growth and differetiation of clonal populations of multipotential mouse clls derived from a transplantable testicular teratocarcinoma. Journal of the National Cancer Institute 1970, 44(5):1001-1014.

Rosner MH, Vigano MA, Ozato K, Timmons PM, Poirier F, Rigby PW, Staudt LM: A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature 1990, 345(6277):686-692.

Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M et al: Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 2000, 24(3):227-235.

Ross-Innes CS, Stark R, Holmes KA, Schmidt D, Spyrou C, Russell R, Massie CE, Vowler SL, Eldridge M, Carroll JS: Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. Gene Dev 2010, 24(2):171-182.

Roux W: Beitrage zur Entwicklungsmechanik des Embryo. Virchows Arch Pathol Anat Physiol Klin Med 1888, 144:113-153.

Rowland BD, Bernards R, Peeper DS: The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene. Nat Cell Biol 2005, 7(11):1074-1082.

Ruberte E, Dolle P, Chambon P, Morriss-Kay G: Retinoic acid receptors and cellular retinoid binding proteins. II. Their differential pattern of transcription during early morphogenesis in mouse embryos. Development 1991, 111(1):45-60.

Rubin GM, Kidwell MG, Bingham PM: The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell 1982, 29(3):987-994.

Rushlow KE, Lautenberger JA, Reddy EP, Souza LM, Baluda MA, Chirikjian JG, Papas TS: Nucleotide sequence analysis of the long terminal repeat of avian myeloblastosis virus and adjacent host sequences. Journal of virology 1982, 42(3):840-846.

Saito S, Strelchenko N, Niemann H: Bovine Embryonic Stem Cell-Like Cell-Lines Cultured over Several Passages. Roux Arch Dev Biol 1992, 201(3):134-141.

Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL: Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell stem cell 2010, 7(1):64-77.

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M: Nucleotide sequence of bacteriophage phi X174 DNA. Nature 1977, 265(5596):687-695.

Saxe JP, Tomilin A, Scholer HR, Plath K, Huang J: Post-Translational Regulation of Oct4 Transcriptional Activity. PloS one 2009, 4(2).

Schaeren-Wiemers N, Andre E, Kapfhammer JP, Becker-Andre M: The expression pattern of the orphan nuclear receptor RORbeta in the developing and adult rat nervous system suggests a role in the processing of sensory information and in circadian rhythm. The European journal of neuroscience 1997, 9(12):2687-2701.

Schnetz MP, Handoko L, Akhtar-Zaidi B, Bartels CF, Pereira CF, Fisher AG, Adams DJ, Flicek P, Crawford GE, Laframboise T et al: CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. PLoS genetics 2010, 6(7):e1001023.

Scholer HR, Dressler GR, Balling R, Rohdewohld H, Gruss P: Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. The EMBO journal 1990, 9(7):2185-2195.

Scholer HR, Hatzopoulos AK, Balling R, Suzuki N, Gruss P: A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. The EMBO journal 1989, 8(9):2543-2550.

Scholer HR, Ruppert S, Suzuki N, Chowdhury K, Gruss P: New type of POU domain in germ line-specific protein Oct-4. Nature 1990, 344(6265):435-439.

Schoorlemmer J, van Puijenbroek A, van Den Eijnden M, Jonk L, Pals C, Kruijer W: Characterization of a negative retinoic acid response element in the murine Oct4 promoter. Mol Cell Biol 1994, 14(2):1122-1136.

Schuldiner M, Eiges R, Eden A, Yanuka O, Itskovitz-Eldor J, Goldstein RS, Benvenisty N: Induced neuronal differentiation of human embryonic stem cells. Brain research 2001, 913(2):201-205.

Selten G, Cuypers HT, Boelens W, Robanus-Maandag E, Verbeek J, Domen J, van Beveren C, Berns A: The primary structure of the putative oncogene pim-1 shows extensive homology with protein kinases. Cell 1986, 46(4):603-611.

Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, Keren L, Yakhini Z, Weinberger A, Segal E: Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat Biotechnol 2012.

Shi Y, Do JT, Desponts C, Hahm HS, Scholer HR, Ding S: A combined chemical and genetic approach for the generation of induced pluripotent stem cells. Cell stem cell 2008, 2(6):525-528.

Shi Y, Kirwan P, Smith J, Maclean G, Orkin SH, Livesey FJ: A Human Stem Cell Model of Early Alzheimer's Disease Pathology in Down Syndrome. Science translational medicine 2012.

Shiotsugu J, Katsuyama Y, Arima K, Baxter A, Koide T, Song J, Chandraratna RA, Blumberg B: Multiple points of interaction between retinoic acid and FGF signaling during embryonic axis formation. Development 2004, 131(11):2653-2667.

Si J, Mueller L, Collins SJ: CaMKII regulates retinoic acid receptor transcriptional activity and the differentiation of myeloid leukemia cells. The Journal of clinical investigation 2007, 117(5):1412-1421.

Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A: Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS biology 2008, 6(10):e253.

Silva J, Chambers I, Pollard S, Smith A: Nanog promotes transfer of pluripotency after cell fusion. Nature 2006, 441(7096):997-1001.

Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, Wray J, Yamanaka S, Chambers I, Smith A: Nanog Is the Gateway to the Pluripotent Ground State. Cell 2009, 138(4):722-737.

Simandi Z, Balint BL, Poliska S, Ruhl R, Nagy L: Activation of retinoic acid receptor signaling coordinates lineage commitment of spontaneously differentiating mouse embryonic stem cells in embryoid bodies. FEBS letters 2010, 584(14):3123-3130.

Singh H, Sen R, Baltimore D, Sharp PA: A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature 1986, 319(6049):154-158.

Singh N, George A, Sharma R, Singla SK, Palta P, Manik R, Chauhan MS, Singh D: Characterization of POU5F1 (OCT4) gene and its promoter in buffalo ESC-like cells identifies multiple transcription start sites and expression of four pseudogenes. Gene 2012, 491(2):165-172.

Singhal N, Graumann J, Wu G, Arauzo-Bravo MJ, Han DW, Greber B, Gentile L, Mann M, Scholer HR: Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. Cell 2010, 141(6):943-955.

Sive HL, Draper BW, Harland RM, Weintraub H: Identification of a retinoic acid-sensitive period during primary axis formation in Xenopus laevis. Gene Dev 1990, 4(6):932-942.

Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M, Rogers D: Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature 1988, 336(6200):688-690.

Smith AG, Hooper ML: Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. Developmental biology 1987, 121(1):1-9.

Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS: Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature 1985, 317(6034):230-234.

Sobue K, Tanaka T, Kanda K, Ashino N, Kakiuchi S: Purification and Characterization of Caldesmon77 - a Calmodulin-Binding Protein That Interacts with Actin-Filaments from Bovine Adrenal-Medulla. P Natl Acad Sci USA 1985, 82(15):5025-5029.

Solt LA, Kumar N, Nuhant P, Wang YJ, Lauer JL, Liu J, Istrate MA, Kamenecka TM, Roush WR, Vidovic D et al: Suppression of T(H)17 differentiation and autoimmunity by a synthetic ROR ligand. Nature 2011, 472(7344):491-U547.

Solter D, Skreb N, Damjanov I: Extrauterine growth of mouse egg-cylinders results in malignant teratoma. Nature 1970, 227(5257):503-504.

Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G: Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells 2009, 27(3):543-549.

Song HL, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, Anton-Culver H, Chang-Claude J, Cramer DW, DiCioccio R et al: A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nat Genet 2009, 41(9):996-U960.

Song K, Nam YJ, Luo X, Qi X, Tan W, Huang GN, Acharya A, Smith CL, Tallquist MD, Neilson EG et al: Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature 2012, 485(7400):599-604.

Souza LM, Komaromy MC, Baluda MA: Identification of a proviral genome associated with avian myeloblastic leukemia. P Natl Acad Sci USA 1980, 77(5):3004-3008.

Spemann H: Die Entwicklung seitlicher und dorso-ventraler Keimhalften bei verzogerter Kernversogung. Z Wiss Zool 1928, 132:105–134

Spits C, Mateizel I, Geens M, Mertzanidou A, Staessen C, Vandeskelde Y, Van der Elst J, Liebaers I, Sermon K: Recurrent chromosomal abnormalities in human embryonic stem cells. Nat Biotechnol 2008, 26(12):1361-1363.

Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K: Role of the murine reprogramming factors in the induction of pluripotency. Cell 2009, 136(2):364-377.

Stadtfeld M, Apostolou E, Akutsu H, Fukuda A, Follett P, Natesan S, Kono T, Shioda T, Hochedlinger K: Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature 2010, 465(7295):175-181.

Stadtfeld M, Nagaya M, Utikal J, Weir G, Hochedlinger K: Induced pluripotent stem cells generated without viral integration. Science 2008, 322(5903):945-949.

Starr TK, Allaei R, Silverstein KA, Staggs RA, Sarver AL, Bergemann TL, Gupta M, O'Sullivan MG, Matise I, Dupuy AJ et al: A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. Science 2009, 323(5922):1747-1750.

Staudt LM, Singh H, Sen R, Wirth T, Sharp PA, Baltimore D: A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature 1986, 323(6089):640-643.

Stevens LC: The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. Developmental biology 1970, 21(3):364-382.

Stewart CL, Kaspar P, Brunet LJ, Bhatt H, Gadi I, Kontgen F, Abbondanzo SJ: Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. Nature 1992, 359(6390):76-79.

Stice SL, Strelchenko NS, Keefer CL, Matthews L: Pluripotent bovine embryonic cell lines direct embryonic development following nuclear transfer. Biol Reprod 1996, 54(1):100-110.

Stojkovic M, Stojkovic P, Leary C, Hall VJ, Armstrong L, Herbert M, Nesbitt M, Lako M, Murdoch A: Derivation of a human blastocyst after heterologous nuclear transfer to donated oocytes. Reproductive biomedicine online 2005, 11(2):226-231.

Strachan T, Read AP. In: Human molecular genetics. 2nd edn. New York; 1999.

Strickland S, Mahdavi V: The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell 1978, 15(2):393-403.

Su Q, Prosser HM, Campos LS, Ortiz M, Nakamura T, Warren M, Dupuy AJ, Jenkins NA, Copeland NG, Bradley A et al: A DNA transposon-based approach to validate oncogenic mutations in the mouse. P Natl Acad Sci USA 2008, 105(50):19904-19909.

Subbarayan V, Kastner P, Mark M, Dierich A, Gorry P, Chambon P: Limited specificity and large overlap of the functions of the mouse RAR gamma 1 and RAR gamma 2 isoforms. Mechanisms of development 1997, 66(1-2):131-142.

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

Sukoyan MA, Vatolin SY, Golubitsa AN, Zhelezova AI, Semenova LA, Serov OL: Embryonic stem cells derived from morulae, inner cell mass, and blastocysts of mink: comparisons of their pluripotencies. Molecular reproduction and development 1993, 36(2):148-158.

Sundqvist J, Falconer H, Seddighzadeh M, Vodolazkaia A, Fassbender A, Kyama C, Bokor A, Stephansson O, Gemzell-Danielsson K, D'Hooghe TM: Ovarian cancer-associated polymorphisms in the BNC2 gene among women with endometriosis. Hum Reprod 2011, 26(8):2253-2257.

Suzuki T, Minehata K, Akagi K, Jenkins NA, Copeland NG: Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice. The EMBO journal 2006, 25(14):3422-3431.

Sylvester I, Scholer HR: Regulation of the Oct-4 gene by nuclear receptors. Nucleic Acids Res 1994, 22(6):901-911.

Szabo E, Rampalli S, Risueno RM, Schnerch A, Mitchell R, Fiebig-Comyn A, Levadoux-Martin M, Bhatia M: Direct conversion of human fibroblasts to multilineage blood progenitors. Nature 2010, 468(7323):521-526.

Szymczak AL, Workman CJ, Wang Y, Vignali KM, Dilioglou S, Vanin EF, Vignali DAA: Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector (vol 22, pg 589, 2004). Nat Biotechnol 2004, 22(12):1590-1590.

Tada M, Takahama Y, Abe K, Nakatsuji N, Tada T: Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr Biol 2001, 11(19):1553-1558.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 2007, 131(5):861-872.

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

Takeda J, Seino S, Bell GI: Human Oct3 Gene Family - Cdna Sequences, Alternative Splicing, Gene Organization, Chromosomal Location, and Expression at Low-Levels in Adult Tissues. Nucleic Acids Res 1992, 20(17):4613-4620.

Tanaka J, Watanabe T, Nakamura N, Sobue K: Morphological and Biochemical Analyses of Contractile Proteins (Actin, Myosin, Caldesmon and Tropomyosin) in Normal and Transformed-Cells. J Cell Sci 1993, 104:595-606.

Taranger CK, Noer A, Sorensen AL, Hakelien AM, Boquest AC, Collas P: Induction of dedifferentiation, genomewide transcriptional programming, and epigenetic reprogramming by extracts of carcinoma and embryonic stem cells. Molecular biology of the cell 2005, 16(12):5719-5735.

Teranishi T, Tanaka M, Kimoto S, Ono Y, Miyakoshi K, Kono T, Yoshimura Y: Rapid replacement of somatic linker histones with the oocyte-specific linker histone H1foo in nuclear transfer. Developmental biology 2004, 266(1):76-86.

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD: New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 2007, 448(7150):196-199.

Thaller C, Eichele G: Identification and spatial distribution of retinoids in the developing chick limb bud. Nature 1987, 327(6123):625-628.

Thomas KR, Capecchi MR: Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell 1987, 51(3):503-512.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM: Embryonic stem cell lines derived from human blastocysts. Science 1998, 282(5391):1145-1147.

Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Becker RA, Hearn JP: Isolation of a primate embryonic stem cell line. P Natl Acad Sci USA 1995, 92(17):7844-7848.

Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Hearn JP: Pluripotent cell lines derived from common marmoset (Callithrix jacchus) blastocysts. Biol Reprod 1996, 55(2):254-259.

Thorel F, Nepote V, Avril I, Kohno K, Desgraz R, Chera S, Herrera PL: Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. Nature 2010, 464(7292):1149-1154.

Tiemann U, Sgodda M, Warlich E, Ballmaier M, Scholer HR, Schambach A, Cantz T: Optimal Reprogramming Factor Stoichiometry Increases Colony Numbers and Affects Molecular Characteristics of Murine Induced Pluripotent Stem Cells. Cytom Part A 2011, 79A(6):426-435.

Tomioka M, Nishimoto M, Miyagi S, Katayanagi T, Fukui N, Niwa H, Muramatsu M, Okuda A: Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Res 2002, 30(14):3202-3213.

Tontonoz P, Hu E, Spiegelman BM: Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor. Cell 1994, 79(7):1147-1156.

Tsai SY, Clavel C, Kim S, Ang YS, Grisanti L, Lee DF, Kelley K, Rendl M: Oct4 and klf4 reprogram dermal papilla cells into induced pluripotent stem cells. Stem Cells 2010, 28(2):221-228.

Tseng H, Green H: Basonuclin: a keratinocyte protein with multiple paired zinc fingers. P Natl Acad Sci USA 1992, 89(21):10311-10315.

Tsonis PA, Del Rio-Tsonis K: Lens and retina regeneration: transdifferentiation, stem cells and clinical applications. Experimental eye research 2004, 78(2):161-172.

Tsubota S, Ashburner M, Schedl P: P-Element-Induced Control Mutations at the R Gene of Drosophila-Melanogaster. Mol Cell Biol 1985, 5(10):2567-2574.

Tsubota S, Schedl P: Hybrid Dysgenesis-Induced Revertants of Insertions at the 5' End of the Rudimentary Gene in Drosophila-Melanogaster - Transposon-Induced Control Mutations. Genetics 1986, 114(1):165-182.

Urbach A, Bar-Nur O, Daley GQ, Benvenisty N: Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells. Cell stem cell 2010, 6(5):407-411.

Utikal J, Polo JM, Stadtfeld M, Maherali N, Kulalert W, Walsh RM, Khalil A, Rheinwald JG, Hochedlinger K: Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature 2009, 460(7259):1145-1148.

Vallier L, Alexander M, Pedersen RA: Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J Cell Sci 2005, 118(Pt 19):4495-4509.

van den Berg DL, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA: An Oct4-centered protein interaction network in embryonic stem cells. Cell stem cell 2010, 6(4):369-381.

van der Wees J, Schilthuis JG, Koster CH, Diesveld-Schipper H, Folkers GE, van der Saag PT, Dawson MI, Shudo K, van der Burg B, Durston AJ: Inhibition of retinoic acid receptor-mediated signalling alters positional identity in the developing hindbrain. Development 1998, 125(3):545-556.

van Eijk MJT, van Rooijen MA, Modina S, Scesi L, Folkers G, van Tol HTA, Bevers MM, Fisher SR, Lewin HA, Rakacolli D et al: Molecular cloning, genetic mapping, and developmental expression of bovine POU5F1. Biol Reprod 1999, 60(5):1093-1103.

VandenDriessche T, Ivics Z, Izsvak Z, Chuah MK: Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. Blood 2009, 114(8):1461-1468.

Vander Heiden MG, Cantley LC, Thompson CB: Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science 2009, 324(5930):1029-1033.

Vanhoutteghem A, Bouche C, Maciejewski-Duval A, Herve F, Djian P: Basonuclins and disco: Orthologous zinc finger proteins essential for development in vertebrates and arthropods. Biochimie 2011, 93(2):127-133.

Vanhoutteghem A, Djian P: Basonuclin 2: an extremely conserved homolog of the zinc finger protein basonuclin. P Natl Acad Sci USA 2004, 101(10):3468-3473.

Vanhoutteghem A, Djian P: Basonuclins 1 and 2, whose genes share a common origin, are proteins with widely different properties and functions. P Natl Acad Sci USA 2006, 103(33):12423-12428.

Vanhoutteghem A, Djian P: The human basonuclin 2 gene has the potential to generate nearly 90,000 mRNA isoforms encoding over 2000 different proteins. Genomics 2007, 89(1):44-58.

Vanhoutteghem A, Maciejewski-Duval A, Bouche C, Delhomme B, Herve F, Daubigney F, Soubigou G, Araki M, Araki K, Yamamura K et al: Basonuclin 2 has a function in the multiplication of embryonic craniofacial mesenchymal cells and is orthologous to disco proteins. P Natl Acad Sci USA 2009, 106(34):14432-14437.

Varum S, Rodrigues AS, Moura MB, Momcilovic O, Easley CA, Ramalho-Santos J, Van Houten B, Schatten G: Energy Metabolism in Human Pluripotent Stem Cells and Their Differentiated Counterparts. PloS one 2011, 6(6).

Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, Wernig M: Direct conversion of fibroblasts to functional neurons by defined factors. Nature 2010, 463(7284):1035-1041.

Vivian JL, Chen Y, Yee D, Schneider E, Magnuson T: An allelic series of mutations in Smad2 and Smad4 identified in a genotype-based screen of N-ethyl-N- nitrosourea-mutagenized mouse embryonic stem cells. P Natl Acad Sci USA 2002, 99(24):15542-15547.

von Baer KE: De ovi mammalium et hominis genesi: Sumptibus L. Vossii; 1827.

von Baer KE: Über Entwickelungsgeschichte der Thier. Beobachtung und Reflexion: Königsberg, Bei den Gebrüdern Bornträger; 1828.

Wang G, Zhang H, Zhao Y, Li J, Cai J, Wang P, Meng S, Feng J, Miao C, Ding M et al: Noggin and bFGF cooperate to maintain the pluripotency of human embryonic stem cells in the absence of feeder layers. Biochem Bioph Res Co 2005, 330(3):934-942.

Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW, Orkin SH: A protein interaction network for pluripotency of embryonic stem cells. Nature 2006, 444(7117):364-368.

Wang T, Chen K, Zeng X, Yang J, Wu Y, Shi X, Qin B, Zeng L, Esteban MA, Pan G et al: The histone demethylases jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. Cell stem cell 2011, 9(6):575-587.

Wang W, Bradley A, Huang Y: A piggyBac transposon-based genome-wide library of insertionally mutated Blm-deficient murine ES cells. Genome research 2009, 19(4):667-673.

Wang W, Hale C, Goulding D, Haslam SM, Tissot B, Lindsay C, Michell S, Titball R, Yu J, Toribio AL et al: Mannosidase 2, alpha 1 deficiency is associated with ricin resistance in embryonic stem (ES) cells. PloS one 2011, 6(8):e22993.

Wang W, Lin C, Lu D, Ning Z, Cox T, Melvin D, Wang X, Bradley A, Liu P: Chromosomal transposition of PiggyBac in mouse embryonic stem cells. P Natl Acad Sci USA 2008, 105(27):9290-9295.

Wang W, Yang J, Liu H, Lu D, Chen X, Zenonos Z, Campos LS, Rad R, Guo G, Zhang S et al: Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. P Natl Acad Sci USA 2011, 108(45):18283-18288.

Wang YJ, Kumar N, Nuhant P, Cameron MD, Istrate MA, Roush WR, Griffin PR, Burris TP: Identification of SR1078, a Synthetic Agonist for the Orphan Nuclear Receptors ROR alpha and ROR gamma. Acs Chem Biol 2010, 5(11):1029-1034.

Warkany J, Schraffenberger E: Congenital malformations induced in rats by maternal vitamin A deficiency; defects of the eye. Archives of ophthalmology 1946, 35:150-169.

Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, Ebina W, Mandal PK, Smith ZD, Meissner A et al: Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell stem cell 2010, 7(5):618-630.

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P et al: Initial sequencing and comparative analysis of the mouse genome. Nature 2002, 420(6915):520-562.

Watson JD, Crick FH: Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953, 171(4356):737-738.

Wei F, Scholer HR, Atchison ML: Sumoylation of Oct4 enhances its stability, DNA binding, and transactivation. The Journal of biological chemistry 2007, 282(29):21551-21560.

Weintraub H, Tapscott SJ, Davis RL, Thayer MJ, Adam MA, Lassar AB, Miller AD: Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. P Natl Acad Sci USA 1989, 86(14):5434-5438.

Weismann A, Parker WN, RÃ¶nnfeldt H: The germ-plasm: a theory of heredity: Scribner's; 1893.

Welstead GG, Creyghton MP, Bilodeau S, Cheng AW, Markoulaki S, Young RA, Jaenisch R: X-linked H3K27me3 demethylase Utx is required for embryonic development in a sex-specific manner. P Natl Acad Sci USA 2012, 109(32):13004-13009.

White JC, Shankar VN, Highland M, Epstein ML, DeLuca HF, Clagett-Dame M: Defects in embryonic hindbrain development and fetal resorption resulting from vitamin A deficiency in the rat are prevented by feeding pharmacological levels of all-trans-retinoic acid. P Natl Acad Sci USA 1998, 95(23):13459-13464.

Whyte WA, Bilodeau S, Orlando DA, Hoke HA, Frampton GM, Foster CT, Cowley SM, Young RA: Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature 2012, 482(7384):221-225.

Widschwendter M, Berger J, Daxenbichler G, Muller-Holzner E, Widschwendter A, Mayr A, Marth C, Zeimet AG: Loss of retinoic acid receptor beta expression in breast cancer and morphologically normal adjacent tissue but not in the normal breast tissue distant from the cancer. Cancer research 1997, 57(19):4158-4161.

Willadsen SM: Nuclear transplantation in sheep embryos. Nature 1986, 320(6057):63-65.

Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA, Gough NM: Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 1988, 336(6200):684-687.

Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH: Viable offspring derived from fetal and adult mammalian cells. Nature 1997, 385(6619):810-813.

Wilson JG, Roth CB, Warkany J: An analysis of the syndrome of malformations induced by maternal vitamin A deficiency. Effects of restoration of vitamin A at various times during gestation. The American journal of anatomy 1953, 92(2):189-217.

Wilson JG, Warkany J: Malformations in the genito-urinary tract induced by maternal vitamin A deficiency in the rat. The American journal of anatomy 1948, 83(3):357-407.

Wilson MH, Coates CJ, George AL: PiggyBac transposon-mediated gene transfer in human cells. Mol Ther 2007, 15(1):139-145.

Wilson PA, Hemmati-Brivanlou A: Induction of epidermis and inhibition of neural fate by Bmp-4. Nature 1995, 376(6538):331-333.

Witte ON, Rosenberg N, Paskind M, Shields A, Baltimore D: Identification of an Abelson murine leukemia virus-encoded protein present in transformed fibroblast and lymphoid cells. P Natl Acad Sci USA 1978, 75(5):2488-2492.

Wolff CF: Theoria generationis: Typis et sumtu Io . Christ. Hendel; 1774.

Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Hamalainen R, Cowling R, Wang W, Liu P, Gertsenstein M et al: piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature 2009, 458(7239):766-770.

Wu H, Zhang Y: Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Gene Dev 2011, 25(23):2436-2452.

Wu SCY, Meir YJJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM: piggyBac is a flexible and highly active transposon as compared to Sleeping Beauty, Tol2 and Mos1 in mammalian cells. P Natl Acad Sci USA 2006, 103(41):15008-15013.

Xie H, Ye M, Feng R, Graf T: Stepwise reprogramming of B cells into macrophages. Cell 2004, 117(5):663-676.

Xiong G, Wang C, Evers BM, Zhou BP, Xu R: RORalpha Suppresses Breast Tumor Invasion by Inducing SEMA3F Expression. Cancer research 2012, 72(7):1728-1739.

Xu C, Rosler E, Jiang J, Lebkowski JS, Gold JD, O'Sullivan C, Delavan-Boorsma K, Mok M, Bronstein A, Carpenter MK: Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium. Stem Cells 2005, 23(3):315-323.

Xu H, Wang W, Li C, Yu H, Yang A, Wang B, Jin Y: WWP2 promotes degradation of transcription factor OCT4 in human embryonic stem cells. Cell research 2009, 19(5):561-573.

Xu RH, Chen X, Li DS, Li R, Addicks GC, Glennon C, Zwaka TP, Thomson JA: BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nat Biotechnol 2002, 20(12):1261-1264.

Xu XC, Liu X, Tahara E, Lippman SM, Lotan R: Expression and up-regulation of retinoic acid receptor-beta is associated with retinoid sensitivity and colony formation in esophageal cancer cell lines. Cancer research 1999, 59(10):2477-2483.

Xu XC, Sneige N, Liu X, Nandagiri R, Lee JJ, Lukmanji F, Hortobagyi G, Lippman SM, Dhingra K, Lotan R: Progressive decrease in nuclear retinoic acid receptor beta messenger RNA level during breast carcinogenesis. Cancer research 1997, 57(22):4992-4996.

Yamaguchi S, Kimura H, Tada M, Nakatsuji N, Tada T: Nanog expression in mouse germ cell development. Gene expression patterns : GEP 2005, 5(5):639-646.

Yamashita A, Takada T, Narita J, Yamamoto G, Torii R: Osteoblastic differentiation of monkey embryonic stem cells in vitro. Cloning and stem cells 2005, 7(4):232-237.

Yang H, Shi L, Wang BA, Liang D, Zhong C, Liu W, Nie Y, Liu J, Zhao J, Gao X et al: Generation of genetically modified mice by oocyte injection of androgenetic haploid embryonic stem cells. Cell 2012, 149(3):605-617.

Yant SR, Wu XL, Huang Y, Garrison B, Burgess SM, Kay MA: High-resolution genome-wide mapping of transposon integration in mammals. Mol Cell Biol 2005, 25(6):2085-2094.

Yao JY, Zhang L, Zhang X, He ZY, Ma Y, Hui LJ, Wang X, Hu YP: H3K27 Trimethylation Is an Early Epigenetic Event of p16(INK4a) Silencing for Regaining Tumorigenesis in Fusion Reprogrammed Hepatoma Cells. J Biol Chem 2010, 285(24):18828-18837.

Yeo S, Jeong S, Kim J, Han JS, Han YM, Kang YK: Characterization of DNA methylation change in stem cell marker genes during differentiation of human embryonic stem cells. Biochem Bioph Res Co 2007, 359(3):536-542.

Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M, Hubner K, Scholer HR: Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development 1996, 122(3):881-894.

Yeom YI, Ha HS, Balling R, Scholer HR, Artzt K: Structure, expression and chromosomal location of the Oct-4 gene. Mechanisms of development 1991, 35(3):171-179.

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

Ying QL, Nichols J, Evans EP, Smith AG: Changing potency by spontaneous fusion. Nature 2002, 416(6880):545-548.

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A: The ground state of embryonic stem cell self-renewal. Nature 2008, 453(7194):519-523.

Yoshida K, Chambers I, Nichols J, Smith A, Saito M, Yasukawa K, Shoyab M, Taga T, Kishimoto T: Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. Mechanisms of development 1994, 45(2):163-171.

Yoshida Y, Takahashi K, Okita K, Ichisaka T, Yamanaka S: Hypoxia enhances the generation of induced pluripotent stem cells. Cell stem cell 2009, 5(3):237-241.

You JS, Kelly TK, De Carvalho DD, Taberlay PC, Liang GN, Jones PA: OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. P Natl Acad Sci USA 2011, 108(35):14497-14502.

You L, Chang D, Du HZ, Zhao YP: Genome-wide screen identifies PVT1 as a regulator of Gemcitabine sensitivity in human pancreatic cancer cells. Biochem Bioph Res Co 2011, 407(1):1-6.

Yu J, Hu K, Smuga-Otto K, Tian S, Stewart R, Slukvin, II, Thomson JA: Human induced pluripotent stem cells free of vector and transgene sequences. Science 2009, 324(5928):797-801.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R et al: Induced pluripotent stem cell lines derived from human somatic cells. Science 2007, 318(5858):1917-1920.

Yu VC, Delsert C, Andersen B, Holloway JM, Devary OV, Naar AM, Kim SY, Boutin JM, Glass CK, Rosenfeld MG: RXR beta: a coregulator that enhances binding of retinoic acid, thyroid hormone, and vitamin D receptors to their cognate response elements. Cell 1991, 67(6):1251-1266.

Yusa K, Horie K, Kondoh G, Kouno M, Maeda Y, Kinoshita T, Takeda J: Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. Nature 2004, 429(6994):896-899.

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

Yusa K, Rashid ST, Strick-Marchand H, Varela I, Liu PQ, Paschon DE, Miranda E, Ordonez A, Hannan NR, Rouhani FJ et al: Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. Nature 2011, 478(7369):391-394.

Zambrowicz BP, Friedrich GA, Buxton EC, Lilleberg SL, Person C, Sands AT: Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature 1998, 392(6676):608-611.

Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P: Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. P Natl Acad Sci USA 1997, 94(8):3789-3794.

Zelent A, Krust A, Petkovich M, Kastner P, Chambon P: Cloning of murine alpha and beta retinoic acid receptors and a novel receptor gamma predominantly expressed in skin. Nature 1989, 339(6227):714-717.

Zelent A, Mendelsohn C, Kastner P, Krust A, Garnier JM, Ruffenach F, Leroy P, Chambon P: Differentially expressed isoforms of the mouse retinoic acid receptor beta generated by usage of two promoters and alternative splicing. The EMBO journal 1991, 10(1):71-81.

Zhang J, Lian Q, Zhu G, Zhou F, Sui L, Tan C, Mutalif RA, Navasankari R, Zhang Y, Tse HF et al: A human iPSC model of Hutchinson Gilford Progeria reveals vascular smooth muscle and mesenchymal stem cell defects. Cell stem cell 2011, 8(1):31-45.

Zhang P, Andrianakos R, Yang Y, Liu C, Lu W: Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. The Journal of biological chemistry 2010, 285(12):9180-9189.

Zhang XK, Hoffmann B, Tran PB, Graupner G, Pfahl M: Retinoid X receptor is an auxiliary protein for thyroid hormone and retinoic acid receptors. Nature 1992, 355(6359):441-446.

Zhang Z, Liao B, Xu M, Jin Y: Post-translational modification of POU domain transcription factor Oct-4 by SUMO-1. Faseb J 2007, 21(12):3042-3051.

Zhao T, Zhang ZN, Rong Z, Xu Y: Immunogenicity of induced pluripotent stem cells. Nature 2011, 474(7350):212-215.

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

Zheng PP, Severijnen LA, van der Weiden M, Willemsen R, Kros JM: A crucial role of caldesmon in vascular development in vivo. Cardiovasc Res 2009, 81(2):362-369.

Zheng PP, Severijnen LA, Willemsen R, Kros JM: Caldesmon is essential for cardiac morphogenesis and function: In vivo study using a zebrafish model. Biochem Bioph Res Co 2009, 378(1):37-40.

Zhou H, Wu S, Joo JY, Zhu S, Han DW, Lin T, Trauger S, Bien G, Yao S, Zhu Y et al: Generation of induced pluripotent stem cells using recombinant proteins. Cell stem cell 2009, 4(5):381-384.

Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA: In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature 2008, 455(7213):627-632.