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

3.1 Introduction

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

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

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

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

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

3.1.2 Identification of Oct4 substitutions that can elicit reprogramming events

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

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

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

3.1.3 Transposon-assisted mutagenesis

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

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

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

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

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

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

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

3.1.4 Chapter Aim

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

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

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

3.2 Experimental Design

3.2.1 Factors taken into consideration for a saturated genetic study

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

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

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

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

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

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

3.2.2 Constructs generated for the execution of the genetic screen

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

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

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

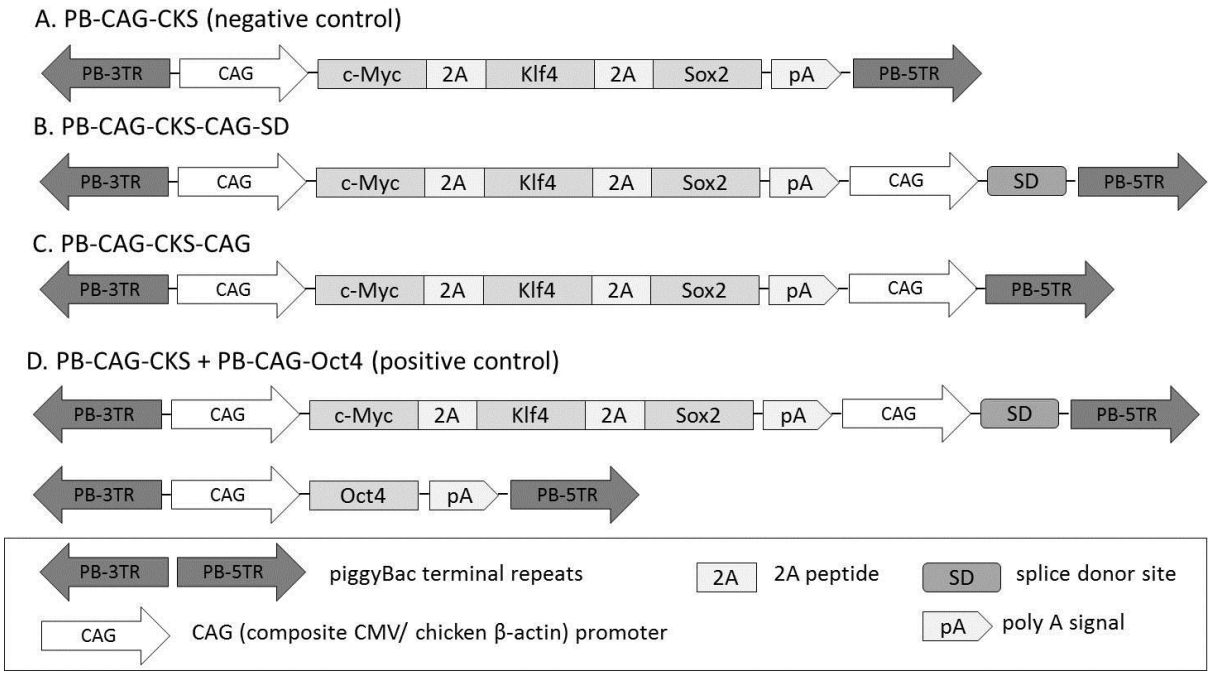

Figure 3.1 Schematic diagram of piggyBac transposon constructs (A) PB-CAG-CKS cassette. The dark arrows flanking both ends depict piggyBac terminal repeats. Between the two boundaries lies a constitutively active CAG promoter driving the expression of c-Myc, Klf4 and Sox2. (B) PB-CAG-CKS-CAG-SD cassette.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

3.3 Results

3.3.1 Generation of mutagenised MEF library

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

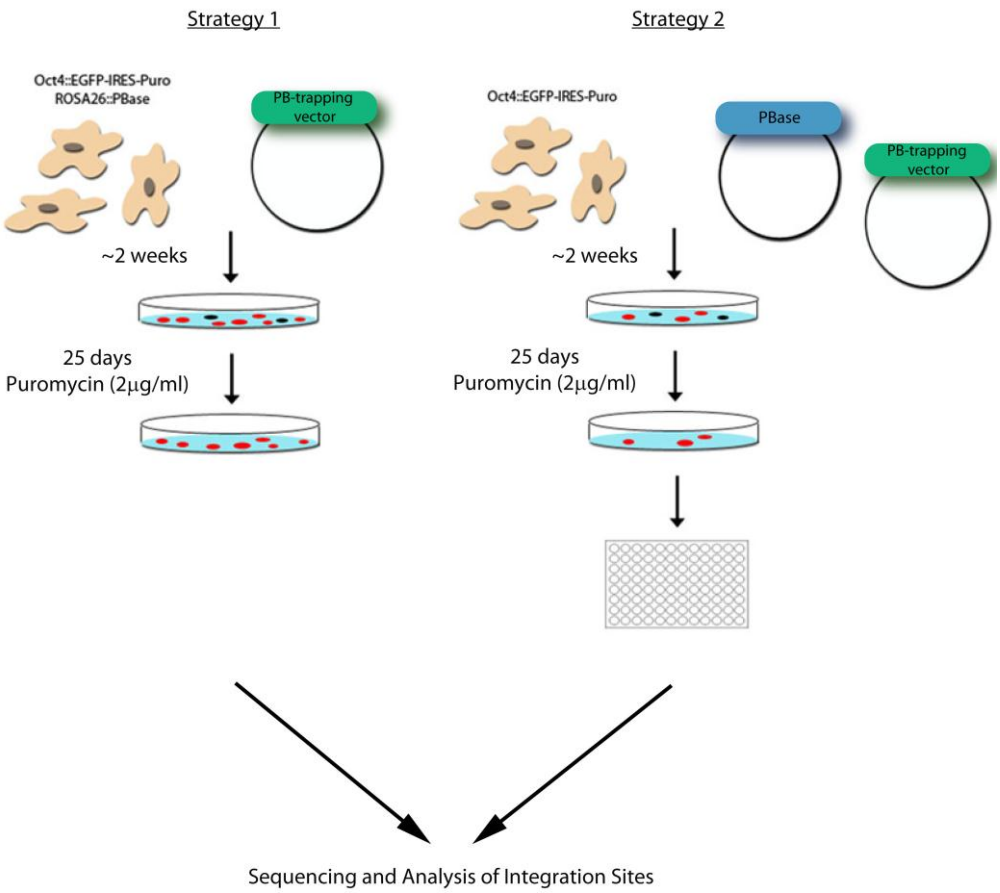

Figure 3.2 Experimental Strategy of Transposon-assisted Genetic Screen On the left: Strategy 1 involving PB-CAG-CKS-CAG-SD. MEFs that harbour two modifications were used as starting material: (i) EGFP expression and Puromycin resistance driven by endogenous Oct4 locus, and (ii) PBase expression controlled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

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

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

3.3.1.2 The PB-CAG-CKS-CAG screen

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

3.3.1.3 Reprogramming MEFs to iPSCs during the genetic screens

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

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

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

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

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

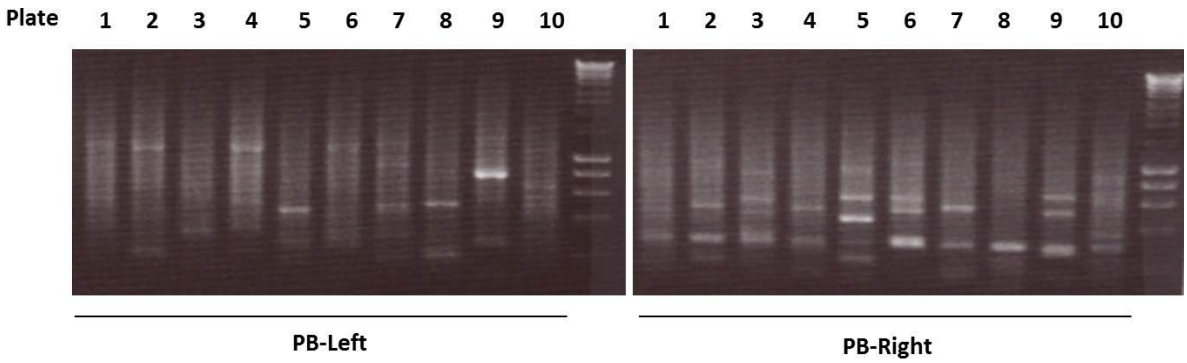

Figure 3.3 Splinkerette PCR of integration sites Gel electrophoresis image of splinkerette PCR products obtained from each of ten 10cm plates with induced pluripotent stem cell colonies as instigated by the presence of PB-CAG-CKS-CAG-SD. Each band represents the presence of an integration site. PB-Left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

3.3.2 Analysis of the PB Integration sites

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

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

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

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

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

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

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

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

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

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

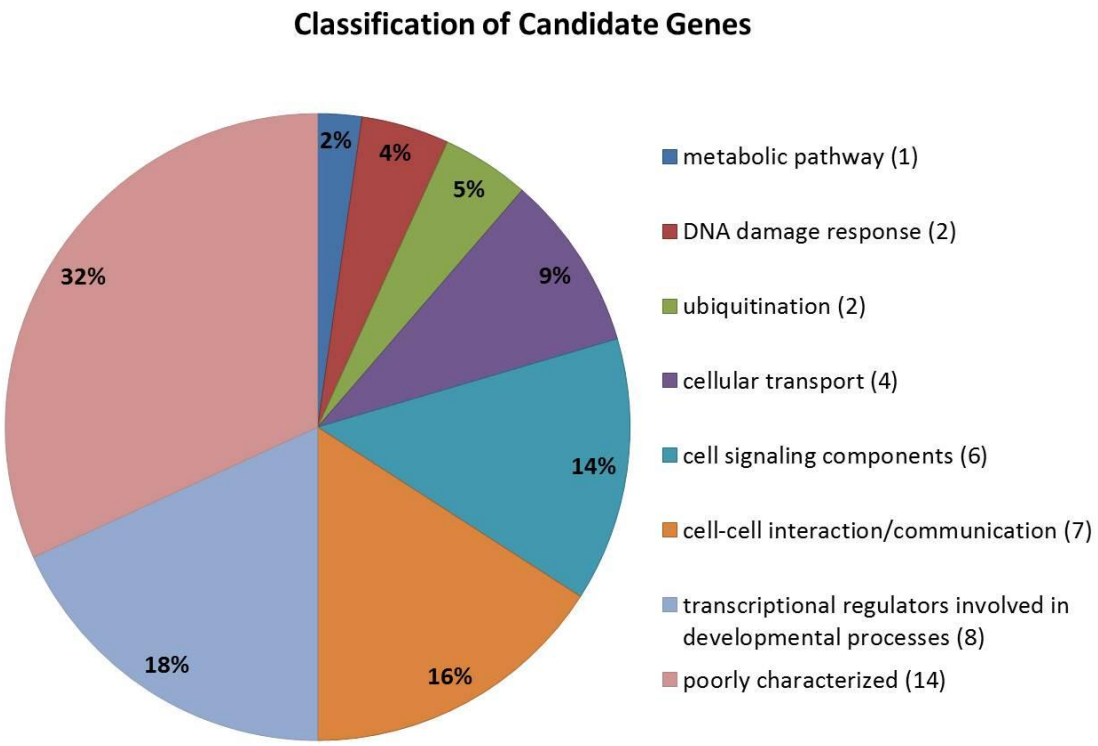

Figure 3.5 Overview of functional classification of candidate genes identified from the PB-CAG-CKS-CAG-SD driven genetic screen Pie chart describing the diverse functions of candidate genes. The colours depict distinct categories, as labelled in the legend. The actual number of genes associated to each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

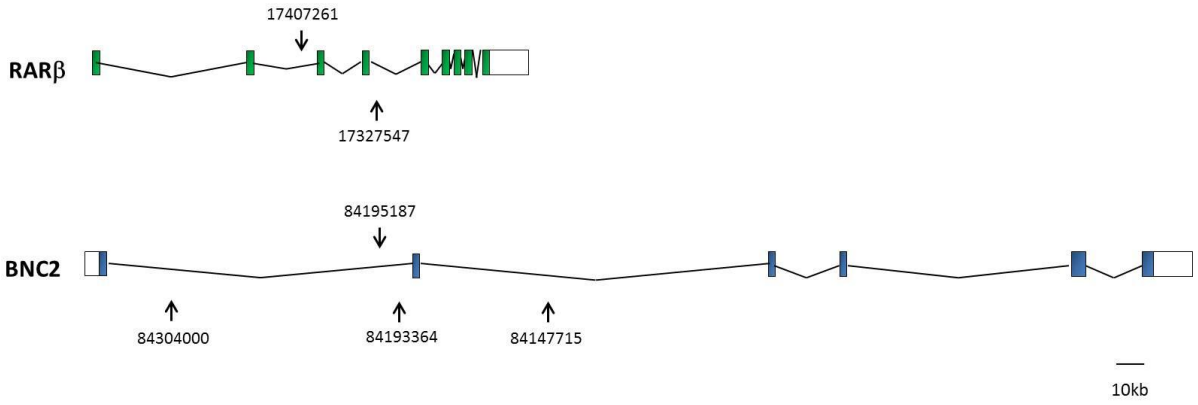

Figure 3.6 Schematic diagram showing transposon integration sites within two candidate genes Integration sites obtained from the PB-CAG-CKS-CAG-SD-assisted genetic screen. Coloured boxes indicate the exons corresponding to the two genomic regions. Arrows depict the location of the integration sites,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

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

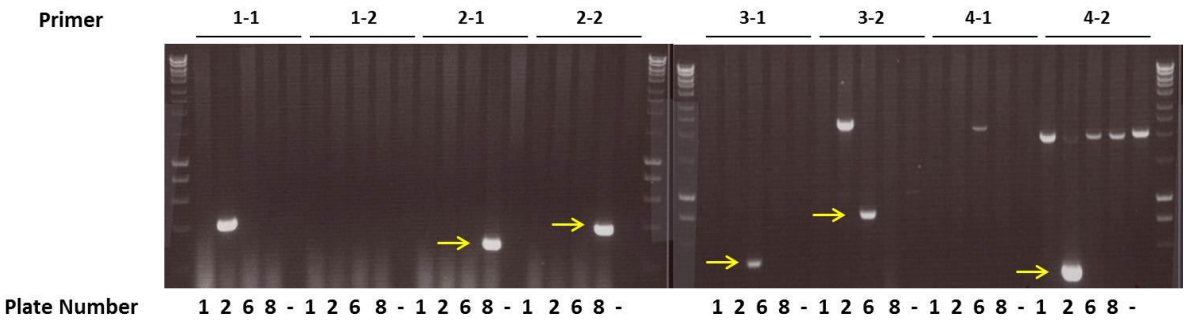

Figure 3.7 Validation of BNC2 insertion sites A pair of primers were designed for each of the 4 insertion sites. Genomic DNA was obtained from each of ten 10cm plates with iPSC colonies. DNA from plates 1, 2, 6, 8 and 10 were used. Insertions in the genomic locus corresponding to BNC2 were observed in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

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

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

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

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

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

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

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

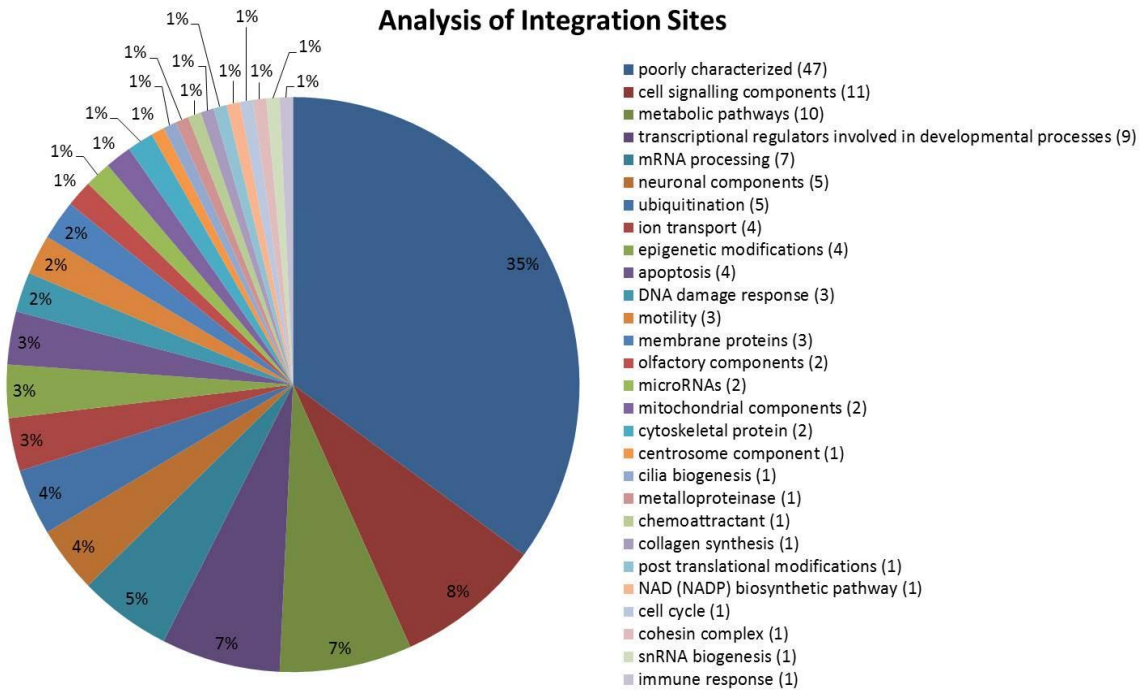

Figure 3.9 Overview of functional classification of candidate genes identified from the PB-CAG-CKS-CAG driven genetic screen Pie chart describing the diverse functions of candidate genes. The colours depict distinct categories, as labelled in the legend. The actual number of genes associated to each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

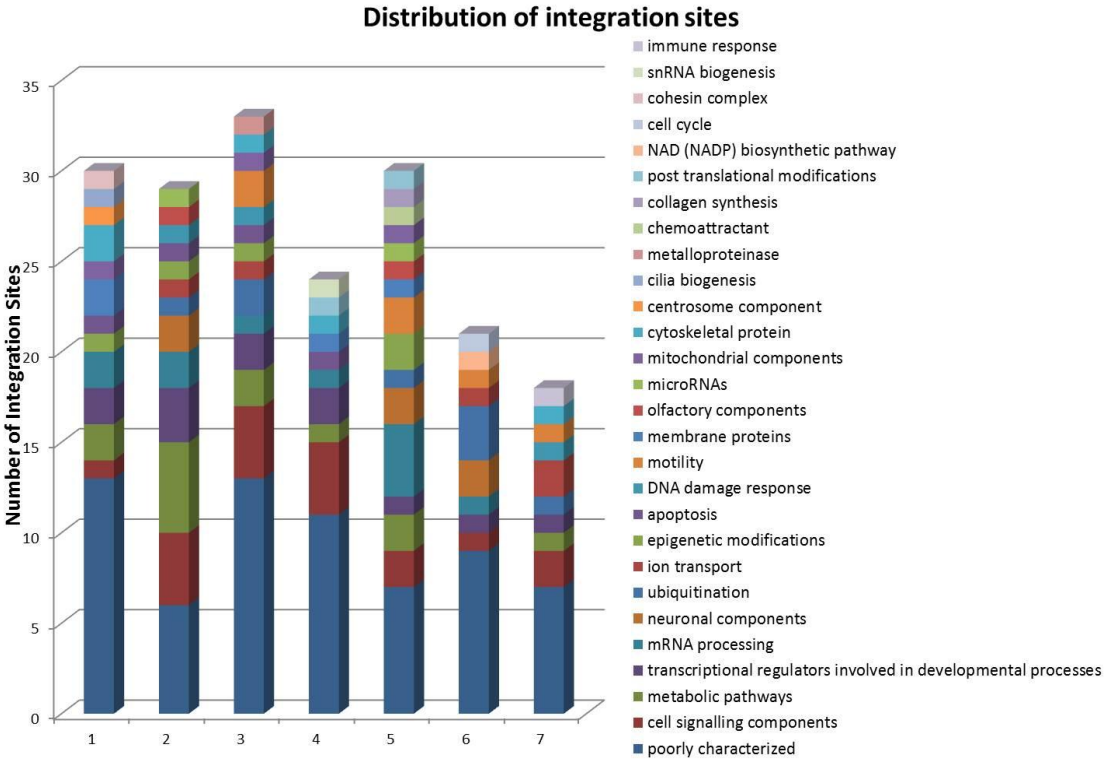

Figure 3.10 Graphical representation of candidate genes identified from the PB-CAG-CKS-CAG driven genetic screen Bar chart displaying the range of mutations in functional groups in each of the seven iPSC colonies. Array of colours depict different categories as listed on the right.

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

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

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

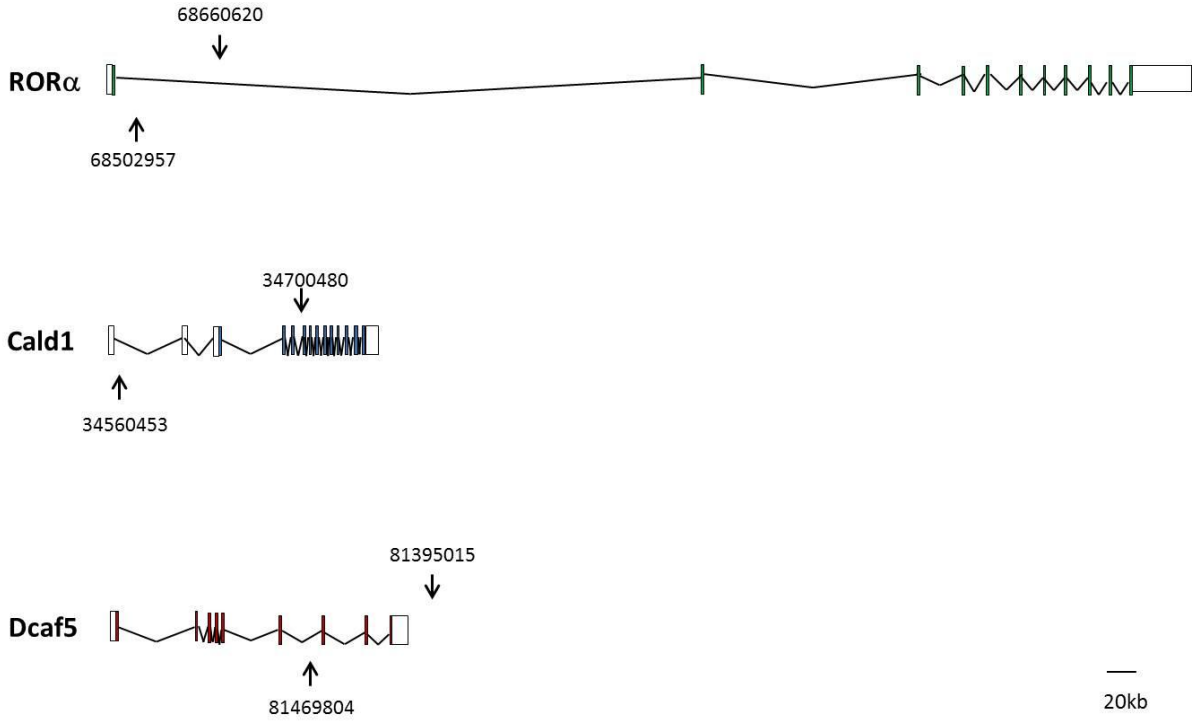

Figure 3.11 Schematic diagram showing transposon integration sites within three candidate genes Integration sites obtained from the PB-CAG-CKS-CAG-assisted genetic screen. Coloured boxes indicate the exons corresponding to the three genomic regions. Arrows depict the location of the integration sites,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

3.4 Discussion

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

3.4.1 Analysis of the ontological classification of candidate genes

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

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

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

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

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

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

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

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

3.4.2 Technical considerations of the screen

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

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

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

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

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

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

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

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

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

3.5 Conclusion

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