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.

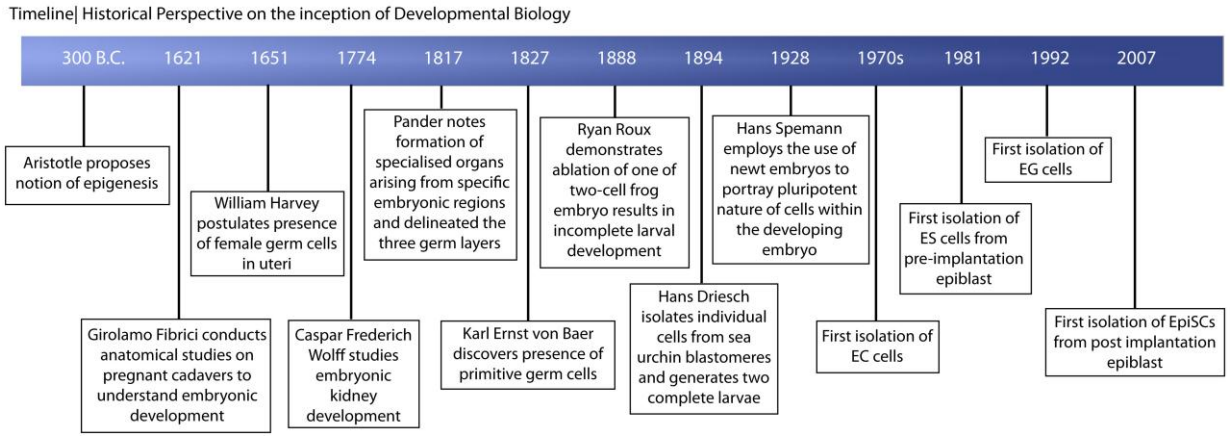

Figure 1.1 Timeline displaying milestones that led to the inception of developmental biology

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*.

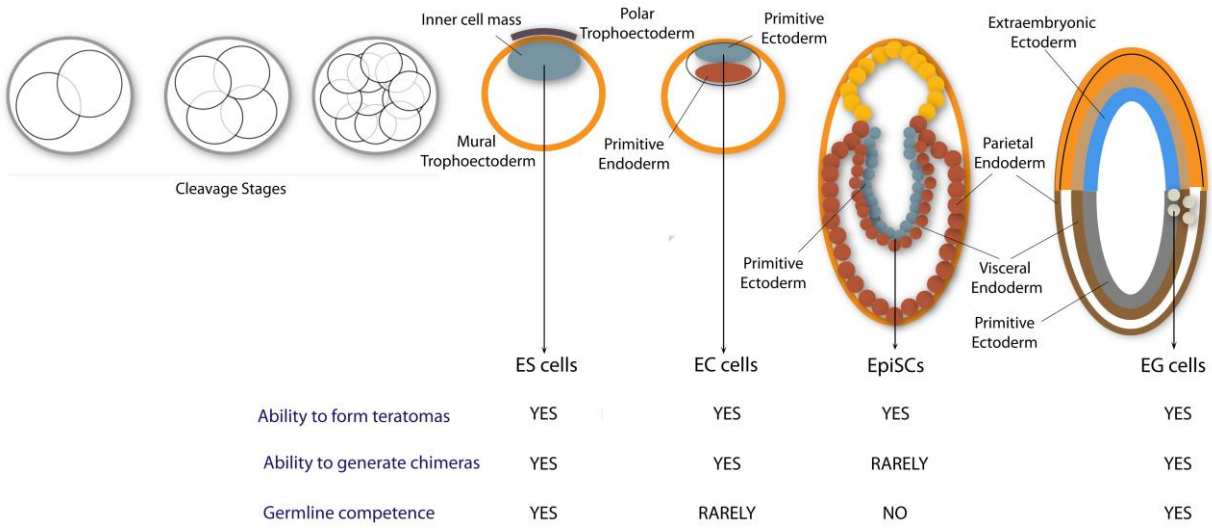

Figure 1.2 Comparison of pluripotent cells established from various stages of embryonic development. Once fertilised, the zygote undergoes several rounds of cell division and forms an inner cell mass, as indicated in grey. Cultivation of the inner cell mass results in the establishment of ESCs. The inner 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 18th 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.

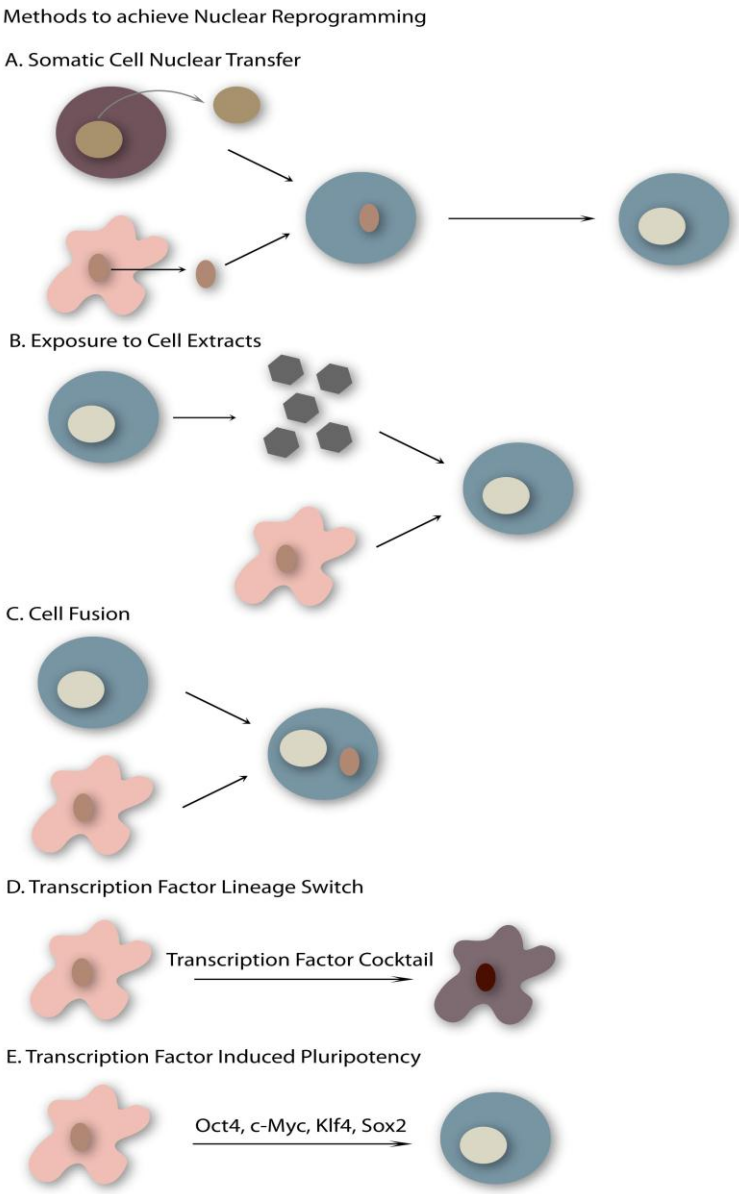

Figure 1.3 Methods to achieve nuclear reprogramming Nuclear reprogramming is defined as the transition between distinct cell types introduced by switches in gene patterns. This can be accomplished through various methods. (A) Somatic cell nuclear transfer entails the transfer of a somatic cell nucleus 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.

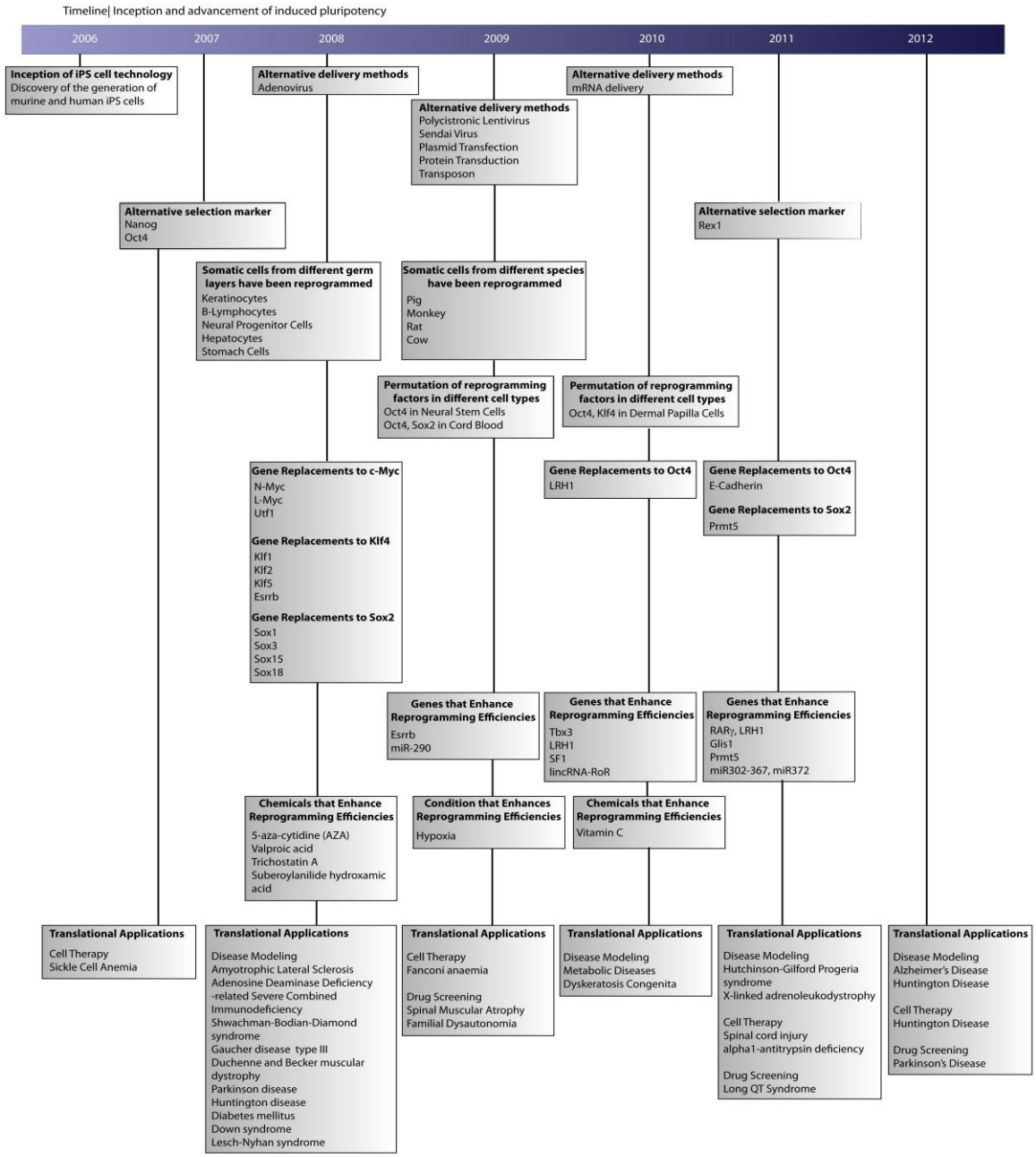

Figure 1.4 Timeline highlighting the milestones achieved since the inception of transcription factor induced pluripotency Transcription factor induced pluripotency was conceived in 2006 through the introduction of four transcription factors, Oct4, c-Myc, Klf4 and Sox2 into mouse fibroblasts. Since its 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).

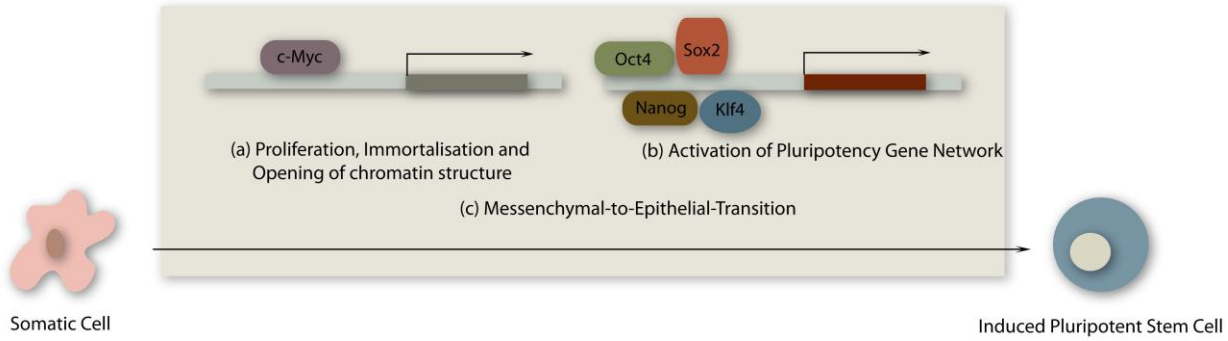

Figure 1.5 Schematic diagram portraying the current knowledge on the molecular mechanisms behind induced pluripotency The conversion of somatic cells to iPSCs involves a radical change in the epigenomic landscape and cellular components. To accomplish this, the cell has to overcome cellular processes 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).

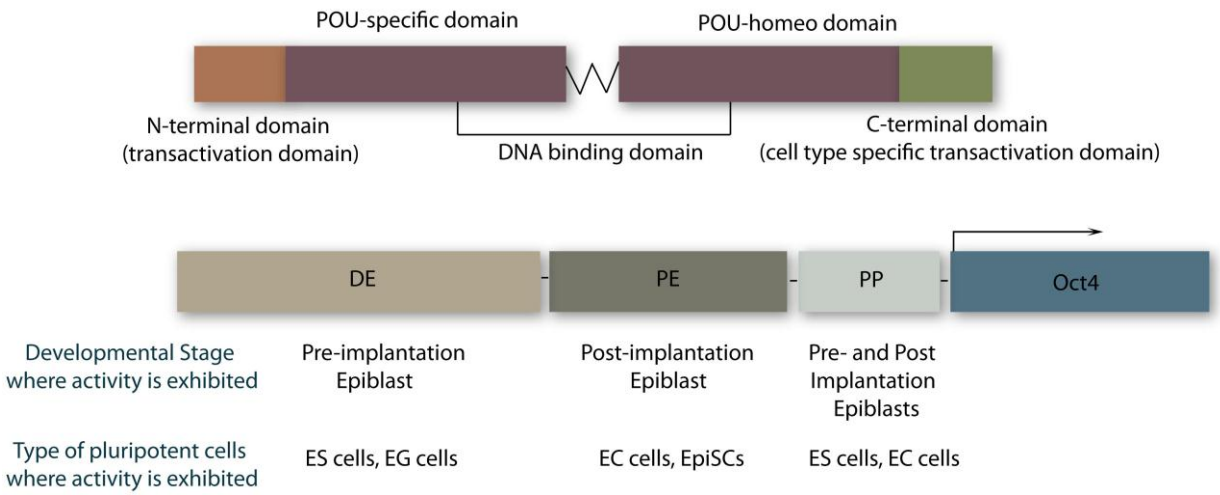

Figure 1.6 Schematic diagram of the structure of Oct4 and its regulatory elements (Top) Oct4 is comprised of N- and C-terminal domains flanking both ends. These regions display high proportions of proline and glycine, suggesting their function as transactivators. Oct4 also bears bipartite POU domains 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).

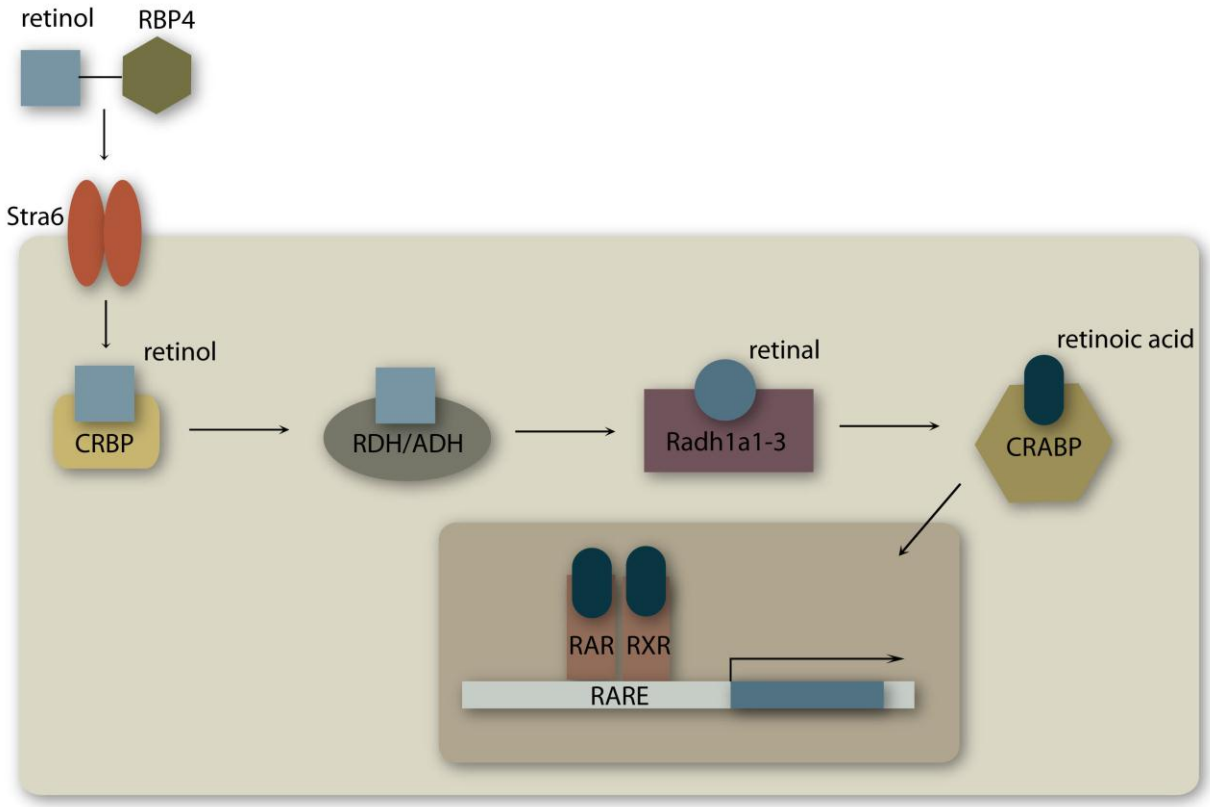

Figure 1.7 Retinoic acid signalling pathway While in circulation, retinol binds to retinol binding protein (RBP4) and is transported across the cell membrane via Stra6. While in the cytoplasm, retinol binds to cellular retinol binding protein (CRBP) until oxidation is facilitated through the action of 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)

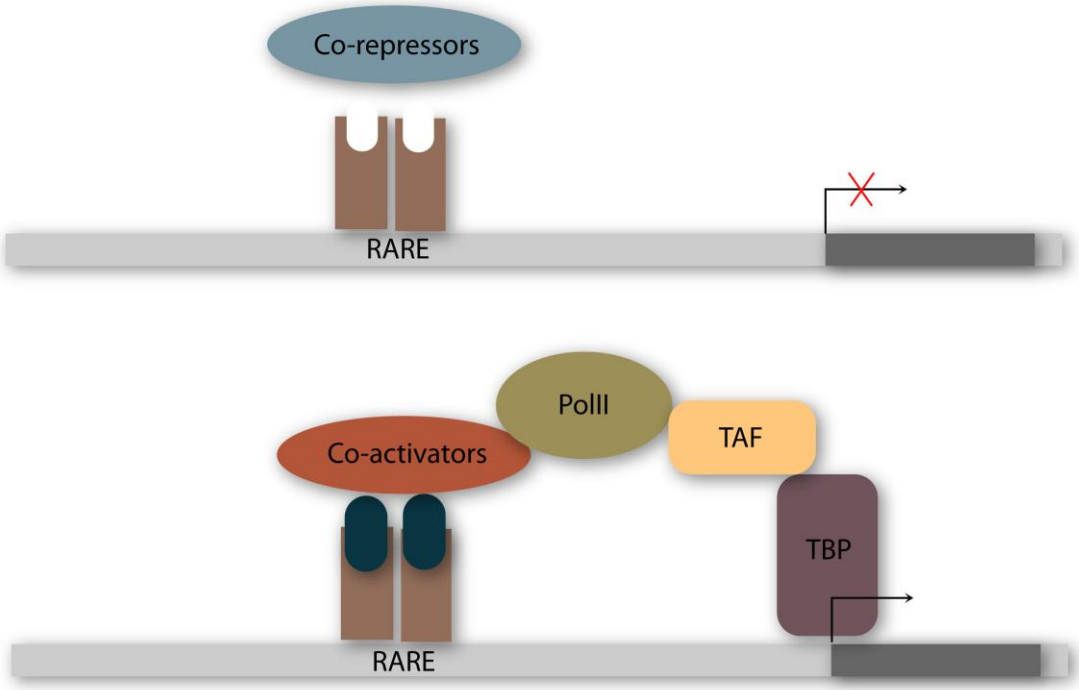

Figure 1.8 Ligand dependent transcriptional control (Top) In the absence of its ligand, Retinoic Acid Receptors (RARs) are able to recognise and bind to Retinoic Acid Response Elements (RAREs). Through its association to co-repressor complexes, transcription at the target gene is inhibited. (Bottom) 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.

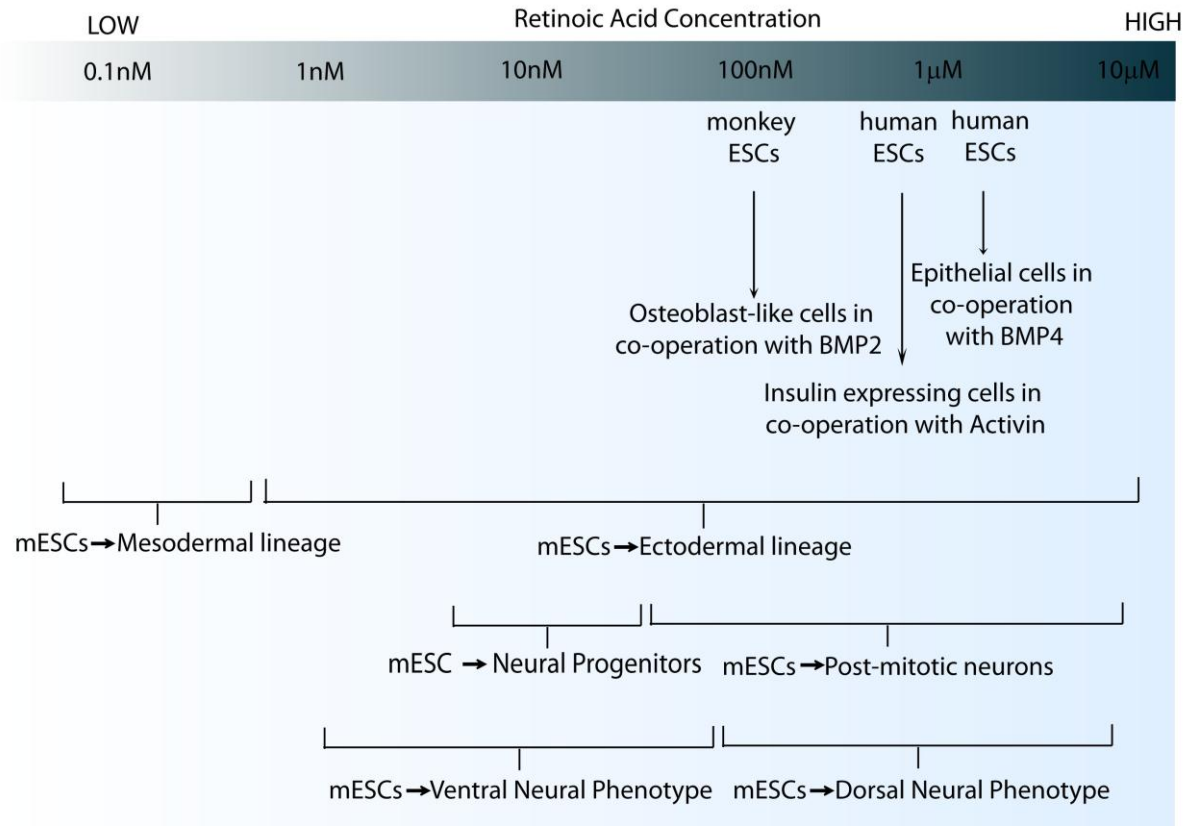

Figure 1.9 Dosage dependence in RA-stimulated differentiation Distinct concentrations of RA can lead to differentiation into different cell types. Overall, introduction of high concentrations of RA to mouse ESCs (mESCs) gives rise to neuronal lineages whereas low concentrations ($<10^{-10}$M RA) leads 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.

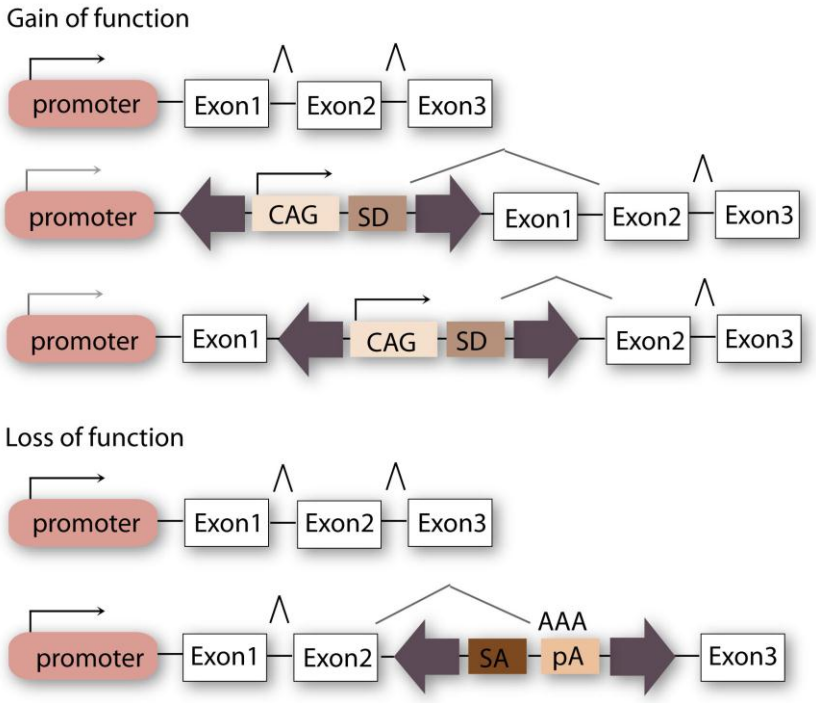

Figure 1.10 Principle behind the execution of transposon mediated genetic screens (Top) Gain of function genetic screens can be conducted through the use of transposons encompassing a strong constitutive promoter, such as CMV early enhancer/chicken beta actin (CAG), preceding a splice donor (SD) site. 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.