CHAPTER 4 : VALIDATION OF CANDIDATE GENES

4.1 Introduction

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

4.1.1 The piggyBac transposon as a reprogramming tool

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

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

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

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

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

4.1.2 Candidate Genes

4.1.2.1 Retinoic Acid Receptor beta (RARβ)

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

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

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

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

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

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

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

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

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

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

4.1.2.4 Basonuclin-2 (BNC2)

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

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

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

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

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

4.1.2.5 Caldesmon-1 (Cald1)

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

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

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

4.1.2.6 DDB1 and CUL4 associated factor 5 (DCAF5)

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

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

4.1.3 Chapter Aim

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

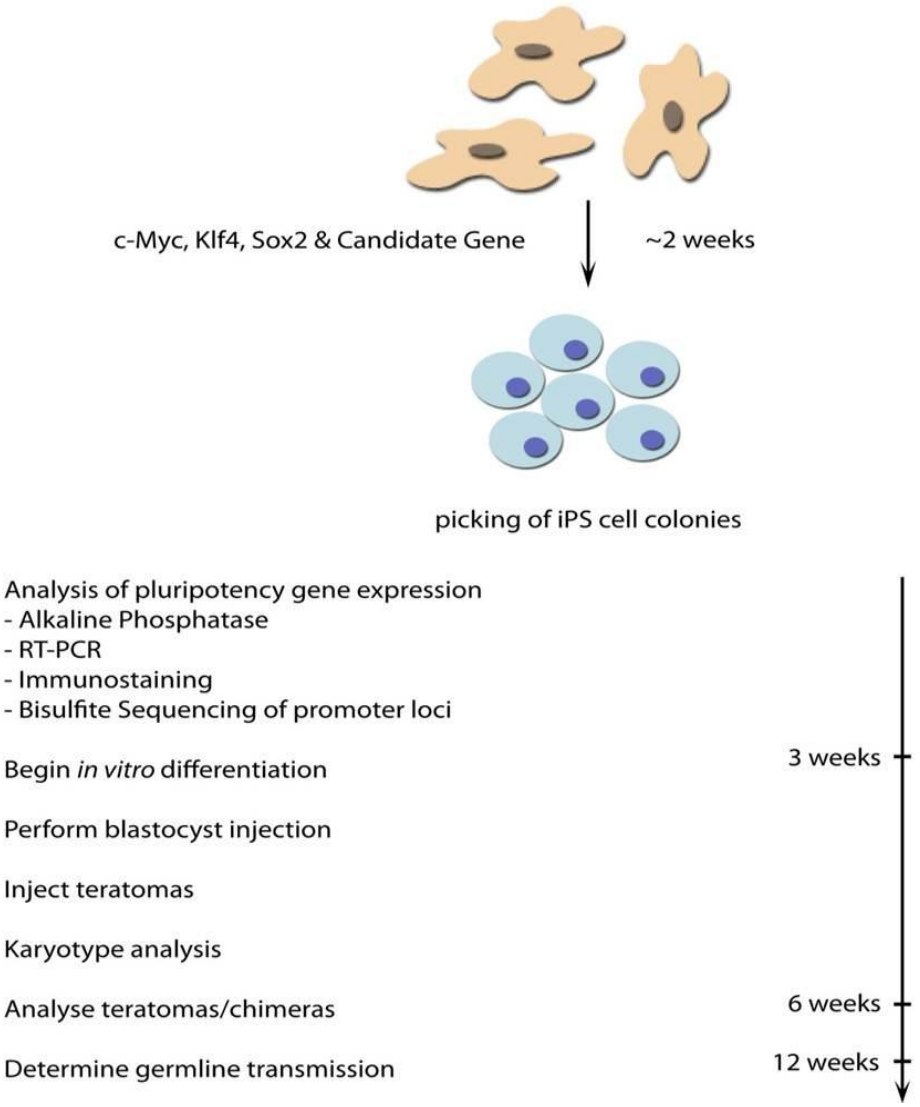

Figure 4.1 Schematic diagram showing strategy to validate pluripotency of iPSCs MEFs transfected with the desired reprogramming cocktail established iPSC colonies in an average of 2 weeks. These colonies were picked and expanded. As the iPSCs were reprogrammed in doxycycline inducible manner, withdrawal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

4.2 Experimental Strategy

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

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

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

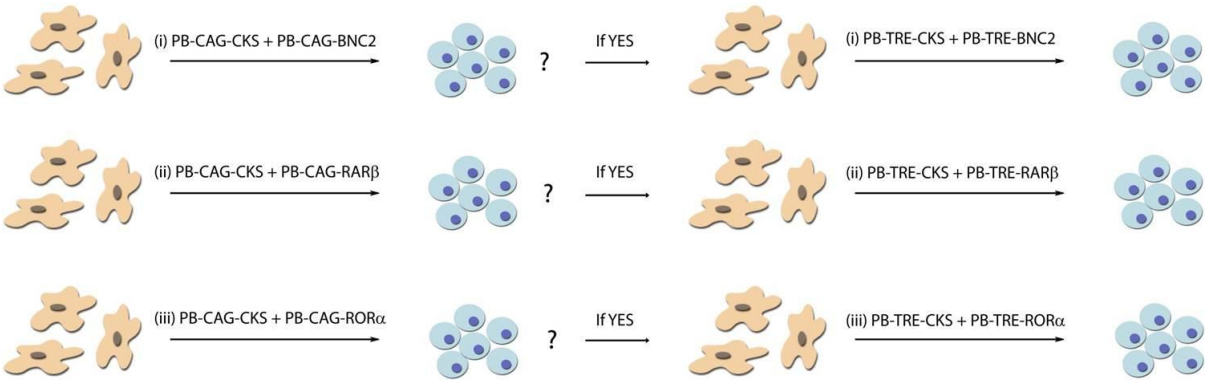

Figure 4.2 Schematic diagram displaying validation strategy (Left) piggyBac transposase was electroporated into MEFs with PB-CAG-CKS and an individual candidate gene. Transfected MEFs were seeded onto a mitotically inactive feeder layer and maintained for 3 weeks. (Right) If colonies were seen, a doxycycline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

4.2.1 Selection of validation candidates

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

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

4.3 Results

4.3.1 Generation of transposons expressing candidate genes

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

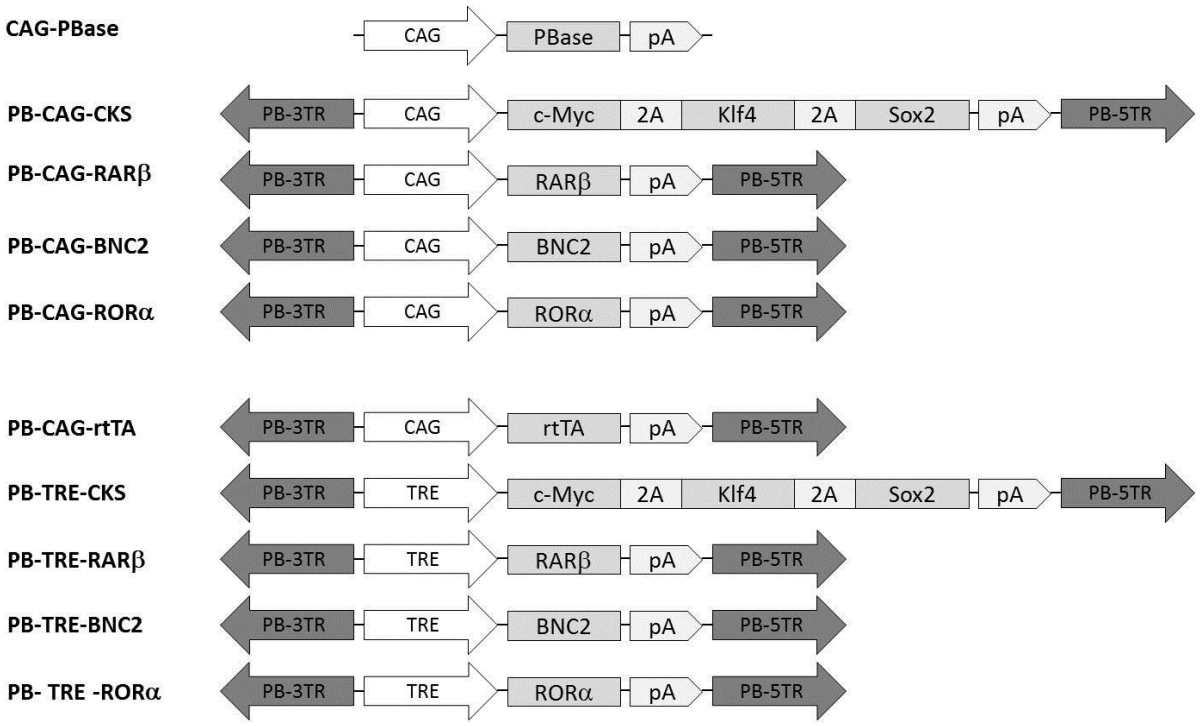

Figure 4.3 Schematic diagram of transgenes employed in the validation process (Top) The expression construct comprised of piggyBac transposase driven by a CAG promoter. (Middle) PB-CAG cassettes. The dark arrows flanking both ends depict piggyBac terminal repeats. Between the two boundaries lies a constitutively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

4.3.2 BNC2 and RARβ can function as Oct4 substitutes

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

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

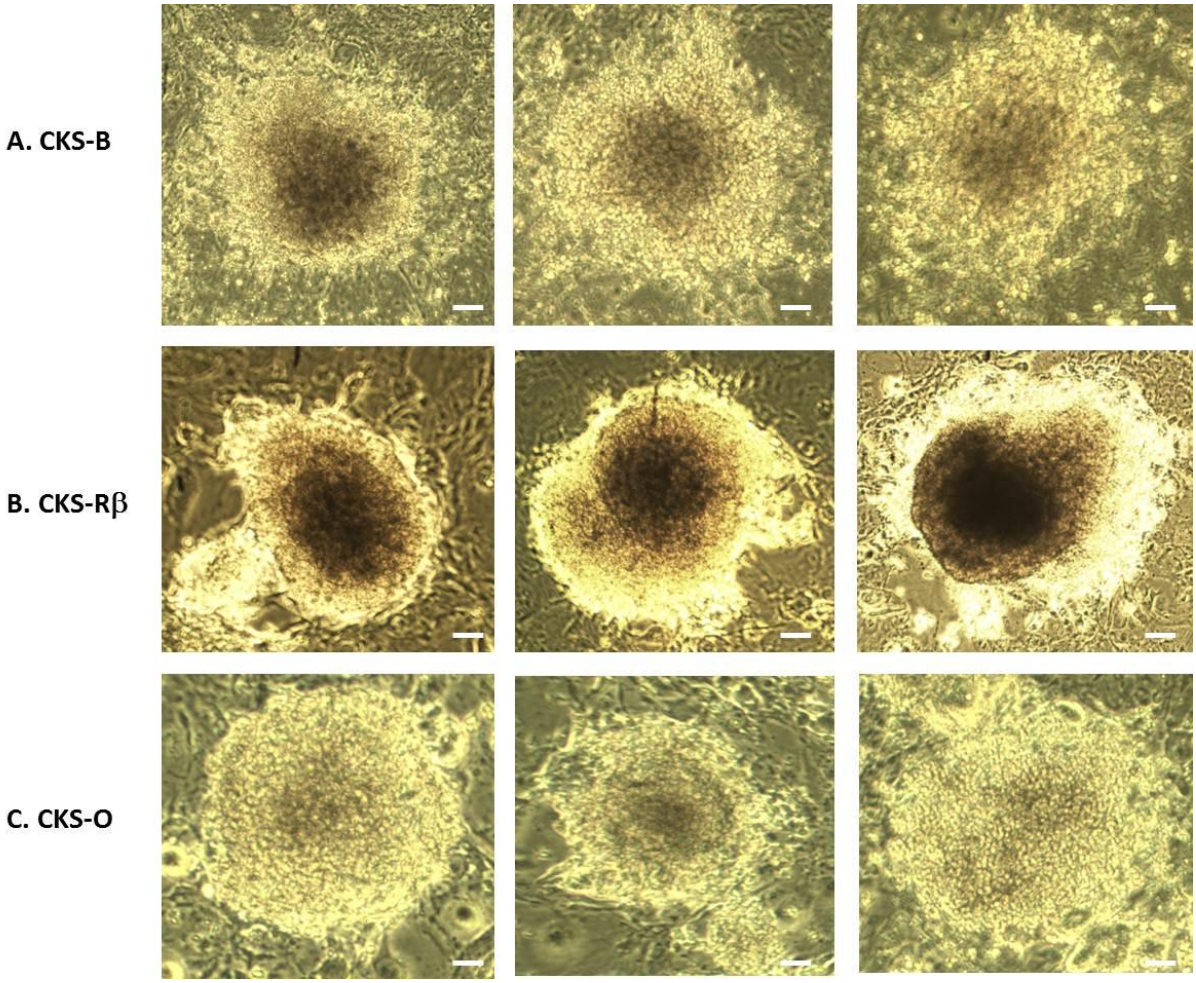

Figure 4.4 Microscope pictures of primary colonies obtained from candidate genes, BNC2 and RARβ MEFs were transfected with 2μg of CAG-PBase, 1μg of PB-CAG-CKS and 1μg of PB-CAG-mFx, where mFx refers to BNC2 (A), RARβ (B) or Oct4 (C). Transfected MEFs were seeded onto a mitotically inactive feeder layer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

4.3.2.1 Comparison of efficiencies between BNC2, RARβ and Oct4

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

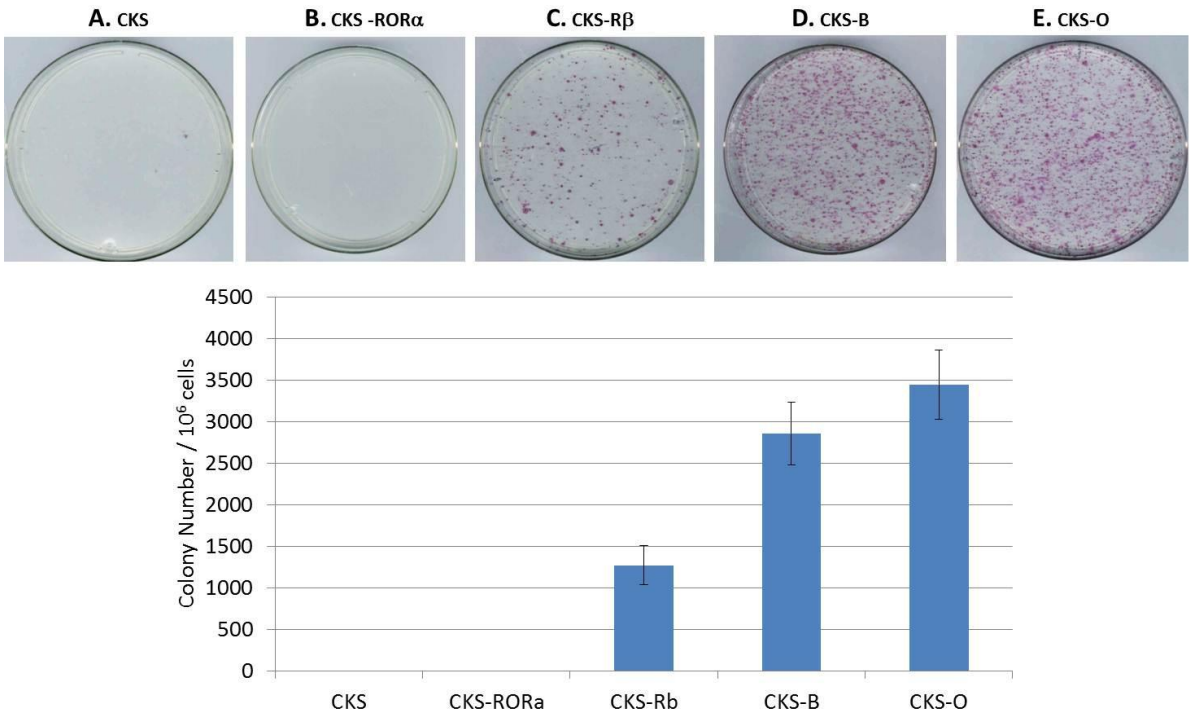

Figure 4.5 Comparison of reprogramming efficiencies between Oct4, BNC2 and RARβ MEFs were transfected with 2μg of CAG-PBase, 1μg of PB-CAG-CKS and 1μg of PB-CAG-mFx, where mFx refers to an empty vector (A), RORα (B), RARβ (C), BNC2 (D), or Oct4 (E). Transfected MEFs were seeded onto a mitotically inactive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

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

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

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

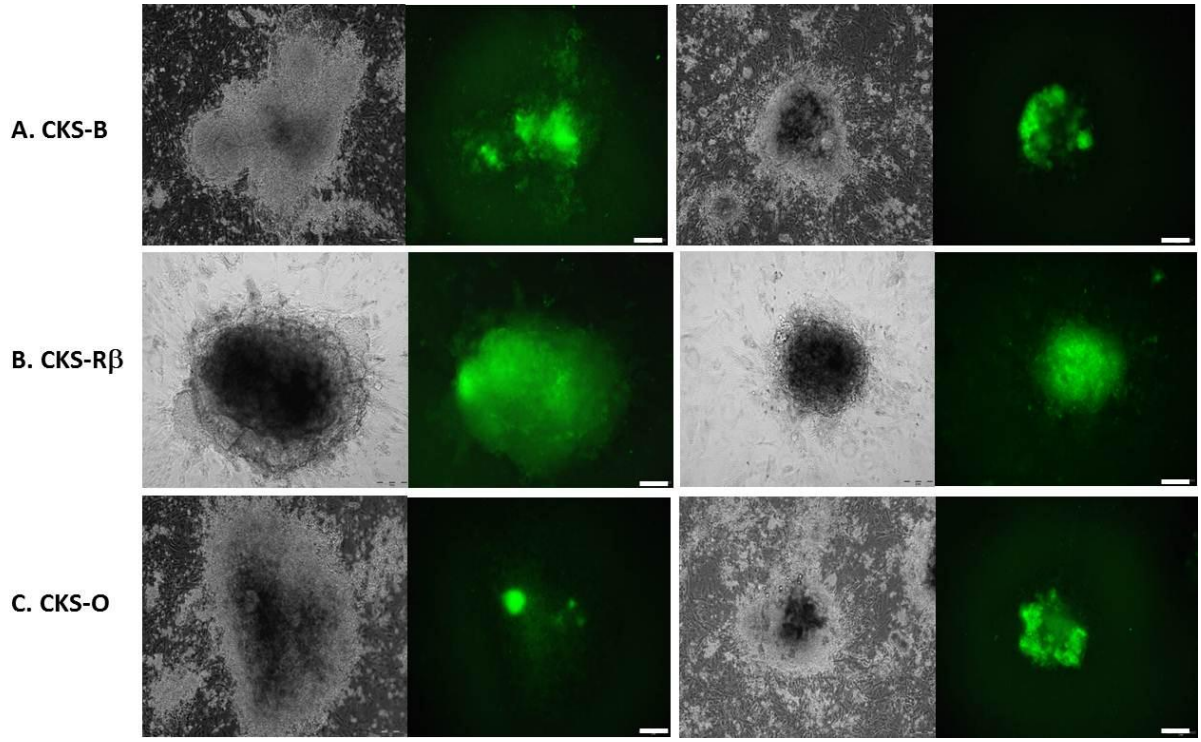

Figure 4.6 Rex1 activation in primary colonies obtained from candidate genes, BNC2 and RARβ Rex1::EGFP-IRES-Puro^{R} MEFs were transfected with 2μg of CAG-PBase, 1μg of PB-CAG-CKS and 1μg of PB-CAG-mFx, where mFx refers to BNC2 (A), RARβ (B) or Oct4 (C). Transfected MEFs were seeded onto a mitotically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

4.3.2.3 Additive effect of BNC2 and RARβ on conventional reprogramming cocktail

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

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

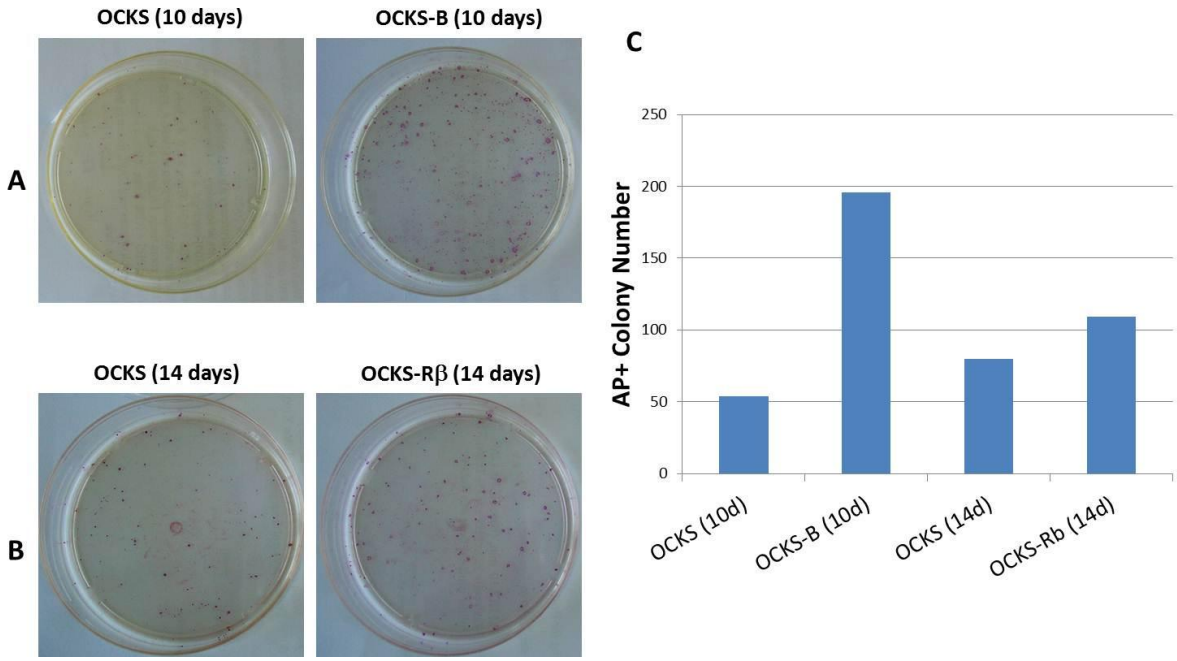

Figure 4.7 Investigation of reprogramming efficiencies in the presence of Oct4 and BNC2 or RARβ MEFs were transfected with 2μg of CAG-PBase, 1μg of PB-CAG-OCKS and 1μg of PB-CAG-mFx, where mFx refers to (A) BNC2 or (B) RARβ, or an empty vector. Transfected MEFs were seeded onto a mitotically inactive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

4.3.3 Generation and characterisation of transgene-independent BNC2-iPSCs

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

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

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

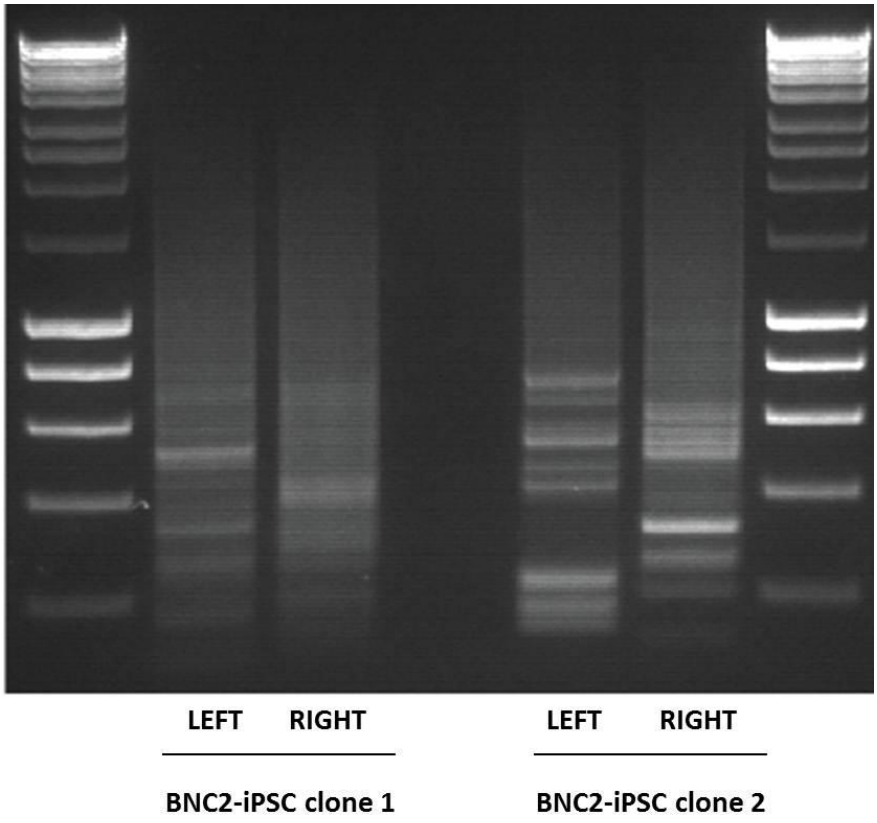

Figure 4.8 Splinkerette Analysis of integration sites in BNC2-iPSCs BNC2-iPSCs were generated using PB-TRE-CKS, PB-TRE-BNC2 and PB-CAG-rtTA. DNA from BNC2-iPSCs was harvested and splinkerette PCR was performed. Left and Right indicate the primers against the respective left or right transposon arm that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

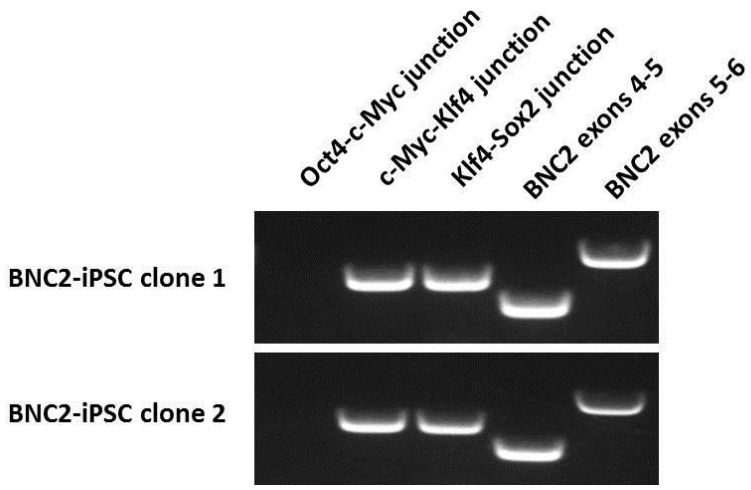

Figure 4.9 Validation of the expression of transposon cassettes Two BNC2-iPSC colonies generated from the introduction of PB-TRE-CKS (1μg), PB-TRE-BNC2 (1μg), CAG-PBase (2μg) and PB-CAG-rtTA (1μg) were expanded and maintained in the absence of doxycycline. RNA was extracted from each clone and converted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

4.3.3.1 Characterisation of BNC2-iPSCs

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

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

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

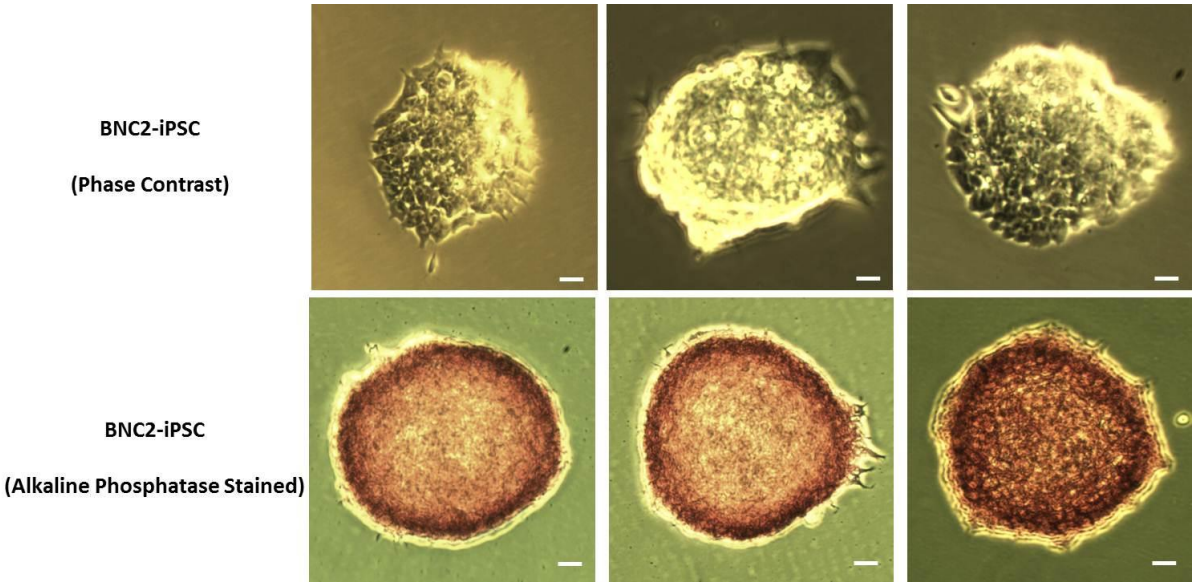

Figure 4.10 Images of doxycycline independent BNC2-iPSC colonies BNC2-iPSC colonies generated from the introduction of PB-TRE-CKS (1μg), PB-TRE-BNC2 (1μg), CAG-PBase (2μg) and PB-CAG-rtTA (1μg) were expanded and maintained in 2i+LIF, in the absence of doxycycline, and on gelatin coated plates. Three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

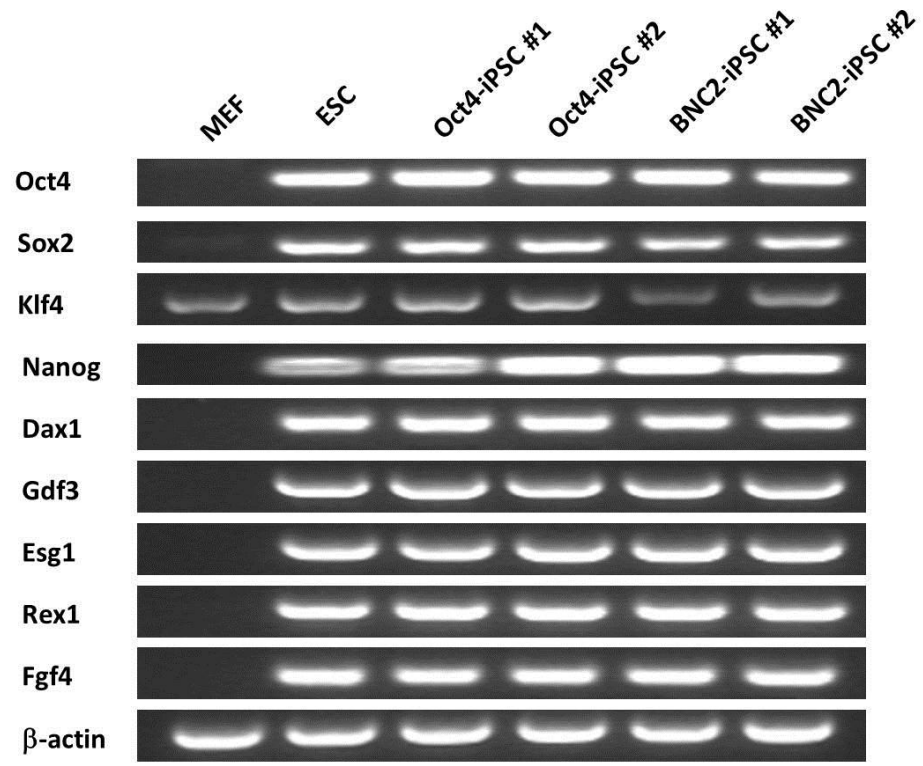

Figure 4.11 Expression levels of pluripotency markers in BNC2-iPSCs RT-PCR of the various pluripotency markers was performed in MEF (lane1), ESCs (lane2), iPSCs generated using c-Myc, Klf4, Sox2 and Oct4 (lanes 3 and 4), and iPSCs generated using c-Myc, Klf4, Sox2 and BNC2 (lanes 5 and 6). Expression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

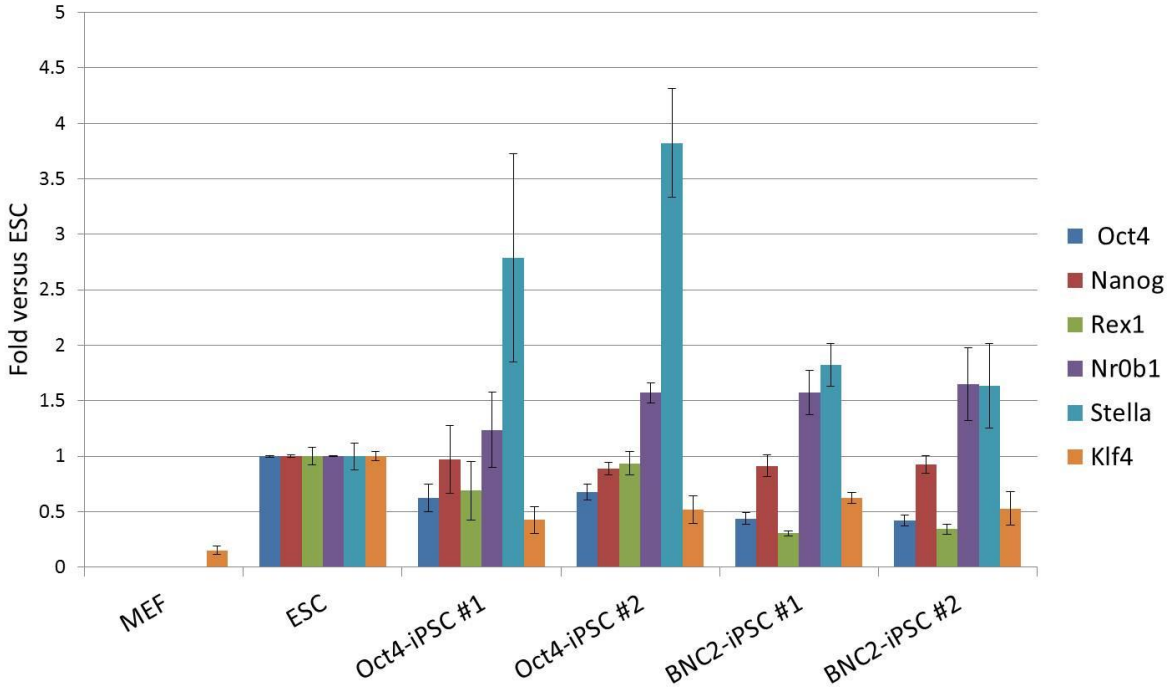

Figure 4.12 Quantitative expression levels of pluripotency markers in BNC2-iPSCs q-PCR of the various pluripotency markers was performed in MEF (lane1), ESC (lane2), Oct4-iPSCs (lanes 3 and 4), and BNC2-iPSCs (lanes 5 and 6). Expression levels of pluripotency markers are normalised to GAPDH amounts and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

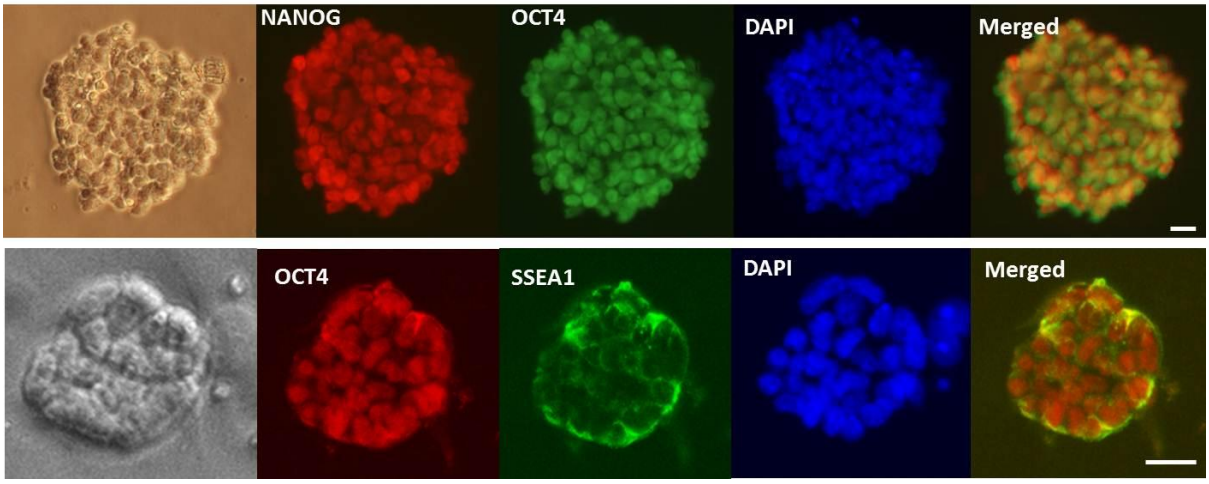

Figure 4.13 Immunostaining of pluripotency markers in BNC2-iPSCs Doxycycline independent BNC2-iPSCs were fixed and stained with antibodies against, (Top) Nanog (red) and Oct4 (green), and (Bottom) Oct4 (red) and SSEA-1 (green). Images of representative colonies are shown, displaying the respective fluorescence,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

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

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

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

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

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

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

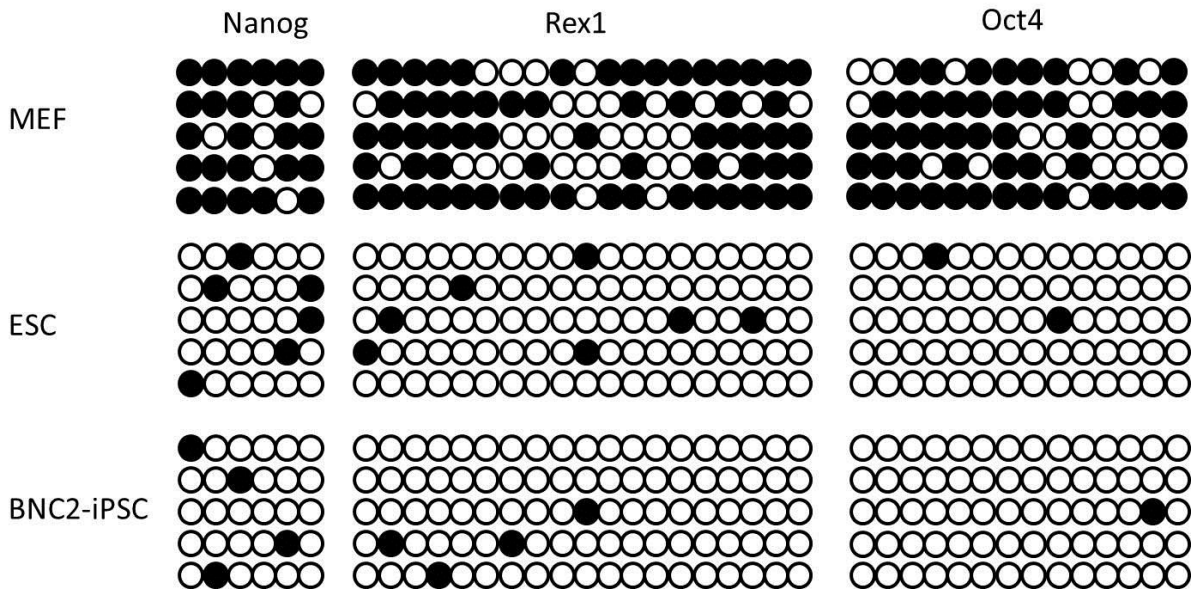

Figure 4.14 Analysis of methylation at promoter loci DNA was obtained from MEF (Top), ESCs (Middle) and doxycycline independent BNC2-iPSCs (Bottom) and treated with sodium bisulphite. PCR directed at promoter loci of Nanog (Left), Rex1 (Centre) and Oct4 (Right) were performed. Amplified products were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

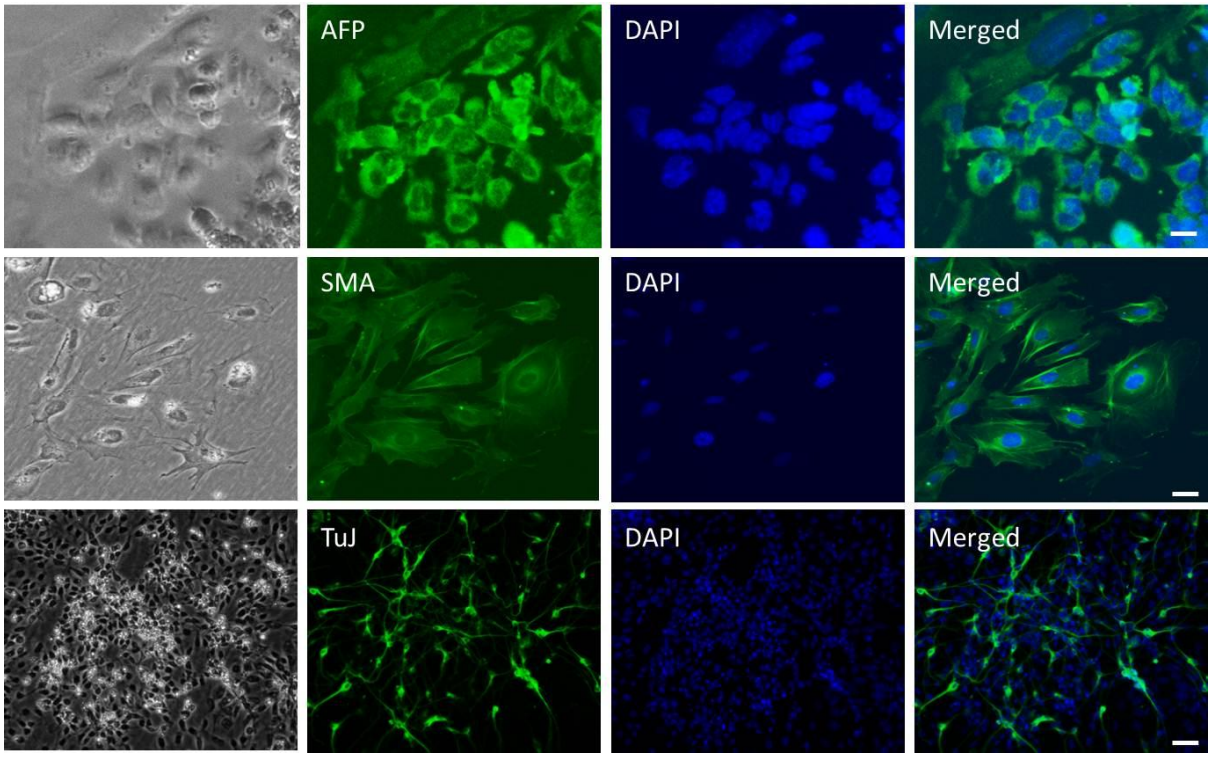

Figure 4.15 BNC2-iPSCs can differentiate into three germ layers *in vitro* BNC2-iPSCs were differentiated into the three germ layers *in vitro* using various culture conditions. Differentiated cells representing the endodermal, mesodermal and ectodermal lineages were immunostained against alpha feto-protein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

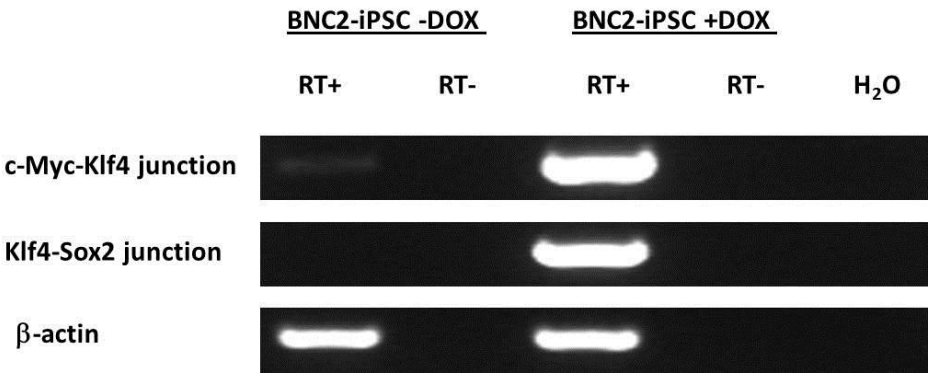

Figure 4.16 Transgenes are silent in the absence of doxycycline BNC2-iPSCs were cultivated in the presence and absence of doxycycline (DOXYCYCLINE). RNA was extracted from these cells and (i) converted to cDNA using reverse transcriptase (RT+) or (ii) tested for contamination with genomic DNA in the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

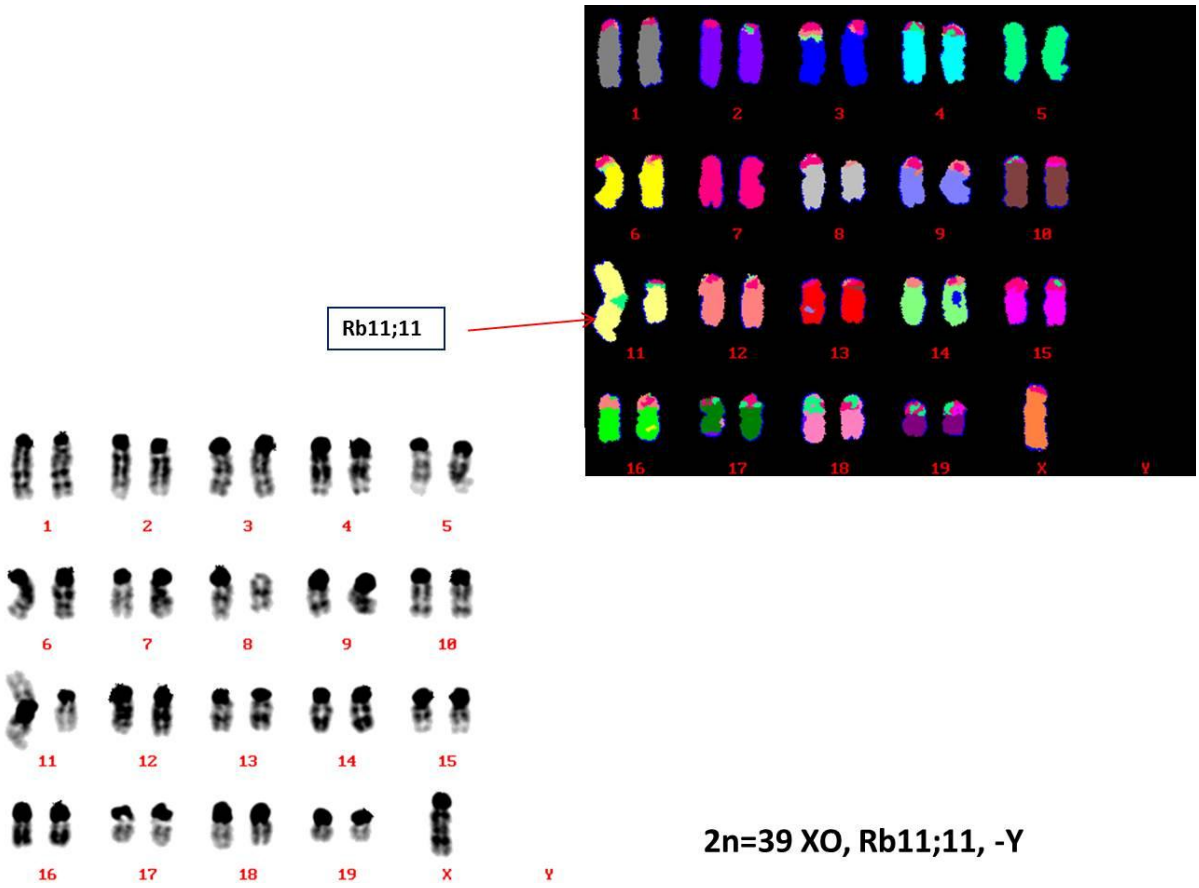

Figure 4.17 Karyotype analysis of BNC2-iPSCs BNC2-iPSCs were mitotically arrested and metaphase spreads were achieved. Using Fluorescence *in situ* hybridisation (FISH), each chromosome was accounted and analysed. The array of colours depicts different chromosomes and the arrow indicates a centromeric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

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

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

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

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

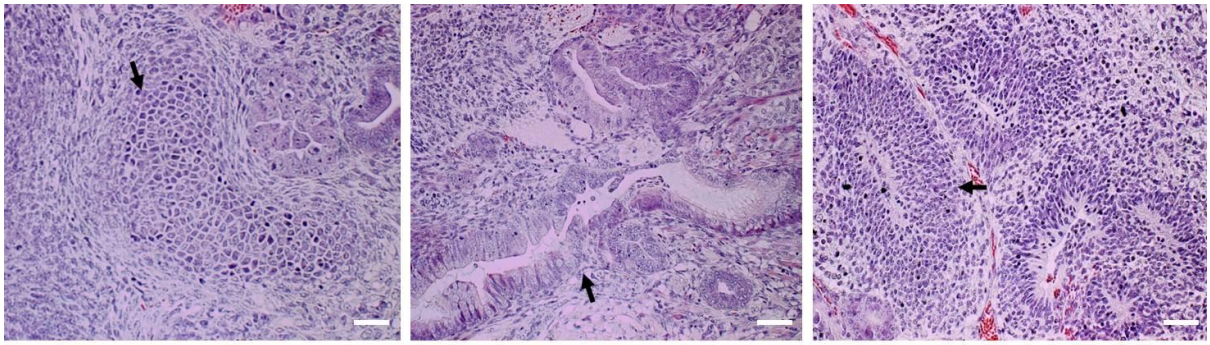

Figure 4.18 BNC2-iPSCs can differentiate into three germ layers *in vivo* BNC2-iPSCs were injected subcutaneously into both flanks of immunocompromised NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice. After 2 weeks, tumours were observed. Histological analyses of sections of these tumours are shown.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

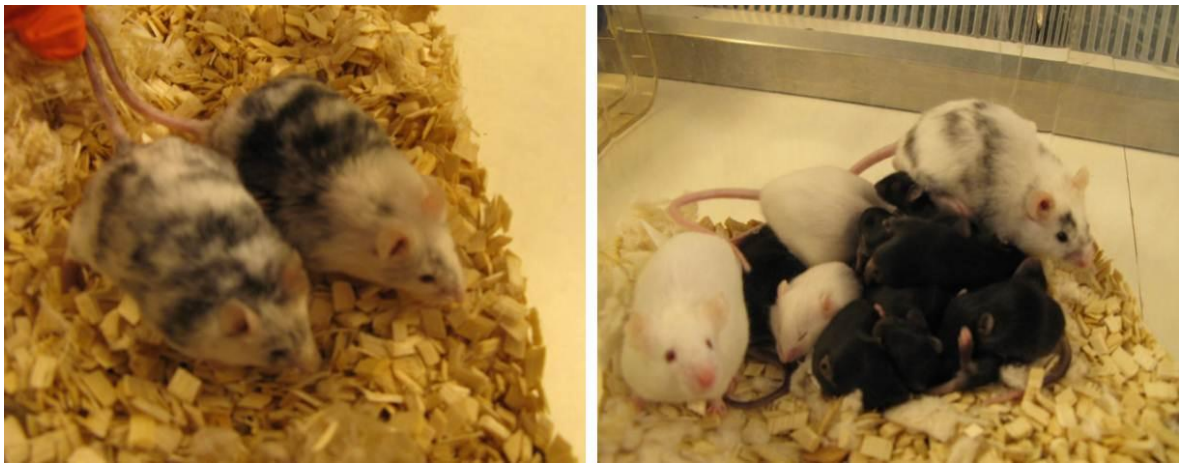

Figure 4.19 BNC2-iPSCs are capable of generating chimeras and contributing to the germline (LEFT) Chimeras (50% and 70%) were generated through the injection of BNC2-iPSCs into mouse blastocysts. (RIGHT) Crossing between a 30% chimera and albino CALB mice resulted in the generation of pups carrying a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

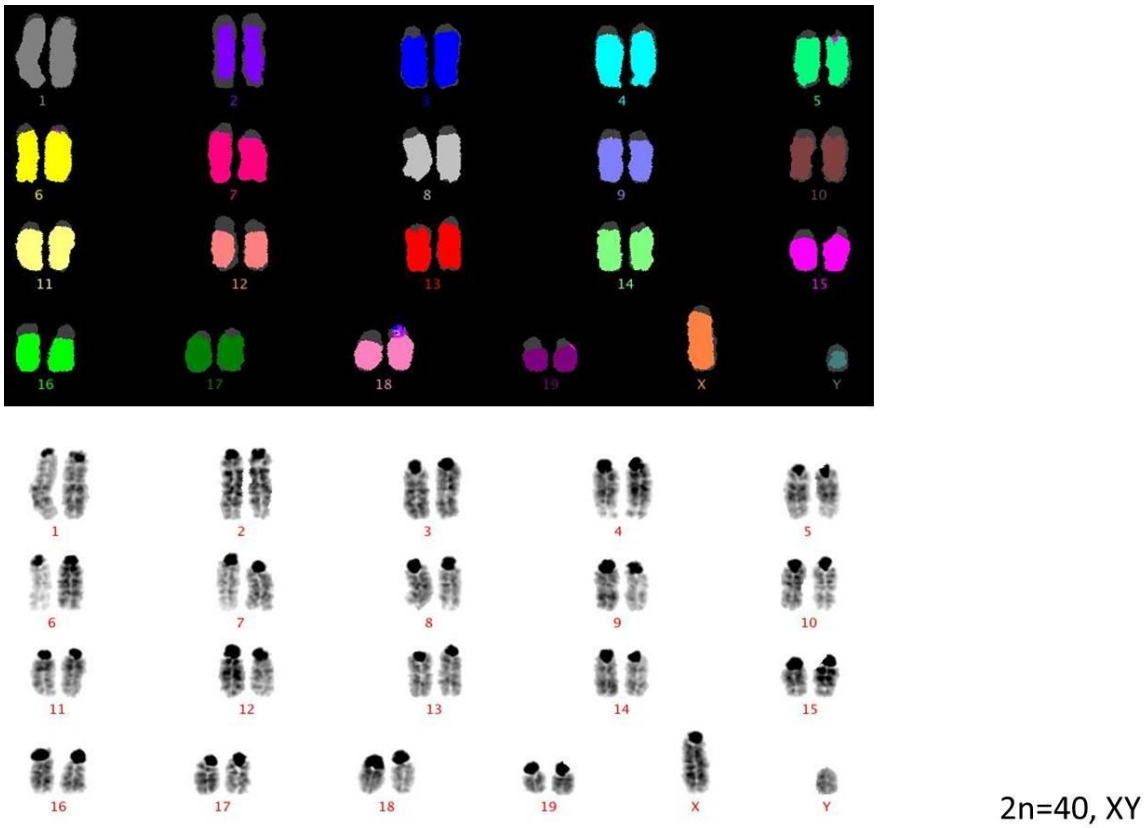

Figure 4.20 Karyotype analysis of splenocytes from a mouse chimera derived using BNC2-iPSCs Splenocytes were extracted from the spleen of a chimeric mouse. Splenocytes were mitotically arrested and metaphase spreads were achieved. Using Fluorescence *in situ* hybridisation (FISH), each chromosome was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

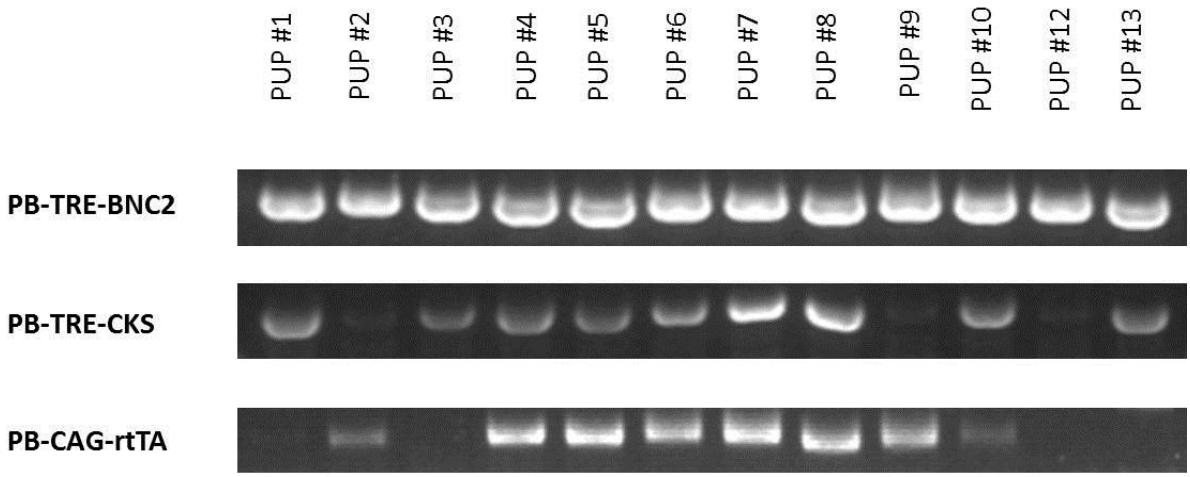

Figure 4.21 Genotyping of mouse colony established from BNC2-iPSCs Pups which displayed germline contribution of BNC2-iPSCs were crossed among themselves to establish a mouse colony. 13 pups belonging to F2 generation were genotyped for the presence of transposon cassettes. DNA was retrieved from ear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

4.3.4 Generation and characterisation of transgene-independent RARβ iPSCs

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

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

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

4.3.4 Characterisation of RARβ-iPSCs

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

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

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

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

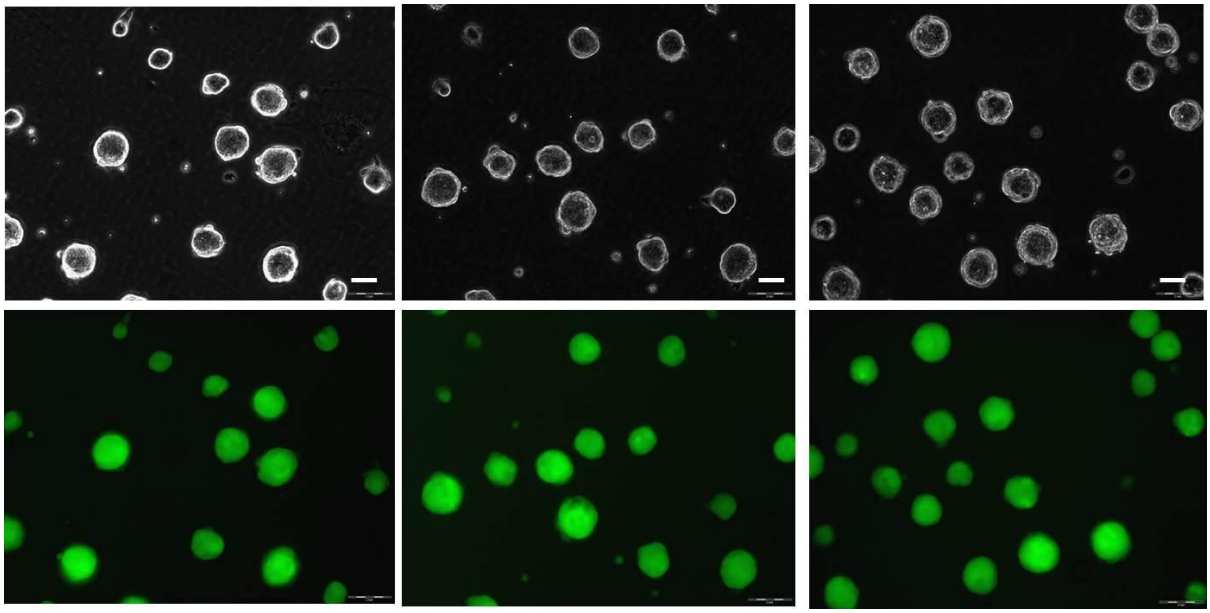

Figure 4.22 Rex1 expressing RARβ-iPSCs RARβ-iPSC colonies generated from the introduction of PB-TRE-CKS (1μg), RARβ (1μg), CAG-PBase (2μg) and PB-CAG-rtTA (1μg) were expanded and maintained in the absence of doxycycline, in 2i+LIF, on gelatin coated plates. As the starting MEFs contain a reporter cassette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

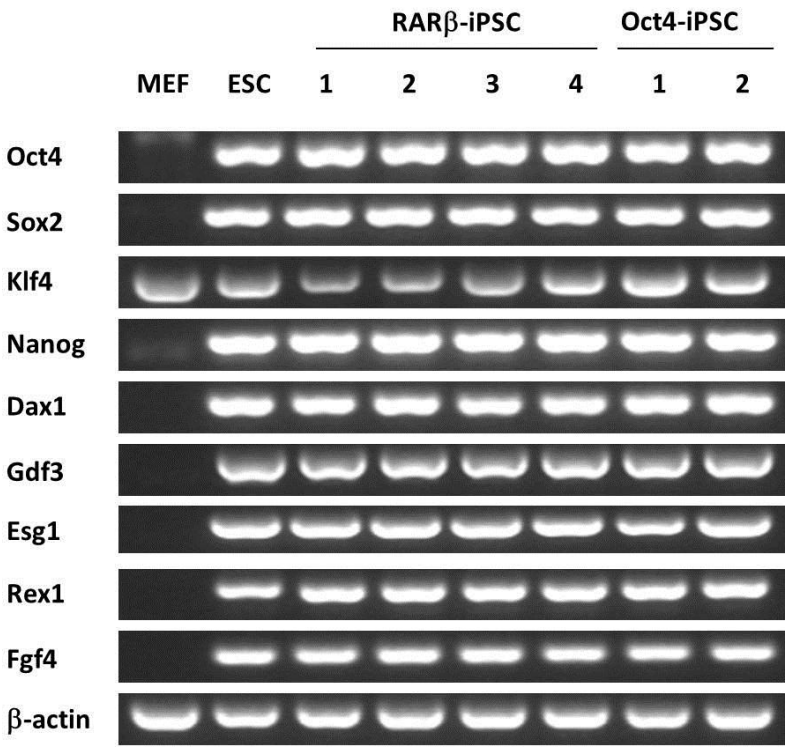

Figure 4.23 Expression levels of pluripotency markers in RARβ-iPSCs RT-PCR of the various pluripotency markers was performed in MEF (lane1), ESCs (lane2), iPSCs generated using c-Myc, Klf4, Sox2 and RARβ (lanes 3-6), and iPSCs generated using c-Myc, Klf4, Sox2 and Oct4 (lanes 7 and 8). Expression levels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

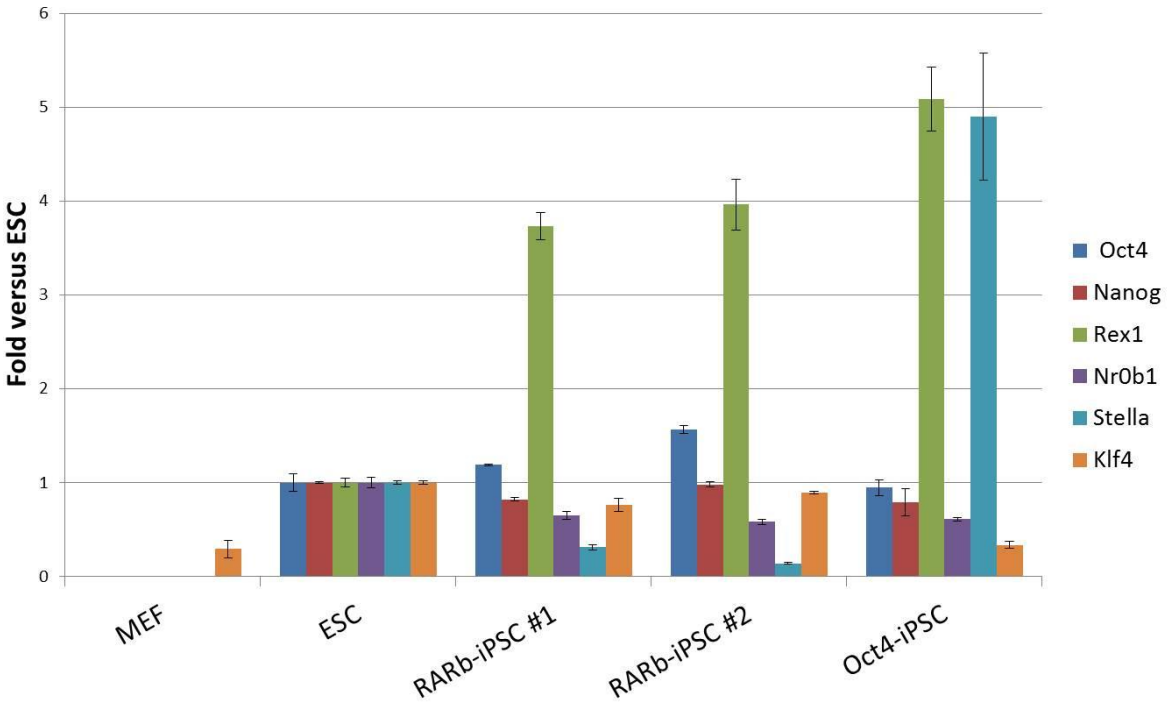

Figure 4.24 Quantitative expression levels of pluripotency markers in RARβ-iPSCs q-PCR of the various pluripotency markers was performed in MEF (lane1), ESCs (lane2), RARβ-iPSCs (lanes 3 and 4), and Oct4-iPSCs (lanes 5). Expression level of each pluripotency marker was normalised to GAPDH amounts and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

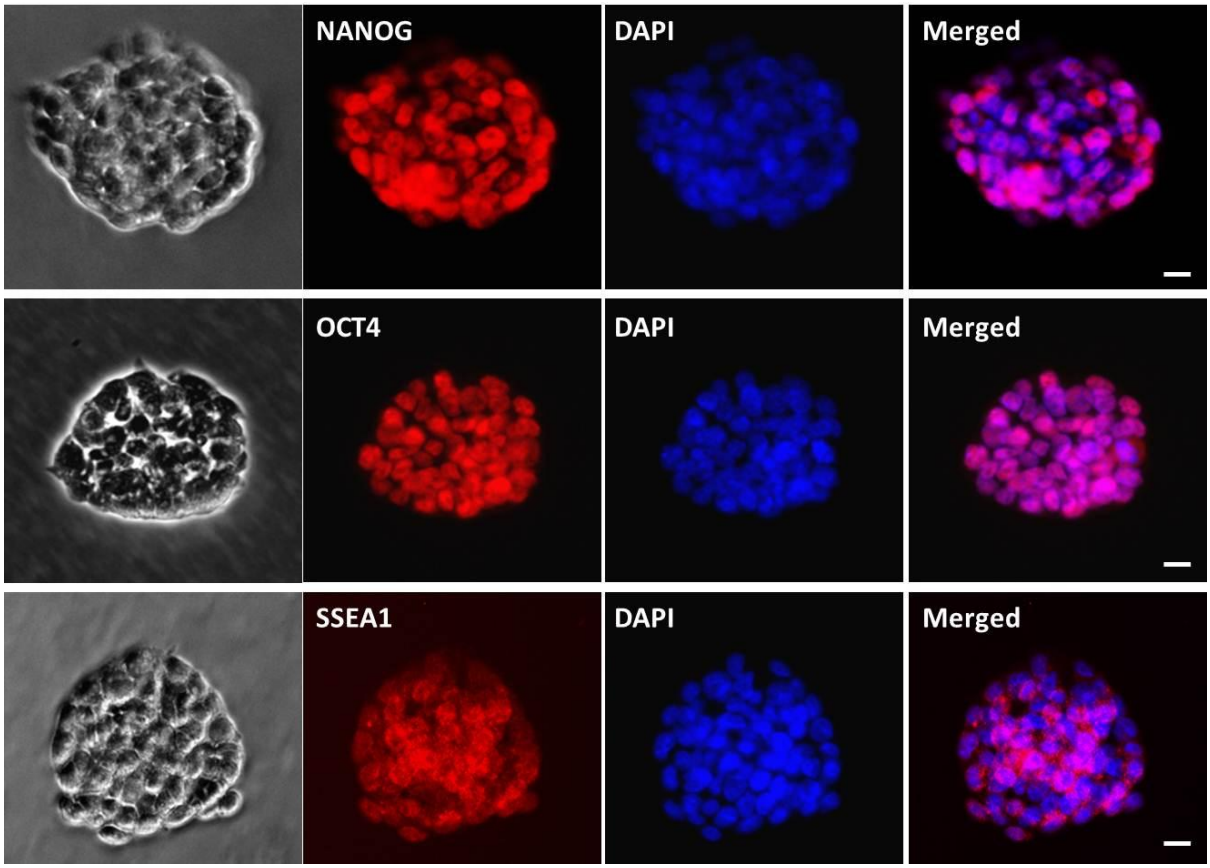

Figure 4.25 Immunostaining of pluripotency markers in RARβ-iPSCs Doxycycline independent RARβ-iPSCs were fixed and stained with antibodies against, Nanog (Top), Oct4 (Middle) and SSEA-1 (Bottom). Images of representative colonies are shown, displaying the respective fluorescence, brightfield and DAPI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

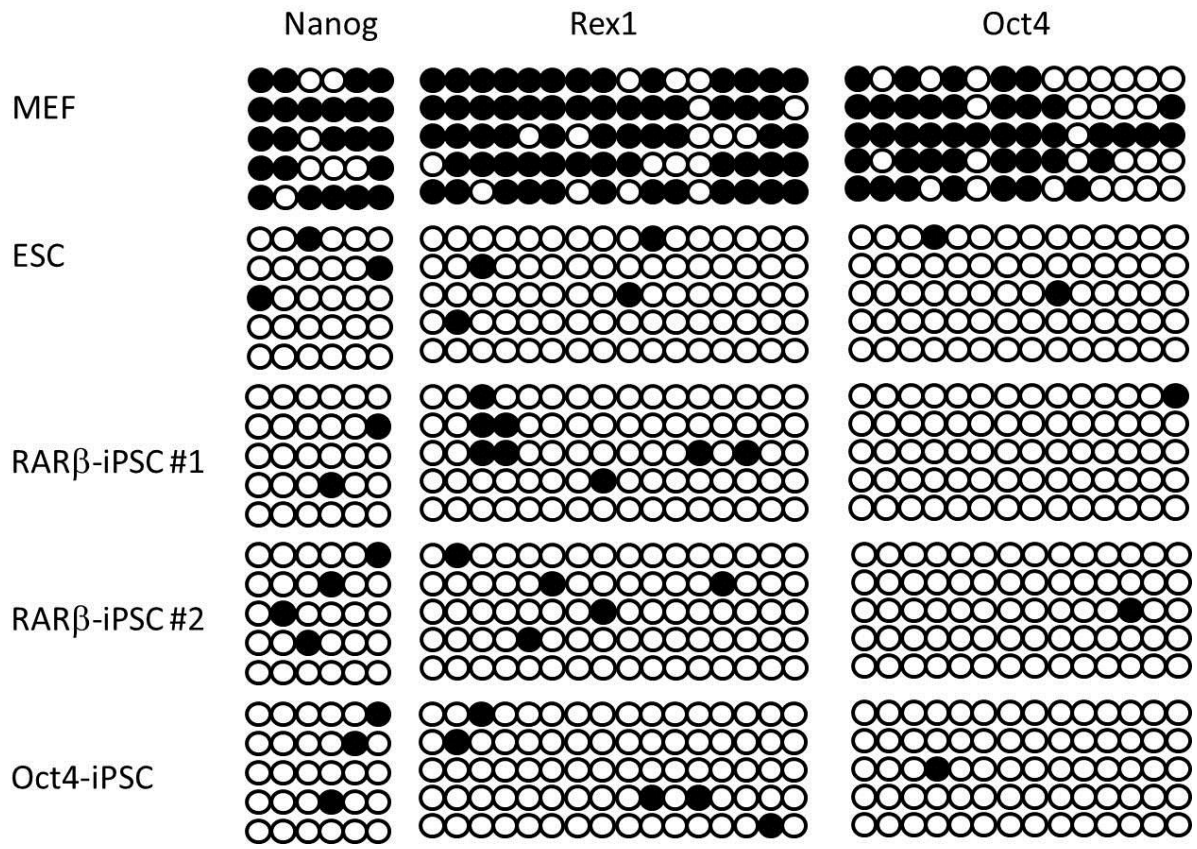

Figure 4.26 Analysis of methylation at promoter loci in RARβ-iPSCs DNA was obtained from MEF, ESCs, two RARβ-iPSC clones and Oct4-iPSCs and treated with sodium bisulphite. PCR directed at promoter loci of Nanog (LEFT), Rex1 (CENTRE) and Oct4 (RIGHT) were performed. Amplified products were analysed and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

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

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

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

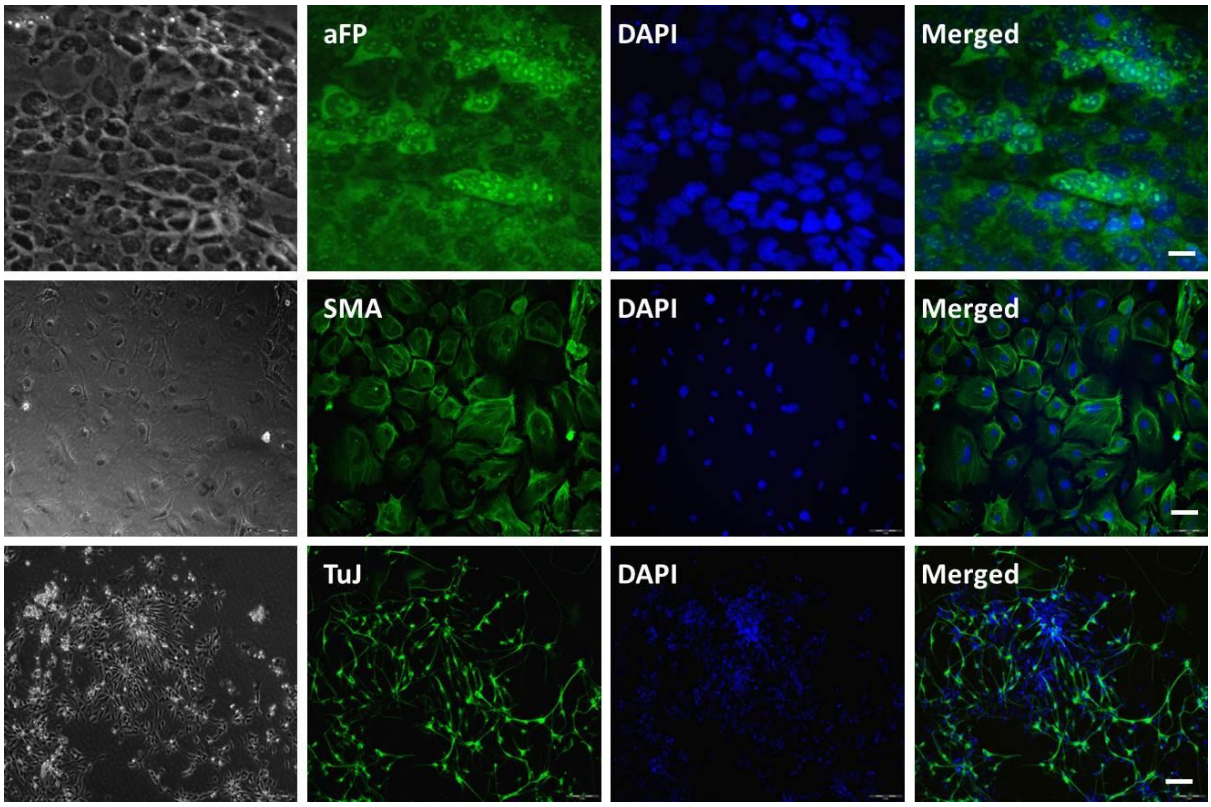

Figure 4.27 RARβ iPSCs can differentiate into three germ layers *in vitro* RARβ-iPSCs were differentiated into the three germ layers in vitro using various culture conditions. Differentiated cells representing the endodermal, mesodermal and ectodermal lineages are immunostained against alpha feto-protein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

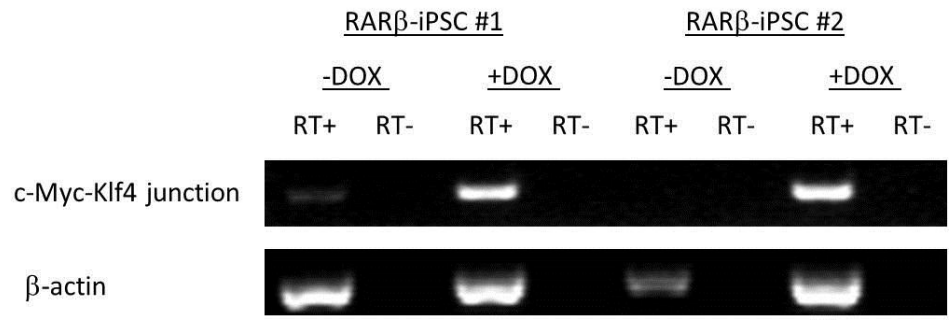

Figure 4.28 Transgenes are silent in the absence of doxycycline Two RARβ-iPSC clones were cultivated in the presence and absence of doxycycline. RNA was extracted from these cells and (i) converted to cDNA using reverse transcriptase (RT+) or (ii) tested for contamination with genomic DNA in the absence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

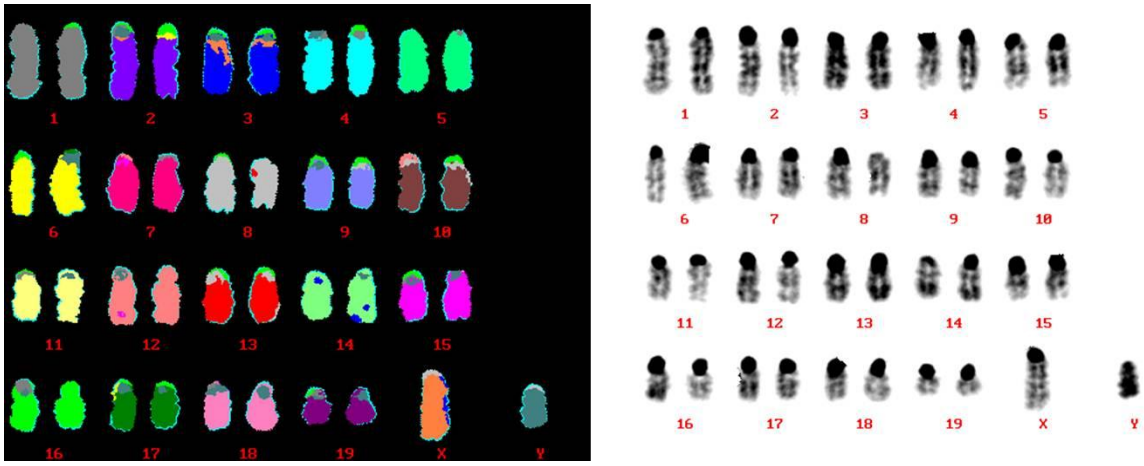

Figure 4.29 Karyotype analysis of RARβ-iPSCs RARβ-iPSCs were mitotically arrested and metaphase spreads were achieved. Using Fluorescence *in situ* hybridisation (FISH), each chromosome was accounted and analysed. The array of colours depicts different chromosomes.

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

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

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

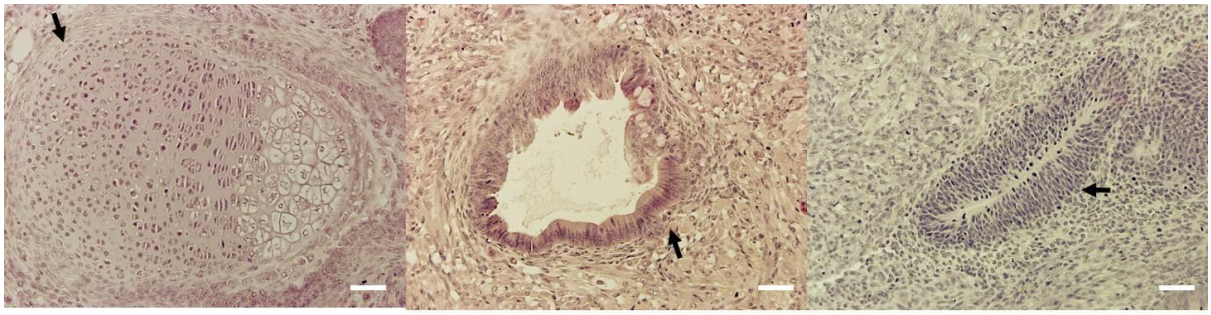

Figure 4.30 RARβ-iPSCs can differentiate into three germ layers *in vivo* RARβ-iPSCs were injected subcutaneously into both flanks of immune-compromised NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice. After 2 weeks, tumours were observed. Histological analyses of sections of these tumours are shown.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

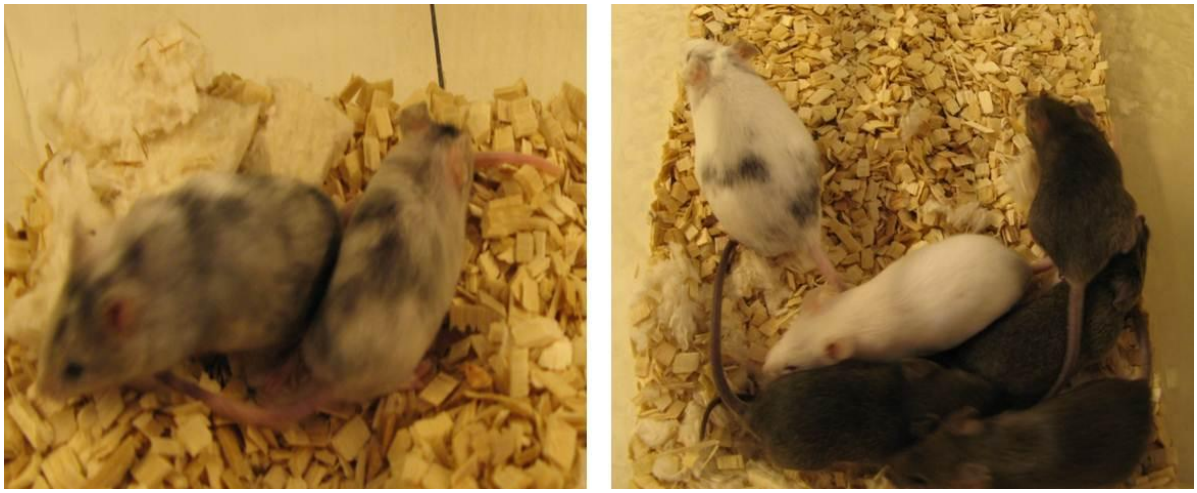

Figure 4.31 RARβ-iPSCs are capable of generating chimeras and contributing to the germline (LEFT) Chimeras (40% and 50%) were generated through the injection of RARβ-iPSCs into mouse blastocysts. (RIGHT) Crossing between a 10% chimera and albino CALB mouse results in the generation of pups carrying a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

4.4 Discussion

4.4.1 Variable reprogramming efficiencies

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

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

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

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

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

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

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

4.4.2 RORα and its inability to trigger pluripotency

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

4.4.3 Genomic instability in BNC2-iPSCs

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

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

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

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

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

4.4.4 Conclusion

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