THE UNIVERSITY
of EDINBURGH

Identifying novel regulators of reprogramming using RNA interference

Sara Brightwell

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2015

Declaration

I declare that the work presented in this thesis is my own, unless otherwise stated, and has not been submitted for any other degree or professional qualification.

_____

Sara Brightwell

Acknowledgements

To Kei and all of the amazing members of my lab, as well as the hard-working staff at the ISCR and SCRM – thank you. I couldn't have done this without all of you, literally!

For Lee and Atlas – because nothing else matters.

Abstract

Since Yamanaka and Takahashi first described the isolation of induced pluripotent stem cells (iPSCs) in 2006, researchers have invested a vast amount of time and resources into trying to understand the process of reprogramming. However, the exact mechanisms underlying the induction of somatic cells to pluripotency is still incompletely understood. With this in mind, a screening approach was undertaken to identify shRNA that enhance the reprogramming process. A retrovirus based system was used to knock down candidate genes during reprogramming of mouse embryonic fibroblasts (MEF) containing doxycycline-inducible reprogramming factors and a Nanog-GFP reporter, which is activated when cells become iPSCs. The initial round of screening with over 150 shRNA vectors successfully identified several shRNAs that enhance reprogramming. One of these shRNA vectors exhibited both faster reprogramming kinetics as determined by activation of the Nanog-GFP reporter 2 to 3 days earlier and increased reprogramming efficiency giving rise to >5 fold more GFP+ colonies when compared with a control. Cell surface marker analysis with flow cytometry demonstrated that changes in CD44 and ICAM1 expression, which occur preceding Nanog-GFP expression, were also accelerated. Validation of this shRNA determined that the enhanced reprogramming phenotype is the result of an unknown off-target effect. Microarray and RNA-sequencing analysis was carried out to identify the off target gene with a view to investigate the functional importance of this knock down and its role in establishing the pluripotency transcriptional network during reprogramming.

Contents

Index of figures

Index of tables

CHAPTER 1 - Introduction

1.1 Changing cellular potential

1.1.1 Reversion to pluripotency through somatic cell nuclear transfer

Early cell development studies focusing on fate changes during differentiation were key in paving the way for progression of modern day stem cell science. In a study published 80 years ago, Hämmerling (1934) used *Acetabularia*, a green algae, to generate heterokaryons and the resulting organism comprised features characteristic of both donor species. This put forward the idea that morphogenetic products derived by the nucleus could transport to the cytoplasm thereby regulating differentiation. Following on from this, later work by Briggs and King (1952) revealed that transplantation of nuclei from *Rana pipiens* (frog) blastulas into enucleated eggs subsequently gave rise to embryos that had normal differentiation capacity. This indicated that during differentiation, the nuclei of cells are not irreversibly changed. Interestingly, when the nuclei from a different species of frog, *Rana catesbeiana*, were transplanted into enucleated eggs of *Rana pipiens*, blastulas formed but died shortly after arrest, which is consistent with the lethal nature of normally produced hybrids of these species. In a subsequent report, the same group carried out serial transplantation of nuclei from late gastrula endoderm in to enucleated eggs and determined that second and subsequent generation clones of nucleated eggs recapitulated development similarly to the first generation i.e. if the first transplanted egg developed normally so did clones generated from serial transplantations and alternatively, if the first transplanted egg arrested at gastrulation subsequent clones behaved similarly and so on (King and Briggs, 1956). This work provided some evidence that differentiation potential of nuclei can be stably inherited, and nuclear changes or defects may not be reset to a normal status with serial transplantation. Similar results were confirmed in *Xenopus laevis* (Gurdon, 1960) and a later report by Gurdon (1962) demonstrated that serial transfer of nuclei isolated from fully differentiated *Xenopus* intestinal epithelium cells into

enucleated eggs could give rise to normal feeding tadpoles. This pioneering work provided evidence for the first time that genetic information is maintained throughout the lifetime of a cell and furthermore, genetic changes incurred during normal development can be reverted back to the original state. These experiments were among the first reports of somatic cell nuclear transfer (SCNT) and almost 50 years later this technique would be used to produce the first mammal to be cloned from an adult somatic cell, a lamb, nicknamed Dolly the sheep to due the mammary gland origin of the somatic nucleus used to clone her (Wilmut et al., 1997). The animal was deemed to be healthy with no genetic abnormalities or otherwise, and in addition went on to produce viable offspring of her own providing conclusive proof that differentiation of cells does not cause permanent or irreversible changes to the genome and that nuclei can be 'reset' to a totipotent embryonic state. Since this achievement, several other animals have been successfully cloned including pigs (Polejaeva et al., 2000), dogs (Lee et al., 2005), horses (Galli et al., 2003) and deer (Berg et al., 2007), among many others but there has been controversy surrounding use of the technique to clone animals due to the low efficiency of the method in producing viable animals in addition to abnormal development commonly occurring in the clones caused by incomplete epigenetic resetting (Rideout et al., 2001).

1.1.2 Embryonic stem cells

Since embryonic stem cells (ESCs) were first described 1981 (Evans and Kaufman, 1981) they have been anticipated to hold great promise for the future of regenerative medicine. ESCs have the characteristic hallmarks of unlimited self-renewal whilst remaining in an undifferentiated state and maintain the pluripotent capacity to differentiate to cells of the three germ lineages, endoderm, mesoderm and ectoderm (Shufaro and Reubinoff, 2004; Nishikawa et al., 2007). ESCs were first derived from *in vitro* culture of mouse blastocysts and identified as proliferating

cells bearing similarities to embryonal carcinoma (EC) cells. These so called EK cells were found to be capable of differentiation both *in vitro* and *in vivo* as teratocarcinomas, and unlike all of the EC cell lines derived by the group they had a normal karyotype ruling out contamination of cultures by EC cells (Evans and Kaufman, 1981). Importantly, this group discovered that EK cells were found to contribute to chimeras capable of germ-line transmission (Bradley et al., 1984). Just a few months following the initial report, another group coined the term "embryonic stem cell" in a report describing a pluripotent cell line derived from the inner cell mass (ICM) of preimplantation mouse embryos (Martin, 1981). Clonal cell lines were found to be capable of differentiating in to an array of cell types through teratoma formation, supporting the previous report. These ESCs were maintained in an undifferentiated state for several passages in a medium condition by EC cells, suggesting that the medium contained some factor either promoting self-renewal or inhibiting differentiation. Indeed it was subsequently determined that leukemia inhibitory factor (LIF) could replicate the effect of "differentiation inhibitory activity (DIA)" in conditioned medium used to propagate ESCs (Smith et al., 1988; Williams et al., 1988). Addition of LIF to the culture medium prevented spontaneous differentiation of ESCs and negated the need for propagation on a fibroblast feeder layer. It was later discovered that bone morphogenic protein 4 (BMP4) in combination with LIF was sufficient to maintain self-renewal capacity of ESCs in the absence of serum in feeder-free conditions (Ying et al., 2003). These early studies paved the way for ESCs to be used as an invaluable model system to dissect the mechanisms of early development, pluripotency and fate decisions determining differentiation of cells.

Human embryonic stem cells (hESCs) with the potential to differentiate into cells representative of all three germ layers were derived 17 years later (Thomson et al., 1998). Despite these cells being established from the ICM of human embryos similarly to mouse ESCs (mESCs), however, their potential to self-renew was not supported by LIF and BMP (Thomson et al., 1998; Reubinoff et al., 2000). Morphologically, hESC colonies appeared flatter than their mouse counterpart and

unlike mESCs, hESCs express markers SSEA-4, TRA-1-81 and TRA-1-60 (Reubinoff et al., 2000; Ginis et al., 2004) and are capable of trophoblast differentiation (Thomson et al., 1998; Odorico et al., 2001; Edwards, 2002). FGF2 (bFGF) and activin/Nodal pathways were later found to support pluripotency and self-renewal of hESCs (Amit et al., 2000; Vallier et al., 2005; Schnerch et al., 2010) and feeder- and serum-free culture conditions were described with transforming growth factor beta (TGF-β) as a key requirement (Amit et al., 2004). Despite the differences between these two model systems, hESCs have also become a useful tool for stem cell research and drug discovery.

More recently, pluripotent stem cell lines have been established from post-implantation mouse embryos, termed epiblast stem cells (EpiSCs). Interestingly, despite expression of typical mESC markers such as SSEA-1, Oct4 and Nanog, these cells are epigenetically distinct and exhibit characteristics similar to hESCs, including flattened colony morphology, incompetence for single cell passaging and dependence on activin/Nodal signalling (Brons et al., 2007; Tesar et al., 2007). Two states of pluripotency have since been designated by Nichols and Smith (2009) to describe these distinct cell types with the ESC-like traits of ICM derived cells termed the "naïve" state and epiblast derived EpiSC-like cells existing in a "primed" state. The naïve state represents cells that possess true pluripotency. That is, they express a full panel of pluripotency markers, both X chromosomes remain active and they can contribute to all germ lineages in blastocyst chimeras. On the other hand, primed cells express core pluripotency factors including Oct4, Sox2 and Nanog but lack expression of many other pluripotency markers, may have limited and potentially biased capacity for differentiation, contain an inactive X chromosome and do not contribute to chimeras in the conventional way without additional manipulation (Nichols and Smith, 2009). However, it has been reported that EpiSCs can contribute to chimera formation either in a permissive environment, with genetic manipulation or by selection of a permissive subpopulation of cells (Huang et al., 2012; Han et al., 2010; Ohtsuka et al., 2012).

1.1.2.1 2i can support pluripotent stem cell derivation and maintenance

One of the most famous pluripotent culture conditions comes from the lab of Austin Smith and is named '2i' after its composition of two inhibitors. Ying et al. (2008) coined the term "ground state pluripotency" as the intrinsic ability for ESC to self-renew regardless of external influence. This was demonstrated by culturing ESCs in a combination of three small molecule inhibitors (3i) CHIR99021 (CHIR), SU5402 and PD184352, which blocked glycogen synthase kinase 3 (GSK3), FGF receptor and ERK signalling, respectively. This defined culture condition maintained ESCs in a self-renewing, pluripotent state in the absence of serum or BMP, in a similar fashion to undefined conventional ESC media, and germline-competent chimeras were obtained even from the non-permissive CBA mouse strain, which has been reported to be difficult to isolate ESCs from in conventional culture conditions (Buehr and Smith, 2003). A few months later, Smith and colleagues reported that manipulation of intracellular signalling by addition of PD0325901 (PD03), a potent ERK inhibitor, and CHIR (termed "2i" conditions) in addition to LIF could drive both neural stem cell (NSC) and MEF derived pre-iPSCs to a fully reprogrammed state as indicated by activation of endogenous Oct4 expression, conversion of X_aX_i to X_aX_a and contribution to chimeras (Silva et al., 2008a).

Previously, only a certain number of mouse strains were deemed "permissive" for ES cell derivation in conventional mESC culture conditions containing LIF and either serum or the growth factor BMP (Buehr et al., 2008), including 129, BALB/C and C57BL/6 (Hanna et al., 2010b). Attempts to establish ESCs from "non-permissive" species or mouse strains, for example, rats and nonobese diabetic (NOD) mice, in mESC conditions has previously been unsuccessful and only EpiSCs could be derived from these animals, in the presence of activin and FGF (Brons et al., 2007; Buehr et al., 2008). This was resolved by Nichols et al. (2009) who derived ESCs from NOD mice using 2i culture conditions and more recently a report by Czechanski et al. (2014) outlined a method by which ESCs could be derived by any mouse strain, including those from which ESCs failed

to be derived. Interestingly, it was subsequently determined that the response to LIF signalling underscored the permissibility of a mouse strains to give rise to mESCs and in permissible strains the JAK-Stat3 pathway was preferentially activated, whereas the MAP kinase pathway was hyperactivated in response to LIF in nonpermissive strains (Ohtsuka and Niwa, 2015).

The increasing availability of pluripotent cell lines, particularly human derived lines, presented great promise for gaining mechanistic insight into developmental questions as well as advancing medical research and drug discovery. However, a major obstacle for both research and clinical use of hESCs is their controversial source since hESCs are usually derived from healthy surplus embryos resulting from fertility treatments. Consequently there is a distinct lack of disease models without the need for further genetic manipulation of the cells. However, embryos obtained from preimplantation genetic diagnosis (PGD) screening represent an alternative source of hESCs with somewhat fewer ethical difficulties that already carry mutations for specific diseases. These hESCs are used for modelling human diseases such as Huntington's disease, cystic fibrosis, Duchenne and Becker muscular dystrophies and fragile X syndrome, among others (Stephenson et al., 2009). In addition to issues surrounding ethics, there is the matter of potential recipient rejection of ESC derived tissue. Nonetheless, progress has been seen in studies of hESCs to treat animal disease models including Parkinson's disease (Ben-Hur et al., 2004) and retinal disease (Lamba et al., 2006). Some years ago the U.S. Food and Drug Administration (FDA) approved the first clinical trial using hESCs to treat patients with spinal cord injury (Couzin, 2009).

1.1.3 Induced pluripotent stem cells

Researchers have long been interested in developing strategies to generate pluripotent cells from differentiated cells which would allow generation of patient specific derived stem cells for clinical use or disease modelling. Methods to generate

such pluripotent cells have included cell fusion of somatic cells with pluripotent cells such as ESCs, SCNT and exposure of somatic cells to ESC extracts (Hakelien et al., 2002; Landsverk et al., 2002; Cowan et al., 2005; Do and Schöler, 2004; Rathbone et al., 2013). Shinya Yamanaka hypothesized that the factors that play important roles in the maintenance of stem cell identity would also play key roles in inducing pluripotency in somatic cells. This was the basis on which Yamanaka and colleagues performed their pioneering experiments leading to successful reprogramming of adult somatic cells to pluripotent cells (Takahashi and Yamanaka, 2006). They used retroviral transfection of four reprogramming factors; c-Myc, Klf4, Sox2 and Oct3/4 (herein Oct4), to reprogram mouse embryonic and adult fibroblasts to a pluripotent state. By using the embryonic specific (but dispensable) gene Fbx15 as a marker of reprogrammed cells, the first attempt to create pluripotent cells generated cells which had ESC-like morphology, Nanog gene expression and were pluripotent by teratoma formation but revealed poorly up-regulated pluripotency gene expression, incomplete DNA demethylation at the Oct4 and Nanog promoters and failed to efficiently silence transgene expression. These cells were termed induced pluripotent stem cells, or iPSCs, as they demonstrated pluripotency despite not being fully reprogrammed. The following year, the same group reported an improvement in the technology, generating germ-line competent iPSCs by selecting for Nanog expression, that exhibited ESC-like DNA methylation patterns at pluripotency and imprinted loci (Okita et al., 2007).

iPS cells share hallmark features of mouse ES cells including pluripotency gene expression, DNA methylation state, activation of both X chromosomes (X_aX_a) in female lines and contribution to germ-line competent chimeras upon injection into blastocysts (Boland et al., 2009; Okita et al., 2007; Wernig et al., 2007). In addition, all-iPSC mice have been generated by tetraploid complementation (Zhao et al., 2009; Kang et al., 2009), suggesting that iPSCs have the ability to re-capitulate normal development compared to ESCs. However, more recently, higher rates of tumorigenesis were reported in all-iPSC mice produced using cells generated with oncogenes (Tong et al., 2011), highlighting the risks involved with using these cells

for clinical medicine. Subsequently, several groups have reported the generation of iPSC lines by using various other transcription factors including Nanog, Lin28, Esrrb and Nr5a2, in both human and mouse systems (Yu et al., 2007; Feng et al., 2009; Heng et al., 2010), with other approaches using synthetic modified mRNA (Warren et al., 2010) or proteins (Kim et al., 2009; Zhang et al., 2012) to generate reprogrammed cells.

It has been reported that there is a temporal requirement for exogenous factor expression during reprogramming. Using a doxycycline (dox)-inducible reprogramming system, Stadtfeld et al. (2008b) reported that exogenous factor dependency is reduced/abolished around day 8 of the reprogramming process; down regulation of transgene expression by removing dox at day 7 or earlier prevented the progression of colonies to a fully reprogrammed iPSC-like state. However, if dox was removed on day 8 or later, some of the colonies present in the culture progressed to fully reprogrammed iPSCs. In addition, it was observed that maintaining transgene expression for 10 days increased the number of colonies that successfully underwent reprogramming. Consequently, two iPSC populations could be identified and defined as 'stable' or 'unstable', depending on their independence or reliance on dox-mediated transgene expression in order to maintain an undifferentiated, ES-like phenotype (Hanna et al., 2009a; Hanna et al., 2010a).

Following the establishment of reprogramming technology by Yamanaka and colleagues, Professor Sir Ian Wilmut made the announcement of his decision not to pursue his work into nuclear transfer as a method of generating human cloned embryos, instead favouring the reprogramming method of generating iPSCs as a more efficient and socially acceptable technique to produce human pluripotent stem cells. Remarkably, iPSC biology saw rapid translation to the clinic when in 2013 plans for the first clinical trial using human iPSCs was approved by the Japanese government to generate retinal pigmented epithelial (RPE) cells for transplantation into patients with macular degeneration. The study, led by Dr. Masayo Takahashi from the RIKEN Center for Developmental Biology, carried out the first transplantations in September 2014, a mere 7 years after the first report on

iPSCs was published. This initial trial is intended to demonstrate safety of iPSC transplantation, rather than the expectation that it will improve the condition, but if successful subsequent stages of the clinical study will endeavour to restore eyesight in patients. Nonetheless this study will pave the way for clinical application of iPSCs, revolutionising regenerative medicine.

1.2 Methods of iPSC generation

iPSC generation is famously inefficient and somatic cells from different tissues and species reprogram differently (Gonzalez et al., 2011). The majority of published studies use fibroblasts for reprogramming experiments, however, mouse embryonic fibroblasts (MEFs) reprogram significantly faster than human dermal fibroblasts (hDFs), taking around 10 or 20 days, respectively. Time is not the only variant when it comes to reprogramming capacities but moreover, the requirement of factors can vary drastically from one cell type to another. For example, CD133+ cells from cord blood and human fibroblasts can be reprogrammed with only Oct4 and Sox2 (Meng et al., 2012; Huangfu et al., 2008b), and MEFs can be reprogrammed with only Oct4 and Klf4 in addition to small molecules, albeit all with lower efficiencies than four factor reprogramming. Since reprogramming technology was first reported a number of different systems have been described which generate iPSCs with varying efficiencies. Some of these reprogramming methods will be described here.

1.2.1 Retrovirus and lentivirus

The first reports of iPSC generation from both mouse and human cells used retroviruses to deliver the reprogramming factors to (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Okita et al., 2007; Wernig et al., 2007; Maherali et al., 2007; Aoi et al., 2008). It is widely know that retroviruses are silenced in ESCs (Macfarlan et al., 2011; Wolf and Goff, 2007; Rival-Gervier et al., 2013; Wolf and Goff, 2009; Hotta and Ellis, 2008) and the use of retroviral based vectors to deliver the reprogramming factors to cells utilises this distinct feature since successful reprogramming relies on cells becoming independent of exogenous factors and establishing endogenous gene expression networks in order to become fully mature iPSCs. A consideration of retroviral-mediated reprogramming is that it requires

dividing cells for transduction of the virus, which was abrogated in later studies using lentiviruses to deliver the reprogramming factors. Although lentiviruses are a subclass of retrovirus, they can infect both dividing and non-dividing cells (Vodicka, 2001) and many groups utilised this virus to successfully reprogram both fetal and adult human cells and mouse cells (Yu et al., 2007; Zhao et al., 2008; Mali et al., 2008; Stadtfeld et al., 2008b). A further improvement was made with the introduction of lentiviral polycistronic vectors capable of expressing the reprogramming factors from a single vector, with translation of individual proteins facilitated by self-cleaving 2A peptide separation of each gene (Shao et al., 2009; Carey et al., 2009; Sommer et al., 2009). This single vector system greatly reduced the extent of potential insertional mutagenesis upon viral integration and made four-factor expression more homogeneous.

1.2.2 Adenovirus reprogramming

Stadtfeld et al. (2008c) reported successful reprogramming of mouse fibroblast and liver cells using adenovirus based vectors to transiently express the reprogramming factors. Because the adenovirus does not integrate into the genome, the iPSCs generated were exogenous factor free, and proved that reprogramming does not require insertional mutagenesis. However, almost a quarter of the iPSC lines analysed were tetraploid, suggesting that adenoviral-mediated reprogramming may induce a cell fusion event rendering cells karyotypically abnormal. Zhou and Freed (2009) later achieved reprogramming of human embryonic fibroblasts, with repeated infection of adenoviral reprogramming factors, after 25-30 days. However, similarly to the former report, despite a high multiplicity of infection (MOI) the efficiency of iPSC generation was extremely low. This was likely due to rapid clearance of the virus from cells with each generation, making this method technically challenging and unlikely that patient-specific iPSCs will be generated in this way for clinical application.

1.2.3 Sendai virus reprogramming

Sendai virus (SeV) is a negative-sense single stranded RNA virus, which replicates in the cytoplasm with no DNA intermediate. Consequently there is no genome integration. This makes SeV a very attractive alternative to other viruses typically used for delivery of reprogramming factors to cells, as reported by Fusaki et al. (2009). This group successfully and efficiently generated human iPSCs at rates of up to 1%, which faithfully up-regulated pluripotency genes and exhibited DNA methylation patterns similar to human embryonal carcinoma cells. Importantly, complete depletion of SeV was confirmed with increasing time and passage number. This method showed such promise that it is now available as the commercially available CytoTune™ reprogramming kit sold by Life Technologies.

1.2.4 *piggyBac* transposon-mediated reprogramming

Several groups reported a virus-free, transgene integration-free method of reprogramming using the *piggyBac* (*PB*) transposon, a mobile genetic element, to deliver the reprogramming factors to cells (Woltjen et al., 2009; Kaji et al., 2009; Yusa et al., 2009). This is an attractive system for use in genome engineering since the transposon requires only two simple conditions to allow for insertion or excision of DNA; transgenes must be flanked by inverted terminal repeats (ITRs) and expression of *PB* transposase is required. *PB* transposase recognises the ITRs catalysing excision of the DNA, which is subsequently inserted into TTAA-specific target sites (Fraser et al., 1996). Woltjen et al. (2009) demonstrated that both mouse and human embryonic fibroblasts could be reprogrammed using doxycycline (dox)-inducible *PB* reprogramming factors and following transposase-mediated excision of the transgenes, 90% of clones achieved seamless excision of the factors and were restored to wild type. Importantly, endogenous expression of the four factors was maintained following transgene removal. Kaji et al. (2009) demonstrated that MEF

could be reprogrammed by transfection of a single polycistronic vector encoding the four reprogramming factors linked by 2A peptides with an efficiency of 2.5% (based on stable transfection efficiency and number of Nanog positive colonies). The presence of two flanking *loxP* sites allowed for transgene removal leaving behind only a small vector footprint, providing an improvement on viral integration methods of reprogramming. This system was then modified, incorporating the *PB* transposon to reprogram human embryonic fibroblasts, providing further evidence that integration-free human iPSCs can be generated. This technology was further improved by Yusa et al. (2009) who employed negative selection to identify transposon-free iPSC clones generated using a *PB* polycistronic vector carrying four or five factors.

1.2.5 Protein-mediated reprogramming

Zhou et al. (2009) reported generation of iPSCs from MEF by repeated transduction of recombinant proteins that contained a fused arginine-tag, allowing the proteins to pass across the cell membrane. The efficiency of reprogramming in this instance was extremely low, with only a few Oct4-GFP+ colonies being obtained per $5x10^4$ cells, and addition of VPA was required for activation of Oct4-GFP, however the successful colonies were shown to be truly pluripotent exhibiting DNA methylation status similar to ESCs at endogenous Oct4 and Nanog promoters, pluripotent gene expression and contribution to tissues of the three germ layers both in vitro and in vivo. The laborious requirement for several applications of the proteins was negated by Cho et al. (2010) who reprogrammed adult mouse fibroblasts with just a single dose of ESC-derived extracted proteins. After 4-7 days only a few colonies were observed and upon subsequent passaging onto feeders numerous secondary colonies were generated. Subsequently, Kim et al. (2009) used direct protein transduction to generate human iPSCs in the absence of small molecules but this system was also reported to be less efficient, by 10-fold, and

taking twice as long to produce colonies as traditional virus-mediated reprogramming. The common denominator of all of these studies is the drastic drop in reprogramming efficiency achieved using a protein-mediated method, nonetheless, these findings demonstrate that iPSCs can be produced without the need for DNA introduction and genomic manipulation of cells, which is attractive in the clinical context.

1.2.6 mRNA-mediated reprogramming

In 2010, Warren et al. and others reported the use of synthetic mRNA as an integration free, efficient method of reprogramming (Warren et al., 2010; Yakubov et al., 2010). Introduction of RNA into cells can trigger an immune response mediated by single stranded RNA (ssRNA) sensors including RIG-I, and PKR, a repressor of protein translation (Pichlmair et al., 2006; Hornung et al., 2006; Nallagatla et al., 2008). For this reason, Warren et al. (2010) used in vitro transcribed mRNAs containing substitutions of cytidine and uridine ribonucleoside bases with naturally occurring modified nucleosides 5-methylcytidine (5mC) and pseudouridine (psi), in addition to B18R, an interferon inhibitor, to avoid invoking the innate immune response in cells, as seen with the use of viral delivery of reprogramming factors. In the most optimum conditions, requiring five reprogramming factors and low oxygen culture, a reprogramming efficiency of over 4% was observed, compared with 0.04% using retrovirus in this study and with the best viral methods reporting around 1% efficiency (Fusaki et al., 2009). A range of fetal, neonatal and adult cells were reprogrammed demonstrating the wide applicability of this method to different tissues, and furthermore, this approach lacks the requirement to screen clones for viral clearance or vector integration making this an attractive reprogramming system in the context of regenerative medicine. However, it involved a labour intensive strategy of daily transfection of mRNA over a course of more than two weeks, detracting from the likelihood of practical application in a

clinical context. The same group later reported an improved process and were able to generate hESC-like colonies with only 6 days of transfections (Warren et al., 2012), and with further refinement this could be a viable and attractive reprogramming strategy with the increasing demand for patient-specific iPSC generation in both research and clinical contexts.

1.2.7 Episomal-mediated reprogramming

Yu et al. (2009) demonstrated an alternative method to generate transgene free iPSCs, taking advantage of the extrachromosomal replication property of Epstein-Barr virus derived oriP/EBNA1 episomal vectors. These vectors are reported to replicate only once per cell cycle and remain non-integrating in 99% of transfected cells. Importantly, upon removal of drug selection, the vector is cleared from cells at a rate of approximately 5% per cell division making this a desirable tool for generation of transgene-free iPSCs. In this study human foreskin fibroblasts required transfection of three episomal vectors encoding seven genes in total, OCT4, SOX2, NANOG, KLF4, SV40 large T (SV40LT) and LIN28, to generate iPSCs at an extremely low efficiency of up to 0.0006%. Subsequently, the same group achieved episomal reprogramming of bone marrow derived mononuclear cells with a greater efficiency 0.035% (Hu et al., 2011). Other modifications to the episomal reprogramming method include successful reprogramming of cord blood mononuclear cells using only a single polycistronic vector to express five factors (Chou et al., 2011) and substitution of C-MYC and NANOG with nontransforming L-MYC in addition to an shRNA targeting p53 has enhanced reprogramming efficiency of human dermal fibroblasts to 0.2% (Okita et al., 2011).

1.2.8 Other methods of reprogramming

1.2.8.1 microRNA-mediated reprogramming

Impressively, reprogramming induced by expression of only micro-RNA 302-367 (miR-302/367) was reported to be more than 100 fold more efficient than traditional four factor reprogramming and Oct4-GFP+ colonies were evident as early as 6 or 7 days post lentiviral transduction of miR-302/367 (Anokye-Danso et al., 2011). This system was reported to be extremely efficient, with almost 80% of colonies expressing Oct4-GFP by day 8, however, certain prerequisites were necessary including low Hdac2 levels, which could be induced with valproic acid (VPA).

Shortly after this report, another group found that direct transfection of mature double stranded miRNAs could reprogram mouse adipose stromal cells (mASCs) (Miyoshi et al., 2011). Candidate miRNAs were determined by comparison of miRNA expression in mouse ESCs and iPSCs with mASCs. miRNAs that were expressed more than 2-fold higher in the ESCs and iPSCs, including members of the mir-200c, mir-302s and mir-369s families, were used in reprogramming experiments. Nanog-GFP+ colonies were observed 15 days after transfection, however the efficiency was extremely low, with only a few colonies generated per $5x10^4$ transfected cells, at an efficiency similar to the original retrovirus-mediated reprogramming report of approximately 0.01% (Takahashi and Yamanaka, 2006). Human ASCs and dermal fibroblasts were also successfully reprogrammed using this method, albeit at an even lower efficiency than observed in the mouse cells, however, the clones generated appeared to be bona fide iPSCs by the standard accepted characterisation of ESC-like morphology, pluripotency marker expression, and teratoma formation (Miyoshi et al., 2011).

Interestingly, these studies have been the only reports of such miRNA-only reprogramming to the best of my knowledge. Other studies have failed to corroborate these findings including a report by Lu et al. (2012) that found the introduction of miR-302/367 was insufficient to generate iPSCs from MEFs, although

it was found that in addition to the traditional reprogramming factors, this and other miRNAs promoted reprogramming. Notably, these studies used two different methods of miRNA expression, with lentiviral and *PB* vectors used, respectively, which could explain the differences in reprogramming success. A second study was unable to reprogram human ASCs using only miRNA-302, but the authors determined that this could be down to the differences in methodology, starting somatic cell type or the absence of miRNA-367 (Hu et al., 2013) which was reported to be key in activating Oct4 by Anokye-Danso et al. (2011). These inconsistencies highlight the need for reproducibility not only due to the extent of variation between both reprogramming systems and individual laboratories but to ensure that high standards are maintained in the quality of peer-reviewed journal publications in order to prevent misleading or false data from being published as far as possible.

Irrespective of the uncorroborated claims reported by Anokye-Danso *et al* (2008), addition of ESC-specific cell cycle-regulating (ESCC) micro RNAs to conventional Yamanaka 4 factor mediated reprogramming has been reported as having a substantial positive effect on reprogramming efficiency in both mouse and human cells. Subramanyam *et al* (2011) added the human orthologues of ESCCs miR-302 (hsa-miR-302b) and miR-291 (hsa-miR-367) to 3- and 4-factor reprogramming and observed a 10 to 15-fold increase in colony number. Soon after, Liao *et al* (2011) reported up to 100-fold enhancement of reprogramming efficiency with the addition of ESCC miRNA clusters miR-106a-363 and miR302-367 to 3- or 4-factor mediated reprogramming of MEFs. These authors proposed that the mechanism underlying this enhancement was through regulation of cell cycle, mesenchymal-to-epithelial transition and epigenetic regulators.

1.2.8.2 Stimulis-triggered factor free reprogramming

Recently, a remarkable phenomenon was reported by Obokata et al. (2014b), who observed that splenic CD45+ cells could revert to a pluripotent state when exposed to low-PH treatment in the absence of any exogenous factors. These lineage

committed somatic cells reprogrammed to stimulus-triggered acquisition of pluripotency (STAP) cells exhibiting hallmarks of pluripotency including pluripotency marker expression, ESC-like DNA methylation signatures at Oct4 and Nanog loci, and contribution to germ-line competent chimeras. However, STAP cells differed from ESCs as they could not self-renew indefinitely in traditional LIF + FBS or 2i containing ESC culture conditions, single-cell dissociation did not give rise to robust ESC-like colonies, Esrrb expression remained low and H3K27me3 dense foci was observed in 40% of female STAP cells indicative of an inactivated X chromosome. These data indicated that STAP cells were a different type of pluripotent cell to ES or iPS cells (Obokata et al., 2014b). Interestingly, STAP cells could be converted to an ESC-like state when cultured with LIF and the peptide hormone adrenocorticotropic hormone (ACTH), a known facilitator of ESC clonal expansion (Ogawa et al., 2004). Further investigation revealed that Oct4-GFP+ STAP cells could be derived from an array of somatic tissue types including brain, lung, muscle, liver and fibroblast cells, at varying efficiencies, and even more remarkably were found to contribute to both embryonic and placental tissues *in vivo* when treated with Fgf4 (Obokata et al., 2014a).

The initial excitement of this novel discovery has since been met with a slew of controversy and criticism, largely triggered by the lack of reproducibility by a number of labs around the world, in addition to some of the published data appearing to be fraudulently reported. Although the authors continued to stand by their findings for a time, the articles were subsequently retracted by Nature in June 2014, with even Obokata being unable to reproduce her own experiment. Accordingly, this method of reprogramming is not likely to be pursued until the method can be independently verified, if indeed the phenomenon of STAP is true.

1.2.8.3 Reprogramming with Mbd3 knock down

Recently, Rais et al. (2013) reported successful elimination of the stochastic heterogeneity of reprogramming by achieving nearly 100% reprogramming efficiency obtained through a "deterministic and synchronous" process. Methyl-

CpG-binding domain protein 3 (Mbd3) is a core member of the nucleosome remodelling and deacetylation (NuRD) complex and mediates repression of genes. Mbd3 depletion in $Mbd3^{fl/-}$ epiblast stem cells (EpiSCs) resulted in more than 90% reversion to ES cells. Astonishingly, the authors went on to describe achievement of 100% reprogramming efficiency by day 8 of reprogramming, as characterised by activation of Oct4-GFP+, when doxycycline inducible four factor, $Mbd3^{fl/-}$ MEFs (Mbd3 depleted) were reprogrammed as single-cell cultures in 2i/LIF conditions. Furthermore, similarly high efficiencies were achieved for reprogramming of a number of other terminally differentiated or progenitor cell types including haematopoietic stem cells, pro-B cells, monocytes and neural precursor cells. This is somewhat consistent with another report that knock down of Mbd3 enhances reprogramming (Luo et al., 2013) (albeit at far lower efficiency than the study by Rais et al. (2013)) and that overexpression of Mbd3 inhibits reprogramming by repressing ESC-specific genes including Oct4 and Nanog However, conflicting results of the Rais et al. (2013) study were later reported by dos Santos et al. (2014) who determined that NuRD function was required for EpiSC reprogramming and in fact up-regulation of Mbd3 facilitated conversion to iPSCs. Furthermore, Mbd3 overexpression was found to increase efficiency in certain reprogramming contexts, including MEF-derived pre-iPSC, in conjunction with Nanog expression. Of note, the $Mbd3^{fl/-}$ cells used in the first study were derived in the lab of Dr. Brian Hendrich, one of the authors of the conflicting paper, yet Rais et al. (2013) reports that Mbd3 levels in these cells is hypomorphic expressing around 20% of wild type levels, whereas dos Santos et al. (2014) find nearly wild type levels expressed. In addition, Onder et al. (2012) found no positive effect on reprogramming with shMbd3. These groups could not confirm the results published by the Hanna group and the findings remain highly controversial.

1.2.9 Secondary reprogramming

Tetracycline responsive promoters have been used for years as a tool for regulated gene expression in mammalian cells (Gossen and Bujard, 1992; Gossen et al., 1995). This system operates through binding of the reverse tetracycline transactivator (rtTA) protein to the tetO operator sequence only in the presence of tetracycline or one of its derivatives, for example doxycycline (dox), activating downstream gene expression. Reprogramming factors whose expression is controlled by the tetO promoter provides a tool by which exogenous factor expression can be initiated by introduction of dox to the culture medium. So-called dox-inducible reprogramming systems have been successfully reported to generate iPSCs in virus- and PB-mediated contexts (Brambrink et al., 2008; Stadtfeld et al., 2008a; Stadtfeld et al., 2008b; Kaji et al., 2009; Woltjen et al., 2009; Hou et al., 2013).

Utilising this technique Wernig et al. (2008) used rtTA-expressing MEFs to generate iPSCs carrying dox-inducible lentiviral reprogramming vectors. These primary iPSCs were used to generate chimeric mice whose MEFs were then isolated to provide somatic cells carrying identical reprogramming factor transgene integrations, controlled by doxycycline inducible promoters. This enabled a genetically homogeneous population of cells to be reprogrammed following treatment with dox to initiate re-expression of reprogramming factors and was termed "secondary reprogramming". Secondary iPSCs from MEFs were generated at a 50-fold increased efficiency of up to 4% using this system, and in addition, many other cell types were successfully reprogrammed including keratinocytes, muscles, neural progenitors and mesenchymal stem cells. This system was subsequently described for human somatic cells, differentiated via teratoma formation, with up to 2-3% reprogramming efficiency being achieved. These studies also provided support for the earlier report in mouse that there is a temporal requirement for transgene expression and variations in transgene re-expression levels correlates with reprogramming efficiency (Hockemeyer et al., 2008; Maherali et al., 2008). Many groups have since utilised secondary reprogramming systems as

a more homogeneous and efficient means to study reprogramming kinetics (Hanna et al., 2010a; Koche et al., 2011; Efe et al., 2011; O'Malley et al., 2013).

1.3 Culture conditions for reprogramming

1.3.1 The role of small molecules in reprogramming

Reprogramming is renowned for being inefficient with initial estimates of reprogramming efficiency being as low as 0.01%. The reason for such low efficiency of conversion of somatic cells to iPS cells is unclear, however, regulation of gene expression in ES cells is tightly controlled and even modest changes in gene expression can result in altered phenotype. For example, in ES cells, Oct4 levels should be maintained within a small range of endogenous levels to maintain an undifferentiated state (Niwa et al., 2000). Therefore it is plausible that the exact levels of each of the four factors required lies within a tight range and that only in a small number of cells does the right combination of expression levels occur enabling reprogramming of this subset. Consequently a number of small molecules have been implicated in supporting the reprogramming process. Many small molecules and other factors known to enhance reprogramming are summarized in Table 1.1, some of which are discussed below.

Factor/Small Molecule	Target/Pathway	Effect on reprogramming	Reprogramming efficiency enhancement	Species	Reference
Valproic acid (VPA)	HDAC inhibitor	Enhances reprogramming efficiency of MEF and enables reprogramming of human fibroblasts with only OCT4 and SOX2	>100-fold	Mouse/ Human	Huangfu et al. (2008a) Huangfu et al. (2008b)
Suberoylanilide hydroxamic acid (SAHA)		Enhances reprogramming efficiency of MEF	10-fold	Mouse	Huangfu et al. (2008a)
Trichostatin A (TSA)		Enhances reprogramming efficiency of MEF	10-fold	Mouse	Huangfu et al. (2008a)
Sodium butyrate (NaB)		Enhances reprogramming efficiency of human fetal or adult fibroblasts. Enhanced reprogramming also observed in the absence of KLF4 and MYC transgenes.	50 to 100-fold	Mouse/ Human	Mali et al. (2010)
5-azazcytidine (5-aza)	DNMT inhibitor	Enhances reprogramming efficiency of MEF and promotes conversion of pre-iPSC to fully reprogrammed iPSC	4-fold	Mouse	Mikkelsen et al. (2008)

PD0325901	Selective MEK/ERK inhibitor	Facilitates rapid and efficient generation of fully reprogrammed human iPSCs	200-fold (with Thiazovivin & SB431542)	Human	Lin et al. (2009)
Thiazovivin	Rho-associated protein kinase inhibitor	Facilitates rapid and efficient generation of fully reprogrammed human iPSCs	200-fold (with PD0325901 & SB431542)	Human	Lin et al. (2009)
SB431542	ALK4/5/7 inhibitor	Facilitates rapid and efficient generation of fully reprogrammed human iPSCs	200-fold (with PD0325901 & Thiazovivin)	Human	Lin et al. (2009)
E-616452		Replaces Sox2 in reprogramming of MEF	30-fold	Mouse	Ichida et al. (2009) Maherali and Hochedlinger (2009)
Ascorbic acid	Facilitates histone demethylation via the JmjC-domain-containing histone demethylase (JHDM) family	Enhances reprogramming efficiency and facilitates efficient conversion of mouse pre-iPSC to fully reprogrammed iPSCs	150-fold	Mouse/ Human	Esteban et al. (2010) Chen et al. (2013)
CHIR99021	GSK3-β inhibitor	Enhances reprogramming efficiency of MEF, blood progenitors and refractory cell populations	5-fold alone or up to 50-fold with ascorbic acid	Mouse	Bar-Nur et al. (2014)
p53 siRNA	p53	Enhances efficiency and kinetics of reprogramming	40-fold (mouse), 120-fold alone or >250-fold with UTF1 (human)	Mouse/ Human	(Zhao et al., 2008) Utikal et al. (2009)
UTF1	ESC-specific transcription factor	Enhances reprogramming efficiency of MEF	12-fold alone, or >250-fold with p53 siRNA	Human	Zhao et al. (2008)
miR302/367	ESC-specific cell cycle-regulating (ESCC) micro RNA cluster with abundant targets playing a role in cell cycle and epigenetic regulation and EMT	Enhances reprogramming efficiency in the absence and presence of reprogramming factors	100-fold (without exogenous factor expression), 30-fold and100-fold with 4 and 3 factor expression, respectively)	Mouse	Anokye-Danso et al. (2011) Liao et al. (2011)
hsa-miR-302b/ hsa-miR-372		Enhances reprogramming efficiency	15-fold (with 3 or 4 factor expression)	Human	Subramanyam et al. (2011)

Table 1.1. List of small molecules and factors reported to have a positive effect on reprogramming.

Two of the most commonly reported small molecules used in reprogramming are the DNA methyltransferase inhibitor 5'-azacytidine (5'-azaC) and the histone deacetylase (HDAC) inhibitor valproic acid (VPA). Huangfu et al. (2008a) showed that both of these drugs have been shown to increase reprogramming efficiency by tenfold and a hundred fold, respectively, with VPA also capable of replacing c-Myc during reprogramming. Some studies focusing on unstable or "non-permissive" cell lines have identified other small molecule drugs that aid reprogramming or replace expression of some or all of the ectopic reprogramming factors, including Wnt3a, CHIR99021 (CH, a GSK3 inhibitor), Kenpaullone (KP, a GSK3β and CDK1/cyclin B inhibitor) and 2i (CH and PD184352, an inhibitor of the ERK cascade) (Ying et al., 2008; Buecker et al., 2010; Hanna et al., 2009a; Shi et al., 2008; Hanna et al., 2010a).

Reports of unstable or so called "metastable" iPSCs have been reported recently in both mouse and human systems and these in addition to 'partially reprogrammed' lines that have been established based on morphology and/or reporter expression during intermediate stages of the reprogramming process (Sridharan et al., 2009) have been useful for investigating the obstacles blocking complete reprogramming. Mikkelsen et al. (2008) previously reported the generation of "partially reprogrammed" cell lines which show up-regulation of only a limited number of stem cell associated genes but incomplete down-regulation of lineage specific factors in addition to DNA hypermethylation at a number of pluripotency associated loci including Nanog, Dppa5 and Rex1. Interestingly, the addition of the DNA methyltrasferase inhibitor 5'-azaC rapidly drove these cells to a stable fully reprogrammed iPS state. In addition, generation of hiPSCs named "hLR5" that can be maintained in conventional mES cell media conditions has been reported, although continuous expression of five exogenous reprogramming factors (Oct4, Sox2, Klf4, c-Myc and Nanog) is required for their maintenance (Buecker et al., 2010). However, these cells failed to re-activate expression of the endogenous reprogramming factors and interestingly, the cells are intrinsically "metastable" and revert to a conventional hiPSC phenotype once ectopic factor expression is

removed. Finally, using doxycycline (dox)-inducible lentiviral vectors, Hanna et al. (2009a) were able to establish pluripotent cell lines (ESCs and iPSCs) from NOD mice (previously considered to be non-permissive for derivation of ESCs) with constitutive expression of Klf4 or c-Myc. Small molecule inhibitors could replace the requirement for Klf4 or c-Myc, including Wnt3a, CH, KP (which has been reported to replace Klf4 during reprogramming) and 2i. The same group later reported the stabilization of Dox-dependant "naïve" human ESCs only in the presence of PD/CH/LIF and with the constitutive expression of Klf4 and Oct4 or Klf4 and Klf2 (Hanna et al., 2010a). These reports suggest that there is an intermediate stage during reprogramming whereby cells are able to maintain a reprogrammed state under certain stabilizing conditions, however, due to incomplete transcriptional remodelling during the reprogramming process this state can only be artificially maintained by additional factors and collapses upon ectopic factor/drug removal.

The first reports of iPSC generation cultured reprogramming cells on a layer of inactivated feeders in conventional ESC medium containing LIF and serum (Takahashi and Yamanaka, 2006; Okita et al., 2007), although now iPSCs are typically grown on gelatin-coated vessels. A number of feeder-free and xeno-free media are now available for maintenance of iPSCs including Knock-Out DMEM with Knock-Out Serum Replacement (KOSR), N2B27, ESF7 and others (Ying et al., 2003; Cheng et al., 2004; Furue et al., 2005; Yamasaki et al., 2013) and over the years a number of small molecules have been reported to enhance reprogramming when added to basal culture media. A few of these will be discussed here.

1.3.1.1 iCD1 chemically defined medium promotes iPSC generation

Chen et al. (2011a) described a chemically defined, serum-free medium comprising over 74 components, which was reported to support highly efficient three factor (OKS) reprogramming of MEFs at an efficiency of 10% by day 8, with few pre-iPSCs generated. Interestingly, c-Myc was found to only increase reprogramming efficiency slightly in the iCD1 culture condition, contrary to other reports, and in addition the authors were able to reprogram with Oct4 alone, albeit

at drastically reduced kinetics and efficiency. Although this medium is defined, it comprises too many components to easily determine those that are responsible for the enhanced reprogramming kinetics. In addition, the medium comprises small molecules known to enhance reprogramming such as vitamin C and CHIR and the later report by Bar-Nur et al. (2014) corroborates these findings, with OSKM + AGi resulting in approximately 10% efficiency by day 7 (see below).

1.3.1.2 AGi promotes homogeneous iPSC generation

More recently, CHIR has been implicated as a component of another inhibitor cocktail found to promote reprogramming. OKSM expression for as little as 48 hours was sufficient to generate iPSCs from granulocyte-macrophage progenitors (GMPs) when treated with a duo of small molecules called AGi, comprising **A**scorbic acid and **G**SK3-β **i**nhibitor CHIR. These iPSCs expressed an Oct4-GFP reporter within 2 days of OKSM expression and Nanog expression was observed 7 days after this (Bar-Nur et al., 2014). Single-cell analysis of GMP reprogramming with AGi revealed almost 100% of clones activated Oct4-GFP after merely 5 days and control conditions (OKSM alone) also reached nearly 100% efficiency, but after 30 days – this is consistent with previous reports that reprogramming is a stochastic process where almost all cells are reprogrammable if given enough time (Hanna et al., 2009b). Interestingly, the authors described that genes from all stages of reprogramming – somatic, transient and pluripotency genes as well as miRNAs – were regulated in the expected manner but with vastly accelerated kinetics through the reprogramming process, providing a possible functional mechanism underlying AGi enhancement and a more synchronous reprogramming culture. Both ascorbic acid (vitamin C) and GSK3-β inhibition have separately been reported to promote generation of iPSCs so it is somewhat unsurprising that treatment of both small molecules exerts an enhancing effect (Li and Rana, 2012).

1.3.1.3 Tgf-β signaling inhibition and reprogramming

In an attempt to uncover small molecules that could replace Sox2 in the reprogramming factor cocktail, an inhibitor of Tgf-β signalling, E-616452, was identified. This chemical, aptly named RepSox, was shown to eliminate the requirement for Sox2 by facilitating Nanog induction in incompletely reprogrammed cells, termed "RepSox responsive" intermediates (Ichida et al., 2009). Microarray data generated by the authors showed a strong up-regulation of Nanog in these cell lines upon treatment with RepSox and in addition, the ability of RepSox to replace Sox2 was abolished with administration of a Nanog shRNA to the cells. Interestingly, RepSox was also shown to replace c-Myc and in fact, inhibition of Tgf-β signalling by alternative chemical or neutralizing antibody measures replicated this result with similar or greater efficiencies of reprogramming than four factors together. Another small molecule screening identified nonsteroidal anti-inflammatory drugs (NSAIDs) and anticancer drugs that can replace c-Myc and/or Sox2 during reprogramming, at least in part by inhibition of MEF-associated gene COX2 (Yang et al., 2011).

Around the same time, Maherali and Hochedlinger (2009) reported similar results using Alk5 inhibitor (Alk5i) treatment in conjunction with a dox-inducible reprogramming system. Here the authors abolished the need for either Sox2 or c-Myc in the presence of Alk5i, and the strongest effect was observed when Alk5i was administered from the earliest stages of reprogramming, suggesting there may be a temporal requirement for suppression of Tgf-β signalling to be effective. In both studies, the effect of Tgf-β signalling inhibition was found to be stronger than that of c-Myc expression on reprogramming efficiency and together these studies highlight the role of the Tgf-β signalling as a barrier to reprogramming.

Taken together, these studies provide evidence that an apparent block to pluripotency resulting from exogenous stimuli can be overcome with simple small molecule inhibition in a defined manner, allowing cells to reach their complete pluripotential, and providing a more defined platform for which to dissect the mechanisms of reprogramming.

1.4 Key events during reprogramming

1.4.1. Early, intermediate and late stages occur during reprogramming

The first obvious signs of reprogramming occur almost immediately with changing somatic cell morphology being the first indicator, concurrent with an increase in cell proliferation (Smith et al., 2010). Down-regulation of somatic cell markers such as Thy1 follows, with subsequent up-regulation of pluripotency related genes such as SSEA-1 (Stadtfeld et al., 2008b). SSEA-1 is a commonly used marker of pluripotent cells in studies of reprogramming, however, incompletely reprogrammed pre-iPSCs express SSEA-1. Moreover pre-iPSCs sorted for SSEA-1 revealed inhibitory DNA methylation at pluripotency loci including Oct4, Nanog, Stella, Dppa5, Utf1 and Rex1 indicating that this marker is merely descriptive rather than indicative of iPSCs (Sridharan et al., 2009; Mikkelsen et al., 2008). Of the four reprogramming factors, c-Myc was found to predominantly function in the early transition of reprogramming to modulate gene expression and in particular has been shown to strongly repress somatic-associated genes such as Thy1. The role for c-Myc as an early effector was supported by observations that expression of c-Myc for only 5 days, with constitutive expression of OSK, gave rise to alkaline phosphatase (AP) positive colonies which only marginally increased with continued expression of c-Myc (Sridharan et al., 2009).

Throughout normal development, cells undergoing differentiation to various cell fates transition from an epithelial cell type to a mesenchymal identity. On the contrary, during the early to intermediate stages of reprogramming of terminally differentiated cells such as fibroblasts, a reversal of this transition is required; a so called mesenchymal to epithelial transition (MET) (Li et al., 2010). This process is indispensible and marks an early transition stage towards pluripotency during reprogramming, marked by distinct morphological changes concurrent with down-regulation of epithelial repressors such as Snail and up-

regulation of epithelial markers such as E-cadherin by Oct4 and Sox2 or Klf4, respectively (Batlle et al., 2000; Cano et al., 2000; Li et al., 2010).

At the cusp of the intermediate to late stages of reprogramming, a subset of progressing cells have been reported as being "partially reprogrammed" (Meissner et al., 2007; Mikkelsen et al., 2008) and these so-called pre-iPSCs are a unique system in which to investigate the transition to the late stage of reprogramming. They are cells which have undergone reprogramming but failed to fully transition to iPSCs and are characterised by ESC-like morphology, often expressing high levels of exogenous factors and somatic genes are usually down-regulated but not all pluripotency genes are up-regulated, particularly those usually expressed at the latter phase of reprogramming (Plath and Lowry, 2011; Mikkelsen et al., 2008; Sridharan et al., 2009). Most commonly Oct4 and Nanog expression is lacking, explained by incomplete resetting of the epigenetic landscape resulting in hypermethylated DNA, particularly within these gene promoters as well as other pluripotency loci including Utf1, Dppa5, Rex1, Gdf3 and Stella (Mikkelsen et al., 2008). This pre-iPSC state can be converted to one of full pluripotency by a number of treatments including administration of vitamin C, overexpression of Nanog, or inhibition of certain pathways with small molecules including 5-aza-cytidine (AZA, inhibiting DNA methyltransferases), E-616452 (RepSox, inhibiting TGF-β signalling), or 2i (inhibiting MAPK and GSK3 signalling) (Mikkelsen et al., 2008; Silva et al., 2008a; Ichida et al., 2009; Esteban et al., 2010; Silva et al., 2009).

Towards the end of reprogramming, ESC-like colonies are clear and the pluripotency gene network is laid down including endogenous expression of reprogramming factors, rendering the iPSCs transgene independent (Plath and Lowry, 2011). During stabilization of the pluripotency transcriptional network, it has been identified that Oct4, Sox2 and Klf4 predominantly target regulators of pluripotency at the later stages of reprogramming, with Oct4 and Sox2 co-targeting genes strongly expressed in ESCs/iPSCs with around half of these targets also occupied by Klf4 (Sridharan et al., 2009). Key regulators of pluripotency, including Nanog and Oct4 have been shown to be ubiquitinated in pluripotent ES cells but not

differentiating ES cells suggesting that protein abundance is tightly and actively regulated in these conditions (Buckley et al., 2012). Endogenous Sox2 activation is a key late event and up-regulation of telomerase and re-activation of the silent X chromosome are characteristic of stable, fully reprogrammed iPSCs (Stadtfeld et al., 2008b; Maherali et al., 2007; Plath and Lowry, 2011; Takahashi and Yamanaka, 2006). Brambrink et al. (2008) described sequential activation of pluripotency markers as identifiers of reprogramming progression with AP expression evident early in reprogramming followed by SSEA-1 up-regulation during the intermediate phase and finally Oct4 and Nanog activation at the late stage. Using a fluorescence activated cell sorting (FACS) approach, O'Malley et al. (2013) used two novel cell surface markers, ICAM-1 and CD44, in combination with a Nanog-GFP reporter, to follow the progression of cells undergoing reprogramming and identified that reprogramming cells followed a systematic route from MEFs to iPSCs. A high-resolution "route map" was established and subpopulation analyses revealed subsets of genes with distinct expression patterns, including transient up-regulation of several genes involved in epidermis/keratinocyte development. This data was supported by several other published data sets, which together provide some evidence that the reprogramming process is more complex than simply a reversion of normal cell development/differentiation.

1.4.2 Exogenous factor expression during reprogramming

Reprogramming is most commonly induced by simultaneous introduction of Oct4, Sox2, Klf4 and c-Myc to somatic cells of interest. However, the timing of addition of reprogramming factors has been reported to influence the progression of the cells towards a pluripotent state. Liu et al. (2013) reported that the number of Oct4-GFP+ colonies observed upon sequential addition of the four factors (OK-M-S) exceeded those resulting from the more traditional practice of simultaneous introduction of factors. Subsequently, the group identified an early epithelial-to-

mesenchymal transition (EMT) indicated by up-regulation of Slug; an event that they suggest boosts reprogramming by giving rise to a more homogeneously mesenchymal, and therefore optimal, fibroblast population capable of efficient mesenchymal-to-epithelial transition (MET). MET analysis revealed that *Snail*, a core mesenchymal gene, is repressed by Sox2, aiding in the collapse of the mesenchymal identity of fibroblasts. Concurrently, the four reprogramming factors work in concert to quell the process of EMT by interfering with TGF-β signalling pathways. Once these initial barriers to reprogramming have been interrupted, expression of *E-Cadherin* through the action of *Klf4* allows cells to gain an epithelial identity before acquiring pluripotency.

Soufi et al. (2012) investigated the initial reprogramming factor binding events that take place in the first 48 hours of reprogramming in human fibroblasts and made several key conclusions. All four factors were found to co-bind to chromatin extensively, across 35% of the genome, upon initial induction of reprogramming compared with only 3% in ES cells. Interestingly, even within this narrow early window, genes associated with progression of reprogramming including GLIS1 and the miR-302/367 were co-targeted by OSKM, in addition to pro-apoptosis genes TP53 and p19 being bound by all four factors or c-Myc alone, respectively, consistent with previous reports in the mouse (Kawamura et al., 2009). A notable finding of this study is that Oct4, Sox2 and Klf4 extensively bind DNaseI resistant domains representing closed chromatin and to a much lesser extent c-Myc. This is indicative of Oct4, Sox2 and Klf4 acting as pioneer factors during reprogramming, allowing subsequent recruitment of other factors to otherwise inaccessible chromatin regions and given that most reprogramming promoting genes lie within closed chromatin this initial pioneer factor activity is essential for progression of cells towards iPSCs. Finally, despite the dynamic activity of pioneer factors, many large genomic regions containing late pluripotency genes were found to remain unbound by OSKM during the initial stage of reprogramming. The authors proposed that this was due to elevated levels of repressive histone mark H3K9me3 in these differentially bound regions (DBRs), and this was supported by

evidence that siRNA-mediated knock down of histone methyltransferases SUV39H1 and SUV39H2 increased Oct4 and Sox2 binding within DBRs and enhanced reprogramming speed and efficiency. The enhancing effect of SUV39H1 inhibition was also reported elsewhere (Onder et al., 2012).

A temporal requirement for reprogramming factor expression was described by Stadtfeld et al. (2008b), using a doxycycline inducible primary reprogramming system to reprogram MEFs carrying an Oct4-GFP reporter. Expression of the four factors was required for at least 7 days before Oct4-GFP+ colonies could be maintained in the absence of doxycycline, with the number of colonies increasing with longer doxycycline treatment. Using a secondary reprogramming system, Hanna et al. (2009b) used pre-B cells carrying a Nanog-GFP reporter and doxycycline inducible reprogramming factors to demonstrate that over time almost all (93%) cells underwent reprogramming to activate expression of Nanog-GFP. Thus the authors contended that reprogramming is a stochastic process and there is merely a temporal requirement for cells to realise their reprogramming potential with enough cell divisions. Furthermore, manipulation of pathways that increased the rate of cell division/proliferation, for example, knock down of the p53 pathway or overexpression of Lin28, resulted in accelerated activation of Nanog-GFP, suggesting that the stochastic nature of reprogramming was proliferation dependant and susceptible to refinement.

1.4.3 Epigenetics of reprogramming

1.4.3.1 Chromatin remodeling and the role of vitamin C during reprogramming

Chromatin and DNA modifications have been purported to play an important role during the reprogramming process. The extent to which epigenetic remodelling plays a role during reprogramming is not yet fully determined, however, even as early as the pioneering studies it was clear that epigenetic reactivation of key pluripotency genes such as Nanog and Oct3/4 was important for

generation of fully reprogrammed, high quality germ-line competent iPS cells (Takahashi and Yamanaka, 2006; Okita et al., 2007). Whilst the first four factor derived iPSCs (iPS-MEF4), using Fbx15 as a reporter, produced cells which were morphologically similar to ESCs and contributed to teratomas and mouse embryonic development, they did not successfully silence the transgenes or re-establish the correct DNA methylation marks at the Oct4 and Nanog promoters and live chimeric mice could not be generated (Takahashi and Yamanaka, 2006). On the other hand, using Nanog as a reporter of reprogramming resulted in iPS cells (Nanog iPS 20D17 and 20D18) showing up-regulation of an extensive set of pluripotency genes, de-methylation of Oct4 and Nanog promoters (similar to that of ESCs) and generation of germ-line competent live chimeras (Okita et al., 2007). In addition to DNA modification, chromatin modification has also been perceived to play an important role during reprogramming. Chromatin of ESCs is widely thought of as being "open" (Orkin and Hochedlinger, 2011); histone acetylation and H3K4 trimethylation (H3K4me3) is associated with "open" euchromatin and active transcription whereas histone deacetylation and H3K27 trimethylation (H3K27me3) is linked to the silent heterochromatin state (Hotta and Ellis, 2008; Maherali et al., 2007). A significant example of chromatin remodelling during reprogramming is reactivation of the inactive X chromosome in female lines, which is thought to provide evidence towards a more reprogrammed state. During iPSC derivation, extensive chromatin reformation must be carried out to remove the distinctive chromatin modifications that are established on the future inactive X for stable silencing (Maherali et al., 2007). The same study carried out genome-wide analysis of K4 and K27 trimethylation using chromatin immunoprecipitation (ChIP) followed by promoter array analysis and revealed that almost 95% of "signature genes" (with differential histone methylation patterns between mouse embryonic fibroblasts (MEFs) and ES cells) in iPS cells exhibited nearly identical methylation patterns to ES cells (Maherali et al., 2007). This indicates that *in vitro* reprogramming can re-establish the epigenome of a differentiated cell to that of an ES cell. Furthermore, microarray gene expression analysis confirmed that iPS cells are

transcriptionally highly similar to ES cells. These data highlight the importance of epigenetic remodelling for complete reprogramming.

Small molecules targeting chromatin-modifying enzymes have long been known to promote reprogramming (Huangfu et al., 2008a; Mali et al., 2010; Liang et al., 2010; Zhang and Wu, 2013). Sodium butyrate, a fatty acid and histone deacetylase (HDAC) inhibitor, has been found to be supportive of ESCs from both mouse and human cells with butyrate treatment being implicated in down-regulation of Tcf3, a suppressor of Nanog in mouse ESCs. In addition, butyrate was found to directly act on epigenetic modifiers with evidence of induction of H3K9 acetylation and DNA demethylation of the Dppa5 promoter being reported (Ware et al., 2009). Several reports have implicated butyrate as a potent enhancer of reprogramming when applied either transiently or at low concentrations with observations of histone H3 acetylation, DNA demethylation at gene promoters and enhancement of pluripotency gene expression all contributing to improved reprogramming (Liang et al., 2010; Mali et al., 2010; Zhang and Wu, 2013). Treatment of reprogramming cultures with valproic acid (VPA), another HDAC inhibitor, reportedly results in reprogramming efficiencies of almost 2.5% (more than 100-fold increase compared with control), and in addition allows for the omission of c-Myc in the reprogramming factor cocktail (Huangfu et al., 2008a).

In human reprogramming, shRNA-mediated knock down of DOT1L, an H3K79 histone methyltransferase was found to increase reprogramming, and replace KLF4 and MYC. Notably, this resulted in an increase in NANOG and LIN28, both of which are known to enhance reprogramming (Onder et al., 2012). The authors found that loss of H3K79me2 in somatic cell specific genes occurs early during reprogramming and suggested that DOT1L functions to facilitate down-regulation of this histone mark usually associated with transcriptionally active genes.

Resetting of the epigenetic landscape within reprogramming cells is a crucial series of events required for robust and complete acquisition of pluripotency. ES cells, which largely contain open, active chromatin, contain bivalent domains that

consist of chromatin decorated with both repressive histone3-lysine27 trimethylation (H3K27me3) and activating histone4-lysine4 trimethylation (H3K4me3) marks. Two key protein complexes, Polycomb group (PcG) and Trithorax group (trxG) complexes underlie these repressive and activating marks, respectively. Early transcriptional events, such as down-regulation of somatic cell specific markers and up-regulation of pluripotency genes, are driven by changes in the chromatin state at these genes, facilitated by the PcG and trxG complexes. Expression of Wdr5, a member of the trxG complex, increases during reprogramming and is required during the early stages to initiate changes to the epigenetic landscape through interaction with Oct4 (Ang et al., 2011). Utx, an H3K27 demethylase also known as Kdm6a, plays a role in regulation of pluripotency induction during reprogramming and absence of Utx in somatic cells causes failure to reprogram by causing deviant regulation of H3K27me3 demethylation directly interfering with activation of pluripotency associated genes including Sall1, Sall4 and Utf1 (Mansour et al., 2012).

Histone3-lysine9 methylation (H3K9me) has been implicated as the key epigenetic block in pre-iPSCs in culture conditions containing serum. Downstream targets of bone morphogenic proteins (BMPs), the H3K9 methyltransferases Suv39h1, Suv39h2 and Setdb1 were shown to play a role in maintaining H3K9 methylation in pre-iPSCs. Induction of H3K9 demethylases belonging to the Kdm family, particularly Kdm4b (also known as Jmjd2b), in pre-iPSCs reduced H3K9 methylation releasing the barrier to reprogramming which was shown to be enhanced by vitamin C (Chen et al., 2013). Vitamin C was found to facilitate removal of the H3K9me3 mark releasing cells from this intermediate state and allowing progression to iPSCs with high levels maintained at key pluripotency genes Dppa3, Nanog and Sox2 sustaining a closed chromatin state and preventing binding of Oct4 (Chen et al., 2013). BMP driven activation of H3K9 methytransferases, Suv39h1 and Suv39h2, was found to support this heterochromatic state in pre-iPSCs but treatment with vitamin C reduced H3K9me through the action of demethylases allowing reprogramming of pre-iPSCs to iPSCs.

Setdb1 knock down alleviated the need for vitamin C treatment supporting a role for Setdb1 as an inhibitor of reprogramming. Strikingly, the authors also showed that knock down of Setdb1 resulted in conversion of almost 100% of pre-iPSCs to Oct-GFP+ iPSCs further emphasising the role of epigenetic regulators in reprogramming. An earlier report by Esteban et al. (2010) corroborated this showing the release of pre-iPSCs to a fully reprogrammed state by addition of vitamin C.

Vitamin C was first implicated as an enhancer of reprogramming by Esteban et al. (2010) where it was identified during investigation of compounds to reduce generation of reactive oxygen species (ROS) during reprogramming. Vitamin C was found to increase reprogramming efficiency partly through modulation of cell proliferation by inhibition of p53 and the downstream target p21. This was thought to be a different mechanism by which Silva et al. (2008a) reported conversion of pre-iPSCs to a fully reprogrammed state by inhibition of Erk signalling in 2i conditions, since total and active Erk levels remained unchanged.

The Jumonji C (JmjC) family of histone demethylases functions in histone demethylation at lysine resides in an iron Fe(II) and 2-oxoglutarate (2OG) dependent manner and vitamin C functions as an electron donor to reduce iron in the absence of substrate, thereby maintaining its catalytic activity (Monfort and Wutz, 2013). Vitamin C has also been reported to facilitate reprogramming by reducing H3K36me3 and when an siRNA targeting the demethylase Kdm3b was used, a reduction was observed in vitamin C–mediated progression of pre-iPSCs to a fully reprogrammed state (Chen et al., 2013; Li et al., 2011; Zhu et al., 2010).

Induction of the H3K36 histone demethylases Jhdm1a and Jhdm1b (also known as which Kdm2a and Kdm2b), which are downstream targets of vitamin C, have been shown to regulate reprogramming in synergy with Oct4 (Wang et al., 2011). Overexpression of these histone demethylases during three-factor reprogramming with Sox2, Klf4 and Oct4 was shown to enhance reprogramming both in the presence and absence of vitamin C in addition to improving reprogramming kinetics as characterised by the earlier appearance of Oct-GFP colonies (Chen et al., 2013). Strikingly, overexpression of Jhdm1b enabled efficient

reprogramming using only Oct4 at rates higher than those previously reported (Chen et al., 2011b; Zhu et al., 2010; Li et al., 2011) but Jhdm1b mutants lacking DNA binding or histone demethylase function were found to diminish Oct4 reprogramming, suggesting that the enhancing effect was owing to demethylation activity and/or binding to Oct4.

MEFs stably expressing the H3K36 histone demethylase Jhdm1b restored three factor (SKO) reprogramming ability in late passage (P6) fibroblasts, which was observed to be completely abolished in control cells due to Ink/Arf induced cellular senescence and addition of vitamin C further increased reprogramming efficiency (Wang et al., 2011). Ink/Arf silencing in MEFs was previously reported by Tzatsos et al. (2009) where Jhdm1b was found to suppress the Ink/Arf locus in MEFs by offsetting the senescence-induced down-regulation of polycomb-repressive complex (PRC) protein Ezh2, through H3K36 demethylation. This resulted in an increase in H3K27me3 at the Ink/Arf locus further driving binding of another PRC component, Bmi1, facilitating silencing of $p16^{Arf}$. Kdm2b promotion of reprogramming was similarly reported by Liang et al. (2012) in a proliferation and Ink4a/Arf independent manner. Instead it was purported to play a role in activation of early genes such as Cdh1, Epcam and Crb3 in collaboration with the reprogramming factors, by maintaining low levels of H3K36me2 at the promoters of these genes.

Strong induction of the ESC specific micro-RNA cluster 302/367 was promoted by a combination of Oct4, Jhdm1b and vitamin C. (Wang et al., 2011). The authors cloned the miRNA 302/367 promoter to a firefly luciferase reporter and found evidence for physical interaction of Oct and Jhdm1b to cooperatively bind the miRNA 302/367 promoter leading to activation. Further ChIP-qPCR and co-immunoprecipitation experiments subsequently verified this interaction, suggesting that histone modifications triggered by Jhdm1b are essential to allow Oct4 binding and target activation during reprogramming.

1.4.3.2 Resetting of DNA methylation and reprogramming

Resetting of methylation status occurs late in reprogramming whereas bivalent domains are laid down steadily following the first wave of transcriptional activity in response to the reprogramming factors (Polo et al., 2012). The implications of imprinting status of iPSCs have been widely discussed. Stadtfeld et al. (2010) reported that a vast majority of iPSCs derived from varying tissues exhibited abnormally low expression of Gtl2 and Rian; maternally expressed imprinted genes located in the Dlk1-Dio3 cluster. This silencing in iPSCs was found to be the result of hypermethylation in an intergenic differentially methylated regions (IG-DMR) located within the Dlk1-Dio3 locus, in which almost all CpGs were found to be methylated in "Gtl2^{off}" iPSCs compared with "Gtl2^{on}" iPSCs or ESCs. Furthermore, these Gtl2^{off} iPSCs did not contribute highly to chimeras and persistently failed to generate "all-iPSC" mice using the tetraploid complementation assay, widely regarded as a gold standard of pluripotency. Administration of valproic acid (VPA) was determined to "rescue" Gtl expression in two Gtl2^{off} clones, one of which subsequently enabled generation of "all-iPSC" mice, however, the pups were deemed non-viable. Another group also reported difficulty in generating all-iPSC mice from some iPSC lines as a result of decreased expression of genes in the Dlk1-Dio3 region (Liu et al., 2010), lending support that proper epigenetic resetting is important during reprogramming to produce high quality iPSCs. Carey et al. (2011) later reported contrasting results claiming that regardless of the imprinting state at this locus, they found no difference in the potential of "Gtl2-ON" or "Gtl2-LOW" iPSCs to contribute to chimeras and so claimed that silencing at this imprinted locus was not strictly definitive of reduced pluripotency. Further to this, these authors determined reprogramming factor stoichiometry to play a critical role in reprogramming, and differences in the expression levels of factors lead to significant differences in the quality of iPSCs derived. A comparison was made between two highly similar reprogramming systems using doxycycline-inducible polycistronic vectors to express the four reprogramming factors, differing only in the sequence in which the factors were expressed (OSKM versus OKSM) and the

linkers used in the vectors. This resulted in OSKM cells with higher levels of Oct4 and Klf4 and OKSM cells with higher levels of Sox2 and c-Myc upon administration of doxycycline to induce reprogramming. Whilst both cell lines were capable of generating iPSCs, the resulting OKSM iPSCs were of poorer quality than those generated with OSKM, as determined by aberrant expression of the imprinted gene Gtl2. However, this could be rectified by overexpression of additional Oct4 and Klf4, and subsequently "all-iPSC" mice could be generated from one of these clones

In a later report, Stadtfeld et al. (2012) determined that addition of vitamin C during reprogramming diminishes aberrant hypermethylation at the IG-DMR of Dlk1-Dio3 locus, and subsequently, all-iPSC mice were generated from iPSCs derived from B cells for the first time. It was determined that vitamin C functions by maintaining histone acetylation and activating histone marks, H3K4me2 and H3K4me3, which is lost in the absence of vitamin C during reprogramming. This report highlighted the effect of variations in culture conditions in the generation of high-quality iPSCs.

1.4.4 Models of reprogramming

Several reports have identified distinct stages of reprogramming with stochastic and deterministic models both being put forward (Yamanaka, 2009; Hanna et al., 2009b; Smith et al., 2010; Samavarchi-Tehrani et al., 2010; Buganim et al., 2012; Golipour et al., 2012). Early on two models were proposed; the elite, or predetermined model, and the stochastic model. The former put forward that only a limited number of predetermined cells were capable of becoming iPSCs, and tissue stem cells, or undifferentiated precursors, were a good candidate. However, evidence that initial reported reprogramming efficiencies could be drastically increased (Nakagawa et al., 2008; Huangfu et al., 2008a; Huangfu et al., 2008b) and reports of iPSCs derived from tissues such as liver, stomach and pancreas (Aoi et al., 2008; Stadtfeld et al., 2008a) challenged the idea that only a few 'elite' cells were

susceptible to reprogramming. The contrasting stochastic model proposes that upon reprogramming factor introduction, most or all cells have the potential to become iPSCs. The initiation stage of reprogramming is driven by BMP signalling in cooperation with reprogramming factor expression, activating the miR-205 and miR-200 families of microRNAs which regulate an immediate MET event (Samavarchi-Tehrani et al., 2010). Polo et al. (2012) observed that successful reprogramming consisted of two "waves" of transcriptional activity with the first wave being driven by c-Myc and Klf4 expression followed by the second wave driven by Oct4, Sox2 and Klf4 expression. Some cells which only initiated succeeded in initiation of the first wave but not the second failed to progress in reprogramming. During the later stages of reprogramming, cells are required to mature and stabilize in order to become transgene independent and maintain their acquired pluripotent state. Golipour et al. (2012) reported that persistent transgene expression hindered the later stabilization of cells and that silencing of exogenous factor expression was required for complete reprogramming, a feature earlier reported by Okita et al. (2007). The maturation and stabilization stages require expression of distinct sets of pluripotency genes including Nanog, Sall4, endogenous Oct4, Rex1 and Esrrb followed by transgene removal and expression of endogenous Sox2, Dppa2 and Pecam1 (Samavarchi-Tehrani et al., 2010; Buganim et al., 2012; Golipour et al., 2012). Interestingly, activation of endogenous Sox2 has been reported to mark a late phase in reprogramming comprising a predictable, step-wise series of gene expression signatures where Sox2 activates successive expression of Sall4, Lin28, Fgf4, Fbxo15 and Dnmt3b, representing a hierarchical model (Buganim et al., 2012).

More recently, a review by Buganim et al. (2013) pulled together these models, discussing the evidence for a two phase model incorporating an early, long lasting, stochastic phase and a later, shorter, hierarchical phase following Sox2 expression. The stochastic phase is initiated by four factor (OSKM) induction and the ensuing dynamics of genome-wide transcriptional regulation, where cells expressing OSKM will immediately undergo cellular fate changes either conducive

to reprogramming or alternatively down refractory pathways such as apoptosis or senescence. The 'stochastic' nature of this phase lies in the unpredictable way in which OSKM expression induces genome wide transcriptional activation and the random probability in which any given cell will experience just the precise genome-wide changes to impart 'reprogrammability'. Indeed, the long latency of this initial phase supports the notion that there is a stochastic element in play and so it follows that time is a key requirement for some cells to lay down the correct gene expression networks to become amenable to reprogramming. Reprogrammable cells will then undergo several defined changes including morphological changes, increased rates of proliferation coupled with transcriptional and metabolic stimulation, mesenchymal to epithelial transition (MET), reorganization of histone modifications and activation of DNA repair pathways. Progression through reprogramming is met by an intermediate stage, which sees activation of early pluripotency markers, laying down the foundations of the pluripotency transcriptional network (Buganim et al., 2012), temporary expression of developmental regulators (Polo et al., 2012) and gradual activation of glycolysis (Hansson et al., 2012). In addition, an unknown bottleneck occurs, contributing to the long delay of cells to progress. Perhaps this is again due to the requirement for certain early predictive genes of reprogramming such as Utf1 and Esrrb to be expressed (Buganim et al., 2012) triggering initiation of the final 'hierarchical' phase in some cells. This 'deterministic' phase progresses in a more predictable manner, marked by the activation of endogenous Sox2 (Buganim et al., 2012). Finally, stabilization of the core pluripotency network via action of Oct4, Sox2 and Nanog allows exogenous factor independence and emergence of bona fide iPSCs (Boyer et al., 2005).

The precise role of each of the reprogramming factors has been widely studied, and yet a precise step-wise sequence of events has not been identified that fully elucidates the transition from a somatic to a pluripotent cell type. This begs the question of whether there is in fact one route to iPS cells, or if the acquisition of pluripotency is a result of the events initiated by the introduction of transcription factors in concert with some perfectly timed yet stochastic events. The answer is

likely to lie somewhere in the middle. It is clear that some very distinct events occur during the course of reprogramming; early on down-regulation of genes associated with the somatic starting cells occurs, followed by mesenchymal-to-epithelial transition (MET) and up-regulation of pluripotency genes, in addition to a whole host of well characterized epigenetic changes which 'reset the genome'. A vast amount of knowledge has been generated about the reprogramming process but the fact of the matter is that efficiencies of reprogramming are still by and large low. Only a few reports of dramatic increases in efficiency have been published, and even so many of these claims remain uncorroborated or controversial, highlighting the importance of independent reproducibility in published works to eliminate potential artefacts of laboratory-to-laboratory variability.

1.4.5 Identifying and overcoming barriers to reprogramming

Senescence has been identified as a major barrier to iPSC generation and 2009 saw a number of reports on the role of p53 and related cell cycle regulating tumour suppressor genes during reprogramming. Up-regulation of p53, $p16^{Ink4a}$ and p21 is triggered upon exogenous reprogramming factor expression (Banito et al., 2009; Kawamura et al., 2009). Knock down of these genes during reprogramming using shRNAs was shown to significantly increase efficiency of pluripotent colony formation demonstrating the importance of cell cycle regulation during reprogramming. Another group reported generation of 3 factor iPSCs was achieved in p53-null ($p53^{-/-}$) MEF and this study identified that overexpression of Mdm2, an E3 ligase and negative regulator of p53 (Vassilev et al., 2004), replicated the positive effect of p53 repression in reprogramming (Hong et al., 2009). When Kawamura et al. (2009) reprogrammed MEF from mice derived with a non-degradable mutant of a related p53 negative regulator, Mdm4, they observed an almost 7-fold increase in Nanog+ colony number providing further evidence for the inhibitory role of p53 in reprogramming. Marion et al. (2009) suggested that this barrier effect caused by p53

during reprogramming was a result of genomic integrity being maintained via the DNA damage response induced by reprogramming factor expression. The authors suggested that whilst depletion of p53 resulted in increased efficiency, that this was at the risk of allowing substandard cells to progress towards iPSCs that normally would not have been able to for reasons of compromised genomic stability; a sort of innate natural selection was being bypassed by abrogation of p53 and related genes. Some evidence was presented to support this theory, as p53-deficient iPSCs were shown to have increased chromosomal abnormalities compared with wild type iPSCs including increased occurrence of chromosomal breaks and end-to-end fusions and in addition, depletion of p53 allowed telomerase deficient cells to reprogram to iPSCs; cells which were otherwise incapable of acquiring pluripotency (Marion et al., 2009). Interestingly, Utikal et al. (2009) showed that p53 -/- cells reprogrammed sub-populations of Thy1+ Thy1- and SSEA-1+ cells had similar reprogramming potential.

The Ink4/Arf locus consists of tumour suppressor genes $p16^{Ink4a}$ and $p19^{Arf}$ (encoded by Cdkn2a) and $p15^{Ink4b}$ (encoded by Cdkn2b) involved in cell cycle and senescence regulation that are known to play a critical role in reprogramming. $p16^{Ink4a}$ regulates the Retinoblastoma (Rb) pathway, downstream of p53 (Sage, 2012) and $p19^{Arf}$ drives expression of p53 by inhibition of Mdm2 mediated degradation (Spike and Wahl, 2011). These genes are expressed at basal levels in differentiated cells and acquire both repressive H3K27me3 and active H3K4me3 marks, known as bivalent chromatin, as they undergo silencing during reprogramming to iPSCs whilst still retaining the ability for reactivation upon differentiation (Li et al., 2009a). $p19^{Arf}$ activation of p53 and p21 in the mouse has been reported as a key roadblock of reprogramming (Spike and Wahl, 2011) and aberration of these pathways has been proven beneficial to reprogramming, enhancing both the kinetics and efficiency of pluripotent colony formation drastically (Li et al., 2009a). Interestingly, these authors showed that by contrast in human cells, INK4a seems to be the inhibitor of reprogramming with shRNA targeting INK4a exhibiting an enhancing effect, whereas ARF shRNA had no effect on iPSC formation.

Another clear barrier to reprogramming potential is the capability of cells to proliferate. Utikal et al. (2009) reported the observation that MEF progressively lose their reprogramming potential following serial passaging; an observation also experienced by our group. An increase in passage number of starting MEF undergoing reprogramming was accompanied by a decrease in alkaline phosphatase (AP) staining associated with pluripotent cells and a concomitant increase in senescence related staining with β-galactosidase. This was also accompanied by an increase in p16, p19 and p21 in agreement with previous reports that these markers obstruct progression towards iPSCs. Furthermore, three immortalized cell lines derived from different somatic tissue origins were used to demonstrate that senescence and limitations in cell-cycle pathways substantially hindered reprogramming capacity of cells, with immortalized cells regularly shown to reach upwards of 40% reprogramming efficiency compared to their non-immortalized counterparts, with near 100% efficiency demonstrated in some cases (Utikal et al., 2009). Smith et al. (2010) used a single-cell imaging approach to retrospectively trace iPSC colonies to their cell of origin and found that these colonies emerged from a subclass of fibroblasts which they termed "fast-dividing". They found that as soon as the first cell division these cells established a higher proliferative rate than normal fibroblasts, and within a few days match that seen in ES cells. This evidence supports the idea that proliferative potential is a key requirement for somatic cells to successfully progress through reprogramming.

Expression of the four reprogramming factors directly affects a number of cluster families of micro-RNAs. p53 regulated miR-34 and miR-145 family members play critical roles in the impediment of reprogramming by promotion of cell cycle arrest, apoptosis and differentiation. miR-34a/b/c function in part by inhibition of key pluripotency genes including Nanog, Lin28, Sox2 and c-Myc and depletion of miR-34a has been shown to enhance reprogramming efficiency and kinetics (Yang and Rana, 2013). miR-145 has been reported to suppress Oct4, Sox2 and Klf4 during differentiation of ESCs although it is repressed by Oct4 in ESCs, suggesting a possible mechanism in reprogramming by which miR-145-mediated endogenous

repression of OSK is abolished by Oct4 repression. MEF enriched miRNAs miR-21, miR-29a and let-7 have been reported as barriers to reprogramming through positive regulation of TGF-β and MAPK pathways or negative regulation of pluripotency genes. c-Myc, which has been shown to play a critical role early in reprogramming, mediates reprogramming initiation in part by repressing these MEF associated miRNAs (Yang and Rana, 2013).

ESC-specific cell cycle (ESCC) miRNAs (Wang et al., 2013c) including miR-290-295, miR-302a-367 and miR-17-92 are known to influence epigenetic status and cell cycle regulation. ESC specific miRNAs, miR-291-3p, miR-294 and miR-295 were found to replace c-Myc during reprogramming, although in the presence of c-Myc there was no effect observed on four factor reprogramming (Judson et al., 2009). This could be explained by the finding that c-Myc binds the promoter of the miR-290 cluster, suggesting that these miRNAs could be targets and downstream effectors of c-Myc, although the exact mechanism remains unclear. A later report determined that human orthologues of these miRNAs, hsa-miR-302a and hsa-miR-372, enhanced reprogramming of both fetal and adult fibroblasts by synergistic repression of multiple targets (Subramanyam et al., 2011). These targets were identified to play roles in cell cycle and epigenetic regulation, signalling and epithelial to mesenchymal transition (EMT), highlighting the fact miRNA-mediated regulation of cell processes is extensive and highly complex making it difficult to determine the exact mechanism(s) by which promotion of reprogramming is conferred.

1.5 RNAi and reprogramming

Reprogramming of somatic cells to a pluripotent state via the ectopic expression of a limited set of transcription factors is now routinely performed by countless research groups, and indeed life science and drug companies, around the world. However, the question still remains as to the exact regulatory mechanism(s) underlying the transition of cells from a specialised, differentiated state to the establishment of pluripotency. Recently, the use of RNA interference (RNAi) has been reported to be a highly useful tool to identify regulators of the reprogramming process, giving some insight in to the mechanism of this inefficient process.

RNA interference (RNAi) is a naturally occurring biological mechanism in cells where gene regulation is controlled by RNA hybridization to target messenger RNAs, resulting most commonly in inhibition of gene expression by way of destruction of mRNA, consequently causing down-regulation of target genes. This innate mechanism is an important part of host-pathogen defence, particularly in response to invasion by viruses, and vast research into the field in the early 1980's saw this mechanism identified in an array of organisms including bacteria (Light and Molin, 1982; Light and Molin, 1983), dictyostelium (Crowley et al., 1985), Xenopus oocytes (Melton, 1985; Harland and Weintraub, 1985), Drosophila (Rosenberg et al., 1985) and plants (Ecker and Davis, 1986) where researchers noticed that transcriptional inhibition was caused by application of anti-sense RNA. Around this time, similar observations were also reported in the mouse (Izant and Weintraub, 1984; Izant and Weintraub, 1985). Since then, the principles and mechanism of RNAi have been extensively studied resulting in the capability of artificially engineered RNAi expression constructs as a tool to investigate gene regulation *in vitro*.

There are two main ways in which RNAi is used to knock down gene expression in mammalian cells *in vitro*. Short interfering RNAs (siRNA) are double stranded RNA fragments of approximately 21bp in length, first described by Elbashir et al. (2001a) which can be designed to target any gene of interest and,

conveniently, can now be artificially synthesized by a number of companies. When introduced to cells, these siRNAs directly interact with the RISC complex to mediate gene repression of a target mRNA containing a complementary sequence to one of the two siRNA strands; the effector strand known as the guide strand (Mittal, 2004; Matzke and Birchler, 2005). This approach may seem straightforward, however, the knock down effect can be weak and/or transient, particularly in rapidly dividing cells and may require serial transfections in order to obtain satisfactory knock down. Alternatively, an expression vector can be created to stably express a short hairpin RNA (shRNA) targeting a gene of interest (Moffat and Sabatini, 2006; Echeverri and Perrimon, 2006). In this case, the sequence is designed to introduce a small loop or hairpin between the complementary RNA strands upon expression, which facilitates endogenous processing of the shRNA in a similar fashion to endogenous microRNAs. Exogenously expressed shRNAs are trimmed by Dicer, an endoribonuclease, to produce smaller fragments similar to siRNA that is similarly incorporated into the RISC complex and involved in gene repression.

Regardless of which RNAi method is used, when investigating functional mechanisms of gene repression, the most important factor to consider is the level of gene repression achieved. This is a very important caveat when considering RNAi as a tool for knock down of gene expression particularly when knock down of many genes is required, for example, in the context of a screen. Extensive studies into optimization of RNAi have led to a considerable list of 'rules' to take into account when designing RNAi sequences or vectors. For example, initial 'first-generation' designs were superseded by 'second-generation' shRNAs that took advantage of a micro-RNA (miRNA)-like backbone design, to aid in more efficient endogenous processing within cells (Silva et al., 2005; Boudreau et al., 2008). In addition, siRNAs should include 2 nucleotide 3' overhangs, mimicking endogenous Dicer cleavage and an increased efficiency of knock down has been reported with UU 3' overhangs although other combinations also work (Elbashir et al., 2001b). Other rules have been suggested regarding optimal region of gene targeting, GC content, concentration of siRNA, specificity of bases at certain positions within the sense

strand as well as many other structural considerations (Semizarov et al., 2003; Mittal, 2004; Reynolds et al., 2004) making it clear that RNAi design is deeply complicated. Despite extensive guidelines within the literature, a definitive formula for designing RNAi sequences that guarantees to achieve robust knock down of a target gene is still undetermined. Furthermore, sequence specificity must be well considered when using RNAi, as it has been widely reported that even a single nucleotide difference in siRNA sequences can drastically abolish gene targeting (Elbashir et al., 2001b; Miller et al., 2003). In addition, off-target effects resulting from incomplete specificity are a significant problem, with Jackson et al. (2003) the first to demonstrate that off-target gene regulation could occur when siRNAs exhibited only partial complementarity. In fact, the authors showed that as few as 11 adjacent complementary nucleotides were enough to elicit an off-target effect in contrast to other reports that siRNA was highly specific in its effect (Jackson et al., 2003; Semizarov et al., 2003; Miller et al., 2003).

Recently, the mesenchymal genes Snai1 and Snai2 have been identified from an RNAi screen to play key opposing roles in reprogramming (Gingold et al., 2014). By inducing partial differentiation in a Nanog-GFP reporter line with administration of retinoic acid (RA), the authors sought out to identify regulators of Nanog expression. Knock down of Snai1 was found to inhibit GFP expression whereas knock down of Snai2 was found to have the opposite effect, with increased GFP expression observed. Subsequent overexpression of Snai1 and Nanog during reprogramming of pre-iPSCs was found to increase the number of fully reprogrammed iPSCs more than 2-fold compared with Nanog alone, with Snai2 overexpression having a negative effect. Additionally, Snai1 was found to facilitate binding of Nanog within the promoter and enhancer regions of the miR-209-295 locus, as well as binding to Lin28, another pluripotency associated gene.

The use of RNAi has been valuable for functional genomics studies, and loss-of-function phenotypes in mammalian cells induced by RNAi screening methods have been successfully generated over the years allowing for the investigation of gene function in a variety of cell types including BJ human foreskin

fibroblasts (Berns et al., 2004), mammary cells (Silva et al., 2008b) and blood cells (Bassik et al., 2013). Previously, whole genome RNAi screens were successfully employed to elucidate genes that govern mouse and human embryonic stem cell identity (Hu et al., 2009; Chia et al., 2010), and siRNA screens have uncovered genes involved in Oct4 modulation and ESC identity, including Paf1C which blocks ESC differentiation following ectopic expression (Ding et al., 2009), the role of chromatin protein complex Tip60-p400 in ESC gene regulation and identity (Fazzio et al., 2008), and MAP kinase phosphatases involved in regulation of ERK and GSK3 signalling pathways as promoters of differentiation in ESCs (Yang et al., 2012). More recently, similar approaches using shRNA libraries have been applied in both mouse and human contexts to identify novel regulators of reprogramming. Yang et al. (2014) uncovered specific genes required at each transitional step during reprogramming and in addition the authors found that the expression of many genes was shown to remain unchanged during reprogramming. Interestingly these were purported to play key roles in controlling transitions of cellular identity during reprogramming, with some genes found to be required and others representing barriers to reprogramming. Qin et al. (2014) generated a global view of reprogramming barriers with genes identified from many major cellular pathways including transcription, ubiquitination, phosphorylation and dephosphorylation, cell adhesion and chromatin regulation emphasising the point that reprogramming of somatic cells to pluripotency is a deeply complex process with cell processes being infinitely interconnected and interdependent. Another approach employing a protein kinase shRNA screen of 734 kinases identified 59 barriers to reprogramming, with remodelling of the cytoskeleton being highlighted as an important modulator of reprogramming. TESK1 and LIMK2, whose inhibition promotes MET, were identified as barriers to this cellular restructuring process and consequently knock down promoted reprogramming in both mouse and human cells (Sakurai et al., 2014).

Taken together these studies demonstrate that RNAi is a powerful tool to identify key players involved in ESC/iPSC pluripotency and self-renewal and gain

novel insight into the mechanisms underlying the highly complex reprogramming process.

1.5.1 CRISPR/Cas9 mediated gene editing

Recently, an exciting new genome editing technology dubbed the 'CRISPR-Cas9 system' has been described, which allows highly efficient gene knock out both *in vivo* and *in vitro* (Swiech et al., 2015; Maddalo et al., 2014). This system was identified in bacteria and archaea as a mechanism of immunity whereby fragments of foreign genetic material from invading phage or plasmid DNA were integrated into the host genome at so-called clustered regularly interspaced short palindrome repeats, or CRISPR, loci (Pourcel et al., 2005; Mojica et al., 2005; Bolotin et al., 2005). Subsequently, the Cas9 protein incorporates both processed CRISPR RNAs (crRNA) and *trans*-activating crRNA (tracrRNA) to form a cRNA-tracrRNA-Cas9 complex which then mediates double strand break at homologous sequences of the invaders (Mali et al., 2013b; Gasiunas et al., 2012). This system has been taken advantage of in an *in vitro* context whereby *in vitro*-transcribed chimeric RNA known as short guide RNA (sgRNA, or guide RNA) takes the place of the crRNA-tracrRNA combination (Jinek et al., 2012). Targeting of loci in human and mouse cells was first demonstrated in recent years, with two groups reporting simple and robust gene targeting of several loci simultaneously using CRISPR-Cas9 systems (Cong et al., 2013; Mali et al., 2013c). Remarkably, simultaneous mutation of up to five genes in mouse ESCs has been demonstrated and in the same study mice were generated with mutations in multiple genes through coinjection of zygotes with Cas9 mRNA and sgRNAs (Wang et al., 2013b). This multiplexed gene editing approach has since had major impact on the accessibility of routine gene editing in laboratories with its straightforward and reproducible protocol, rendering difficult and lengthy gene targeting approaches redundant.

The original Cas9 protein causes a double strand break in the target DNA, which can be repaired either by non-homologous end joining (NHEJ) which is error prone, or homology directed repair (HDR). This potential drawback was quickly overcome by the use of mutant Cas9 which converted the nuclease function of wild type Cas9 to a nickase function instead (Cong et al., 2013). Subsequently, a "double nicking" system was describe whereby a pair of offset guide RNAs targeting opposite DNA strands were expressed with the mutant Cas9, mediating targeted double strand breaks. Because the mutant Cas9 is unable to cause double strand breaks on its own, this system has increased specificity and fidelity by reducing off-target activity (Mali et al., 2013a; Ran et al., 2013).

This gene editing technique has quickly become the front-runner as the go-to gene knock out tool, and has already revolutionised reverse genetics investigation of both single and multiple genomic loci. At least three groups have employed lentiviral-based gRNAs in conjunction with Cas9 (either as part of the same construct or using Cas9 constitutively expressing cells) for genome wide targeting in the mouse and human (Koike-Yusa et al., 2014; Shalem et al., 2014; Wang et al., 2014) and researchers have also demonstrated use of the CRISPR system as a potential method of gene therapy to treat cystic fibrosis and blood disorders (Xie et al., 2014; Schwank et al., 2013). These reports demonstrate the robust and reproducible nature of this technology.

1.5.2 Summary

In summary, although a vast amount of knowledge has been gained about the reprogramming process in a relatively short period of time since Takahashi and Yamanaka (2006) first reported generation of iPSCs, the exact mechanism(s) underlying the reprogramming process remains largely elusive. One of the key issues that remains to be fully understood is why reprogramming is so consistently inefficient regardless of the system used. A number of groups have claimed

surprisingly high reprogramming efficiencies, with some reports even declaring almost 100% reprogramming efficiency. However, these reports are rarely corroborated and offer little in the way of mechanistic insight to the findings. For the work presented in this thesis, I used an RNAi screening strategy to identify novel regulators of reprogramming with the objective of gaining some mechanistic insight into the reprogramming process in the context of enhanced or diminished reprogramming efficiency.

CHAPTER 2 – Materials and Methods

2.1 Mammalian cell culture

2.1.1 Cell culture and reprogramming reagents

ESC and iPSC medium:

Glasgow Minimal Essential Medium (GMEM, Sigma G5154) supplemented with:

10% Foetal Calf Serum (FCS, Invitrogen, 10270 Batch 40F0240K)

Non-essential amino acids (1X NEAA, Gibco 11140-035)

L-Glutamine (2 mM, Invitrogen, 25030-024)

Sodium pyruvate (1 mM, Invitrogen, 15140-122)

2-mercaptoethanol (100 μM, Life Technologies 31350010)

Leukemia inhibitory factor (LIF, 100 units/ml, homemade human recombinant)

Penicillin-Streptomycin (P/S, 50 units/ml (P) or 50 μg/ml (S), Sigma, P4333)

MEF medium:

ESC and iPSC medium

bFGF/FGF2 (5 ng/ml, Peprotech 100-18-B)

Heparin (1 ng/mL, Sigma)

Reprogramming medium:

ESC and iPSC medium

Ascorbic acid/Vitamin C (10 μg/ml, Sigma A4034)

Doxycycline hyclate (dox, 300 ng/ml, Sigma, D9891-1G)

Cell freezing medium:

Dimethysulfoxide (DMSO, 10%, VWR International, 23500.260)

FCS (90%)

Cell culture selection drugs:

Blasticidin S hydrochloride (10 ug/ml, Sigma, 15205)

Puromycin dihydochloride (5 ug/ml, Sigma, P8833)

Transfection reagents:

Calcium chloride hexahydrate (2 M, Sigma, 442909)

HBS (2X, made in house by adding 8 g NaCl + 0.2 g Na2HPO4-7H2O + 6.5 g HEPES with volume brought up to 500 ml with dH2O and pH adjusted to 7.0)

Polybrene (10 mg/ml, Merck Millipore, TR-1003-G)

FuGene HD transfection reagent (Promega, E2311)

Opti-MEM I Reduced Serum Medium (OptiMEM, Life Technologies, 31985-062)

Miscellaneous cell culture reagents:

Dulbecco's Phosphate Buffered Saline (PBS, Sigma D8537)

Gelatin (0.1% in PBS, Sigma G5154)

Trypsin (0.25%, Gibco 15090-046)

Ethylenediaminetetraaceticacid (EDTA, 0.1%, Sigma 03620)

2.1.2 Cell lines used in this study

E14 ES cell line

TNG MKOS ESC line containing a doxycycline-inducible MKOS-ires-mOrange transgene and a Nanog-GFP reporter

MEF primary culture from 129 mice

TNG MKOS MEFs derived from chimeric embryos generated with TNG MKOS ESC line, (referred to as 'transgenic MEF' herein)
Platinum-E (Plat-E) retroviral packaging cell line

2.1.3 ESC culture technique

ESCs were cultured in ESC medium with trypsin/EDTA passaging every 2-3 days or when the cells reached approximately 80% confluency. For passaging, cells were washed with PBS, incubated with trypsin/EDTA for 2-3 minutes at 37°C, harvested in complete GMEM medium then centrifuged at 300g for 3 minutes. The cell pellet was resuspended in complete GMEM medium and cells seeded to a new gelatin-coated tissue culture vessel as appropriate and incubated in a humidified incubator set to 37 °C/7.5% CO_2.

2.1.4 Generation of chimeric embryos

ESCs were cultured as above and 48 hours prior to use a doubling dilution series was generated from $2x10^6$ cells in 2 ml of complete GMEM medium in a gelatin-coated 6-well tissue culture plate. From this, single cells or colonies of an appropriate size (5-8 cells) were harvested and aggregated with morulae of C57Bl/6 mice. All embryo manipulation techniques were carried out by Transgenic Unit staff at the Scottish Centre for Regenerative Medicine.

2.1.5 Mouse embryonic fibroblast isolation, cryopreservation and cell culture technique

Mouse embryonic fibroblasts (MEFs) were taken from embryonic day 12.5 (E12.5) mouse embryos. Pregnant mice were culled and dissected to isolate embryos that were decapitated, eviscerated and then dissociated in 0.25% trypsin/0.1% EDTA using a 21 gauge needle. Cells from individual embryos were then resuspended in 10 ml of MEF medium, large pieces allowed to sediment and ~9.5 ml of MEF cell suspension plated in a 10 cm^2 tissue culture dish then incubated in a humidified incubator set to 37 °C/7.5% CO_2.

When cells reached confluency, they were harvested for cryopreservation as follows: cells were washed with PBS, harvested by trypsin/EDTA incubation then collected and counted prior to centrifugation at 300g for 3 minutes. The cell pellet was resuspended in freezing medium at a concentration of 2-5x10^6 cells/ml and 1 ml aliquots cryopreserved with initial storage at -80 °C for 24 hours followed by transfer to liquid nitrogen (LN_2) for long term storage.

To recover cryopreserved MEF, vials were briefly thawed in a water bath set to 37 °C, resuspended in MEF medium followed by centrifugation at 300 g for 3 minutes. Cells were resuspended in MEF medium then counted and seeded in tissue culture vessels as required.

2.1.5.1 Quantification of transgenic cell contribution

In this study we used transgenic MEF carrying dox-inducible four reprogramming factors with a mOrange reporter. To quantify this, following MEF isolation a small aliquot of the MEF (approximately 1/20 of the isolated MEF) were plated in a 12-well tissue culture dish in reprogramming medium for 48 hours and the total percentage of transgenic cells was determined based on mOrange reporter expression by flow cytometry (BD Fortessa).

2.1.6 Reprogramming from MEF

In general reprogramming experiments were carried out in a 6-well format, with $1x10^5$ MEFs (wild type or transgenic, depending on the experiment) seeded per well. MEFs were recovered from cryopreservation approximately 2 days prior to seeding for experiments. Cells were either seeded directly from the initial recovered culture for experiments (passage 2, P2) or passaged/expanded to a new vessel up to one time prior to seeding, with cells only being used up to P3 to maintain good proliferation conditions.

2.1.6.1 Transfection of MEF for primary reprogramming experiments

Wild type MEFs were seeded in gelatinized 6-well plates at a density of $1x10^5$ cells per well and incubated for 24 hours before transfection. Spent medium was replaced with 2 ml of fresh MEF medium without heparin or P/S and the following ratio of vectors prepared per transfection in Eppendorf tubes: 0.5 μg PB-TAP-IRI-2LMKOSimO (O'Malley et al., 2013), 0.5 μg CAG-rtTA, 0.5 μg hyperactive *piggyBac* transposase (hyPBase, Yusa et al. (2011)). Depending on the experiment, 1 μg of PB-overexpression vector and/or 1 μg of PB-shRNA vector were also added. Fugene HD was used at a ratio of 1:4 for total DNA (μg) : Fugene HD reagent, so if DNA vectors totaled 3.5 μg in 3.5 μl then 14 μl of Fugene HD was used. A 100 μl transfection volume was used per 6 well, so the total volume of DNA and Fugene HD reagent was calculated and subtracted from 100 μl and this volume, e.g. 82.5 μl, of OptiMEM Medium was added to the DNA mix. The Fugene HD reagent was then added carefully below the fluid surface, avoiding contact with the plastic tube. The mixture was incubated at room temperature for 5 minutes and then 100 μl of the transfection mixture was added drop-wise to each well of MEFs in 2 ml of MEF Medium without P/S. After overnight incubation at 37 °C, the spent medium was replaced with fresh reprogramming medium to induce reprogramming factor expression, and this was taken as day 0 of reprogramming. Nanog-GFP+ colonies

were counted around day 10 to quantify any differences observed compared with control wells.

2.1.6.2 Preparation of shRNA retrovirus supernatant for reprogramming

Platinum-E (Plat-E) retrovirus packaging cells were seeded at a density of 2.5-3x10^6 cells per 10 cm plate in 10 ml of MEF medium containing blasticidin (10 ng/µl) and puromycin (2 ng/µl). After 24 hours, with Plat-E reaching confluency of around 60%, individual shRNA vectors were transfected by calcium chloride transfection: 63 µl of 2 M $CaCl_2$ was added to 15 µg of vector in 437 µl of distilled water (dH_2O), mixed and added dropwise to 500 ul of 2X HBS in an Eppendorf tube. Approximately 950 µl of this transfection solution was added dropwise to a 10 cm plate of Plat-E cells in 10 ml of MEF medium while swirling the plate to ensure even distribution, then incubated overnight at 37 °C. This process was repeated for each shRNA transfection. The following day, the transfection medium was replaced with fresh MEF medium and incubated for 24 hours at 32 °C to maintain stability of the virus. The following day, the virus supernatant was removed from the cells and filtered using a 0.45µm filter and syringe in to a 50 ml tube, and polybrene was added at a final concentration of 6-8 µg/ml. 2 ml per 6-well was distributed on to the MEF cultures replacing the MEF medium, as required. Viral titre was not calculated and a pMXs-DsRed vector was used as a transfection/infection control.

2.1.6.3 Reprogramming transgenic MEF with shRNAs

For all reprogramming experiments involving "topping up" with shRNA, MEFs were seeded at a density of 1x10^5 MEFs per 6 well in MEF medium, constituting 3% transgenic MEFs (based on mOrange expression quantification outlined above) diluted with 97% wild type 129 MEFs. The following day after MEF plating, cells were infected with virus as detailed above and the cultures were incubated at 32 °C for 4 hours and then transferred to 37 °C incubation overnight. 24 hours later, virus supernatant was replaced by reprogramming medium and this

was counted as day 0 for all reprogramming experiments. Medium was replenished every 2 days.

2.1.6.4 Primary reprogramming with shRNAs

For primary reprogramming with shRNAs, $1x10^5$ wild type MEFs were seeded in 6 wells and transfected with a polycistronic vector containing the four reprogramming factors MKOS with a mOrange reporter (PB-TAP-attP2LMKOSimO) using Fugene HD transfection reagent as per the manufacturers instruction (see below section). The day after transfection, virus supernatant was applied to cells as above and 24 hours later this was replaced with reprogramming medium (d0). Efficiency of transfection could be estimated by eye after 48 hours by observation of mOrange reporter expression.

Where other factors were introduced that were not shRNAs but were contained in virus vectors, the virus supernatant was produced and applied in exactly the same way as for the shRNA vectors above.

2.1.6.5 Colony counting experiments

For end-point analysis, Nanog-GFP+ colonies were counted at various time points either manually by eye using a fluorescent microscope, or for later experiments using the Celigo cell cytometer instrument.

2.1.6.6 Time course analysis

For time-course analysis by flow cytometry, whole 6-well reprogramming cultures were harvested by trypsin/EDTA at each required time point, cells counted and centrifuged at 300 g then re-suspended in PBS at a concentration of approximately $1x10^6$ cells/ml. The cell suspension was then aliquoted into round bottom 96-wells for antibody staining, as required.

2.1.6.4 Transgenic MEF reprogramming for Q-PCR analysis

To determine changes in gene expression in cells undergoing reprogramming, >94% transgenic MEFs were reprogrammed and samples taken at various time points, as required. For these experiments, >94% transgenic MEFs were seeded in gelatinized 6-well plates at a density of $2.5x10^4$ cells per well in MEF medium. The following day, the medium was replaced with reprogramming medium, and this was used as day 0 of reprogramming. For each time point, whole well samples were harvested with trypsin/EDTA and RNA extraction and cDNA synthesis were carried out (see Sections 2.2.2 and 2.2.3), followed by Q-PCR analysis using the Roche LightCycler® 480 and Universal Probe Library systems (see Section 2.2.4).

2.2 Molecular Biology Techniques

2.2.1 Plasmid vectors used in this study

pENTR-2B2	Entry vector for cloning of gene(s) of interest (GOI) with Gateway recombination sites, enabling easy cloning of DNA into compatible vectors. Kanamycin resistant.
pMXs-gw	Destination vector for pMXs-GOI plasmids. Contains Gateway recombination sites flanked by retrovirus LTRs. Ampicillin resistant.
pBabe-dual	Retroviral vector containing dual converging U6 and H1 RNA polymerase III promoters, published by Li et al. (2006) and available on Addgene. Ampicillin resistant.
pRetroSuper-Hyg	Plasmid containing a multiple cloning site (MCS) flanked by retrovirus long terminal repeats (LTR) and pUC origin of replication. Ampicillin resistant.
CAG-rtTA	Plasmid constitutively expressing reverse tetracycline transactivator (rtTA), a protein that binds the TetO operator sequence when bound by doxycycline. Required for activation of doxycycline inducible vectors.
pCMV-hyPBase	Plasmid encoding a hyperactive *piggyBac* transposase, which catalyzes the integration and excision of

	piggyBac transposons into DNA, under control of a CMV promoter. Ampicillin resistant.
PB-TAP-attP2LMKOSimO	Plasmid vector expressing doxycycline inducible *piggyBac* transposon encoding four reprogramming factors c-Myc, Klf4, Oct4 and Sox2 with mOrange reporter. Ampicillin resistant.
pMXs-Dmrtc2	Retroviral vector containing the open reading frame (ORF) of Dmrtc2 (note this vector is lacking the 3'-UTR). Ampicillin resistant. Used for overexpression experiment.
pMXs-hCD2	As above but containing the human CD2 ORF. Ampicillin resistant. Used in overexpression experiment.
pRS-Hyg-U6H1-shRNA	Modified pRetroSuper-Hyg vector containing dual converging RNA Polymerase III promoters U6 and H1. All shRNA expression vectors created using this plasmid by ligating shRNA oligos in to BbsI site.

2.2.2 RNA isolation

Cells in culture were washed with PBS and 1 ml of TRIzol reagent (Life Technologies, 15596-026) applied directly to cells per 6 well and processed according to manufacturer instruction. Briefly, cells in TRIzol were homogenized to ensure complete lysis then transferred to a 1.5 ml Eppendorf tube and incubated at room temperature for 5-10 minutes. 200 μl of chloroform was added and the tube

vigorously shaken then samples were centrifuged for 15 minutes at 12,000 g. The colourless upper aqueous phase was transferred to a new 1.5 ml Eppendorf tube (approximately 400 µl) then 400 µl of 100% isopropanol was added and mixed well. After incubation at room temperature for 10 minutes, centrifugation at 12,000 g was repeated for 10 minutes and the supernatant was removed leaving the RNA pellet behind. The pellet was washed with 1 ml of 75% ethanol, vortexed briefly then centrifuged at 12,000 g again for 5 minutes. The supernatant was removed and pellet dried at room temperature for 5-10 minutes then resuspended in nuclease free water. The RNA was either used immediately or stored at -80 °C. All centrifugation steps were carried out at 4 °C.

2.2.3 cDNA synthesis

MML-V reverse transcriptase (Life Technologies, 28025-013) was used for cDNA synthesis according to manufacturer instruction but briefly 20 µl reactions were prepared in 2 steps. Firstly 1 µg of total RNA was added to a nuclease-free microcentrifuge tube with oligo dT (100 µM, T24) and 100 mM of each of four deoxyribonucelotides dATP, dTTP, dCTP and dGTP (Life Technologies, 10297-018) totaling 10 µl. This was incubated at 65 °C for 5 minutes then placed on ice. Next, a mastermix of 5X First-Strand Buffer, 0.1 M DTT, MML-V reverse transcriptase (Life Technologies, 28025-013) and RNaseIN (NEB, M03141) was made and 10 µl added to the initial reaction mixture. Samples were then incubated for 1 hour at 37 °C, 10 minutes at 90 °C then cooled and stored at 4 °C until use. If not being immediately used, samples were placed at -20 °C for short-term storage or -80 °C for long-term storage.

2.2.4 Quantitative PCR analysis

Quantitative real time PCR (Q-PCR) was carried out using the Roche Universal Probe Library (UPL) hydrolysis probe-based system with a Roche LightCycler® 480 Real-Time PCR instrument. An 8 μl Q-PCR reaction was carried out as follows: a master mix was made up containing 4 μl of 2X Probes master mix, 0.08 μl of UPL probe and 1.92 μl of 20 μM forward and reverse primers for each reaction, and this was loaded on to a 384-well LightCycler® 480 Real-Time PCR plate. cDNA made from 1 μg of total RNA was diluted 1:10 with nuclease-free water and 2 μl of this was loaded in to each reaction. The plate was centrifuged at 300 g for 1 minute to collect the reactions at the bottom of the wells and the plate loaded into the LightCycler® 480 instrument using the UPL program recommended by the manufacturer. A standard curve was included in the reactions to allow for relative quantification to be calculated by the LightCycler® 480 software. Q-PCR reactions were carried out in duplicate and normalized using Tbp housekeeping gene.

2.2.5 Bacterial transformation and plasmid DNA preparation

DH5α E. coli bacteria were routinely transformed to generate plasmid DNA stocks for standard high copy plasmids. Typically <10 ng of stock plasmid, or 5-10 μl of ligation reaction was used per transformation as follows: 50-100 μl frozen aliquots of bacteria were removed from -80 °C storage and thawed on ice. DNA was added to the bacteria, in an appropriate concentration/volume, and the mixture was incubated on ice for 5 minutes followed by heat shock at 42 °C for 30 seconds then immediate transfer back on to ice for 2 minutes. 500 μl of LB broth was added to the tube and incubated at 37 °C for 1 hour. 10-500 μl of the mixture was spread on to LB agar plates, with antibiotics as appropriate, and plates were incubated for 16-24 hours at 37 °C. Individual colonies were picked from plates into 5ml of LB broth,

containing antibiotics as appropriate, after being stabbed on to new LB agar plates as a master plate. This plate was again incubated for 16-24 hours at 37 °C while the inoculated broth was incubated with 250 rpm agitation at 37 °C for 16-18 hours. Following incubation, the bacterial broth was centrifuged at 4000 g for 10 minutes and the supernatant removed. Plasmid DNA was purified from the bacterial cell pellet using the QIAprep Spin Miniprep kit (Qiagen, 27106). This scale of plasmid preparation was usually used for initial ligation reaction validation, and once confirmed, larger stocks were generated by inoculation of 50-100 ml of LB broth, using the QIAprep Spin Midiprep kit (Qiagen, 12243) for plasmid DNA purification.

2.2.6 Restriction enzyme digestion and purification of DNA fragments

Restriction enzyme digestion was primarily used either during the course of cloning new plasmid vectors or for confirmation of correct plasmids following ligation reactions, for example. For cloning new vectors, the amount of DNA digested varied but was typically >1 μg and for confirmation of plasmid vectors typically 100-200 ng was digested. All restriction enzymes and buffers were supplied by NEB with the specific quantities/concentrations used according to manufacturer instruction. Digestion reactions were carried out in 20 μl reactions and incubated for 1 hour at 37 °C unless otherwise recommended. After DNA digestion an OrangeG loading buffer (NEB, made up with 40% sucrose) was added to samples and reactions were loaded onto a 0.8-2% (w/v) agarose gel for electrophoresis separation of DNA fragments. A 1Kb Plus DNA ladder (Life Technologies, 10787-018) was loaded alongside samples and gels were run at 100-120 V for ~45 minutes. GelRed nucleic acid stain (Cambridge BioScience) was incorporated into gels to visualize DNA bands with UV illumination. DNA fragments of the correct size were extracted from gels using a scalpel and DNA was purified using the Zymoclean Gel DNA Recovery kit (Zymo Research) according to manufacturer instruction, with DNA eluted in 6-8 μl of sterile water.

2.2.7 shRNA expression vector construction

21 bp shRNA sequences for candidate genes were determined using the Life Technolgies online tool BLOCK-iT™ RNAi Designer. For vector construction, BbsI restriction sites were included at either end of the oligo duplex and a loop sequence was inserted between the sense and antisense sequences resulting in a 55-nucleotide oligo. An example is shown below, where red indicates the sequence for insertion into a BbsI restriction enzyme site, and blue indicates the short hairpin loop sequence:

Forward oligo:

5′-aaagTATGTAGGTTTCTGTAAGCAAttcaagagaTTGCTTACAGAAACCTACATA-3′

Reverse oligo:

5′-aaaaTATGTAGGTTTCTGTAAGCAAtctcttgaaTTGCTTACAGAAACCTACATA-3′

The sequences are orientated in the 5′ to 3′ direction with the first black sequence representing the 'sense' shRNA sequence and the 'antisense' shRNA sequence following the blue loop sequence – it is the antisense sequence which we expect to be employed as the guide strand of each shRNA. Note the red 'G' in the forward oligo is due to the preference of the U6 promoter to initiate transcription at a 'G' nucleotide (Ma et al., 2014a). Three shRNAs were designed for each candidate gene of interest and the vectors constructed as follows: 9 µl of 100 µM forward and reverse oligos were added to 2 µl of 10X T4 DNA ligase buffer. Annealing of the oligos was carried out in a PCR machine set to 98°C and programmed to reduce the temperature slowly to 20 °C. The annealed oligos were diluted to 100 nM and 4 µl added to 50 ng of purified pRS-Hyg-U6H1 digested with BbsI, 10X T4 DNA ligase buffer and 1 µl of T4 DNA ligase (NEB, M0202L). The samples were made up to 10

μl with dH_2O and incubated at room temperature for 1.5-2 hours, then transformed into DH5α E. coli bacteria (see Section 2.2.5 above).

2.3 Flow cytometry

2.3.1 Flow cytometry materials

All flow cytometry analyses were carried out using the BD LSR Fortessa cell analyzer instrument with harvested cells resuspended in flow cytometry buffer (FB, 2% FBS in PBS) for antibody staining in 96-well U bottom plates, then aliquoted into FACS tubes (BD Falcon, 3520) for analysis.

2.3.2 Antibodies used in this study for flow cytometry analysis

Antibody	Clone	Host & Isotype	Working dilution	eBioscience catalogue number
Anti-Mouse CD54 (ICAM-1) Biotin	YN1/1.7.4	Rat IgG2b, kappa	1/100	13-0541
Anti-Human/Mouse CD44 APC	IM7	Rat IgG2b, kappa	1/300	17-0441
Streptavidin PE-Cy7	-	-	1/1500	25-4317
Anti-Human/Mouse SSEA-1 647	eBioMc	Mouse IgM	1/50	51-8813
Anti Mouse/Rat CD90.1 (Thy1.1) APC	HIS51	Rat IgG2b, kappa	1/100	17-0900
Anti-CD324 (E-cadherin) Biotin	DECMA-1	Rat IgG1	1/100	13-3249

2.3.3 Antibody staining technique for flow cytometry analysis

Harvested cells were counted then centrifuged at 300g and resuspended in FB at a concentration of $1x10^6$ cells/ml. Primary antibodies were added and samples were incubated on ice for 15-30 minutes. Cells were centrifuged, primary antibodies aspirated and cells washed in FB. After centrifugation, the wash step was repeated and cells were resuspended in FB. Secondary antibodies were then added and samples incubated on ice for 5-10 minutes. The wash steps were repeated as before and cells were finally resuspended in FB at a concentration of $2x10^6$ cells/ml and transferred to FACS tubes on ice, ready for analysis using the BD LSR Fortessa analyzer.

2.3.4 Instrument settings for flow cytometry analysis

Flow Cytometry with BD LSR Fortessa		Laser excitation line		
		488 nm	561 nm	640 nm
Band pass (BP) filter	530 ± 30	GFP		
	582 ± 15		mOrange	
	780 ± 60		PE-Cy7	
	670 ± 30			APC

Table 2.1 BD LSR Fortessa instrument settings for flow cytometry analyses. The laser excitation line indicates the operational wavelength of each laser and the band pass filter indicates the range of wavelength of light detected by the instrument.

2.3.5 Flow cytometry gating and data analysis strategies

Gating strategy used for flow cytometry analyses is detailed in O'Malley et al. (2013), specifically Supplementary Figure 4, but briefly described here. Cells were initially gated according to forward- and side-scatter (FSC, SSC), broadly correlating with cells size and granularity, respectively, to enrich for 'live' cells while disregarding dead/dying cells and debris. Early in time course experiments this gating was broad, in line with the heterogeneous nature of reprogramming cultures containing a mixture of both large fibroblast-like cells and progressively smaller reprograming cells. Transgenic cells could be identified from mOrange expression originating from the four-factor polycistronic vector, thus indicating cells specifically undergoing reprograming. In addition to the antibody-specific gene expression profile of interest, depending on the experiment, these cells could also be gated for Nanog-GFP+ or GFP- expression. Correction of any overlap between fluorophores used in these experiments during acquisition (also known as compensation) and data analysis was carried out using Diva or FlowJo software, respectively (BD, TreeStar).

2.3 Genome-wide gene expression analysis

2.3.1 Microarray analysis

Reprogramming samples retrovirally infected with shLacZ, MUT 1, MUT 5 and shDmrtc2 for microarray were collected and RNA isolated as per Section 2.2.2. The samples were then processed using the Illumina® TotalPrep RNA Amplification Kit (Life Technologies, AMIL1791) to produce biotinylated cRNA. The quality of the samples was determined using an Agilent Bioanalyzer, and the samples were subsequently used for microarray analysis using the Illumina MouseWG-6 Gene Expression BeadChip. Microarray analysis was carried out by Louise Evenden at the Wellcome Trust Clinical Research Facility located at the Western General Hospital in Edinburgh.

Data analyses for the microarray was carried out separately by Rafal Gumienny and Alexander Kanitz from Mihaela Zavolan's lab at the University of Basel, Florian Halbritter from Simon Tomlinson's lab at the University of Edinburgh.

2.3.2 RNA-sequencing analysis

Samples of MEF infected with shLacZ, MUT 9, MUT 11 and shDmrtc2 were harvested with trypsin/EDTA and aliquoted 25% and 75% into separate Eppendorf tubes. The tubes were centrifuged at 300 g for 3 minutes, supernatant removed and pellets were snap frozen in liquid nitrogen. These samples were shipped to Afzal Syed in Mihaela Zavolan's lab at the University of Basel for RNA-sequencing (RNA-seq) and Northern Blot analyses. Briefly, the RNA-seq technique involved mRNA isolation directly from frozen cell pellets using the Dynabeads® mRNA DIRECT™ Kit (Life Technologies, 610.11). 100 ng of mRNA was fragmented with alkaline hydrolysis buffer and then cleaned up using the RNeasy MinElute Cleanup Kit

(Qiagen, 74204). The mRNA ends were repaired by dephosphorylation then phosphorylation and the clean up step was repeated. 3'- sequencing adapter ligation was carried out overnight, clean up step repeated followed by 5'- sequencing adaptor ligation overnight. A final clean up step was performed and cDNA prepared. A pilot PCR was carried out followed by the final PCR, and the PCR products from this were cleaned up using the Agencourt AMPure XP kit (Beckman Coulter, A63880). The samples then forwarded to the sequencing facility. RNA-seq analysis was carried out by R. Gumienny using a biophysical modeling approach, MIRZA, to predict putative shDmrtc2 target sites (Khorshid et al., 2013; Gumienny and Zavolan, 2015).

CHAPTER 3 – An RNAi screening to identify novel regulators of reprogramming

3.1 Introduction

The use of reverse genetics to disrupt genes and determine the phenotype has been long used in biology. In the context of stem cell biology, several groups have reported the use of RNA interference (RNAi) to identify genes involved in stem cell self-renewal, pluripotency regulation, chromatin modulation and signaling pathways required to maintain ESC identity (Hu et al., 2009; Chia et al., 2010; Ding et al., 2009; Fazzio et al., 2008). In addition genome-wide and custom knock down screens have been used as a tool to dissect the reprogramming process and distinguish genes and associated pathways that are important during each stage of the transition of somatic cells to pluripotent iPSCs (Yang et al., 2014; Qin et al., 2014; Samavarchi-Tehrani et al., 2010; Sakurai et al., 2014). However, the complex and dynamic nature of reprogramming has made it difficult to pinpoint the exact determinants and sequence of cellular events for a somatic cell to successfully become an iPSC and the low efficiency at which the majority of reprogramming systems operate has hindered this further.

The foundation of the work within this thesis stemmed from the published work of another member of our lab, James O'Malley, who identified two cell surface markers, ICAM1 and CD44, that could be used to track the movement of cells undergoing reprogramming using flow cytometry. Using a secondary reprogramming system with transgenic fibroblasts which carried the four reprogramming factors containing a mOrange reporter and a Nanog-GFP reporter under control of the endogenous Nanog promoter, O'Malley et al. (2013) developed a novel high-resolution system allowing for reprogramming cells to be visualized at a single cell level during reprogramming by following their expression of ICAM1 and CD44. They determined that fibroblasts prior to reprogramming expressed

broad levels of ICAM1 and high expression of CD44. As reprogramming was initiated ICAM1 repression was followed by downregulation of CD44 with iPSCs emerging from the low CD44 population with up-regulation of ICAM1. By sorting 6 different cell populations at day 10 of reprogramming, based on the distinct series of population changes that occurs as determined by ICAM1/CD44 and Nanog-GFP expression, the authors were able to determine global gene expression profiles concurrent with the changing populations during reprogramming, using RNA-sequencing. From the gene expression analysis, five distinct patterns of gene expression were identified (O'Malley et al. (2013), Figure 3A, Groups A to E) and it is clear from a number of other published data sets that distinct gene expression patterns occur during reprogramming (Samavarchi-Tehrani et al., 2010; Sridharan et al., 2009). Group B, that is genes that displayed low expression in fibroblasts and iPSCs with transient up-regulation during reprogramming, was the focus at the foundation of this thesis. This group of genes was particularly interesting as we hypothesized that the transient up-regulation of gene expression that was observed was either required for cells to successfully progress through reprogramming to iPSCs (and therefore overexpression would enhance reprogramming), or this was an aberrant transient up-regulation that hindered efficient reprogramming (and therefore using shRNA to knock down/suppress gene expression would enhance reprogramming). By designing shRNAs targeting the group B genes, those that had a positive or detrimental effect on reprogramming could be easily identified by using the flow cytometry technique described by O'Malley et al. (2013) in addition to simpler assays such as colony counting based on Nanog-GFP expression. Using this strategy I sought to identify novel genes that play a role in the reprogramming process in an attempt to further dissect the mechanism(s) and pathways underlying the successful progression of a somatic cell to an iPSC.

3.1.1. Aims of this chapter

The aims of this chapter are threefold. Firstly, I analysed RNA-seq and microarray data generated both within our lab and others to compile a list of candidate genes for targeted knock down. Secondly, in parallel, I established an efficient and effective RNA interference (RNAi) system to knock down target genes during reprogramming. Finally, I tested the effect of knock down of candidate genes during reprogramming and identified shRNAs targeting candidate genes that gave rise to an enhanced reprogramming phenotype.

3.2 Results

3.2.1. Candidate gene selection for RNAi screening

The starting point for this study was established from earlier work in our lab performed and published by James O'Malley. Detailed in the O'Malley *et al.* (2013) publication, RNA-sequencing data was generated using samples of cells that were at early, intermediate and late phases of reprogramming based on cell sorting using novel cell surface markers ICAM1 and CD44 together with a pluripotency marker Nanog-GFP. J. O'Malley identified five groups of genes based on distinct patterns of gene expression during reprogramming. One of these groups, "group B", contained genes that exhibited low levels of expression in MEF, transiently up-regulated then down-regulated expression when cells reached an iPSC stage. This group of genes we termed "UP-DOWN" genes (Figure 3.1a and b) and this data was the main basis for candidate gene selection for this project. It was unknown as to whether this transient up-regulation was required for reprogramming to progress or simply aberrant up-regulation triggered by strong induction of the four reprogramming factors. We hypothesized that some of these genes may represent the former situation and if so knocking down gene expression early in reprogramming, thereby preventing transient (aberrant) up-regulation, may enhance reprogramming by abolishing this barrier. On the other hand, knock down of genes that are required to transiently up-regulate during reprogramming would result in loss of reprogramming potential. Thus this approach has the potential to identify both barrier and essential genes for reprogramming. A principal component analysis generated from the RNA-sequencing data (Figure 3.1c) indicated that the cells take a detour to reach a pluripotent state (red arrow), instead of taking the shortest route towards iPSCs (blue arrow). I aimed to make a 'shortcut' by inhibiting transient up-regulation of inhibitory genes.

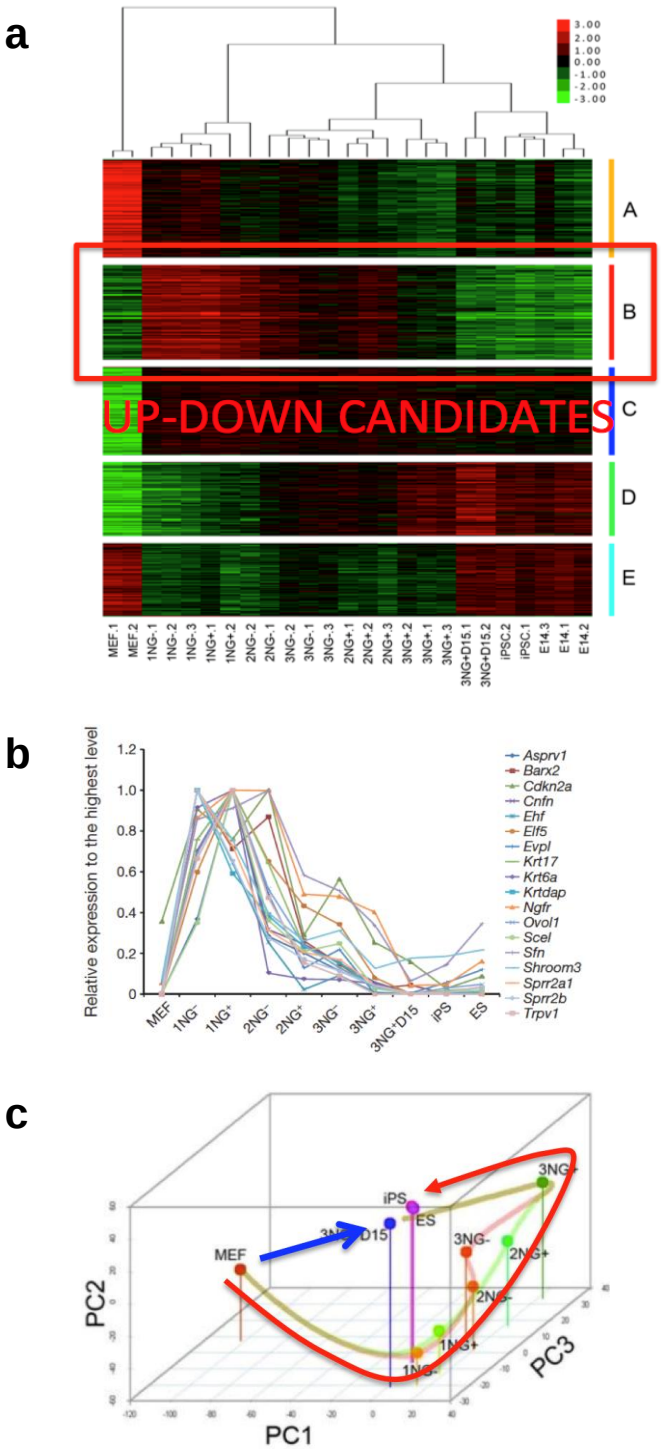


Figure 3.1. Gene expression profile of cells undergoing reprogramming. Five distinct mRNA expression profiles were determined for cells at early, intermediate and late stages of reprogramming (a). An “UP-DOWN” group of genes was identified as having low expression in MEF and iPSCs/ESCs with transient up-regulation occurring during reprogramming (a, b). Principal component analysis demonstrates the usual route of reprogramming MEF to iPSCs (red arrow) with a hypothesized “short-cut” proposed (blue arrow). Adapted from O'Malley et al. (2013).

The UP-DOWN group contained 706 genes, many more genes than could be screened manually as was intended for this project rather than using a high-throughput approach. As a means to narrow down the candidate list to a more practical number, an additional data set published by Samavarchi-Tehrani *et al.* (2010) was considered in conjunction with the UP-DOWN gene list. Briefly, the authors used an RNAi screening approach using short interfering RNA (siRNA) targeting over 4000 genes in order to identify regulators of the early initiation phase of reprogramming. After 5 days of reprogramming, alkaline phosphatase (AP) staining was used as a read out of reprogramming cells and the area of AP staining was calculated for each well. By putting these results in order of AP staining area I was able to narrow down my candidates based on the genes which had a higher AP staining value than p53 on the siRNA list since it is well documented that knocking down p53 during reprogramming leads to enhanced reprogramming efficiency (Banito et al., 2009; Kawamura et al., 2009; Hong et al., 2009). This approach identified 47 "barrier" candidate genes and I reasoned that knock down of these genes during reprogramming might enhance reprogramming as well as or better than knock down of p53. Similarly, 15 "essential" candidate genes were identified as having lower AP staining values than Oct4, which has been reported to decrease reprogramming efficiency when knocked down (Heng et al., 2010; Samavarchi-Tehrani et al., 2010), and I reasoned that if knock down of one of these genes impeded reprogramming more than Oct4 then expression of that gene may be required during reprogramming. This could subsequently be tested with overexpression during reprogramming. In addition to siRNA data, the Samavarchi-Tehrani study also included a microarray time course analysis of gene expression during reprogramming. This data set was analyzed in conjunction with the O'Malley RNA-seq data resulting in elimination of 8 and 6 genes from the "barrier" and "essential" candidate lists, respectively, as they only exhibited the UP-DOWN pattern in one data set. Thus, from this approach 39 potential "barrier" gene candidates and 4 potential "essential" gene candidates were determined.

In addition to these 43 candidates, I made use of unpublished microarray data that I generated during previous work investigating a highly stable iPSC line, C7s2.11, which had lost the capacity to differentiate after it was established from reprogramming of MEF. I reasoned that genes which were strongly up-regulated or down-regulated in this cell line compared with ESCs might represent genes that play an important role in the acquisition and/or maintenance of pluripotency and consequently it would be interesting to including them in my knock down screening. This strategy gave rise to 11 additional candidates. In total, an initial candidate gene list of 54 genes was determined for testing in an RNAi knock down screening during reprogramming (Table 3.1).

9930023K05Rik (Rik)	Mxi1	Taf1b
Adrb2	Nagk	Tcfap2a
Asprv1	Nfe2l3	Tlx2
Bhmt2	Ovol	Fosb
Bmp8b	Peg3	Foxj2
Cldn4	Perp	Jun
Dmkn	Phox2a	Nfatc2
Dmrtc2	Phox2b	Aldh3a1
Drp2	Plcd3	Dkkl1
Elf4	Prx	Ephx1
Elf5	Rapgef4	Fetub
Hand1	Rhox6	Hmgn3
Hat1	Rhox9	Pnpla3
Krtdap	Scel	Rac3
Lgals7	Sfn	Tgm2
Map3k6	Smyd1	Dclk2
Map3k8	Spink2	Dok2
Mid1	Stk19	Emp2

Table 3.1. List of candidate genes identified from published and unpublished data sets. Candidate genes identified from the O'Malley et al. (2013) and Samavarchi-Tehrani et al. (2010) data sets are grouped according to identification as UP-DOWN genes >AP+ than p53 siRNA genes (orange) or UP-DOWN genes <AP+ than Oct4 siRNA genes (blue). Genes identified from unpublished microarray data are grouped according to identification as highly expressed in C7s2.11 iPSCs compared with ESCs (green) or highly expressed in ESCs compared with C7s2.11 iPSCs (yellow).

3.2.2. A retrovirus-based shRNA expression vector for efficient knock down of target genes

Reprogramming is a notoriously inefficient and slow process. In addition, using transient siRNA transfection for knocking down genes during reprogramming can also be ineffective at reducing gene expression to a suitable level. With these drawbacks in mind it was important to design a highly efficient and practically straightforward system for screening the knock down of candidate genes during reprogramming. In order to do this, several points were considered:

(1) Reprogramming is inefficient and typically less than 1% of cells become iPSCs.

(2) Reprogramming is highly heterogeneous and many factors influence the success of reprogramming including the reprogramming system used, specific reagents used, the condition of starting cells etc. and this can make interpretation of reprogramming data difficult from one experiment to the next.

(3) For this screening, cells must contain all four reprogramming factors in addition to the knock down vector; both induction of reprogramming and stable transfection of shRNA must be efficient.

(4) RNAi knock down systems are complicated and can be ineffective at reducing gene expression to a sufficient level to induce a phenotype and additionally different genes require different levels of knock down to have an effect.

We use a robust and reproducible reprogramming system taking advantage of transgenic MEF carrying a polycistronic cassette encoding the four

reprogramming factors, c-Myc, Klf4, Oct4 and Sox2, under control of a tetO doxycycline-inducible promoter, and constitutively expressed reverse tetracycline transactivator (rtTA) (Figure 3.2). The four factors are separated with self-cleaving 2A-peptides and followed by ires-mOrange, allowing clear visualisation of expression of the transgene cassette. In addition the MEFs harbour a GFP reporter under control of the endogenous Nanog promoter, a gene that is expressed towards the end of reprogramming when cells acquire pluripotency. We use the expression of Nanog-GFP as a read out of reprogrammed cells/colonies (referred to as 'colony counting' herein). This system limits some of the heterogeneity of reprogramming by ensuring all transgenic cells, at least in theory, have equal capacity to express the four factors at similar levels (this reprogramming system will be referred to as 'transgenic (Tg) reprogramming' herein).

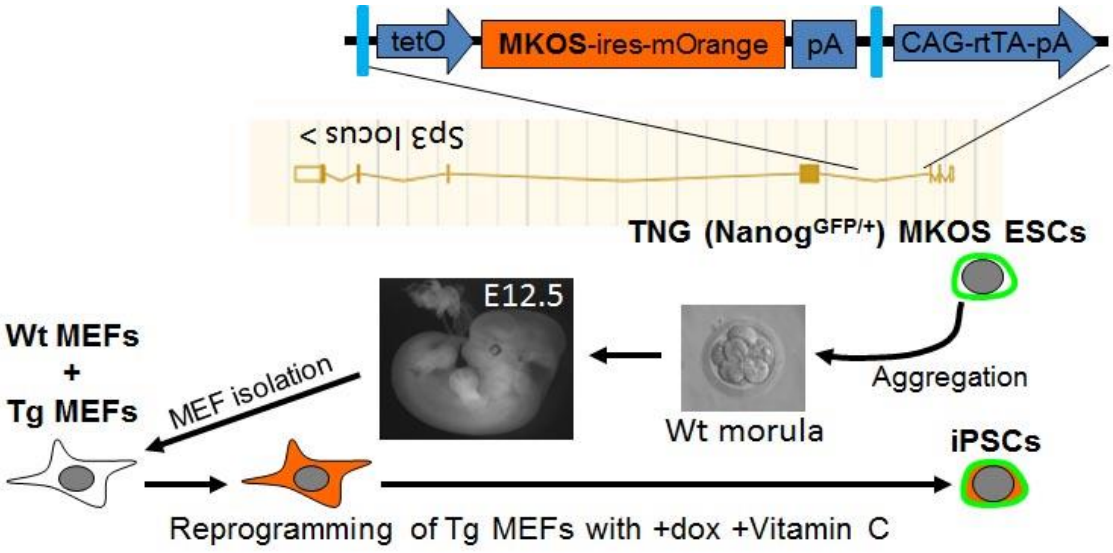


Figure 3.2. Generation of transgenic MEFs for reprogramming. A doxycycline-inducible vector containing MKOS and mOrange reporter was targeted in ESCs containing a Nanog-GFP reporter. These cells were used to generate chimeric mice from which transgenic MEFs were isolated. Expression of MKOS was induced with administration of dox and visualized with mOrange expression and Nanog-GFP+ iPSCs were generated.

To knock down my candidate genes, I designed a retrovirus based short hairpin RNA (shRNA) expression vector, building on the pRetroSuper plasmid backbone, containing dual converging RNA polymerase III promoters, U6 and H1. This dual promoter system enabled the shRNA to be expressed in both directions thereby ensuring high expression of shRNA could be achieved from a single vector (Zheng et al., 2004; Chen et al., 2005). In addition, a BbsI restriction enzyme site was introduced between the two promoters, enabling highly efficient cloning of any shRNA sequence of interest. A virus packaging cell line, PlatE, was transfected with each shRNA vector to produce virus supernatant containing retrovirus-carrying shRNAs. This supernatant was then used to infect our transgenic MEFs for 24 hours and replaced with dox containing medium to initiate reprogramming. The reprogramming cultures were monitored every day for appearance of iPSC colonies and Nanog-GFP+ colonies were counted from day 10 onwards (Figure 3.3).

One of the main drawbacks of using an shRNA knock down system is that designing oligos targeting your gene of interest is based on computational algorithms following a set of 'guideline rules'. The sequences are not experimentally validated and as such are not guaranteed to efficiently target a gene resulting in knock down. Although it is possible to scan the literature to find sequences that have been reported to efficiently knock down specific genes, this is not feasible for a larger scale screen, particularly involving genes that have not been widely studied or well characterized. Consequently, it is advised to test more than one shRNA sequence for each gene of interest in order to increase the likelihood that knock down can be achieved. For this reason, I chose to design three shRNAs per gene of interest (denoted P1, P2 or P3), resulting in construction of over 150 vectors representing the candidate genes. For an unknown reason, cloning of some of the shRNAs was difficult and could not be completed for all shRNAs targeting each of the 54 candidate genes as persistent mutations occurred. As a result, shRNA vectors were generated for only 44 of the 54 original candidate genes, with some candidates represented by less than 3 vectors, in addition to control vectors targeting LacZ. A retroviral vector expressing DsRed (pMXs-DsRed) was used as a virus infection

control and by FACS analysis I could reproducibly achieve 50-75% infection efficiency as determined by DsRed positive cells.

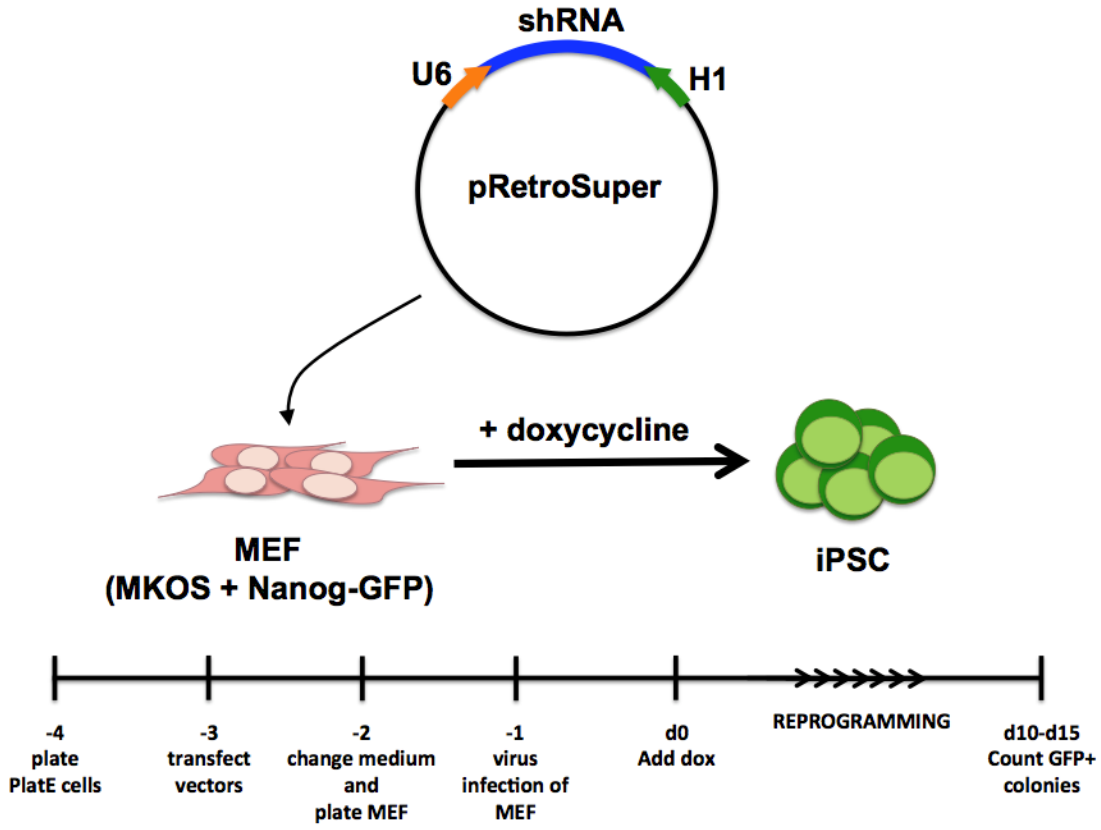


Figure 3.3. Schematic of shRNA knock down of candidate genes during reprogramming. pRetroSuper shRNA vectors were cloned for each candidate gene and used to make virus supernatant with which transgenic MEF carrying doxycycline-inducible four reprogramming factors were infected. Dox was administered to initiate reprogramming and Nanog-GFP+ colonies were counted from day 10 onwards.

3.2.3 A custom shRNA screening to identify novel regulators of reprogramming

Five initial rounds of reprogramming with individual shRNAs were carried out as a first pass screening. Approximately 24 to 38 shRNAs were tested in each experiment with all shRNAs representing any given gene tested together within the same experiment to enable side-by-side comparison. However, it is important to note that since viral titre was not determined for each shRNA in any of the experiments, and only DsRed was used as an infection control, it is highly likely that the viral titre produced for each shRNA differed. As such, these preliminary experiments were being used to determine several things; 1) To confirm that method of shRNA vector transfection could successfully produce viral supernatant, 2) To confirm that any virus supernatant produced could successfully infect MEF, 3) To determine whether the transgenic MEF could still undergo reprogramming upon retroviral infection and finally 4) To determine if the addition of any of the candidate shRNAs to reprogramming cells enhances or hindered the generation of Nanog-GFP+ iPSCs, when compared with the control. From experience of using this reprogramming system within our lab, we know that Nanog-GFP+ colonies typically establish by around day 10 and therefore Nanog-GFP+ colonies were counted on days 10, 13 and 15 for this preliminary screen (Figure 3.4).

For all experiments 5% transgenic MEFs were plated in 6 wells, with wild type MEF used to dilute the cells to the appropriate ratio. This equates to approximately 5000 transgenic cells per well that are capable of reactivating the reprogramming factors upon dox treatment and undergoing reprogramming. With this in mind, it should be noted that our Tg reprogramming system is highly efficient, as can be seen from the experiments where uninfected MEF were reprogrammed (Figure 3.3). In these experiments, the reprogramming efficiency varied between 2-7%. An initial observation found that infection of MEF with retroviral shRNAs significantly hindered reprogramming of transgenic cells, as evident from a substantial decrease in Nanog-GFP+ colony number in DsRed and shLacZ conditions when compared with uninfected reprogramming conditions. It

seems that viral infection and/or expression of exogenous shRNA is somewhat toxic to the cells. Despite this, even with reduced efficiency the control shLacZ condition continued to display reprogramming efficiencies of typically 1-3%, which is higher than many reported reprogramming systems.

When shRNAs targeting candidate genes were tested, several generated an increased number of Nanog-GFP+ colonies when compared to shLacZ control on day 15, and a few shRNAs even generated more colonies than the uninfected condition including shRNAs targeting Aldh3a1, Rac3, Emp3, Nagk and Dmrtc2. Interestingly even as early as day 10 many shRNAs were found to give several fold increase in Nanog-GFP+ colonies, indicating that knock down of these genes might enhance reprogramming kinetics in addition to enhanced efficiency. The most significant of these included shRNAs targeting Rhox9, Hand1, Mxi1, Nfe2l3, Phox2a, Phox2b, Prx, Rapgef4, Aldh3a1, Dkkl1, Rac3, Emp2, Elf4 and Dmrtc2. Notably, an shRNA targeting Dmrtc2 (shDmrtc2 P3) was determined as a significant enhancer of reprogramming, and this was confirmed twice in these preliminary experiments.

Figure 3.4. Effect of candidate shRNAs during reprogramming. Five initial reprogramming experiments reveal shRNAs that enhance or inhibit generation of Nanog-GFP+ colonies during reprogramming of transgenic MEF. Bars represent the average number of Nanog-GFP+ colonies counted from 2 independent wells for each shRNA.

A number of shRNAs also resulted in a decrease in colony number, representing a number of potential "essential" gene candidates for reprogramming. These included shRNAs targeting Ovol1, Sfn, Fosb and Tgm2. Interestingly, Ovol1 has been implicated in driving MET in cancer (Roca et al., 2013), Sfn (also known as 14-3-3σ) plays a role in proliferation of ESCs through binding of GSK-3β (Chang et al., 2012), Fosb has been identified as a reprogramming factor for generating induced haematopoietic stem cells from HUVECs (Lucas and Frenette, 2014) and Tgm2 is an enzyme that has been implicated with diverse roles in cell adhesion, proliferation and apoptosis depending on it's locality within the cell (Nadalutti et al., 2011). Thus, some of these genes might feasibly play a role as positive regulators or enhancers of reprogramming, although the specific knock down of these genes was not validated. The effect of these shRNAs during reprogramming was not followed up with overexpression experiments to see if expression of their cDNA together with Yamanaka factors could enhance reprogramming. This was because reduction of reprogramming efficiency can be caused many reasons and even control shRNA vector infection causes reduced reprogramming efficiency compared to a non-infection control. Less iPSC colonies by an shRNA may be due to higher off-target toxicity of the shRNA. Moreover, this initial screen already identified several potential 'barrier' candidates. We imagined an enhanced reprograming phenotype was less likely to be caused by off-target effects. Nonetheless, further investigation of the "essential" candidates mentioned above is a potential avenue of further study.

Going forward, 20 shRNAs representing 17 genes were chosen for further validation during further rounds of reprogramming. This second round of screening enabled exclusion of many candidates as there was no significant difference compared with control in the number of Nanog-GFP+ colonies at day 10 or later. However, several shRNAs gave rise to a 5 to 10-fold increase in the number of Nanog-GFP+ colonies at day 10, compared with control. These included shRNAs targeting Emp2, Rac3, Rhox9, Dkkl1 and Elf4 (Figure 3.5). Additionally, this experiment confirmed the earlier observation that one shDmrtc2 P3 in particular

induced a significantly enhanced reprogramming phenotype. Remarkably, at day 10 around a 45-fold increase in Nanog-GFP+ colony number was observed and by day 14 a 10-fold increase was maintained when this shRNA was expressed. This is at a time point when we know that most colonies have become iPSCs and express Nanog-GFP. This finding indicated that not only did shDmrtc2 P3 enhance reprogramming efficiency in terms of the total number of colonies obtained at the end of reprogramming, but also accelerated reprogramming kinetics as demonstrated by the 45-fold increase in colony number at day 10.

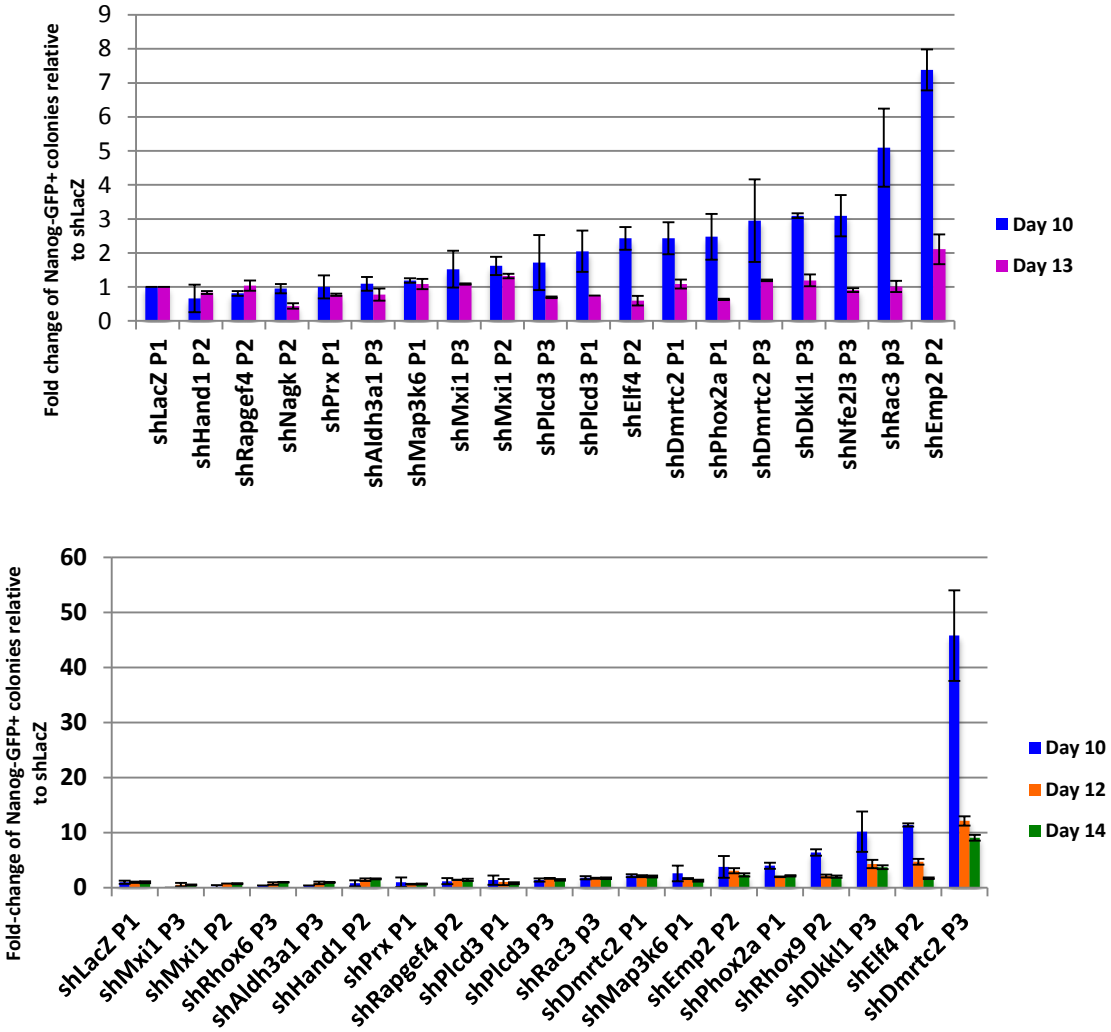


Figure 3.5. Further validation of candidate shRNAs during reprogramming. Most shRNAs could be excluded from these results for further testing as no significant difference in Nanog-GFP+ colony number was observed when compared with control. However, several shRNAs were identified as enhancers of reprogramming, including those targeting Emp2, Rac3, Rhox9, Dkkl1, Elf4 and most significantly Dmrtc2.

In addition to assessing the number of Nanog-GFP+ colonies during reprogramming I also took advantage of a technique developed in our lab using fluorescence activated cell sorting (FACS) analysis to track cells as they undergo reprogramming using novel cell surface markers ICAM1 and CD44 (O'Malley et al., 2013). During reprogramming, cells express ICAM1 heterogeneously and are positive for CD44 expression. As reprogramming progresses, ICAM1 is down-regulated followed by down-regulation of CD44 and finally, as cells become iPSCs they once again express high levels of ICAM1. This ICAM1+/CD44- position is where iPSCs and ESCs are found (Figure 3.6a). This technique allows for changes in the 'normal' or 'typical' FACS profile to be easily identified, for example if the cells move through the 'reprogramming route' at a faster rate, and thus is an ideal tool to identify if an added factor has a positive effect on reprogramming kinetics. This analysis also takes advantage of the Nanog-GFP reporter carried by the cells, enabling the percentage of reprogramming (transgenic) cells that express Nanog-GFP to be determined at any time point throughout the experiment. Again, this can be used as further validation if an added factor has a positive effect on reprogramming.

I chose to perform FACS analysis for three top shRNAs based on the preliminary Nanog-GFP+ colony count data; shDmrtc2 P3, shDkkl1 P3 and shEmp2 P2. I predicted that if any of these shRNAs did indeed enhance reprogramming efficiency then I would observe an increase in the percentage of Nanog-GFP+ cells and if there was a positive effect on the timing or kinetics of reprogramming then I would observe the cells moving through the ICAM1/CD44 profile faster and/or expressing Nanog-GFP earlier. Indeed this is what I observed; as expected, the shLacZ control and uninfected samples exhibited similar FACS profiles throughout the time course as reported by O'Malley et al. (2013). Albeit there were more Nanog-GFP+ cells in the uninfected sample (28%) compared with shLacZ (2%) at day 10 (Figure 3.6b). This is unsurprising given that infection with retrovirus of the shRNAs seems to be somewhat toxic as mentioned previously, however, by day 13 both conditions give rise to a similar proportion of Nanog-GFP+ cells. In contrast, all

three shRNAs had a positive effect on reprogramming. shDkkl1 P3 gave rise to 56% Nanog-GFP+ cells at day 13 compared with 36% for shLacZ. In addition, the appearance of Nanog-GFP+ cells was already evident around day 7 of reprogramming compared with day 10 for shLacZ. Even more strikingly, shEmp2 P2 gave rise to 69% Nanog-GFP+ cells at day 13 and by day 10 had already matched the percentage observed in the uninfected condition. The most pronounced result, however, was observed for shDmrtc2 P3. By day 10, 58% of cells were already positive for the pluripotency marker Nanog-GFP, increasing to 77% by day 13. Even more remarkably, there was a clear acceleration in the rate of reprogramming observed with most cells already having down-regulated CD44 and many up-regulating ICAM1 by day 7. This is consistent with two to three day acceleration in the rate of reprogramming compared to control. This striking result led me to focus my investigation on shDmrtc2 P3 (referred to as shDmrtc2 herein).

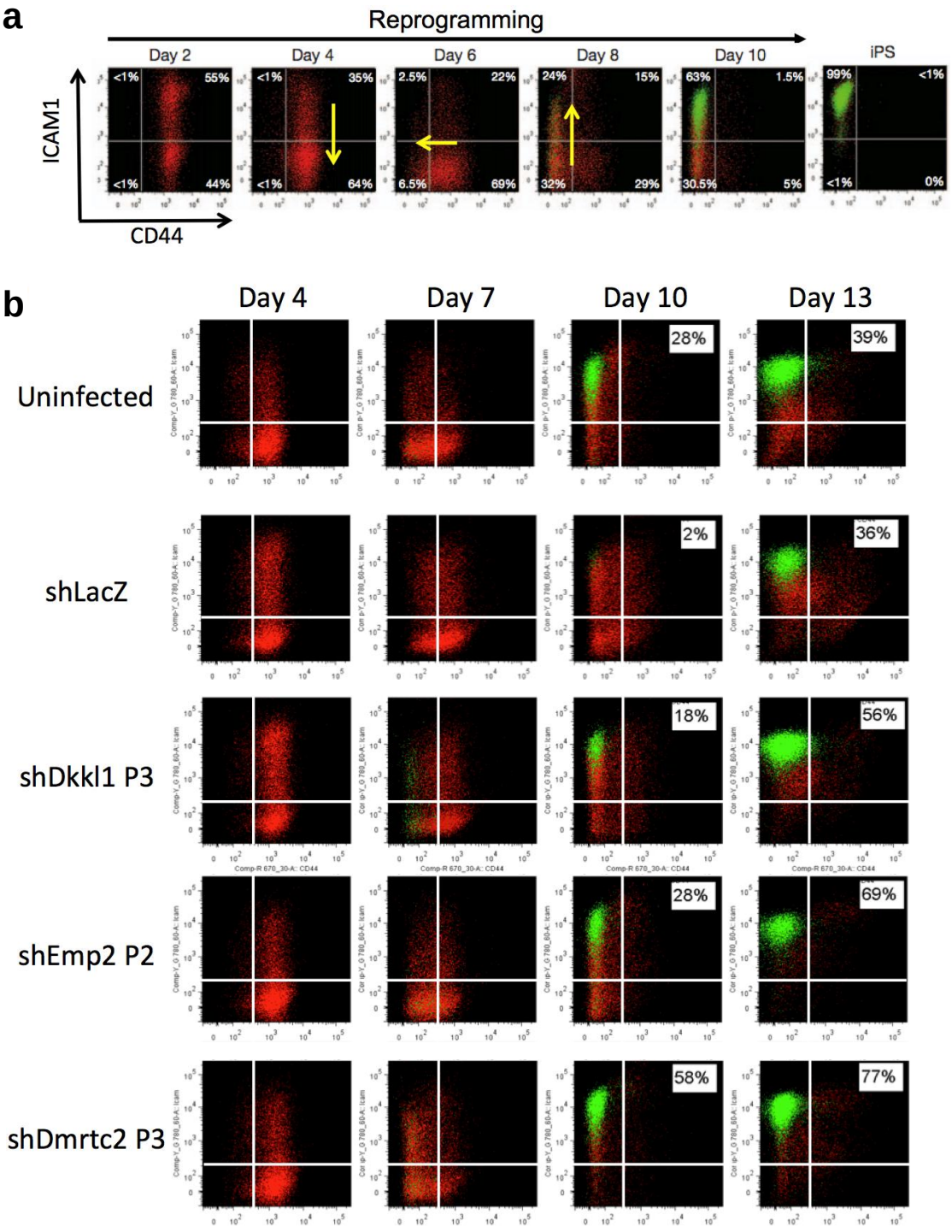


Figure 3.6. FACS analysis of ICAM1, CD44 and Nanog-GFP expression in cells undergoing reprogramming. (a) Cells undergoing reprogramming follow a specific "route to iPSCs" determined by ICAM1 and CD44 expression (adapted from O'Malley et al. (2013)). (b) Addition of shDkkl1 P3, shEmp2 P2 and shDmrtc2 P3 give rise to enhanced reprogramming efficiency with shDmrtc2 P3 showing significantly faster ICAM1/CD44 kinetics as early as day 7, and a majority of cells expressing Nanog-GFP by day 13.

3.3 Discussion

3.3.1 shRNA is a powerful tool to identify novel regulators of reprogramming

These initial experiments have provided evidence to support published studies that shRNA can be a powerful tool to knock down gene expression, and in this case to successfully uncover novel roles for genes during reprogramming. shRNA expression vectors containing dual RNA Polymerase III promoters, U6 and H1, were used in this study over the option to purchase synthetic siRNA for several reasons. Firstly, siRNA can be a costly option, particularly when many siRNAs are required to knock down many genes, for example in the case of an siRNA library with prices usually upwards of several thousands of pounds. In contrast, the shRNA vectors used in this study were constructed at a low cost, facilitated by the fact that the pRetroSuper backbone was able to be propagated indefinitely, and the shRNA cloning technique was highly efficient resulting in quick and easy validation. In addition, contrary to the transient nature of siRNA, shRNA expression vectors allowed for persistent and high expression of shRNAs within the reprogramming cells, alleviating the requirement for serial transfection as is required with siRNA which is depleted in cells over time unless replenished by further transfections. The fact that reprogramming cells form colonies may even hinder or abolish the possibility of efficiency repeated siRNA transfection during reprogramming. Furthermore, since I had devised a relatively small candidate list, the screen was practically simple and specific shRNAs targeting the genes of interest could be tested individually, abrogating the need to carry out pooled shRNA experiments which can make interpretation and validation of results more complicated and time consuming. In other words, positive or negative hits could be easily traced to specific genes without the need for further screening to link a phenotype to one of many pooled shRNAs. One potential drawback of both siRNA and shRNA-mediated knock down in vitro has been identified in reports that overload of ectopic si/shRNAs can compete with endogenous miRNAs for

processing machinery and lead to toxicity as a result (Grimm et al., 2006). However, designing Ago-2 specific RNAi is one way to alleviate this miRNA perturbation (Ma et al., 2014b). It seems that the shRNA expression vectors used in this study were somewhat toxic to cells when compared with an uninfected control and it was unknown whether this was due to virus infection of cells or shRNA expression itself. However, infection of cells with a pMXs-DsRed vector gave rise to similar results as shLacZ infection, suggesting that it was virus infection in general rather than shRNA expression that is toxic. Since several shRNAs in our screening gave rise to enhanced reprogramming even when compared with the uninfected control, this phenomenon didn't seem to be a major factor in this context. In fact comparison of candidate shRNAs to the uninfected control ensured that only shRNAs that gave rise to a greatly enhanced reprogramming phenotype were taken forward in the study as these were predicted to be the most likely to have reproducible and robust effect.

Another point for consideration is our use of a dual promoter vector using converging U6 and H1 promoters to express shRNAs from both strands of the same DNA template. Having had limited experience or expertise within our lab of using RNAi at the start of this study, we theorized that a dual promoter system would be more likely to ensure high expression of shRNA than a single promoter system. Indeed it has been widely reported that a U6/H1 dual promoter RNAi expression vector can mediate efficient gene knock down (Kaykas and Moon, 2004; Zheng et al., 2004; Chen et al., 2005). Early experiments comparing single U6 promoter vectors with dual U6/H1 converging promoter vectors in our hands gave inconclusive results as to whether one system was more efficient than the other (not shown), and as such we opted to proceed with the dual promoter vector with the reasoning that it had the potential to produce a higher level of shRNA expression.

As already briefly mentioned, the reprogramming system we use in our lab is generally robust and highly efficient. However, as can be seen clearly in the data presented within this chapter, the number of iPSC colonies generated during each reprogramming experiment is greatly variable, even for the same batch of

transgenic cells reprogrammed in the same way. Another difficulty comes with additional layers required for reprogramming experiments, for example, the necessity for virus infection in these experiments provides a further source of variability to the results. Nonetheless, results were easily reproducible in terms of the phenotype observed, however, consideration must be made to take into account the interpretation and presentation of these data.

When these considerations are taken together, this shRNA screening proved to be a powerful tool to identify several novel regulators of reprogramming, with shDmrtc2 being found to have a significant effect on reprogramming efficiency and kinetics.

3.3.2. FACS analysis of ICAM1 and CD44 is a useful tool for studying reprogramming kinetics

It is commonplace in the reprogramming field to quantify reprogramming efficiency as the number iPSC-like colonies generated from an initial number of reprogramming cells at a certain arbitrary time point. Furthermore, although many studies have identified factors or small molecules as enhancers of reprogramming efficiency, few studies have reported drastic or reproducible improvement in the kinetics, or timing, or reprogramming. And since reprogramming is widely known to be a slow and inefficient process, the identification of shDmrtc2 P3 as both an enhancer of reprogramming efficiency and kinetics is an exciting prospect. In conjunction, the ICAM1/CD44 FACS analysis technique developed in our lab has provided a unique tool with which investigation of reprogramming kinetics is easy and clear. We can determine, literally at a glance, whether any condition/factor/small molecule has a significant effect on reprogramming using this method, and when used in conjunction with additional markers such as Nanog-GFP, a more complete picture of the route by which any given cells undergo reprogramming can be produced. In addition to enhancement, aberrations or

barriers can also be easily identified, and when investigating several factors (in this case several different shRNAs), a meaningful comparison can be made and robust conclusions drawn since single cells are analysed and visualised, as opposed to using somewhat arbitrary or subjective methods such as colony counting to determine reprogramming efficiency.

3.3.3. The importance of experimental controls

Since these reprogramming experiments were carried out in conjunction with retroviral transduction (to express the shRNA vectors), it was important to determine what impact, if any, viral transduction would have on reprogramming efficiency and furthermore to choose an appropriate shRNA control for which to compare all candidate gene shRNAs with. Notably, it was clear that viral transduction had a detrimental effect on reprogramming, as several fold fewer Nanog-GFP+ colonies were observed when cells were transduced either with a retroviral shRNA or non-shRNA control vector. When initially designing control shRNA vectors, a number of vectors were tested side by side and shLacZ was determined at the time to be the most appropriate shRNA control for a number of reasons. Firstly, an shRNA targeting LacZ was thought to have few or no off-targets since LacZ is not a gene expressed in the mouse, and therefore it was reasoned that this vector should produce shRNA without a target (and therefore have little or no effect on reprogramming). Secondly, when tested alongside a non-shRNA retroviral vector (pMXs-DsRed), a consistently similar number of Nanog-GFP iPSC colonies were observed in experiments. This suggested that the decrease in reprogramming efficiency when compared with the uninfected control was probably due to viral transduction rather than toxicity of shRNA production. Thirdly, another shRNA targeting a different sequence of LacZ (shLacZ P2) was also tested, however, this gave rise to a similar number of Nanog-GFP+ colonies as the uninfected control. One reason for this could be that viral production/infection was inefficient with this

vector so it was excluded as a reliable control. Based on these observations it was determined that shLacZ would be an appropriate control vector. In hindsight, it would have been useful to have a positive shRNA control, for example targeting p53. Knock down of p53 is well reported to enhance reprogramming efficiency and this could have been a useful control to determine the extent of any positive effects observed with transduction of group B candidate gene shRNAs on reprogramming. Although I am confident that I have identified several shRNAs that enhance reprogramming efficiency and kinetics in this chapter, by including an shRNA that is known to enhance reprogramming would only serve to strengthen my results.

CHAPTER 4 - Investigation of shDmrtc2 as a novel positive regulator of reprogramming

4.1 Introduction

Dmrtc2, also known as Dmrt7, is a protein-coding gene belonging to the doublesex and mab-3-related transcription factor (Dmrt) family. This family of genes contains a characteristic cysteine-rich DNA binding motif known as the DM domain. This facilitates DNA binding through a highly intertwined structure that chelates zinc, allowing binding to the major groove of DNA. Outside of this DM domain, the genes have very little sequence similarity. Interestingly, Dmrt7 and Dmrt8 are only found in mammals and no other vertebrate species, suggesting they are mammalian specific Dmrt genes (Hong et al., 2007). Dmrtc2 is expressed in embryonic gonadal tissue, specifically the ovary or testes with higher abundance found in the female rather than male gonads. However, Dmrtc2 becomes expressed in a male-specific manner postnatally and is required for spermatogenesis. Dmrt7 knock out mice have been generated by gene targeting and were found to be developmentally indistinguishable from littermates and could grow into adulthood suggesting Dmrt7 is dispensable for embryonic development. However mutant mice were infertile and no sperm could be detected in the epididymis of $Dmrt7^{-/-}$ males (Kawamata and Nishimori, 2006). This is reportedly due to an arrest of spermatogenesis at the pachytene stage of meiosis (Kim et al., 2007). Interestingly, there has been very little reported on this gene in recent years.

RNAi can be a very useful but imperfect system for knock down of gene expression. Since RNAi systems utilize endogenous machinery to process si/shRNA, there are a number of factors that must be considered and validated when using such systems to ensure not only robust knock down of genes of interest but also to ensure confidence that the results obtained are a direct consequence of specific knock down of a gene and not some other secondary cause or effect which can be a

common problem with RNAi strategies (Singh et al., 2011; Jackson et al., 2003; Echeverri et al., 2006). Indeed, published studies usually provide functional validation of genes identified through RNAi screens (Yang et al., 2014; Qin et al., 2014). At the very least, a hit obtained through RNAi studies should be validated by further RNAi targeting the proposed gene, and where possible, a rescue experiment should be carried out to confirm that overexpression of the gene, in a form that cannot be targeted by the RNAi, abolishes the phenotype observed with knock down (Kittler et al., 2005).

4.1.1. Aims of this chapter

In the previous chapter an shRNA targeting Dmrtc2 was identified to significantly enhance Nanog-GFP+ colony number and kinetics when applied during reprogramming of transgenic MEF. Following on from this the aims of this chapter are to characterize further the effect of shDmrtc2 during reprogramming, to validate that shDmrtc2 does target Dmrtc2 causing knock down of this gene at an mRNA level as expected and to further validate whether it is the knock down of Dmrtc2 that is responsible for the positive phenotype observed during reprogramming by performing a rescue experiment.

4.2. Results

4.2.1. Validation of shDmrtc2 as an enhancer of reprogramming efficiency and kinetics

Following on from identification of shDmrtc2 as a positive regulator of reprogramming, several repeat experiments were conducted with shDmrtc2 in order to confirm that the strong positive phenotype observed was in fact true and reproducible. This was particularly important because of the variability of reprogramming which can make consistency in reprogramming data difficult, not least between labs but even within the same hands. To overcome this, I carried out several additional experiments to be certain that the phenotype observed using shDmrtc2 during reprogramming was reproducible.

Firstly I confirmed that the increased reprogramming efficiency conferred by shDmrtc2 during reprogramming could be replicated. Indeed, upon repeating several more colony counting experiments, I could faithfully reproduce my previous results; addition of shDmrtc2 to reprogramming resulted in over 40-fold increase in Nanog-GFP+ colonies compared to shLacZ (Figure 4.1a). Although the total colony number varied between experiments I consistently observed increased relative colony numbers in the shDmrtc2 sample compared with both the uninfected and shLacZ samples indicating that this is a real phenotype due to the shDmrtc2 vector (Figure 4.1b).

Since addition of shDmrtc2 in reprogramming resulted in many Nanog-GFP+ colonies appearing by day 10 and FACS analysis indicated the presence of Nanog-GFP+ cells as early as day 7 I sought to determine the earliest time point at which Nanog-GFP+ colonies emerged during reprogramming with shDmrtc2. To do this I monitored the appearance of colonies during reprogramming and tracked individual colonies throughout their progression to iPSCs (Figure 4.2). Within only a few days of reprogramming induction, a change of cell morphology was clearly evident in the shDmrtc2 condition, with cells becoming more compact and rounded

in shape, and clear formation of small foci in the cultures. Strikingly, I observed the appearance of bright green Nanog-GFP+ colonies as early as day 5 of reprogramming with shDmrtc2. By comparison, the first faint green colonies appeared on day 7 in the shLacZ control condition (Figure 4.2). This confirmed that the addition of shDmrtc2 to reprogramming resulted in at least a 2 to 3 day acceleration of the appearance of Nanog-GFP+ colonies compared with control. Given that reprogramming is a notoriously slow process, this was a particularly encouraging and exciting result.

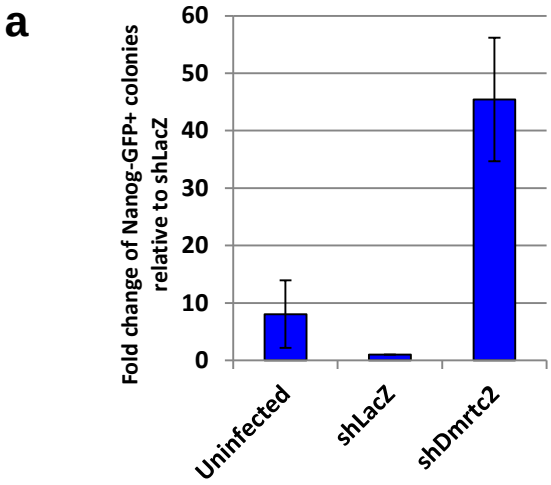
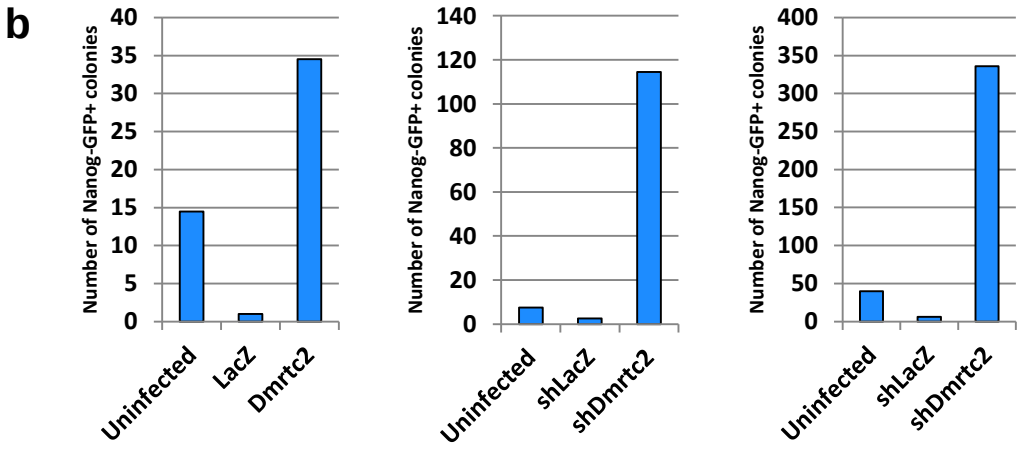


Figure 4.1 Validation of shDmrtc2 during reprogramming. Addition of shDmrtc2 during reprogramming reproducibly enhances the number of Nanog-GFP+ colonies by day 10 compared with control, represented as fold-change relative to shLacZ (a) or absolute colony number from 3 independent experiments (b).



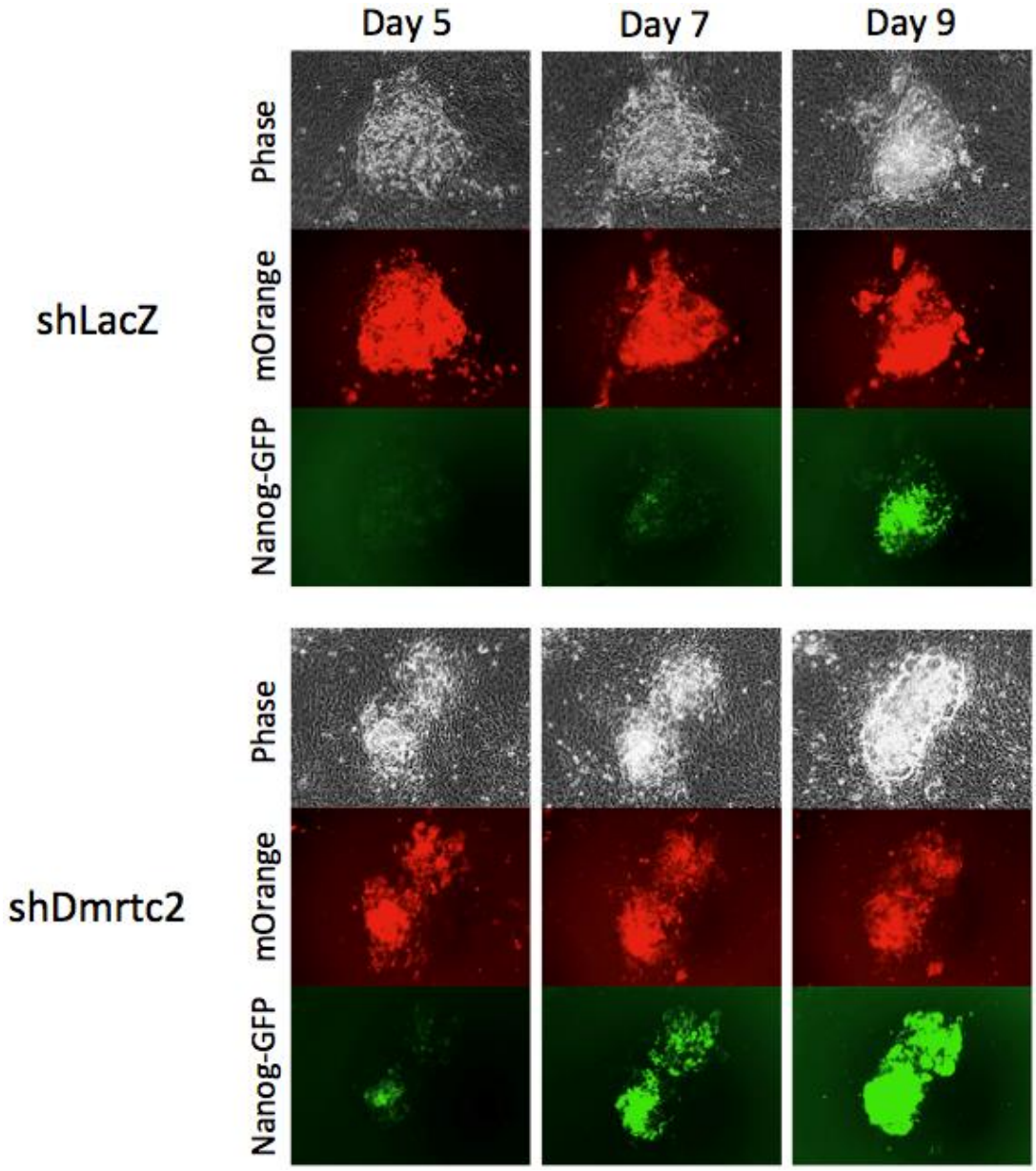


Figure 4.2. Colony tracking reveals accelerated appearance of Nanog-GFP+ colonies with shDmrtc2 during reprogramming. Bright green Nanog-GFP+ colonies appeared as early as day 5 of reprogramming with shDmrtc2. By comparison, there was a 2 to 3 day delay in the appearance of Nanog-GFP+ colonies with shLacZ control.

Small molecule inhibition of Tgfβ receptor Alk5 (Alki) has been reported to enhance reprogramming and this has been confirmed in our lab (Li et al., 2009b; Maherali and Hochedlinger, 2009). The addition of Alki and vitamin C to reprogramming cultures is routinely used within our lab as a 'best condition' for reprogramming. That is, the most efficient reprogramming system in our hands. As such, we sought to find a novel factor that enhances reprogramming even further in these conditions and tested shDmrtc2 as a potential candidate. When Alki was added to the shRNA reprogramming cultures, an increased number of Nanog-GFP+ colonies were observed in the control conditions on day 9, as expected (Figure 4.3). Strikingly, in this enhanced condition shDmrtc2 resulted in a 5 to 8-fold increase in Nanog-GFP+ colonies in the uninfected and shLacZ controls, respectively. When reprogramming in the absence of Alki was carried out in parallel, Alki was surprisingly found to have a negative effect on reprogramming with shDmrtc2. In fact, in the absence of Alki, shDmrtc2 gave rise to approximately 14 to 100-fold more colonies than in both control conditions. Therefore, shDmrtc2 was found to greatly promote reprogramming in both the presence and absence of Alki.

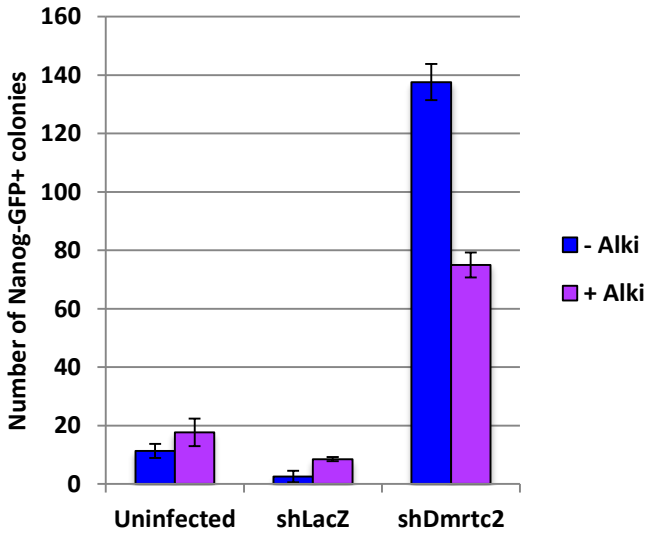


Figure 4.3. The effect of Alki on reprogramming with shDmrtc2. Addition of Alki enhanced reprogramming efficiency in the uninfected and shLacZ control conditions but hindered reprogramming with shDmrtc2. Despite this, shDmrtc2 generated many fold more Nanog-GFP+ colonies than controls in all conditions.

After successfully confirming the phenotype in our transgenic reprogramming system, it was important to replicate this experiment using a different system to be confident that this result was not simply an anomaly of the reprogramming system used. Accordingly, I used a primary reprogramming system (see Chapter 2 – Materials and Methods) to test shDmrtc2. Briefly, to carry out primary reprogramming we use *piggyBac* transposon to deliver a polycistronic cassette carrying the four factors by co-transfection with piggyBac transposase into MEF constitutively expressing rtTA from the Rosa 26 locus (Rosa rtTA), and carrying a Nanog-GFP reporter. Again, four factor induction could be monitored by mOrange expression and Nanog-GFP+ colonies are counted from day 10. shDmrtc2 was delivered by retrovirus infection as before, after transfection of the four factor cassette. Using this primary reprogramming approach I was able to successfully replicate the enhanced reprogramming phenotype (Figure 4.4), with over 30-fold more Nanog-GFP+ colonies counted with shDmrtc2 compared with shLacZ control, providing evidence that this phenotype is a robust and true result of addition of shDmrtc2 during reprogramming.

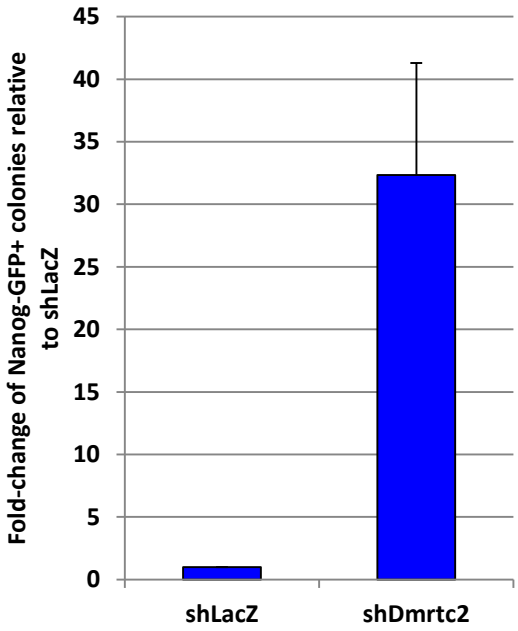


Figure 4.4. Enhanced reprogramming with shDmrtc2 is reproducible in a piggyBac-mediated primary reprogramming system. When shDmrtc2 was applied to primary reprogramming, a 30-fold increase in Nanog-GFP+ colonies was observed on day 10, compared with shLacZ control.

Furthermore, consistent with previous data, when individual colonies were monitored, I could observe a 4 day acceleration in the appearance of Nanog-GFP+ colonies with shDmrtc2 compared with shLacZ in the primary reprogramming context (Figure 4.5). Bright green colonies were evident by day 10 during reprogramming with shDmrtc2 with the equivalent type of colony appearing in the control condition by day 14. Notably these time points are several days later than those observed in the secondary reprogramming system but this is likely due to the more efficient nature of secondary reprogramming compared with primary reprogramming. Importantly the phenotype is conserved.

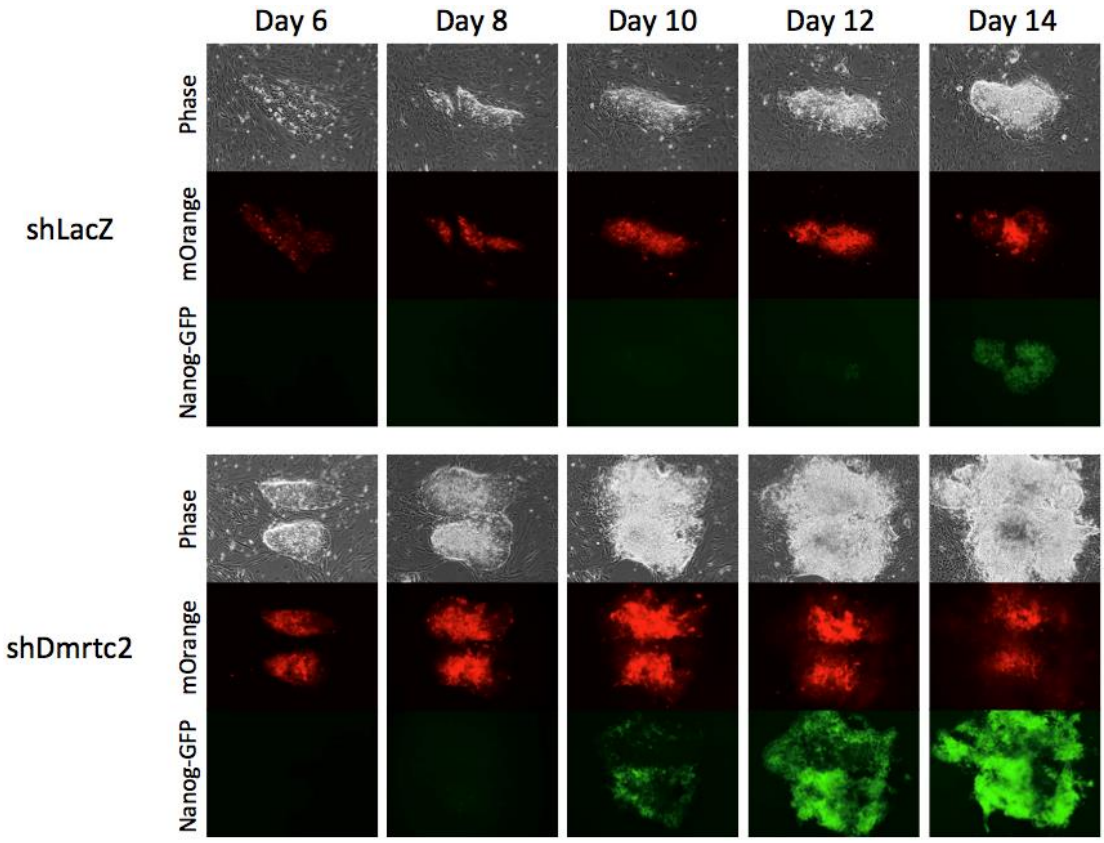


Figure 4.5. The shDmrtc2 effect on reprogramming is conserved between reprogramming systems. Addition of shDmrtc2 during primary reprogramming results in accelerated appearance of Nanog-GFP+ colonies by approximately 4 days when compared with control, supporting the data obtained in the secondary reprogramming context.

Following on from these results, I wanted to take advantage of the presence of a puromycin selection cassette contained within the Nanog-GFP reporter, allowing for Nanog expressing cells to be maintained in the presence of puromycin. I theorized that since shDmrtc2 accelerated the appearance of Nanog-GFP+ colonies during reprogramming, cells in this condition could be more amenable to surviving puromycin selection through earlier activation of Nanog and if so, I was interested to know how soon after administration of dox that puromycin selection could be applied while still enabling Nanog-GFP+ colonies to emerge during reprogramming. I carried out these experiments using our transgenic reprogramming system, plating approximately $2.5x10^4$ transgenic cells (94% transgenic) per 6-well with dox remaining in the culture medium throughout. Remarkably, when 1 μg/ml of puromycin (puro) was administered as early as 1 day after initiation of reprogramming with shDmrtc2 and subsequently cultured for an additional 9 days in the presence of puro and dox, more than 450 puro resistant colonies survived demonstrating Nanog-GFP+ expression, compared with less than 30 Nanog-GFP+ colonies with shLacZ,. This was more than a 16-fold increase suggesting that shDmrtc2 induces accelerated activation of Nanog (Figure 4.6). Furthermore, if puro was added 3 or more days after the start of reprogramming, up to 2ug/ml of puro could be applied to give a similar result (not shown).

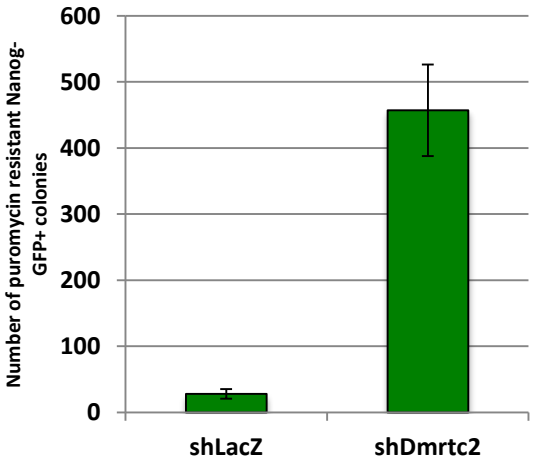


Figure 4.6. Accelerated activation of Nanog-GFP by shDmrtc2 confirmed by puromycin selection. shDmrtc2 enabled puromycin selection as early as 1 day after initiation of reprogramming demonstrating accelerated Nanog activation compared with control. Nanog-GFP+ colonies were counted on day 10 of reprogramming.

It should be clarified that these data provide evidence that cells become puromycin resistant before obtaining detectable GFP expression (cells were not GFP+ at day 1), while both transgenes connected by ires were expressed under the control of the endogenous Nanog promoter. It is probably because different numbers of molecules are required for the detection of GFP and puromycin resistance. In other words fewer numbers of molecules may be required by cells to confer puromycin resistance than the molecules required for detection of GFP, hence the cells appear to be puromycin resistant at an earlier time point than when Nanog-GFP+ cells are detected. Nonetheless, these data demonstrate a significant difference in the ability of reprogramming cells to acquire puromycin resistance with shDmrtc2 compared with control, which is consistent with the observation that Nanog-GFP+ colonies are detected days earlier with shDmrtc2. Since dox was administered throughout these experiments to day 15, we don't know if shDmrtc2 enables cells to become dox-independent (and therefore transgene independent) at an earlier time point. This would be an interesting experiment, and I predict that shDmrtc2 would accelerate transgene independence of reprogramming cells.

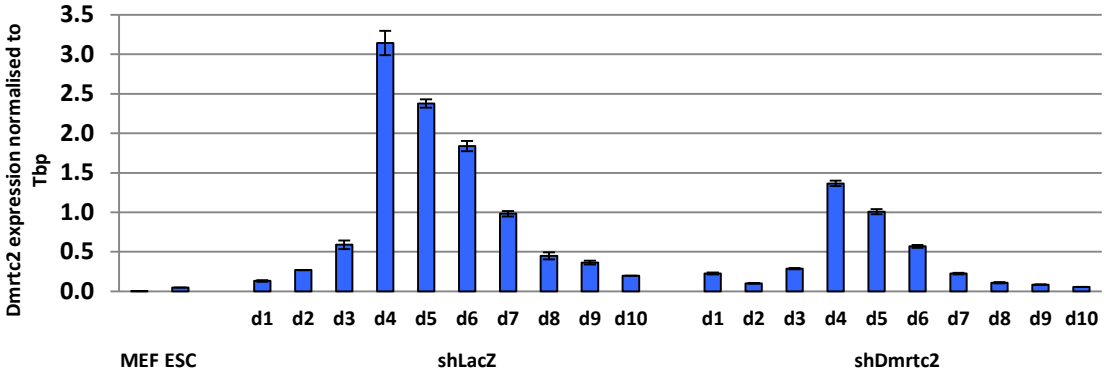
4.2.2. shDmrtc2 knocks down Dmrtc2 expression at an mRNA level

The most important experimental validation required when using RNAi is to determine whether the shRNA does in fact target the gene of interest, in this case Dmrtc2. To do this I used quantitative RT-PCR (Q-PCR) to assess the expression level of Dmrtc2 in the knock down condition compared with control. As a reminder, Dmrtc2 expression is low in MEF, is up-regulated during the first days of reprogramming and subsequently down-regulated where it remains low in iPSCs/ESCs. Because of this transient up-regulation during reprogramming it was important to check expression levels for several time points to ensure that gene expression was being knocked down even at the time points when endogenous levels usually increased. I chose to analyze Dmrtc2 expression every day for the first

10 days of reprogramming for two reasons; 1) to confirm the UP-DOWN pattern of expression observed in the initial published data by O'Malley et al. (2013) and determine at what time point expression levels peaked and 2) to determine if Dmrtc2 expression levels were knocked down throughout reprogramming or for only a certain period.

Since Q-PCR determines mRNA expression from a bulk sample, rather than single cells, it was important to use samples in which all or most of the cells were undergoing reprogramming. Therefore, I chose to carry out reprogramming using high contribution transgenic MEFs where >94% of cells were found to activate mOrange expression when dox was administered. Using these cells, I could initiate reprogramming with dox and use bulk reprogramming cultures at each time point for Q-PCR, negating the need for further manipulation of cells by flow cytometry. Samples were taken every day for 10 days of reprogramming for Q-PCR analysis to provide a high-resolution picture of Dmrtc2 expression during reprogramming with or without shDmrtc2.

I confirmed that shDmrtc2 expression was transiently up-regulated during reprogramming, as reported, and peaked at around day 4 followed by gradual down-regulation to the level found in iPSCs/ESCs by day 10 (Figure 4.7). Furthermore, shDmrtc2 was found to knock down expression of Dmrtc2 at an mRNA level as expected by approximately 60-80% compared with shLacZ and this knock down was maintained throughout the first 10 days of reprogramming (Figure 4.6b). Interestingly, all time points with the exception of d1 exhibited knock down; given that the cells are infected with the shRNA viral supernatant 24 hours before reprogramming is induced, this result suggests that it took approximately 48-72 hours for the shRNA to exert an effect. This is particularly notable since the first Nanog-GFP+ colonies in the shDmrtc2 condition could be detected by day 5, highlighting that the effect of shDmrtc2 is likely occurring within a very short time frame of just a few days.



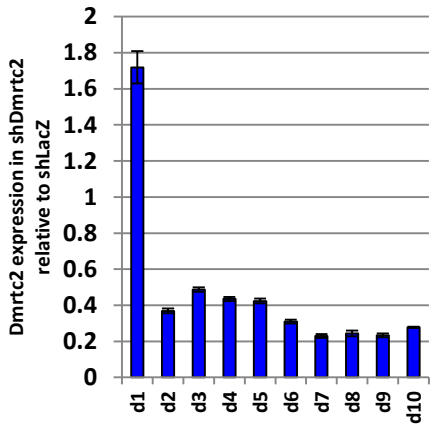


Figure 4.7. Dmrtc2 expression during reprogramming. Transient up-regulation of Dmrtc2 is observed during reprogramming (top panel, left). This is knocked down by shDmrtc2 (top panel, right and bottom panel). Bars represent the average of 2 independent samples treated with the same shRNA virus supernatant batch, and are representative and consistent with Dmrtc2 expression observed in other experiments using different virus supernatant batches.

4.2.3. Additional shRNAs targeting Dmrtc2 do not replicate the phenotype

Following confirmation of Dmrtc2 knock down by shDmrtc2, it was important to test further shRNAs targeting Dmrtc2. This was to determine if the phenotype could be replicated by knock down of Dmrtc2 by additional shRNAs targeting different sequences of Dmrtc2 mRNA, validating that the phenotype was specific to Dmrtc2 knock down and not resulting from non-specific knock down of another gene. I designed and tested 5 additional shRNAs (P1, P2, P4, P5 and P6) targeting independent sequences within Dmrtc2. As before, reprogramming was carried out with addition of these shRNAs and control shRNA, and Nanog-GFP+

colony numbers were counted at different time points. Additionally, the extent of Dmrtc2 knock down in the presence of each new shDmrtc2 shRNA was determined.

Interestingly, none of the additional shRNAs enhanced reprogramming efficiency or kinetics based on Nanog-GFP+ colonies counted. Addition of P1, P5 and P6 in reprogramming gave rise to a moderate 2-fold increase of Nanog-GFP+ colonies at day 12 compared with control, however, this was not comparable to the 6 to 7-fold increase demonstrated by P3 (Figure 4.8a). Furthermore, none of the additional shRNAs gave rise to a significant number of early appearing Nanog-GFP+ colonies by day 8, as observed with P3. Surprisingly, when the expression level of Dmrtc2 for each condition was checked by Q-PCR, three out of the five additional shRNAs knocked down expression comparable to P3 (Figure 4.8b). P1 and P5 resulted in no knock down even though both of these shRNAs induced a moderate increase in Nanog-GFP+ colonies by day 12. P2, P4 and P6 on the other hand reduced mRNA levels by approximately 50%, (similar to P3) but despite this, none of these shRNAs gave rise to the increased efficiency or kinetics demonstrated by P3. This unexpected result gave an indication that knock down of Dmrtc2 may not in fact be responsible for the phenotype observed by addition of shDmrtc2 P3 during reprogramming and that an off-target effect might be in play.

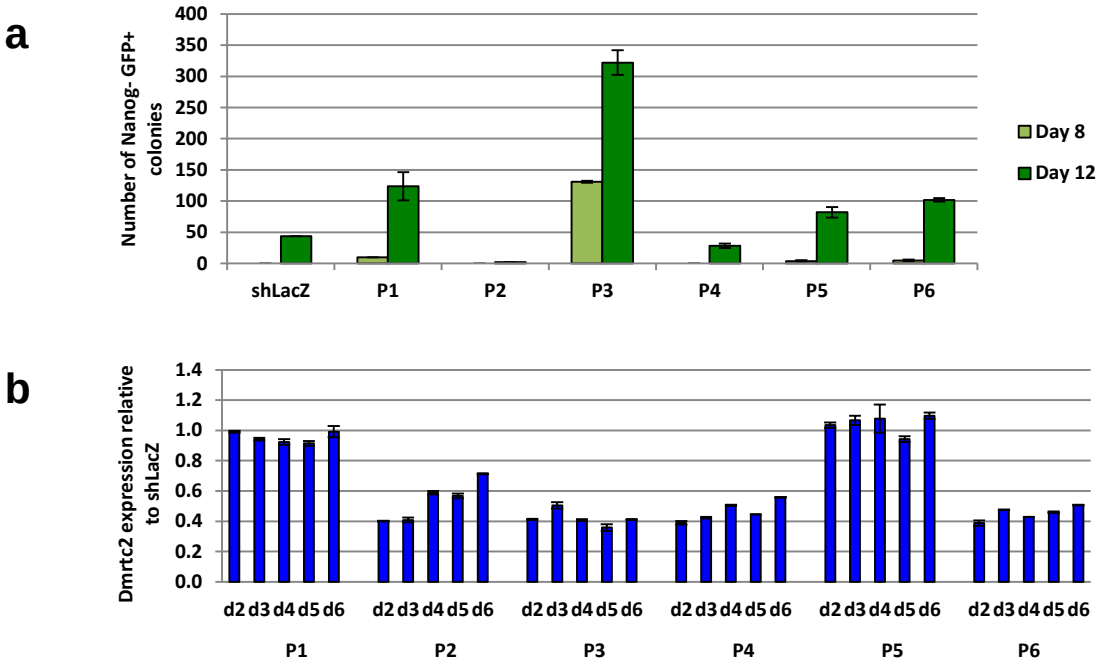


Figure 4.8. Additional shRNAs do not replicate the reprogramming phenotype observed with shDmrtc2 P3. No enhancement of reprogramming is observed with addition any alternative shRNA targeting Dmrtc2 (a), however shRNAs P2, P4 and P6 knock down Dmrtc2 expression to similar levels as P3 (b).

4.2.4 Overexpression of Dmrtc2 does not rescue the phenotype

A rescue experiment was carried out following these unexpected results, with a view to gain clarification of the suspicion that an off-target effect may be underlying the phenotype. If knock down of Dmrtc2 mediated by shDmrtc2 P3 was responsible for the enhanced reprogramming then I expect that overexpression of Dmrtc2 (which lacks target sequence of shDmrtc2 P3) would compensate for the effect of the knock down and 'rescue' or reverse the positive phenotype. In other words I would anticipate no effect on reprogramming in this context if knock down of Dmrtc2 was not responsible for the phenotype. This is a commonly used

validation experiment within the RNAi field (Kumar et al., 2006). To induce ectopic overexpression of Dmrtc2 in reprogramming cells, I created a retrovirus overexpression vector encoding the cDNA of Dmrtc2 (pMXs-Dmrtc2) and co-transfected this with shDmrtc2 P3 during reprogramming. It is important to note here that the targeting sequence of shDmrtc2 is complementary to a region within the 3'-UTR of the Dmrtc2 mRNA. pMXs-Dmrtc2 was constructed using the Dmrtc2 cDNA, which does not include the 3' UTR, and so it is refractory to targeting by shDmrtc2. Therefore, only endogenously expressed Dmrtc2 should be knocked down, with the overexpression vector compensating for this. A human CD2 overexpression vector (pMXs-hCD2) was used as a control for the Dmrtc2 overexpression vector and Nanog-GFP+ colonies were counted on day 9 of reprogramming. The control conditions (shLacZ with either pMXs-hCD2 or pMXs-Dmrtc2) gave rise to similar numbers of colonies and perhaps unsurprisingly, shDmrtc2 with either pMXs-hCD2 or pMXs-Dmrtc2 gave approximately 3 to 6-fold more Nanog-GFP+ colonies than the controls (Figure 4.9), indicating that overexpression of Dmrtc2 did not abolish the enhanced phenotype bestowed during reprogramming by shDmrtc2.

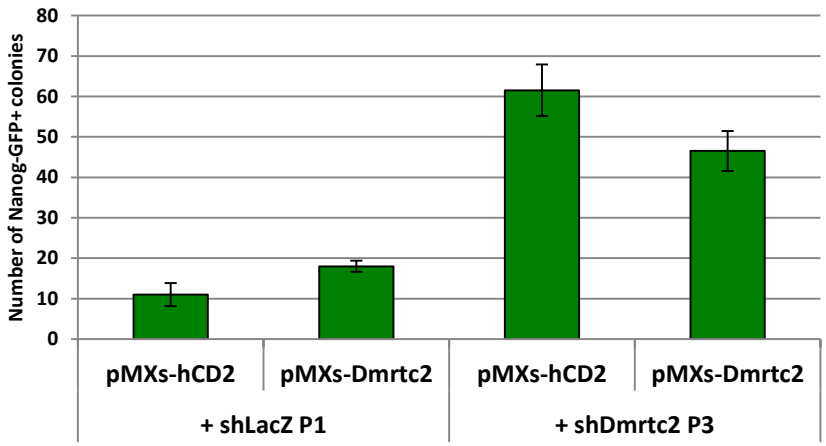


Figure 4.9. Overexpression of Dmrtc2 does not rescue the enhanced reprogramming phenotype induced by shDmrtc2. Despite overexpression of Dmrtc2, an increased colony number was still observed in the presence of shDmrtc2 P3 during reprogramming.

In parallel, the mRNA level of Dmrtc2 was checked by Q-PCR to ensure that overexpression by pMXs-Dmrtc2 was in fact occurring. In the context of overexpression by pMXs-Dmrtc2 in conjunction with knock down by shDmrtc2 during reprogramming, Dmrtc2 expression levels were found to be at least 10 to 100-fold higher compared with the control shLacZ condition, including at day 4 when Dmrtc2 expression levels reached their highest (Figure 4.10).

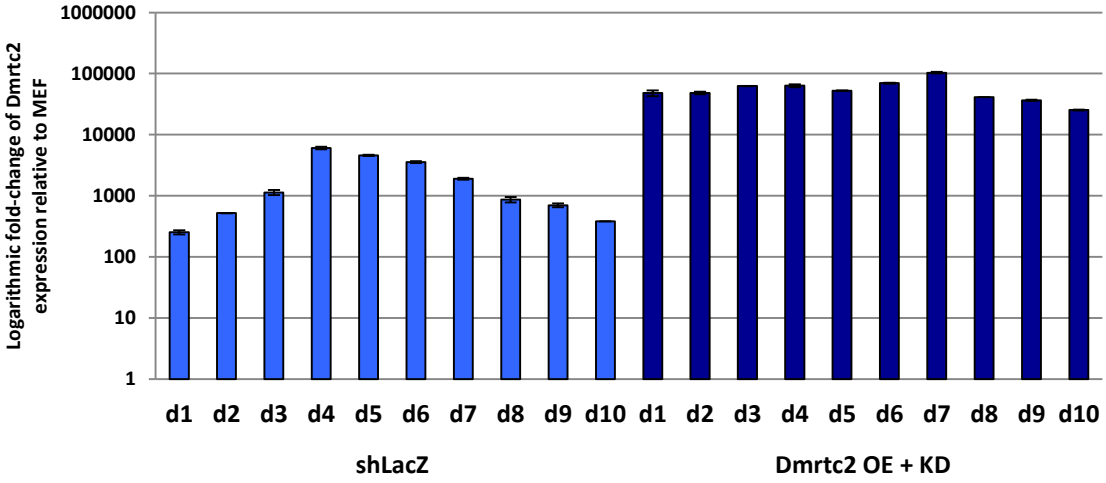


Figure 4.10. Q-PCR analysis of Dmrtc2 mRNA levels during reprogramming in control and overexpression conditions. Overexpression (OE) of Dmrtc2 with pMXs-Dmrtc2 in the presence of shDmrtc2 knock down (KD) shRNA confirms elevated expression of Dmrtc2, compared with levels observed in control conditions. Note the values are represented on a logarithmic scale.

Taken together, these data further supported the notion that regulation of Dmrtc2 by the knock down vector shDmrtc2 was not responsible for the enhanced reprogramming phenotype.

4.3 Discussion

4.3.1 The importance of knock down validation when using RNAi

As demonstrated in this chapter, it is not enough to simply determine that a gene is specifically knocked down to have confidence that this is the cause of any phenotype observed. In addition to this basic validation, the phenotype should be replicated by knocking down the same gene with additional shRNAs targeting different sequences of the gene. In this study, 5 additional shRNAs targeting various regions of the Dmrtc2 gene were tested. When using RNAi for knock down experiments, it is typically advised that 3 to 6 si/shRNAs are tested for each gene of interest in order to reasonably expect at least one of them to faithfully reduce gene expression. Indeed, this was demonstrated here since I observed knock down of Dmrtc2 to similar levels in 4 out of the 6 shRNAs. As a result, I could be reasonably confident that the effect of shDmrtc2 was not due to repression of Dmrtc2, since all of the additional shRNAs tested failed to reproduce either the enhanced efficiency or acceleration of reprogramming observed with shDmrtc2, even moderately. Further to this, a rescue experiment showed that even in the presence of excessively high levels of exogenous Dmrtc2, the phenotype exerted by shDmrtc2 was able to persist, providing support that knock down of Dmrtc2 was not the mechanism by which reprogramming efficiency and kinetics was enhanced. Taken together, these validation results gave me confidence that an off-target event was the causal link between the addition of shDmrtc2 and positive effect on reprogramming.

4.3.2 Gene specific knock down can be achieved with shRNA

The basic validation following discovery of a phenotype during reprogramming involved firstly confirming by Q-PCR that shRNAs used in this study knocked down genes in a specific manner (Josefsen and Lee, 2011). Q-PCR

data representing mRNA expression levels for Dmrtc2 confirmed that when compared with the normal levels of Dmrtc2 observed throughout reprogramming (in this case in the shLacZ condition), the addition of shDmrtc2 resulted in a knock down of Dmrtc2 mRNA by up to nearly 80%. Even at the time points at which Dmrtc2 was maximally up-regulated in the control, a 60% knock down was achieved.

Knock down efficiencies vary greatly between RNAi systems used and in addition it is not difficult to imagine that different genes potentially require different levels of knock down to result in a phenotypic change (Mantei et al., 2008; Yang et al., 2014; Qin et al., 2014; Gingold et al., 2014; Berns et al., 2004; Paddison et al., 2002; Chen et al., 2005). Furthermore, validation by Q-PCR to assess knock down at the transcript level is not exhaustive. Since it is more often proteins and not mRNA that are the "end-point" functional effectors within cells, it is good practice to determine whether any knock down at mRNA level is reflected in a knock down at the protein level by Western Blot analysis. This can provide useful information, particularly in a case where knock down is seen at the mRNA level, but no phenotype results. In this case, investigation of the protein output for the gene of interest may indicate that there is no reduction of protein levels, as may be the case for particularly stable or long lasting proteins, and hence no phenotype has resulted from mRNA knock down. It has been suggested that for some genes, a knock down of more than 95% at the transcript level is required before changes in the protein product can be seen (Moffat and Sabatini, 2006) so it is important in some cases to verify a reduction in protein levels in addition to Q-PCR validation. With substantial ongoing research into the mechanisms of RNAi, more and more is being discovered about RNAi processing machinery and the importance of RNAi design to facilitate increasing knock down efficiencies (Schwarz et al., 2003). However, with new gene editing technologies such as CRISPR and TALEN becoming more accessible (Bogdanove and Voytas, 2011; Shalem et al., 2014; Wang et al., 2014; Wang et al., 2013a), with which gene knock out can easily be achieved, it is likely

that the use of RNAi will become less commonplace as a championed gene interference technology.

In addition to variance in knock down efficiency, it is important to make a point here about the variability of reprogramming and the subsequent implications on data. A good example of this can be seen in the colony counting experiments for shDmrtc2; although I could very faithfully reproduce the phenotype inferred by shDmrtc2, and thus I am confident that it is a true result, the variable nature of reprogramming very often results in wildly different numbers of colonies being produced between experiments, despite the fact that the same materials and reagents are used for each experiment. This is exemplified in Figure 4.1b; the results for three independent reprogramming experiments are shown and although they all confirm that shDmrtc2 gives rise to several fold more colonies than control, the absolute numbers of colonies in each experiment are drastically varied. This is also evident in the original screening experiments carried out in Chapter 3, Figures 3.3 and 3.4. This can make representing data difficult; if I wanted to represent the data using absolute numbers of colonies then the standard deviation and subsequently error bars would be large. In order to circumvent this issue, I have represented the data as a fold-change value comparing shDmrtc2 to shLacZ (Figure 4.1a).

4.3.3 Off-target effects are a potential consequence of RNAi

One of the main drawbacks of using RNAi in gene knock down studies is the phenomenon of off-target effects. That is, the unspecific action of an si/shRNA on an unintended target causing an unwanted effect (Jackson and Linsley, 2010). Off-target effects are a well-known side effect of using RNAi, and have been found to increase with increasing concentration of siRNA and sequence specificity (Jackson et al., 2003; Jackson et al., 2006). Off-target silencing can create problems for interpretation of data, particularly if several off-target events occur (see Chapter 6 – Discussion for more on this topic). As a general rule, off target silencing is an

unwanted phenomenon, but in the case of this study it appears that there is a significant and reproducible positive effect on reprogramming as a result of an off-target. As such, my investigation of this reprogramming phenotype continues with a view to determine the off-target gene, to enable further characterization and functional importance in the context of reprogramming.

CHAPTER 5 – An off-target effect of shDmrtc2 underlies enhanced reprogramming efficiency and kinetics

5.1 Introduction

There has been much research on the topic of off-target effects of RNAi and with a multitude of sources being reported as potential causes of this unintended silencing, pinpointing the exact root is often complex and difficult. Indeed in some cases there may be a combination of factors in play. There is always the possibility of the guide strand finding imperfect pairing with some region of the 3′-UTR of an off-target mRNA, causing either cleavage or inhibition of translation which is known as miRNA-like off-targeting, since miRNAs primarily target 3′-UTRs for gene regulation. In particular, off-target effects of this type are commonly mediated by sequence complementarity of the 3′-UTRs of unintended targets to the si/shRNA seed region (Jackson et al., 2006). Interestingly, it was found that base mismatches in the seed region abolished silencing of some off-targets, but these were simply replaced with silencing of new unintended targets with complementarity to the resulting seed sequence (Jackson et al., 2003). Another unintended effect of ectopic shRNA expression is the competition with endogenous miRNAs for RNAi processing machinery. For example, it has been shown that overexpression of shRNA can saturate Exportin-5 (Exp5), a mediator of nuclear export required by shRNAs and pre-miRNAs, which comes at the expense of endogenous miRNA processing. This could be rescued with overexpression of Exp5 (Yi et al., 2003; Yi et al., 2005). It has also been reported that in the case of shRNAs, the positioning of the loop structure is crucial for correct Dicer cleavage into subsequent siRNA, and a deviant cleavage site can result in increased off-target effects due to promiscuous Dicer cleavage and incorrect shRNA processing (Gu et al., 2012). Thus, off-target effects are a common occurrence and it is unlikely that a perfect si/shRNA can be designed to completely abolish this phenomenon. The most that can be hoped for is

that any resulting off-target effects do not interfere with the ultimate goal of the RNAi being used, and that any phenotypes observed are true of the specific on-target.

During the previous chapter I provided evidence to suggest that the positive effect on reprogramming observed by the addition of shDmrtc2 was due to an off-target effect of the RNAi; that is, shDmrtc2 acting on an unknown and unintended target caused the phenotype. Since the effect on reprogramming was substantial, I chose to move forward with my investigation of shDmrtc2 with a view to identifying the off-target gene responsible for the enhanced reprogramming phenotype observed.

5.1.1 Aims of this chapter

The aim of this chapter is to use single and double nucleotide mutants of shDmrtc2 to identify shRNAs that can or cannot replicate the phenotype observed with shDmrtc2 in reprogramming. Subsequently these vectors will be used to carry out genome-wide microarray and mRNA-sequencing analysis with the aim of identifying candidates that exhibit differential gene expression between shDmrtc2 and its mutants. Through validation of these candidates, the aim is to identify the off-target responsible for the positive phenotype of shDmrtc2 in the context of reprogramming.

5.2 Results

5.2.1. Strategy to identify off-target candidates of shDmrtc2

It is clear that the efficiency of knock down by RNAi is highly dependent on sequence specificity between the target mRNA and antisense siRNA sequences. Even as little as a single nucleotide substitution in a si/shRNA sequence has been shown to be sufficient to abolish the resulting knock down of a target gene (Jackson et al., 2003; Elbashir et al., 2001b; Martinez et al., 2002). With this in mind, I sought to generate mutants of shDmrtc2 with double or single nucleotide substitutions, with a view to identify vectors which could either abolish the reprogramming phenotype observed with shDmrtc2, or enhance reprogramming similarly to shDmrtc2 (so called "negative mutant" or "positive mutant", respectively). I reasoned that a negative mutant would be a useful control to investigate off-target knock down, particularly if it maintained Dmrtc2 knock down, since the sequence would be almost identical to that of shDmrtc2 but have no effect on reprogramming. On the other hand, a positive mutant would be advantageous as an additional positive control allowing for candidates to be refined based on the differential expression of negative controls (shLacZ and negative mutant) and positive samples (shDmrtc2 and positive mutant). The strategy is summarized in Figure 5.1.

a

Reprogramming phenotype	Off-target KD of causal gene	Knock down of Dmrtc2?	Other off-target KD?	Comment
Enhanced	Yes	Yes	Yes	Original shDmrtc2 vector
No effect	No	Yes	Yes	Negative control
Enhanced	Yes	No	Yes	Positive control
No effect	No	No	Yes/No	No siRNA production or knock down of targets with no role in reprogramming?

b

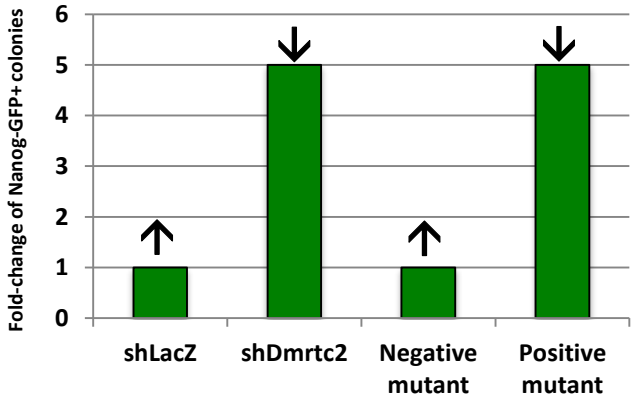


Figure 5.1. Strategy to determine causal off-target gene promoting enhanced reprogramming phenotype. (a) By comparing differential gene expression during reprogramming between samples that enhance (yellow) or have no effect (blue, and shLacZ control) on reprogramming I sought to identify the off-target gene (green) underlying the enhanced reprogramming phenotype. (b) Expected reprogramming result of negative or positive shDmrtc2 mutants based on Nanog-GFP+ colony counting. Arrows represent the expected gene expression of off-target gene 'X' responsible for the enhanced reprogramming phenotype caused by shDmrtc2 in each sample: high expression in shLacZ control and negative mutant samples (up arrow) or low expression/knock down in shDmrtc2 and positive mutant samples (down arrow).

5.2.2. Mutants of shDmrtc2 can abolish or maintain the phenotype

Mutant vectors were created in two sets; initially, 6 mutant vectors were created containing identical shRNA sequences to shDmrtc2 with the exception of two individual nucleotide substitutions (Figure 5.2a, MUT 1-6). Mutants 1 to 4 contained substitutions of nucleotides within the seed region, whereas mutants 5 and 6 contained one substitution within the seed region and one substitution out with the seed region. Although reports suggested that a single nucleotide substitution is enough to affect knock down efficiency of RNAi (Jackson et al., 2003), I chose to include two nucleotide substitutions in these initial vectors to optimize the chance of finding a vector that would abolish the phenotype.

Reprogramming was carried out with the addition of mutants 1 to 6 to determine the effect, if any, of the double nucleotide mutations on the phenotype observed using shDmrtc2. Strikingly, all six mutant vectors abolished the positive phenotype observed with shDmrtc2, giving rise to fewer Nanog-GFP+ colonies at a number similar to shLacZ (Figure 5.2b). In order to validate these results, it was important to determine the levels of Dmrtc2 transcript in the presence of each mutant during reprogramming. Interestingly, despite all of these mutants having little or no effect on reprogramming compared with shLacZ, mutants 1 and 5 were determined to knock down expression of Dmrtc2 at a level similar to that observed with shDmrtc2 (Figure 5.2c). This gave further evidence that the positive phenotype conferred by shDmrtc2 during reprogramming was not due to the knock down of Dmrtc2 and must be the result of some off target effect. From these data I proposed that when considering global gene expression I would expect the gene or genes responsible for the positive reprogramming phenotype resulting from shDmrtc2 would exhibit lowered gene expression in shDmrtc2 when compared with shLacZ, MUT 1 and MUT 5.

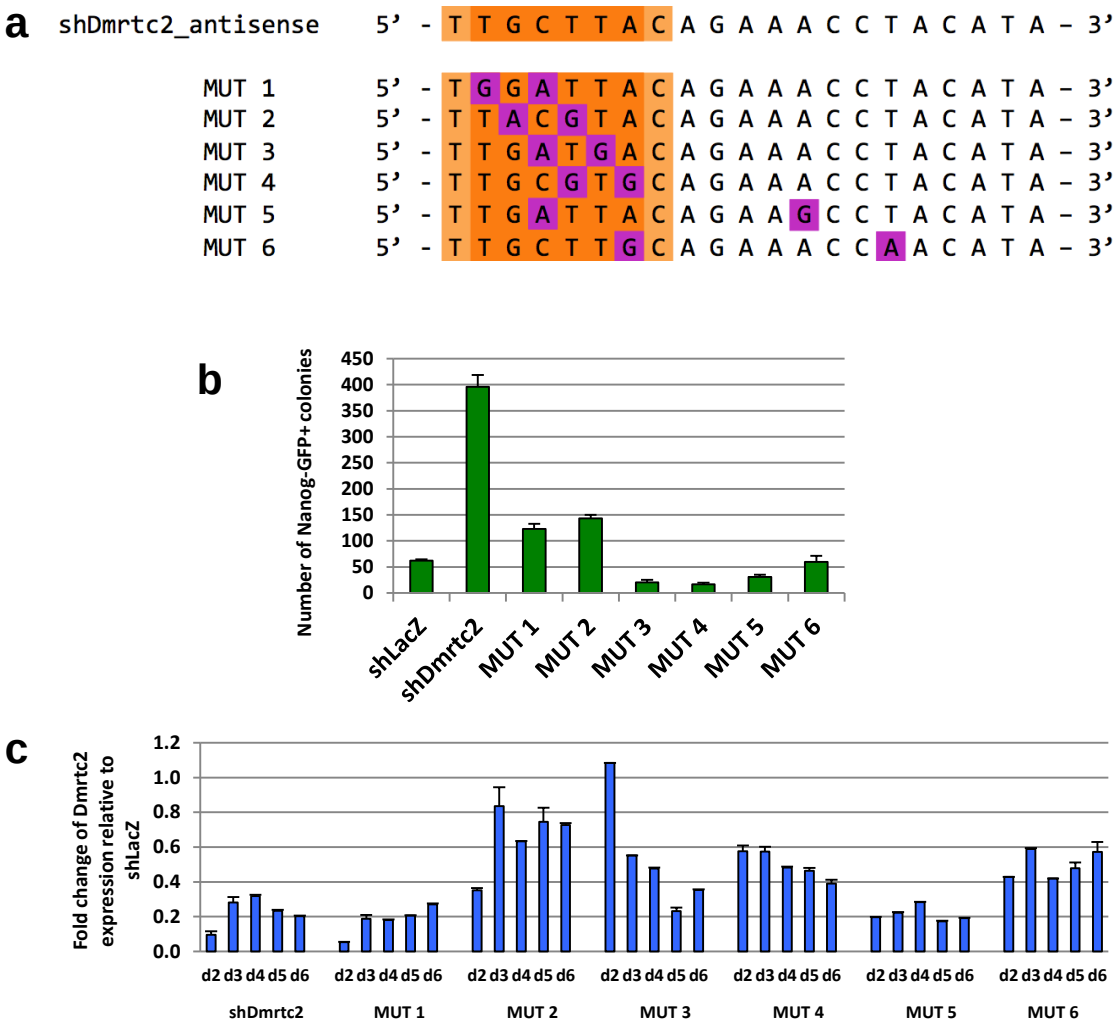


Figure 5.2. shDmrtc2 double nucleotide mutants during reprogramming. (a) The 21bp sequence of shDmrtc2 is antisense, or complementary, to a sequence within the Dmrtc2 3'-UTR region (top line). 6 mutants with double nucleotide mutations were generated. Seed region is shown in orange and mutations are shown in purple. (b) The effect of shDmrtc2 double nucleotide mutants on reprogramming was quantified by Nanog-GFP+ colony count on day 12. (c) Q-PCR of Dmrtc2 expression was carried out for samples on days 2 to 6 of reprogramming. MUT 1 and MUT 5 knocked down Dmrtc2 expression to similar levels as shDmrtc2 but did not enhance reprogramming.

When flow cytometry analysis for ICAM1/CD44 was carried out on samples reprogrammed with MUT 1 and MUT 5, these gave similar profiles to that of shLacZ with cells in the shDmrtc2 sample clearly advancing at an earlier time point, around day 7, as described previously in Figure 3.6 (Figure 5.3).

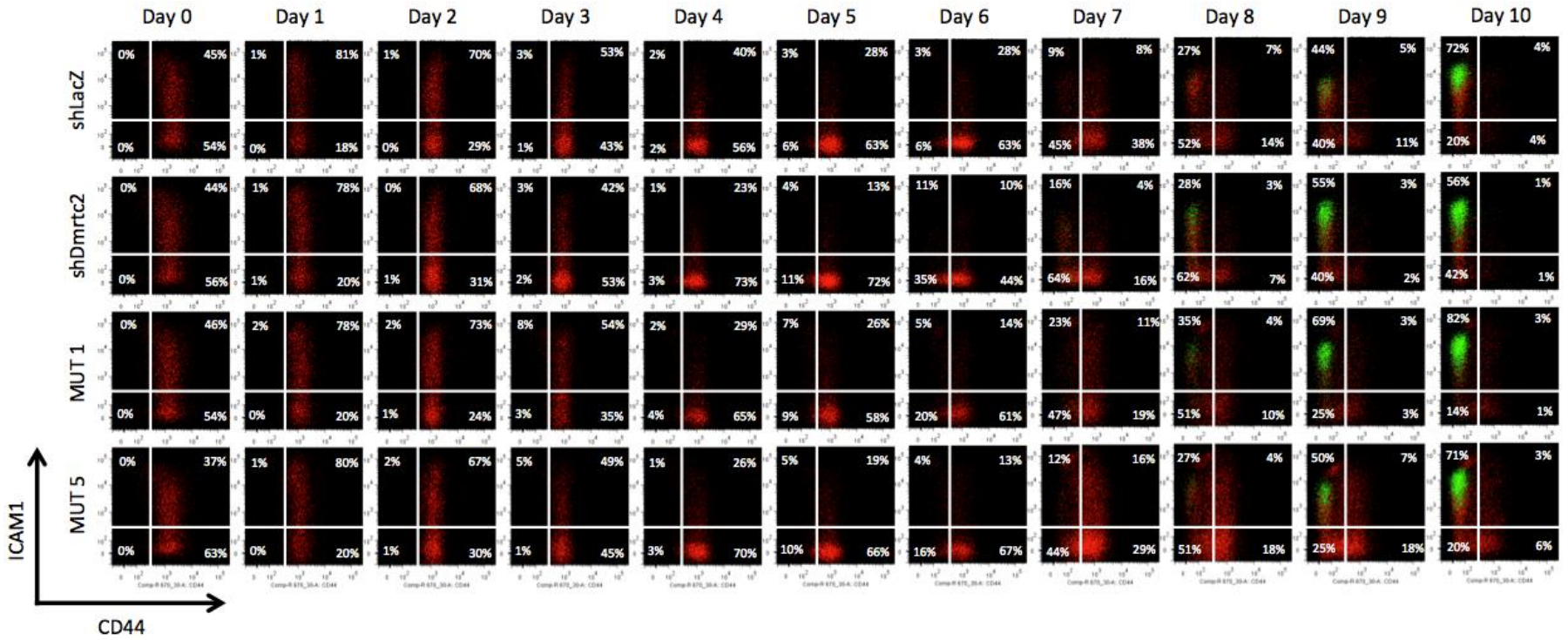
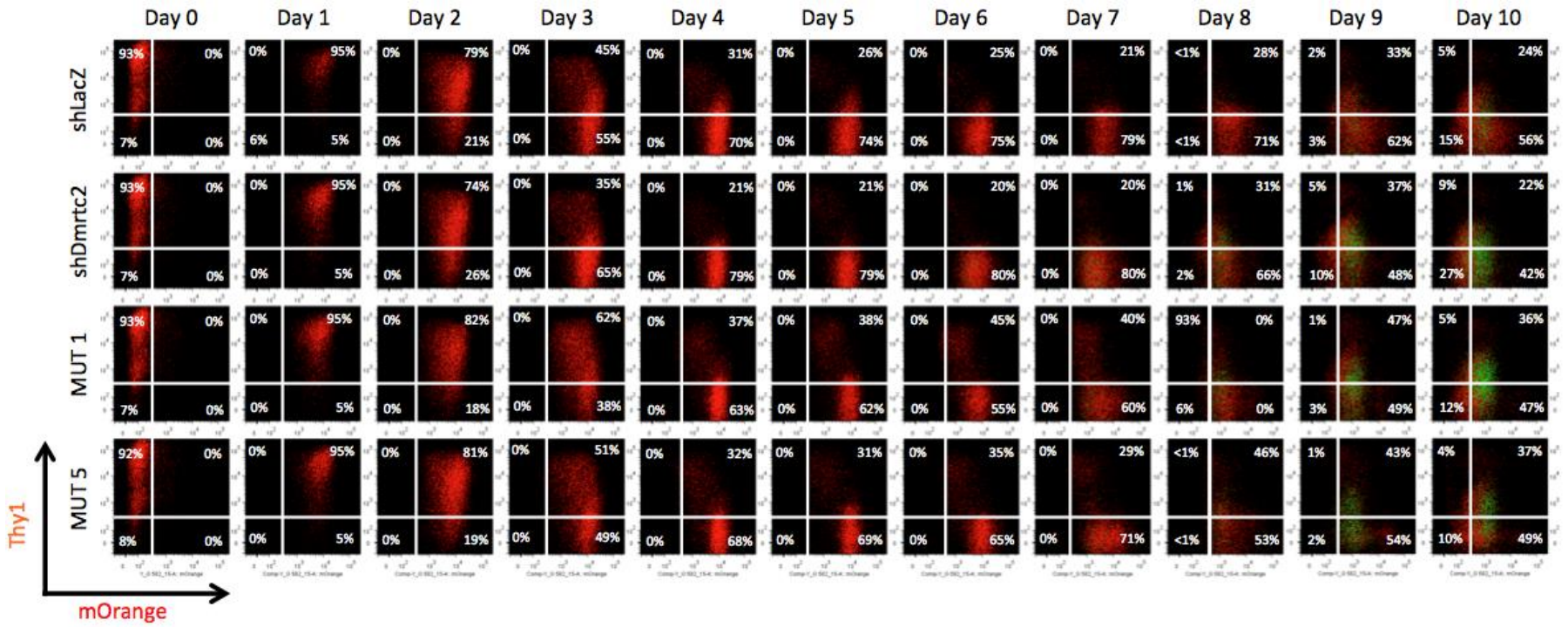
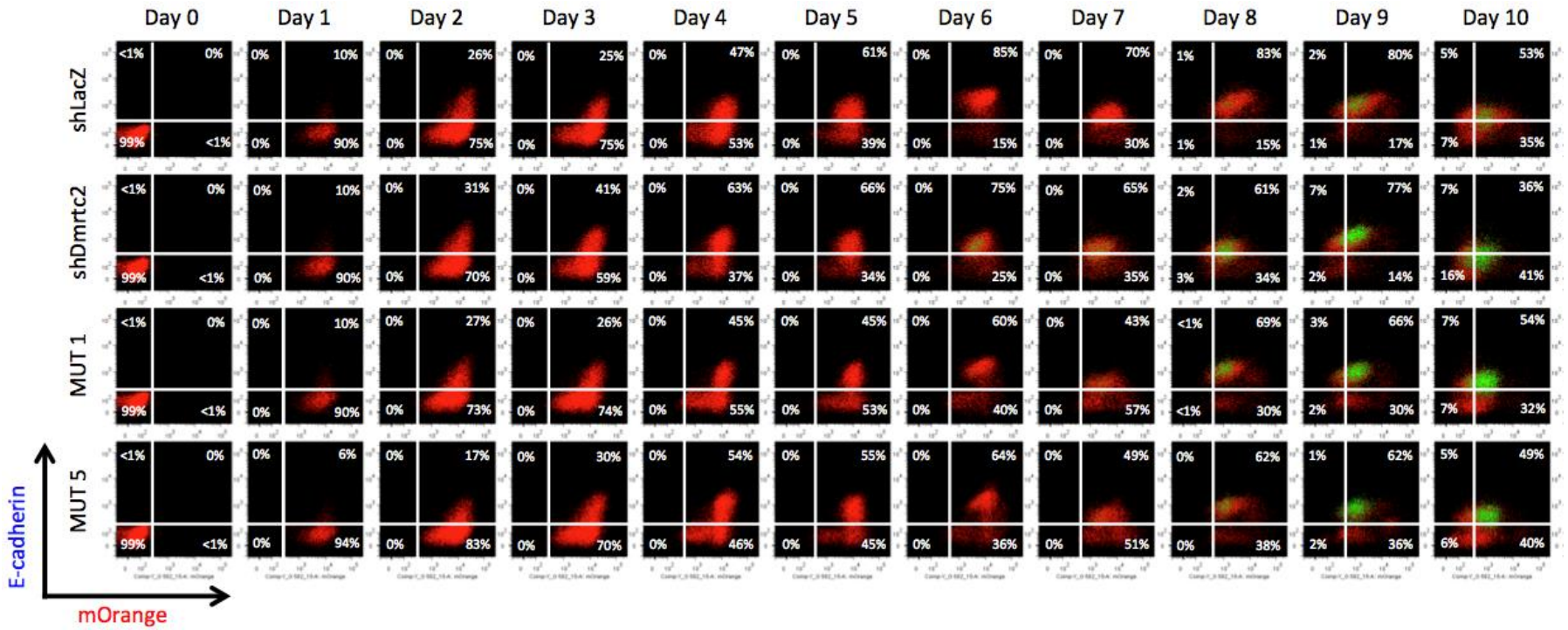


Figure 5.3. ICAM1/CD44 FACS of shDmrtc2 double nucleotide mutants 1 and 5. MUT 1 and MUT 5, which did not increase Nanog-GFP+ iPSC colony numbers, also failed to accelerate reprogramming kinetics observed in shDmrtc2 samples (days 6 and 7, highlighted in red box).

Flow cytometry analysis was carried out for three markers commonly used to characterize the progression of cells undergoing reprogramming; Thy1, E-cadherin and SSEA-1 (Figure 5.4). During reprogramming, expression of a MEF related marker, Thy1, decreases and expression of E-cadherin, a cell-adhesion molecule that is required for the maintenance of pluripotency, increases. These changes in gene expression are characteristic of a mesenchymal to epithelial transition (MET) during reprogramming. It was interesting to determine whether the positive phenotype exhibited by shDmrtc2 was concurrent with advanced timing of MET. When these two markers were analysed by flow cytometry shDmrtc2 exhibited acceleration of Thy1 down-regulation and E-cadherin up-regulation on days 3 and 4. However, analysis of a pluripotency marker, stage-specific embryonic antigen 1 (SSEA-1), demonstrated no marked difference in the shDmrtc2 sample compared with controls suggesting that there is no acceleration of acquisition of this intermediate pluripotency marker with shDmrtc2. SSEA-1 is commonly used as a marker of pluripotency but we see no earlier expression of this marker in the shDmrtc2 sample as expected given the enhanced reprogramming phenotype. In addition, these flow cytometry data demonstrate that Nanog-GFP+ cells can reside within both the SSEA-1 positive and negative populations and so care must be taken not to consider this a definitive marker of pluripotency.


Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Day 9
Day 10
shLacZ
shDmrtc2
MUT 1
MUT 5
Thy1
mOrange
93% 0% 7% 0%
0% 95% 6% 5%
0% 79% 0% 21%
0% 45% 0% 55%
0% 31% 0% 70%
0% 26% 0% 74%
0% 25% 0% 75%
0% 21% 0% 79%
<1% 28% <1% 71%
2% 33% 3% 62%
5% 24% 15% 56%
93% 0% 7% 0%
0% 95% 0% 5%
0% 74% 0% 26%
0% 35% 0% 65%
0% 21% 0% 79%
0% 21% 0% 79%
0% 20% 0% 80%
0% 20% 0% 80%
1% 31% 2% 66%
5% 37% 10% 48%
9% 22% 27% 42%
93% 0% 7% 0%
0% 95% 0% 5%
0% 82% 0% 18%
0% 62% 0% 38%
0% 37% 0% 63%
0% 38% 0% 62%
0% 45% 0% 55%
0% 40% 0% 60%
93% 0% 6% 0%
1% 47% 3% 49%
5% 36% 12% 47%
92% 0% 8% 0%
0% 95% 0% 5%
0% 81% 0% 19%
0% 51% 0% 49%
0% 32% 0% 68%
0% 31% 0% 69%
0% 35% 0% 65%
0% 29% 0% 71%
<1% 46% <1% 53%
1% 43% 2% 54%
4% 37% 10% 49%


Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Day 9
Day 10
shLacZ
shDmrtc2
MUT 1
MUT 5
E-cadherin
mOrange

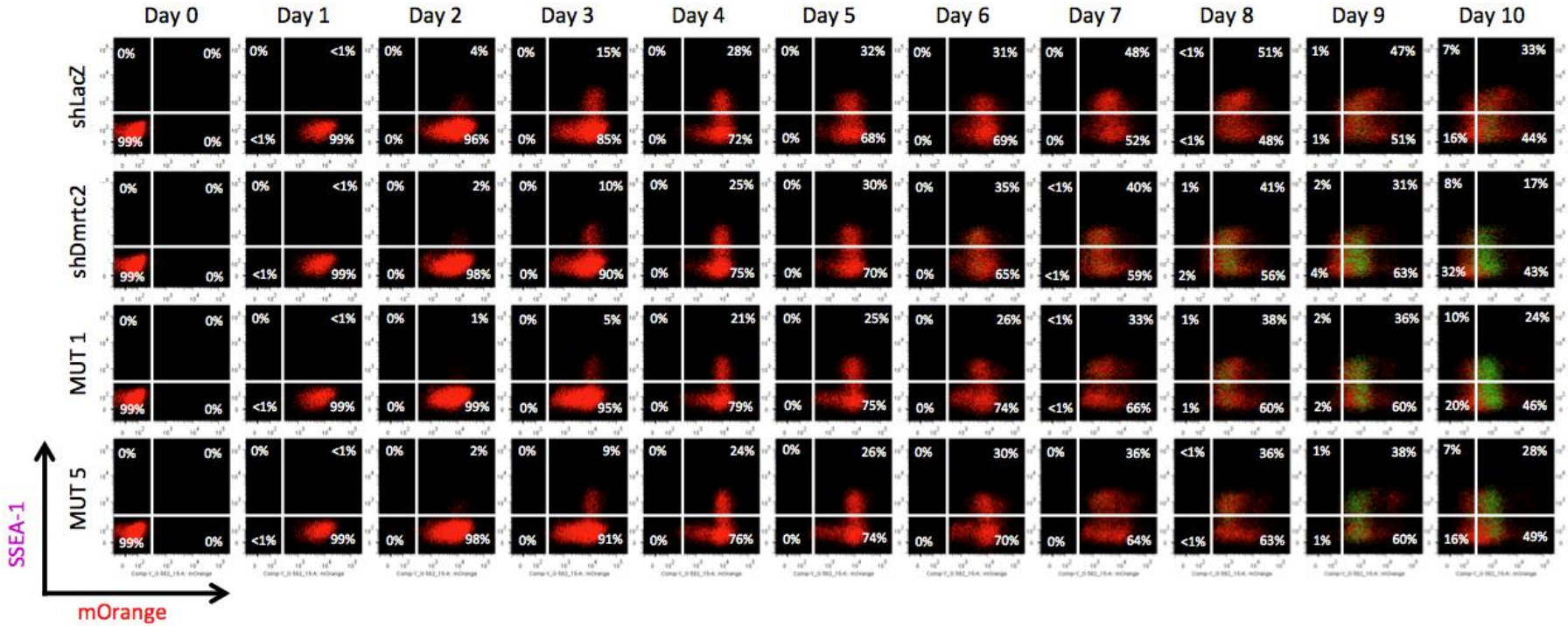


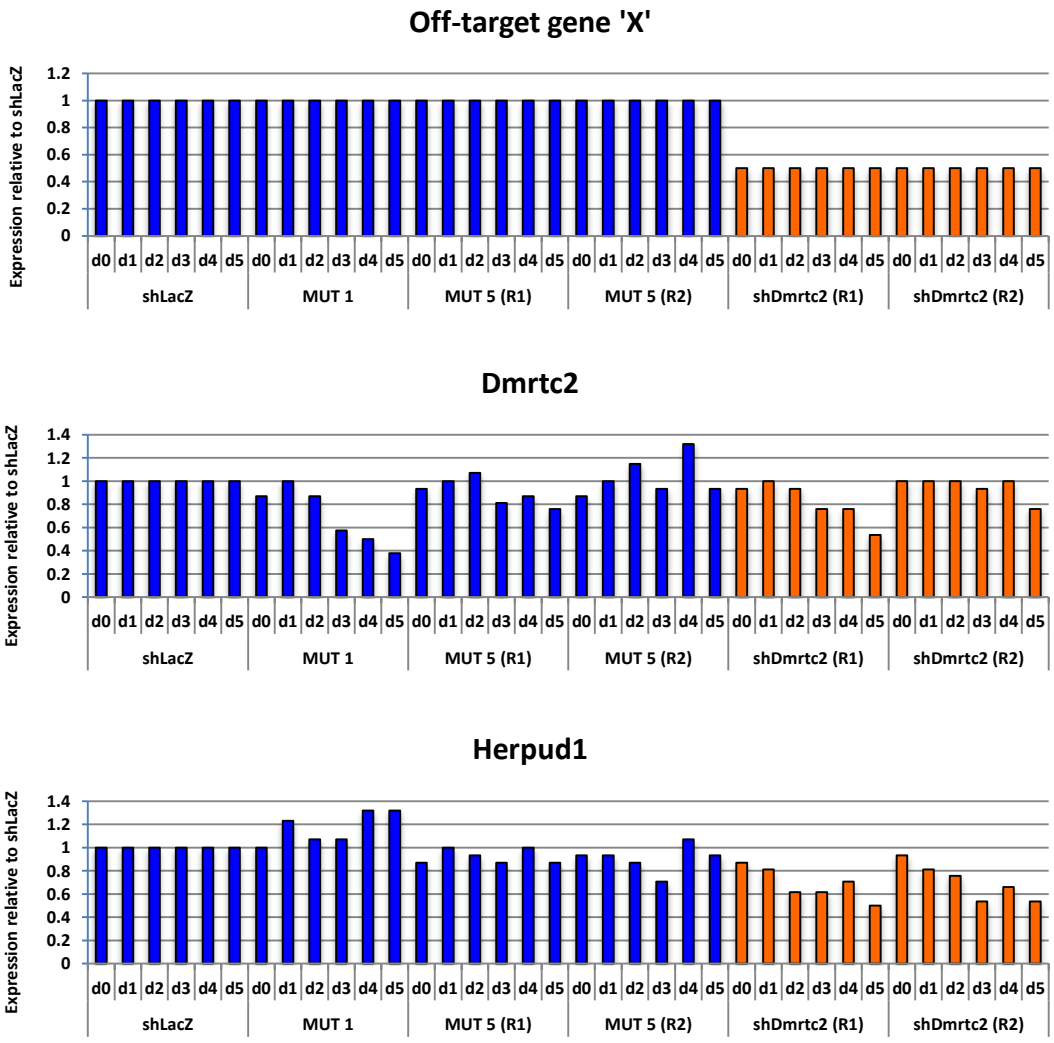
Figure 5.4. Flow cytometry analysis for markers Thy-1, E-cadherin and SSEA-1. The flow cytometry profiles generated from all conditions demonstrated no remarkable differences in the gene expression profiles for these markers between shDmrtc2, control and mutants.

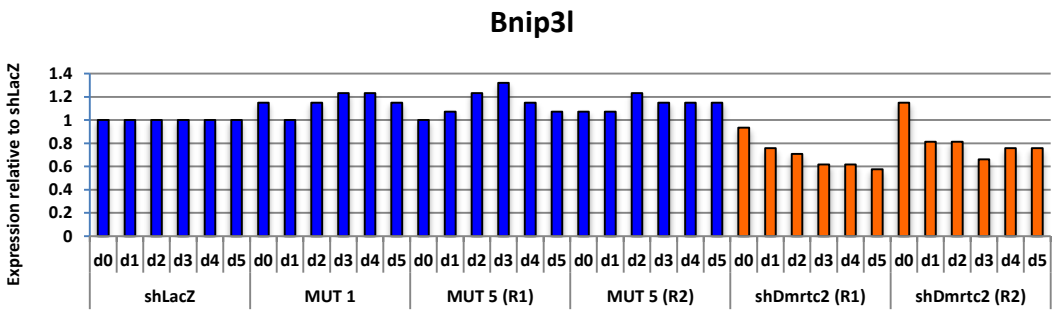
5.2.3 Microarray analysis reveals off target candidates of shDmrtc2

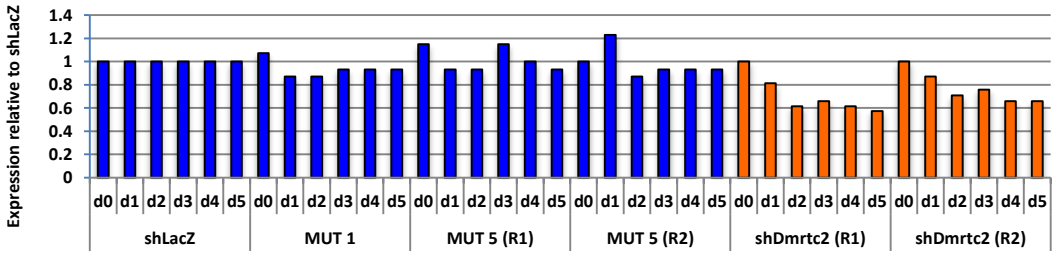
A robust and reproducible difference both in the colony number and timing of Nanog-GFP+ colonies between shDmrtc2 and shLacZ was abolished by only two mutations in the sequence of shDmrtc2 in MUT 1 and MUT 5. Despite this, both mutant shRNAs could knock down Dmrtc2 transcript levels indicating efficient production of siRNA but altered off-targets. To identify an essential off-target knock down by shDmrtc2 that enhances reprogramming I took shDmrtc2, shLacZ, MUT 1 and MUT 5 samples forward for microarray analysis using the Illumina MouseWG-6 v2 BeadChip. In total, six samples of cells undergoing reprogramming from days 0 to 5 were taken for six conditions consisting of one each of shLacZ and MUT 1 and two replicates each of shDmrtc2 and MUT 5 (R1, R2), totalling 36 samples for analysis. If knockdown of off-target gene X is the cause of the phenotype by shDmrtc2, I expected the expression pattern of X to be like Figure 5.5 (top graph).

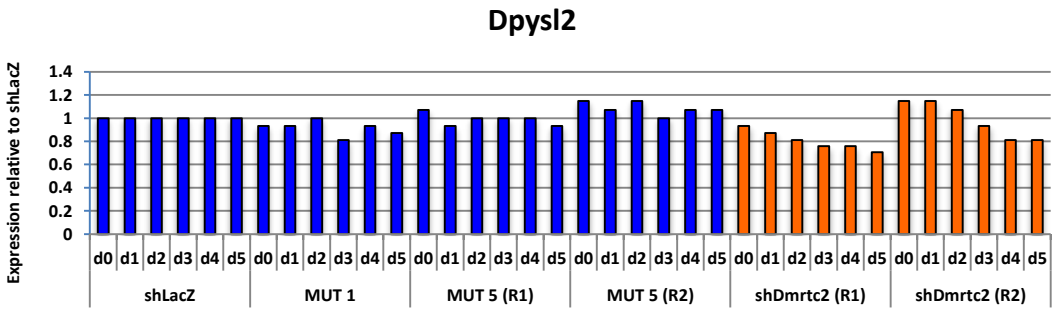
The results from this microarray approach were somewhat unclear and careful analysis was carried out in order to draw some conclusions. Firstly, when expression of Dmrtc2 was checked, all shRNA conditions (shDmrtc2, MUT 1 and MUT 5) were expected to indicate a down-regulation of Dmrtc2 compared with shLacZ, as determined from experiment previously described in Section 5.2.2. However, this was only clearly the case for MUT 1 and shDmrtc2 (R1). Both replicates of MUT 5 and shDmrtc2 REP2 showed little or no knock down of Dmrtc2. This was a first indication that the microarray approach might not be sensitive enough to detect off target candidates, unless they were strongly repressed by shDmrtc2. However, despite this potential drawback, at first glance several candidate genes could be identified from the microarray as having lower expression in the shDmrtc2 replicates compared with controls, with the clearest candidates shown in Figure 5.5. These included Herpud1, Bnip3l, Pak3, Dpysl2 and Qdpr.

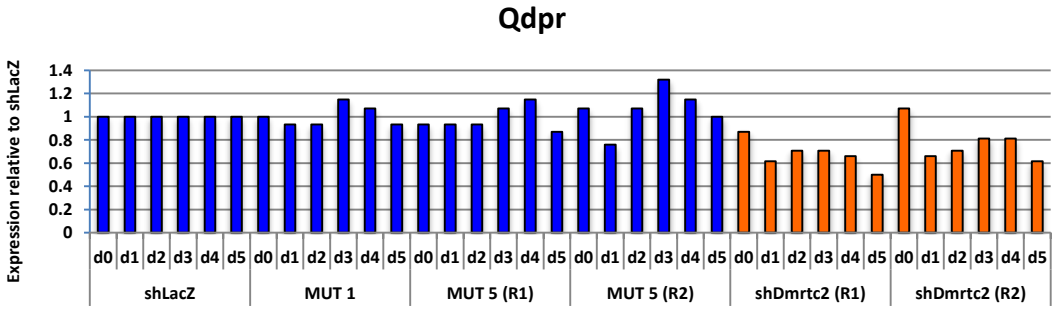
Figure 5.5. Microarray analysis identifies genes with differential expression with shDmrtc2 during reprogramming. The expected expression of off-target gene 'X' is shown (top panel). Top hits of genes with differential gene expression comparing shDmrtc2 samples with shLacZ or negative mutant samples MUT1 and MUT 5 were Herpud1, Bnip3l, Pak3, Dpysl2 and Qdpr.




Bnip3l
Expression relative to shLacZ
1.4
1.2
1
0.8
0.6
0.4
0.2
0
d0 d1 d2 d3 d4 d5
shLacZ
MUT 1
MUT 5 (R1)
MUT 5 (R2)
shDmrtc2 (R1)
shDmrtc2 (R2)


Pak3
Expression relative to shLacZ
1.4
1.2
1
0.8
0.6
0.4
0.2
0
d0 d1 d2 d3 d4 d5
shLacZ
MUT 1
MUT 5 (R1)
MUT 5 (R2)
shDmrtc2 (R1)
shDmrtc2 (R2)


Dpysl2
Expression relative to shLacZ
1.4
1.2
1
0.8
0.6
0.4
0.2
0
d0 d1 d2 d3 d4 d5
shLacZ
MUT 1
MUT 5 (R1)
MUT 5 (R2)
shDmrtc2 (R1)
shDmrtc2 (R2)


Qdpr
Expression relative to shLacZ
1.4
1.2
1
0.8
0.6
0.4
0.2
0
d0 d1 d2 d3 d4 d5
shLacZ
MUT 1
MUT 5 (R1)
MUT 5 (R2)
shDmrtc2 (R1)
shDmrtc2 (R2)

Notably, of the genes that did exhibit differential gene expression between shDmrtc2 and controls, a maximal knock down of up to 50% was observed at some time points for some candidates, with many genes reaching a knock down of 40% or less. Although it cannot be known for any given gene the required extent of knock down to elicit a phenotype, it is probably unlikely that a 40-50% decrease in transcript level would result in such a striking phenotype as induced by shDmrtc2.

In order to approach the data analysis in a more systematic way, we worked in collaboration with the lab of Dr. Mihaela Zavolan. R. Gumienny used bioinformatics approaches including sequence target prediction and seed region count prediction to identify candidates that were predicted to be direct targets of shDmrtc2. That is, targets which had some degree of complementary sequence specificity to the antisense sequence of shDmrtc2. A biophysical model designed to identify miRNA targets and published by the Zavolan lab, called MIRZA, was used to determine direct target candidate genes (Khorshid et al., 2013). We hypothesized that a miRNA target prediction model can be applied to identify siRNA off-target since siRNAs act like miRNAs when a target does not have 100% complimentarity. This model takes into account several parameters underlying the miRNA-mRNA interaction facilitated by the Argonaute protein including sequence specific position-dependent energy parameters that may impose constraints on this interaction. In simpler terms, a MIRZA score can be calculated giving an indication of the likelihood that a specific sequence (in this case the sequence of shDmrtc2, particularly the 7 base pair seed region) can target and interact with any given (gene) sequence based on the sequence complementarity and other parameters outlined in the MIRZA model. The higher the MIRZA score, the more likely an interaction. When this model was applied for shDmrtc2, a list of genes was identified with purported potential to be targeted by shDmrtc2 (Figure 5.6, MIRZA score).

MIRZA SCORE	GENE	SEQUENCE MATCH /21	
390817	Dmrtc2	TATGTAGGTTTCTGTAAGCAA	21
499.136	Herpud1	TTTCATAGGCACTGTAAGCAG	12
612.312	Qdpr	ATTCTTTGGTGATGTAAGCAA	13
172.383	Cpne8	GAATCCTGATTTTGTAAGAAA	12
		GGATGGAGATGATGTAAGAGT	8
		ACTGTTCTTAAGAGTAAGTTT	9
127.122	Bnip3l	TTTGAGACCTATTGTAAGCAT	12
		AGAATCTTTATTTGTAAGGAA	11
		CCCTTTTTGGCCTGTAAGACA	9
		TCTTAGGAAGAAAGTAAGTCT	8
501.701	Mtap	TAAGCTGTTGATTGTAAGCAA	14
		CCCATCTCTGGCTGTAAGCAG	11
39.1379	Pak3	AAGCCAAGACTATGTAAGCCA	12
		ATGTAATTTATTTGTAAGCCT	10
		GTTTGGTTTTTAAGTAAGTTA	10
		CATTATACCATTAGTAAGGCA	9
		TTCATTAGTGATGGTAAGATC	9
		GTCTCCAGGATGAGTAAGACT	7
		AAAATGTCCCGAGGTAAGAAT	7
182.334	Nupr1	GAAATTAGGAGTTGTAAGCAG	11
186.095	Mllt11	CATGAGGGACCCTGTAAGTAG	13
		CTCTACCAGGCCAGTAAGCAA	9
		ATAAGACCGTGAAGTAAGCAG	9
86.3097	Dynlt3	AGAACTGAATATTGTAAGCAG	10
128.263	Aifl1	CAGTGAGATTTCAGTAAGCAC	14
		TTTTTTTCTCTATGTAAGCAC	13
		TCAAAAGATGTTTGTAAGGTA	12
464.235	Prpf19	CTTGCAGGGTTTTGTAAGCAG	15
		GCTCTAGACTGTTGTAAGCAG	13
		AAGAGAGCTGGTAGTAAGCTA	11
		CTCTGAGGTTGAGGTAAGTGA	11
		TTTCAACAACTAAGTAAGAGA	10
		CTGCTGTCCAAGGGTAAGGAA	8
488.091	Ccny	CATGTCCATTGTTGTAAGCAA	14
		ATCTGTGTGCTCTGTAAGAAA	11
		AATTACTAGGCATGTAAGTAT	9
		TTCTTCCTCACAAGTAAGATG	7
76.9604	Ssr3	TACGAAGTGTTCAGTAAGACA	14
		TGTGATGACCTTAGTAAGCAC	12
111.78	Rbms1	TCTAGTGCACCTTGTAAGCAG	11
		TTTCACATAGCGTGTAAGGTA	9
		GTGAAAGCTAAAAGTAAGGAT	9
1275.54	Rrm2b	GGTTGTGGTTCTTGTAAGCAA	14
		AATTTGTATAATAGTAAGTCC	9
292.1	Wisp1	GATATGATTGTTTGTAAGCAC	13
		CGTTAAAGAGTTTGTAAGCTA	12
		CATAAATGGATTTGTAAGTAG	11
		TCTCATCTTAACAGTAAGAAA	11
		GGCTGTGAATGCTGTAAGATA	10
		AACAGCTAACTCTGTAAGAAC	10
		CTCCTGCCATTTTGTAAGTGA	10
		TAGCCCCTTTCACGTAAGAGG	9
		ATCATCAGCAACAGTAAGGCT	8
		GTTGTCTCAAAAAGTAAGATG	8
		CCCAGTGACTGTGGTAAGGCC	7
153.012	Cd9912		
134.998	Zbtb41	CTTACATATTTTTGTAAGAAA	13
		TATCTACAGATGAGTAAGGAA	13
		CGTGGTTGGTGGTGTAAGAAG	11
		AAATGTGATATTTGTAAGAAA	11
		CAATGTGATGTTTGTAAGAAG	11
		TCATTTGAACCTAGTAAGCAA	11
		GCCTTTTAATGCTGTAAGCTG	10
		AGTTTCCAAAAATGTAAGAAC	9
23.8914	Calu	AGAGTCCTGTGCTGTAAGACA	11
		ACTTGAAATCACTGTAAGGCC	10
		GCTTGGAAAGATTGTAAGTAA	9
NA	Socs5	CTATTGAGTTTATGTAAGCAT	13
		TTATCTCTTTGTTGTAAGCTC	10
977.318	Zfp148	TGTTGTACTGTATGTAAGCAA	13
		TTTCTTGTACAGTGTAAGGGT	10
		TAGGGTTCATCGAGTAAGGAT	10
		AAAAAAATCAATAGTAAGAGA	8
666.658	Dpys12	TTTACAGGTCCCTGTAAGCAG	15
		TTCCGCAGTCCGTGTAAGCAT	11
NA	Prkd1	TGATAGTGGTGTTGTAAGCAA	12
		TATTTAAACTCTAGTAAGTGC	11
		GTCTGCCAGTTCTGTAAGAAG	10
282.774	Cd109	GTTTTTCTTTTTTGTAAGCAC	13
		GGCCAGAGCTTGTGTAAGCAC	11
		GCCCAGTGATGCTGTAAGTCA	10
		AGCAACGTAGAAAGTAAGCAG	8
		TAGAATGTAGGTAGTAAGTTT	8
		TTCACCTCTCCTAGTAAGACG	7
885.921	Trp53inp1	TTTAATCTTTTGTGTAAGCAA	14
		ACTGAAAGTACAAGTAAGAAG	11
		GTCTTTAAGTACAGTAAGGTA	9
		CACACCACCTATAGTAAGTGG	7
113.685	Acox3		
299.791	Celf2	TATGCTGTTTTTTGTAAGACA	15
		GAGAAAGGCTACTGTAAGCTG	13
		CCTAGTGGAAACTGTAAGACC	10
		AGTGGTTGGCCATGTAAGAGG	9
		TTAAAATACAGCGGTAAGTTT	8
696.384	Limch1	GTTTCTCTGTTATGTAAGCAA	12
		TTCTTGAGAGGCTGTAAGGAG	10
		TACGAGGAGGAGCGTAAGATA	10
		GCTCGGAGATGCAGTAAGCGG	10
116.793	Pde12	ACTTTTTATAGATGTAAGCAT	11
		AAATCTCTATATTGTAAGGGA	9
		TACTCTTAGGTTAGTAAGTCT	8
		CCGCAGCTGTTCAGTAAGTTG	8
171.013	Tomm22	TTGTGGGAAAGAAGTAAGTAC	9
		ATGTCTAGGACATGTAAGCAC	9

Figure 5.6. Computational and MIRZA analysis of microarray data in conjunction with analysis of gene expression profiles of reprogramming samples determines shDmrtc2 candidates. Candidate genes were determined based on MIRZA analysis and the number of seed counts identified (green = 1, orange = 2) by our collaborators in the Zavolan lab. Annotation of these genes included identification of sequences anywhere within the transcript of candidate genes with complementarity to at least 5 consecutive nucleotides in the shDmrtc2 seed region (red). The total number of nucleotides within the adjacent 21bp region of this sequence that matched the shDmrtc2 sequence was determined (number on the right). Candidates are listed (from top, left to bottom, right) in order of top hits based on transcript knock down from microarray data. Candidates highlighted in blue represent the genes containing a sequence with the highest number of complementary nucleotides to shDmrtc2 (15/21). Dmrtc2 is highlighted in purple.

The first thing to note is that Dmrtc2 has a score of 390817. This reaffirms that shDmrtc2 should indeed target Dmrtc2. Surprisingly the gene with the closest MIRZA score to this was Rrm2b with a score of 1275; two orders of magnitude lower than that of Dmrtc2. Indeed, most of the candidate genes had much lower scores than this, indicating that no genes were an obvious off target from which to start further analyses based on the MIRZA score alone. In addition to the MIRZA score, several additional annotations were made in order to narrow down a candidate list from which functional validation experiments could be carried out. This included a 'count of seeds' score (either 1 or 2) illustrating the number of times a sequence was found within the 3'-UTR of a given candidate gene that was complementary to the seed region of shDmrtc2 (as highlighted in red in Figure 5.6). Surprisingly, when I manually searched for sequence matches within candidate gene sequences I discovered that for many of the candidates there were several sequences that maintained sequence complementarity to at least 5 consecutive bases within the seed region of shDmrtc2, and further complementary bases within the surrounding 21bp region. In fact, 3 of these candidate genes (Prpf19, Dpysl2 and Celf2) had as much as 15/21 nucleotide complementarity in at least one region of the

transcript, which I hypothesized could reasonably be the most likely off-target candidates as it has been reported that as few as 11 to 14 base pairs can induce gene silencing (Dorsett and Tuschl, 2004). Interestingly, a few of the candidate genes that demonstrated some level of reduction in gene expression from the microarray were not denoted MIRZA scores but this down-regulation could be due to a secondary/indirect effect of shDmrtc2. Since the MIRZA score is calculated from several parameters, including seed region and whole 21bp sequence complementarity as well as nucleotide position dependent energy parameters between the shRNA and target gene, not all candidate genes generated a MIRZA score. In addition, there were some apparent inconsistencies with genes having a MIRZA score but no apparent sequence homology (Acox3 or Cd99l2 for example), however, this was put down to differences in specific transcripts used in the various analyses carried out on these data. This work was analyzed using early versions of prediction tools developed by our collaborators, which have recently published following optimization of the system (Gumienny and Zavolan, 2015).

With a candidate gene list generated I worked through validating some of these candidates by Q-PCR, as described previously, in shLacZ and shDmrtc2 conditions (Figure 5.7). Firstly, I confirmed the knock down of Dmrtc2 with shDmrtc2 by approximately 60-70%, which was consistent with previous data. Subsequently, I was able to confirm knock down by at least 50% of the majority of candidate genes in the shDmrtc2 reprogramming compared with shLacZ. Several of these genes including Qdpr, Herpud1, Rbms1, Nupr1, Bnip3l and Pak3 were consistently down-regulated over the time course, whereas others including Dpysl, Prpf19 and Ccny gave inconsistent results.

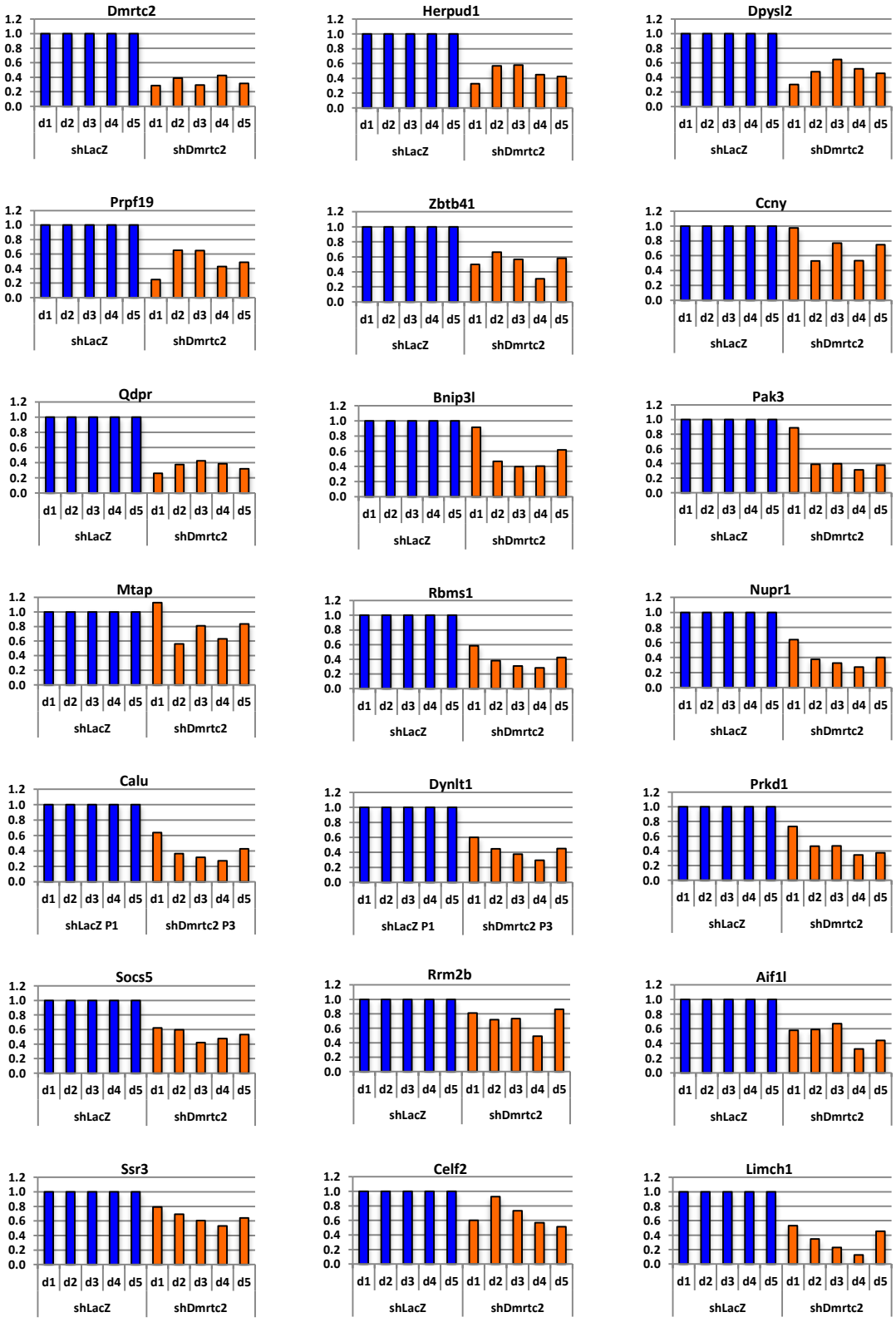


Figure 5.7. Q-PCR validation of candidate genes. Data represented as gene expression in shDmrtc2 sample relative to shLacZ sample.

This led us to wonder whether the microarray technique was sensitive enough to determine differential expression of the elusive off-target gene. In addition, there was no way to determine whether any observed down-regulation of genes was due to knock down by shDmrtc2 or if it was simply a consequence of reprogramming, particularly in the cases where down-regulation during reprogramming was the usual expression profile observed for a gene as is the case for Herpud1, Pak3 and Bnip3l among others (data not shown).

5.2.4 mRNA-sequencing reveals direct and indirect target candidates of shDmrtc2

Following investigation of the double nucleotide mutants, I sought to determine if single nucleotide mutants had any effect on reprogramming. A further 9 mutant vectors were created containing single nucleotide substitutions at different positions throughout the length of the shRNA sequence (Figure 5.8a, MUT 7-15). I was particularly interested to know if any of these mutants were able to recapitulate the phenotype observed in shDmrtc2. When added during reprogramming, mutants 9, 10, 12 and 15 had little or no effect on Nanog-GFP+ colony numbers. Mutants 7, 8, 13 and 14 gave rise to a moderate increase in the number of Nanog-GFP+ colonies observed compared with control, but strikingly, mutant 11 generated over 14-fold more Nanog-GFP+ colonies than shLacZ, almost fully reproducing the result observed with shDmrtc2. Therefore, I successfully identified a mutant of shDmrtc2 that promoted reprogramming in a similar way. I was surprised to find that out of the 15 double or single mutants that I tested in total only a single mutant, with only a single nucleotide difference, enhanced reprogramming to a similar extent as shDmrtc2. However, this supports the notion that shRNA is highly specific with even a single nucleotide mutation being sufficient to abolish target specificity.

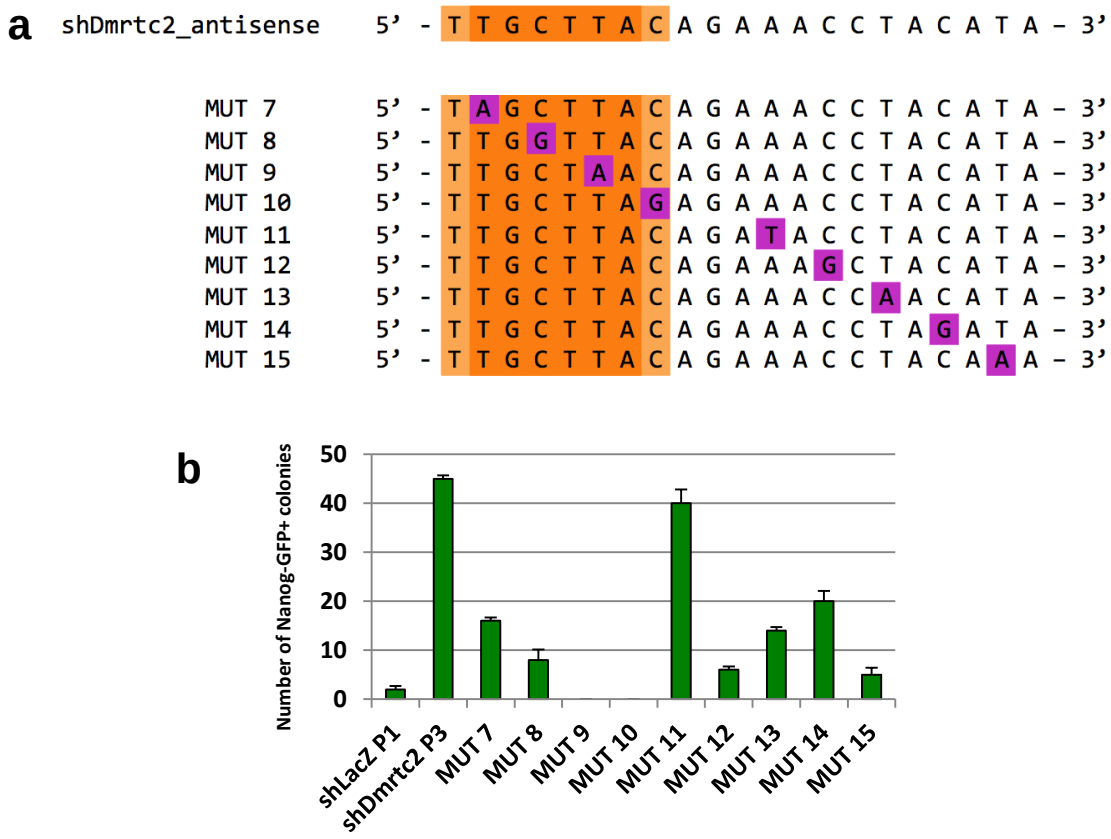


Figure 5.8. shDmrtc2 single nucleotide mutants during reprogramming. (a) 9 mutants with single nucleotide mutations were generated. Seed region is shown in orange and mutations are shown in purple. **(b) Mutant vectors were added during reprogramming and** Nanog-GFP+ colonies were counted on day 14. Mutant 11 gave rise to a similar number of Nanog-GFP+ colonies as shDmrtc2, while all other mutant vectors had little or moderate effect on reprogramming.

As a more sensitive approach to determining differential gene expression between samples we decided to use mRNA-sequencing (RNA-seq) to further the investigation. Additionally, instead of using reprogramming samples with shRNAs, we thought that MEFs infected with shRNAs would give us a clearer idea of genes that are down-regulated as a direct result of the shRNAs, rather than the reprogramming process. Briefly, to prepare samples MEF were plated and exposed

to a double infection of viral supernatant containing shRNAs for 24 hours each. Samples were collected 3 days after initial infection for RNA-seq analysis. Afzal Syed, from the Zavolan lab, carried out RNA-sequencing sample preparation following Northern Blot analysis. Initially, in addition to shLacZ and shDmrtc2, several samples were considered based on the results from the single nucleotide mutant experiment. A. Syed carried out Northern Blot analysis for these samples to determine whether small RNAs, i.e. shRNAs, could be detected (Figure 5.9). A probe against miR-199a was used as a positive control in wild type MEF and the shLacZ condition, as this is highly expressed in MEF. Indeed a band could be detected in both of these conditions (Figure 5.9, small RNA lanes 8 and 9).

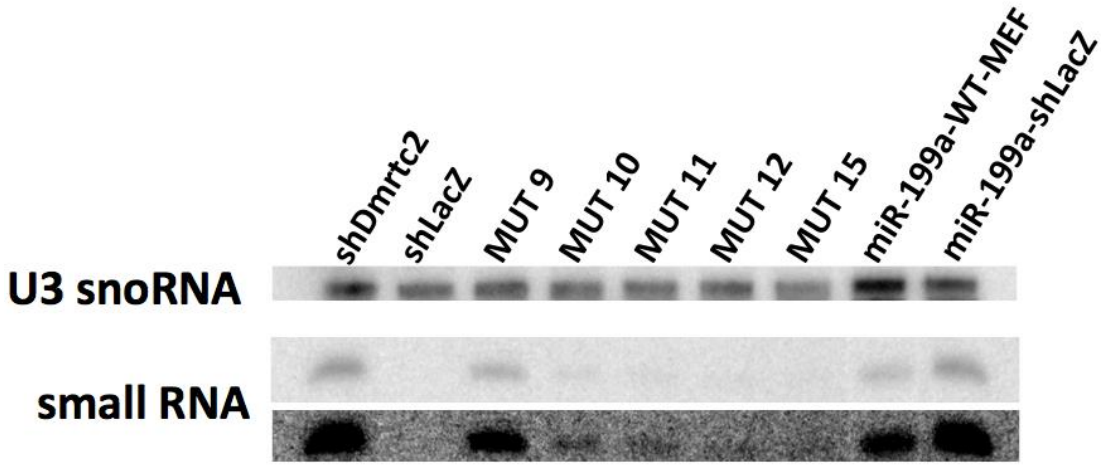


Figure 5.9. Northern Blot of samples for RNA-sequencing. U3 snoRNA was used as a loading control and miR-199a was used as a positive control (highly expressed in MEF). Bottom panel represents resulting blot with altered brightness/contrast. Northern Blot was performed by A. Syed.

Interestingly, while shDmrtc2 expressed small RNA well, MUT 11 did not. Given that MUT 11 replicates the enhanced reprogramming phenotype similarly to shDmrtc2 this was very surprising. Moreover, MUT 9, which was found to have little or no effect on reprogramming efficiency, expressed the shRNA well, with a

faint signal visible for MUT 10 also. However, when the image brightness/contrast was enhanced, a faint signal could be detected for all shRNAs.

In addition to the Northern Blot analysis, reprogramming was carried out using the same viral supernatant as used to infect the MEFs for the Northern Blot and RNA-seq samples. This was to ensure that the enhanced reprogramming phenotype was confirmed using the same virus/shRNAs. Indeed, as expected shDmrtc2 and MUT 11 gave rise to significantly more Nanog-GFP+ colonies than all other conditions when quantified on day 14 of reprogramming (Figure 5.10).

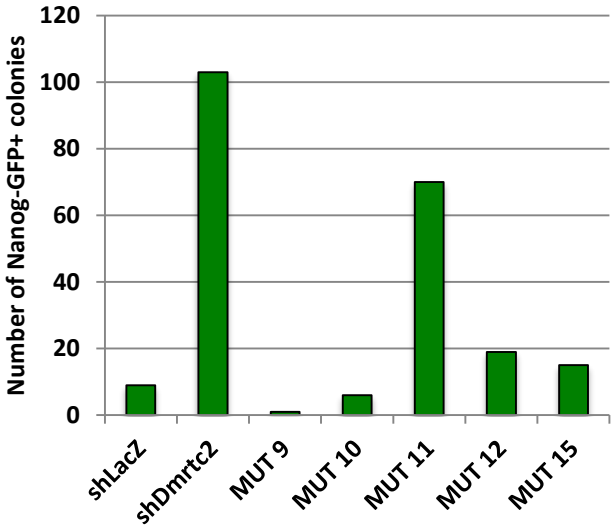


Figure 5.10. Confirmation of enhanced reprogramming phenotype using RNA-seq shRNA virus supernatant. Nanog-GFP+ colonies were counted on day 14 of reprogramming.

Following these results, the samples for RNA-seq were chosen; MUT 9 was used as an additional "negative" control with shLacZ and MUT 11 was chosen as an additional "positive" control with shDmrtc2. With these samples we theorized that the off-target gene would be more highly expressed in shLacZ and MUT 9 than in both shDmrtc2 and MUT 11 samples. However, this is with the caveat that the same off-target gene is responsible for the phenotype observed in shDmrtc2 and MUT 11. If a different mechanism were operating in these two conditions, the off-target gene may not be down-regulated in both samples.

A. Syed performed the RNA-seq and R. Gumienny performed computational data analysis from which P-values and target prediction scores were generated for genes when the two negative controls (shLacZ, MUT 9) were compared with the two positive conditions (shDmrtc2, MUT 11). By sorting the genes in order of p-value, candidate genes were determined by assessing the target prediction score and expression across all samples. The target prediction score was calculated from a combination of parameters based on a match to the 7-nucleotide seed region and the probability that this is a functional site, scaled to a threshold. For the top hits, the differential expression was considered for the positive samples compared with the controls and 10 direct targets were determined. In addition, we considered if an indirect target could be identified that may act downstream of the direct off-target by analyzing all differentially expressed genes, as opposed to only those with a predicted target score. Top hits that were the most differentially expressed between controls and positive samples were identified and interestingly, when a known and predicted protein interaction tool called STRING was used, 5 of these were suggested to interact each other (Figure 5.11). These were taken forward for further validation in addition to the 10 direct candidates identified.

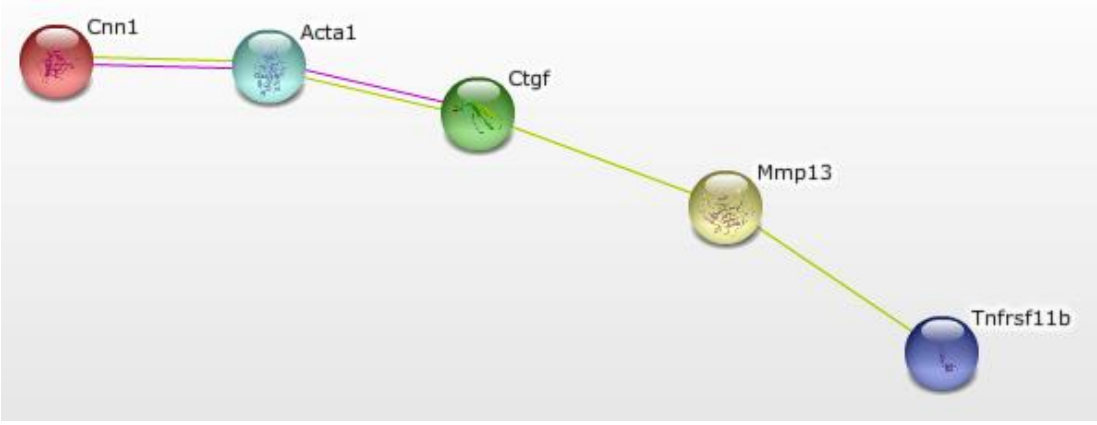


Figure 5.11. Indirect shRNA off-target candidate interaction identified by STRING tool. Putative protein-protein interactions were identified between five genes Cnn1, Acta1, Ctgf, Mmp13 and Tnfrsf11b using online tool STRING. (Image taken from website string-db.org, version 9.1, entering the gene names in the 'search multiple names' function. Interactions were identified by STRING from experimental or text mining data).

A workflow for the bioinformatics and candidate target selection is shown in Figure 5.12 and the 15 direct and indirect candidates with associated RNA-seq data analysis are summarized in Table 5.1. It was reasoned that the best way to analyze the RNA-seq data would be to first list genes based on their target prediction score (denoted by our collaborators using bioinformatics tools). This would give rise to the top hits for which putative targeting by shDmrtc2 was predicted. These targets were further narrowed by identifying those with a low P-value in combination with the most significant differential expression when comparing the positive and negative samples. In this way, the top 10 candidate targets that were putative direct targets of shDmrtc2 were determined. In addition, it was reasoned that including potential indirect targets would also be useful in determining potential pathways involved in the enhanced reprogramming phenotype induced with shDmrtc2. To this end, in order to determine potential indirect target candidates the entire data set was listed in order of genes that exhibited the most significant differential expression when negative and positive samples were compared irrespective of any other consideration. This gave rise to a vast list of genes, and in order to choose a number of genes to put forward for validation it was reasoned that genes which were involved in the same pathway or could be identified as having some functional link would be the best candidates to take forward. Indeed, 5 genes were determined to have putative protein-protein interactions as outlined in Figure 5.11.

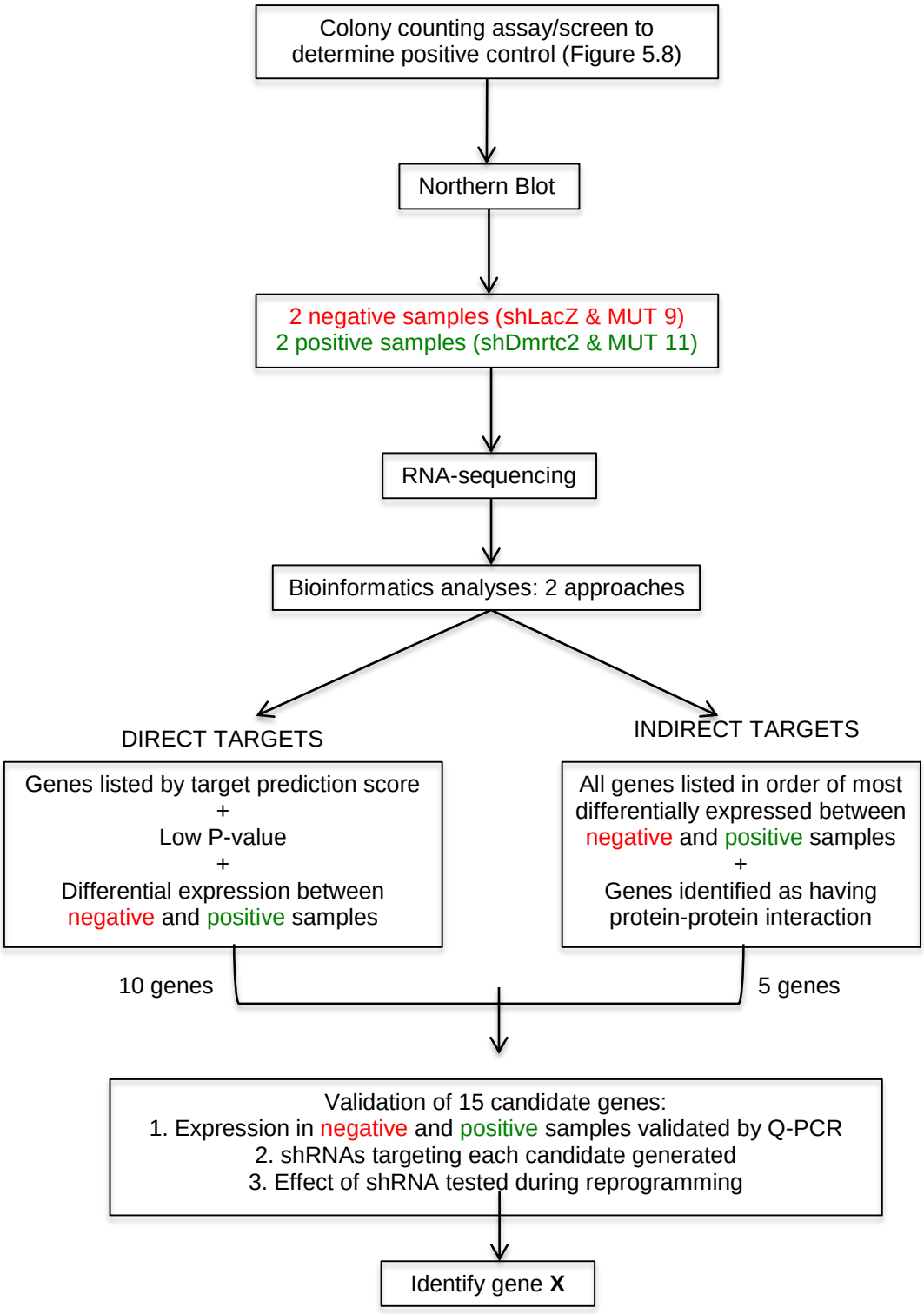


Figure 5.12. Workflow of bioinformatics and candidate target selection from RNA-seq analysis.

	Count of seeds	Adjusted P-value	RefSeq	Name	Predicton Score	nreads_ ShLacZ	nreads_ MUT 9	nreads_ shDmrtc2	nreads_ MUT 11	shD / shL	shD / MUT 9	MUT 11 / shL	MUT 11 / MUT 9
DIRECT CANDIDATES	3	0.00095	**NM_199476**	Rrm2b	0.87981	385	419	267	219	0.69259	0.63566	0.56783	0.52116
	2	0.00043	**NM_153098**	Cd109	0.83315	8782	9011	6959	6092	0.79236	0.77229	0.69365	0.67608
	1	0.00057	**NM_013822**	Jag1	0.63948	2291	2469	1819	1666	0.79400	0.73683	0.72719	0.67484
	1	0.01996	**NM_024236**	Qdpr	0.49665	483	477	325	327	0.67244	0.68059	0.67745	0.68566
	2	0.07047	**NM_029766**	Dtl	0.45635	556	543	394	397	0.70784	0.72437	0.71392	0.73059
	1	0.00052	**NM_181070**	Rab18	0.32939	4472	4506	3497	3197	0.78193	0.77602	0.71492	0.70952
	1	0.00588	**NM_008301**	Hspa2	0.31192	547	618	393	395	0.71774	0.63488	0.72196	0.63861
	1	0.00052	**NM_144731**	Galnt7	0.30063	762	805	501	533	0.65758	0.62233	0.70002	0.66250
	1	0.00095	**NM_021451**	Pmaip1	0.25696	3041	3202	2423	2120	0.79690	0.75667	0.69725	0.66205
	1	0.05838	**NM_008800**	Pde1b	0.18584	249	231	163	134	0.65612	0.70743	0.53815	0.58024
INDIRECT CANDIDATES	NA	NA	**NM_008607**	Mmp13	0.29984	225	303	169	169	0.75021	0.55606	0.75128	0.55685
			NM_008764	Tnfrsf11b	NaN	4425	5120	3467	3021	0.78358	0.67728	0.68261	0.59001
			NM_009606	Acta1	NaN	6462	6672	5343	4778	0.82680	0.80076	0.73932	0.71603
			NM_010217	Ctgf	NaN	49917	54863	41651	38704	0.83441	0.75918	0.77536	0.70546
			NM_009922	Cnn1	NaN	18069	18662	15282	13333	0.84573	0.81887	0.73786	0.71442

Table 5.1. Direct and indirect candidates from RNA-sequencing analysis. Candidates are listed in order of their prediction score. nreads = total number of reads assigned to the gene, count of seeds = number of shDmrtc2 seed sequences detected within the gene sequence.

5.2.5 Validation of direct and indirect target candidates of shDmrtc2

Following on from candidate selection from RNA-seq analysis, I sought to functionally test whether shRNAs targeting these genes had any effect on reprogramming. Although RNA-seq technology is more sensitive than microarray in terms of transcript detection, it was clear that none of the candidate genes were strongly knocked down with shDmrtc2 or MUT 11 (Figure 5.13). A maximum of 40-50% knock down was observed for a few of the candidates compared with the controls, including Rrm2b, Qdpr and Galnt7, however, a majority of the candidates only demonstrated a knock down of approximately 20-30%. Nonetheless, these candidates represented the top hits from the RNA-seq data analysis, and shRNAs were once again designed to test targeted knock down of these genes during reprogramming. As before, 3 shRNAs were designed and cloned for each of the 15 candidate genes. However some of the vectors were not cloned successfully during the first round of vector construction, and due to time limitations only the shRNAs generated during this round of cloning were tested during reprogramming.

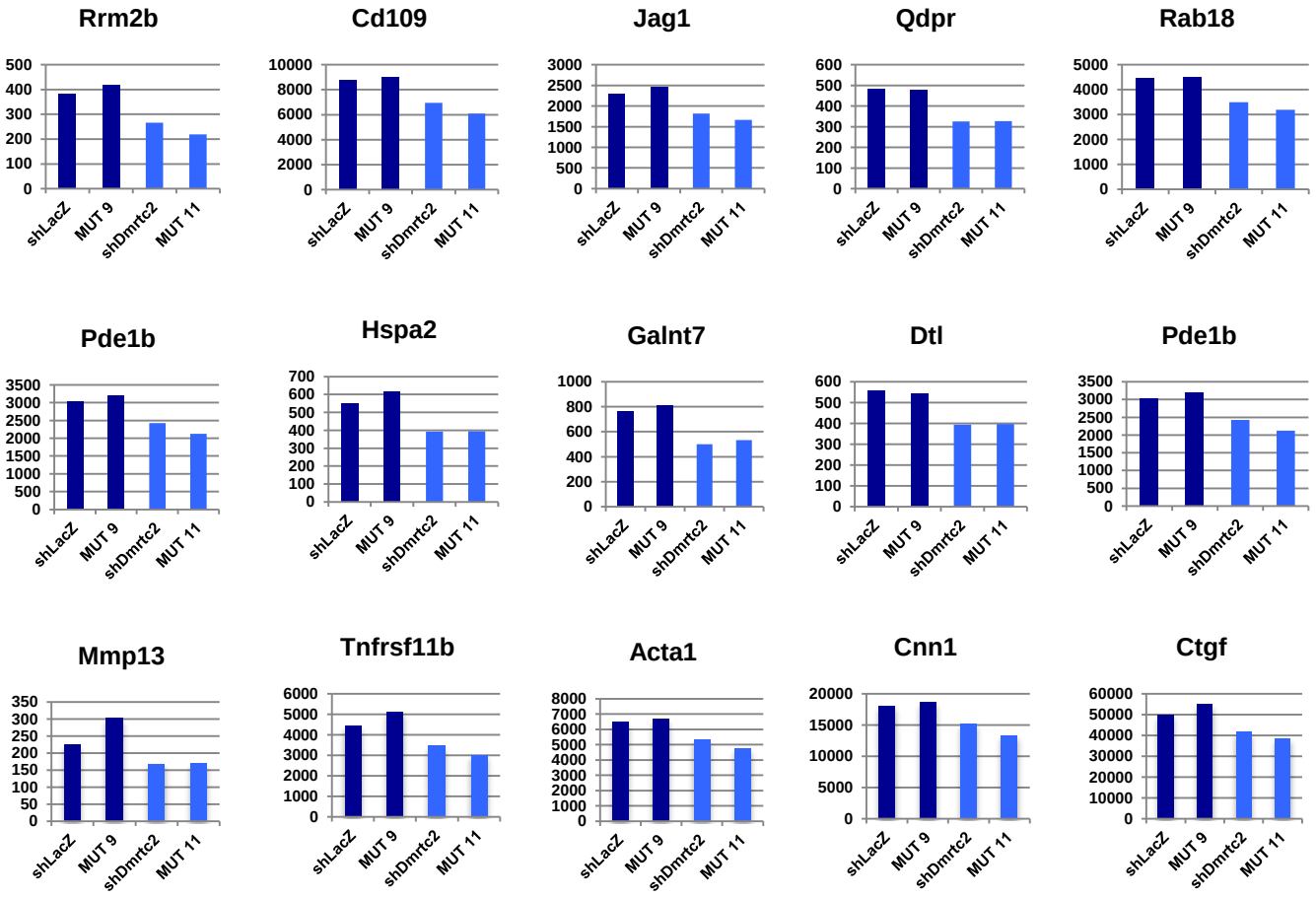
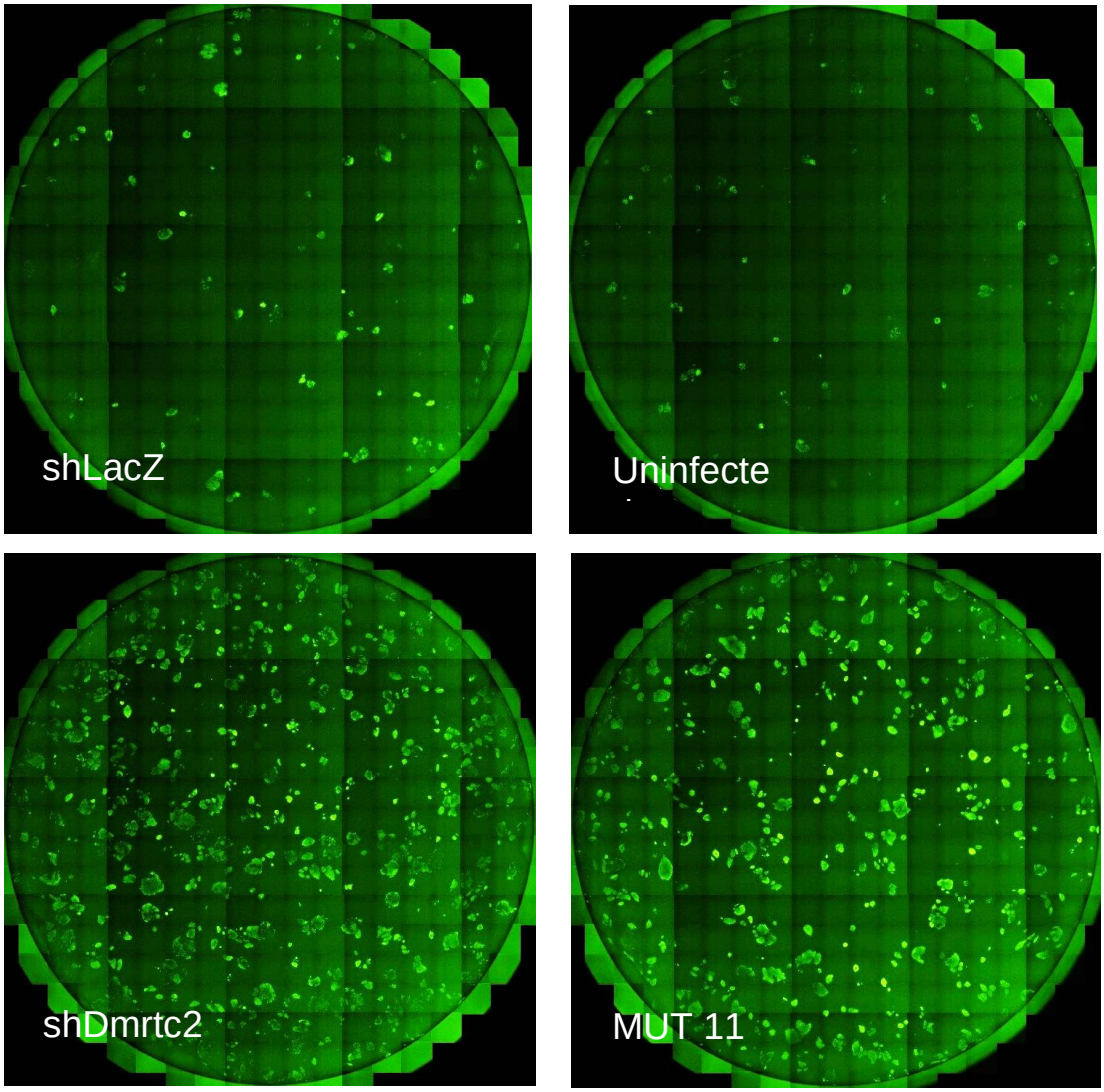


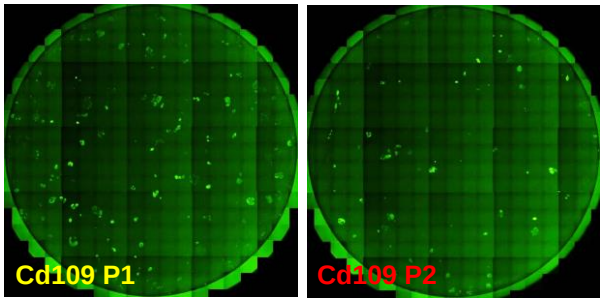
Figure 5.13. RNA-seq results for direct and indirect candidates. The number of total reads assigned per gene is represented (Y-axis) for each candidate gene. Direct candidates (first and second rows) and indirect candidates (bottom row) are shown. Dark blue bars indicate controls (shLacZ, MUT 9) and light blue bars indicate enhanced reprogramming samples (shDmrtc2, MUT 11).

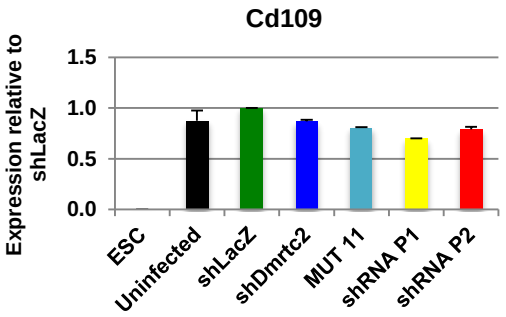
Transgenic MEF were infected with retrovirus supernatant carrying the candidate shRNAs, and reprogramming was induced by addition of dox and Nanog-GFP+ colonies were imaged and counted on day 14 using the Celigo cell cytometer instrument. Appearance of Nanog-GFP+ colonies was observed in the shLacZ and MUT 9 control conditions and as expected shDmrtc2 and MUT 11 gave rise to an abundance of bright green colonies by this time point (Figure 5.14). Unfortunately, no acceleration of reprogramming or increased efficiency was observed with the addition of any shRNA targeting any of the candidate genes (Figure 5.14). In parallel, samples of shRNA-infected MEF were taken to assess the knock down efficiency of the candidate shRNAs. Firstly, shDmrtc2 and MUT 11 conditions were analysed to determine whether expression of candidate genes was reduced compared with controls, as determined by the RNA-seq. Surprisingly, knock down of most of the candidates was not observed in shDmrtc2 and MUT 11 conditions. Galnt7, Pde1b and Ctgf were knocked down comparably to the levels observed from RNA-seq data, however, even this knock down was modest. The remaining candidates showed no difference in expression between controls and "positive" samples despite accelerated reprogramming being maintained in shDmrtc2 and MUT 11. This gave a hint that down-regulation of the candidate genes was not responsible for the phenotype. In addition, many of the shRNAs did not appear to knock down their reciprocal candidate gene, which suggest difficulty in designing efficient shRNAs. However, one or more shRNAs induced reasonable (>50%) knock down of some genes but resulted in no enhanced reprogramming phenotype providing some clue that these candidates may not be the off-target. These included Dtl (P2, P3), Galnt7 (P1, P3), Pde1b (P3), Pmaip1 (P3), Rrm2b (P2, P3) and Tnfrsf11b (P1). Nonetheless, even for these shRNAs there was not strong enough knock down to confidently rule out any of these candidates as the off-target. This further highlighted the difficulty of the shRNA system. Note that of the original 15 candidates, Qdpr was omitted from this analysis and Mmp13 knock down was not analysed. This is due to unsuccessful cloning of Qdpr shRNAs, and

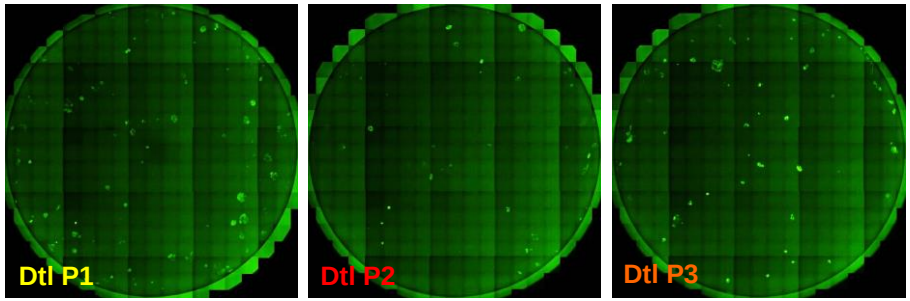
failure of Mmp13 primers during for Q-PCR. As such, only reprogramming data was generated for Mmp13.

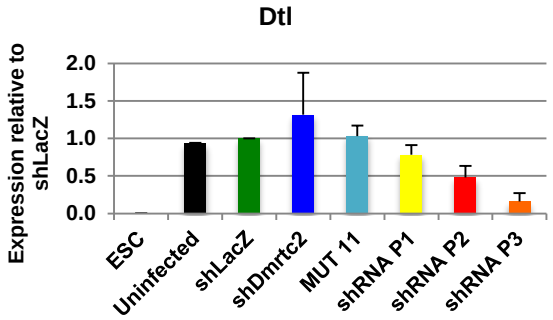
Figure 5.14. Whole well Nanog-GFP imaging of reprogramming with shRNA. The Celigo system was used to image whole 6 wells on day 14 of cells undergoing reprogramming with shRNAs targeting candidate genes. shLacZ and MUT 9 were used as controls and shDmrtc2 and MUT 11 demonstrate significantly enhanced Nanog-GFP+ colony number by comparison. Knock down of gene expression is shown with fold change of Nanog-GFP+ colonies represented relative to shLacZ.

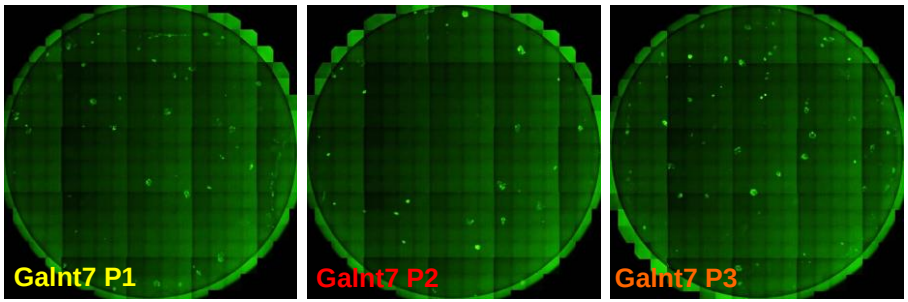


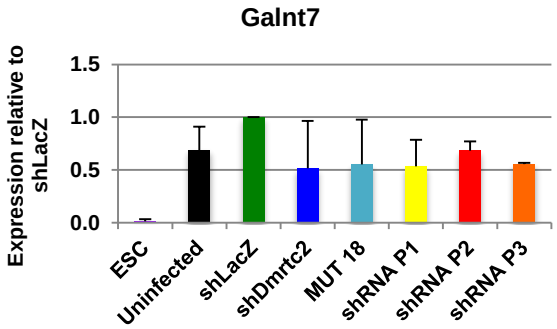

Cd109 P1
Cd109 P2

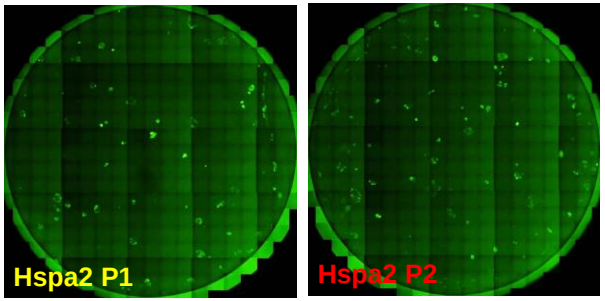

Cd109
Expression relative to shLacZ
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ
shDmrtc2
MUT 11
shRNA P1
shRNA P2

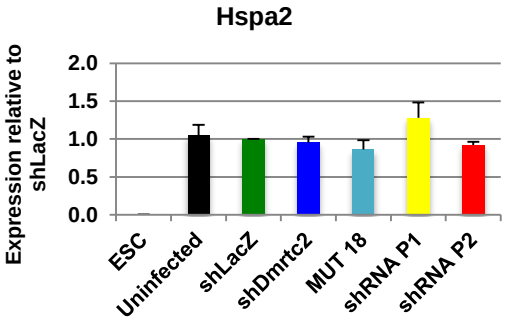

Dtl P1
Dtl P2
Dtl P3

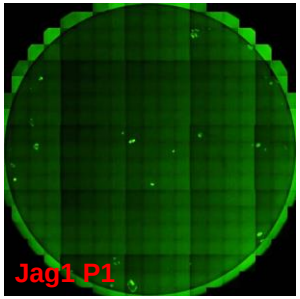

Dtl
Expression relative to shLacZ
2.0
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ
shDmrtc2
MUT 11
shRNA P1
shRNA P2
shRNA P3

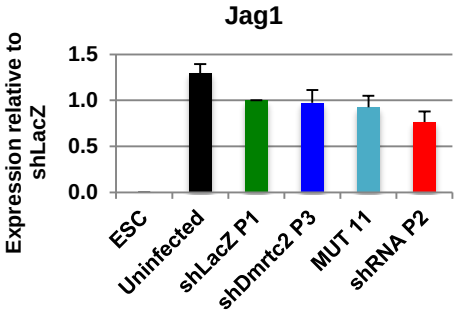

Galnt7 P1
Galnt7 P2
Galnt7 P3

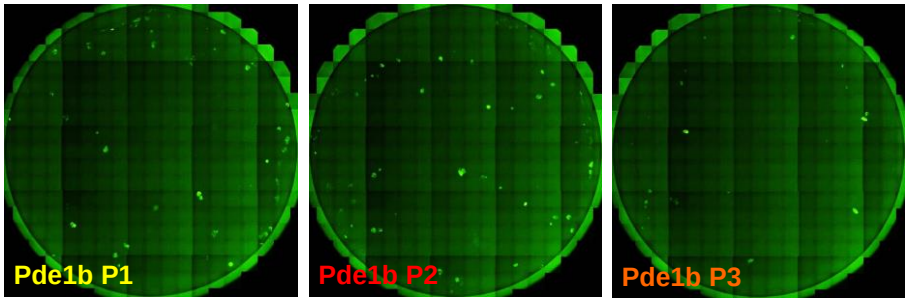

Galnt7
Expression relative to shLacZ
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ
shDmrtc2
MUT 18
shRNA P1
shRNA P2
shRNA P3

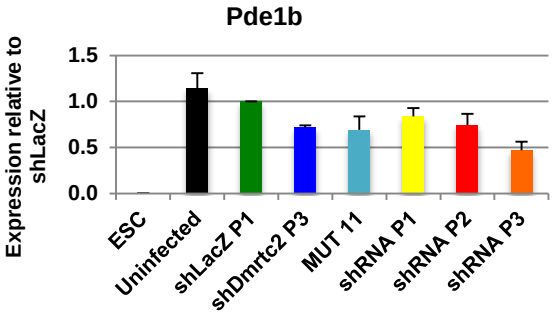

Hspa2 P1
Hspa2 P2


Hspa2
Expression relative to shLacZ
2.0
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ
shDmrtc2
MUT 18
shRNA P1
shRNA P2

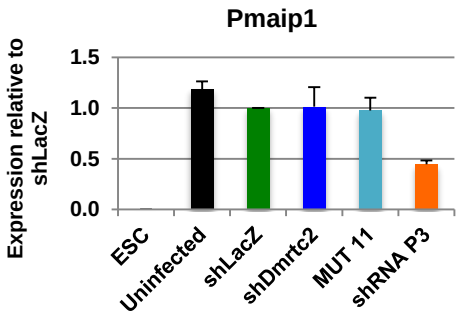

Jag1 P1

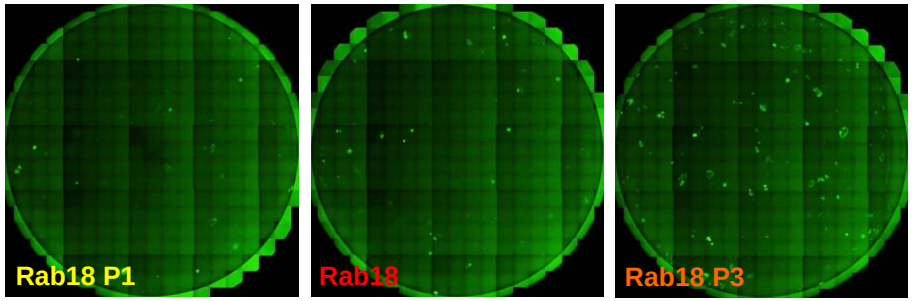

Jag1
Expression relative to shLacZ
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ P1
shDmrtc2 P3
MUT 11
shRNA P2

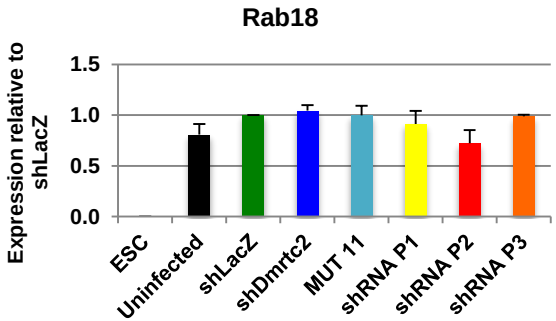

Pde1b P1
Pde1b P2
Pde1b P3

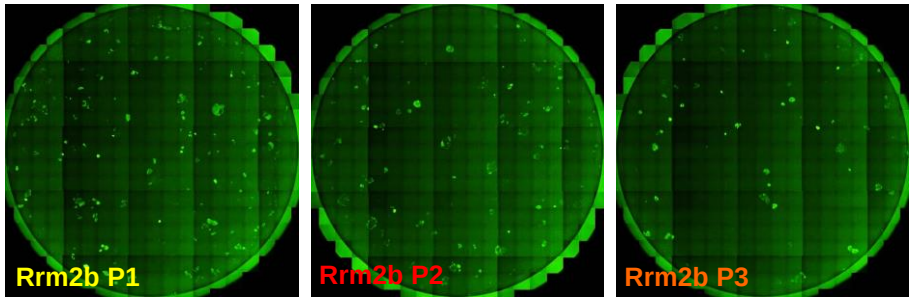

Pde1b
Expression relative to shLacZ
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ P1
shDmrtc2 P3
MUT 11
shRNA P1
shRNA P2
shRNA P3

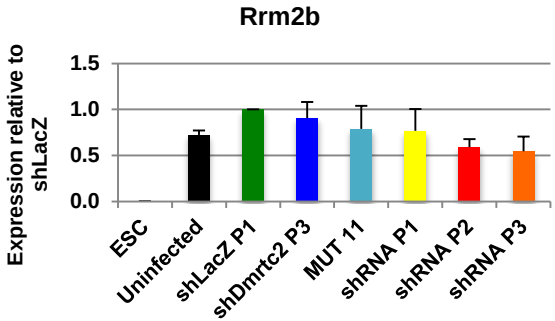

Pmaip1 P3

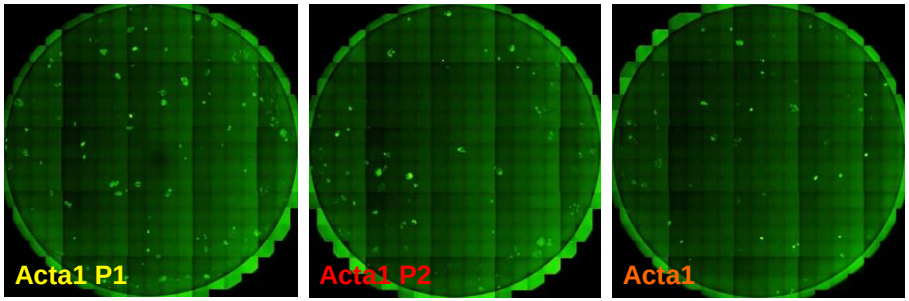

Pmaip1
Expression relative to shLacZ
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ
shDmrtc2
MUT 11
shRNA P3

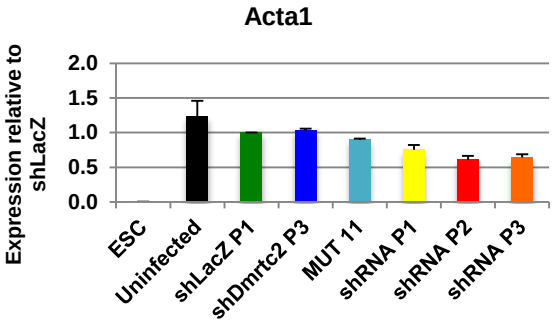

Rab18 P1
Rab18
Rab18 P3

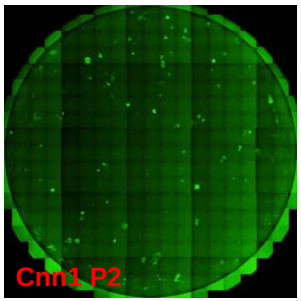

Rab18
Expression relative to shLacZ
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ
shDmrtc2
MUT 11
shRNA P1
shRNA P2
shRNA P3

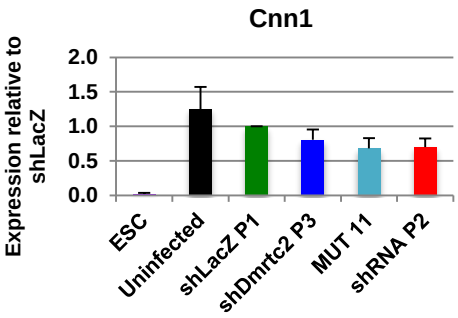

Rrm2b P1
Rrm2b P2
Rrm2b P3

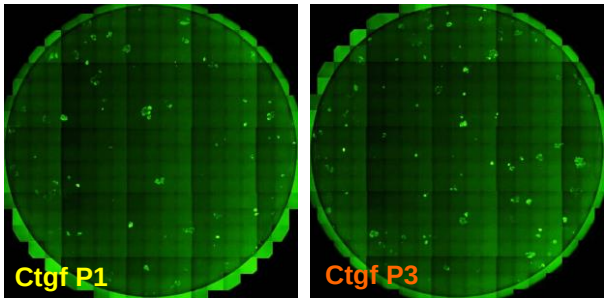

Rrm2b
Expression relative to shLacZ
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ P1
shDmrtc2 P3
MUT 11
shRNA P1
shRNA P2
shRNA P3

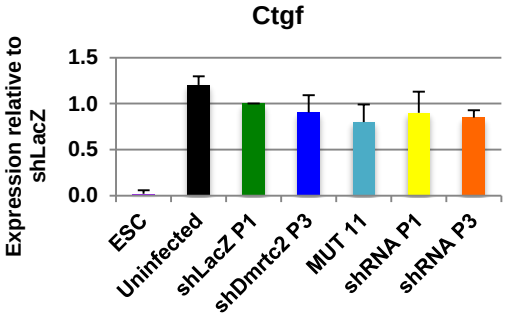

Acta1 P1
Acta1 P2
Acta1

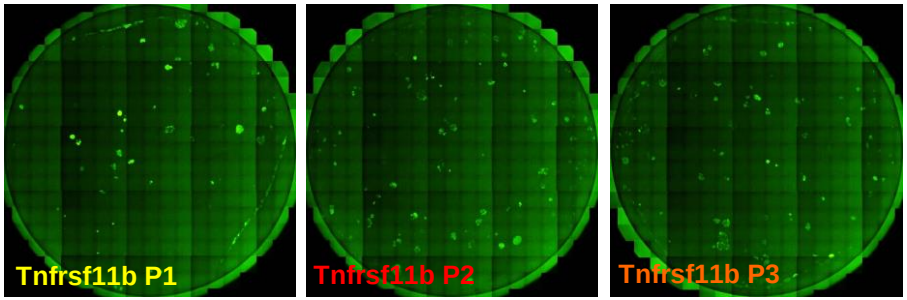

Acta1
Expression relative to shLacZ
2.0
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ P1
shDmrtc2 P3
MUT 11
shRNA P1
shRNA P2
shRNA P3

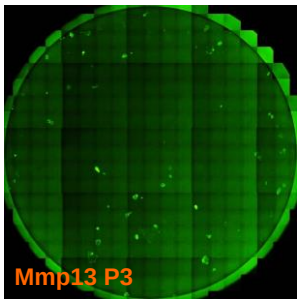

Cnn1 P2

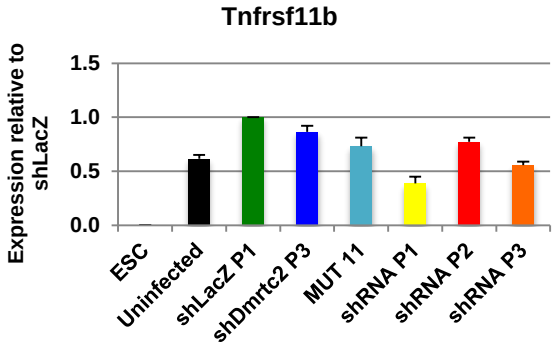

Cnn1
Expression relative to shLacZ
2.0
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ P1
shDmrtc2 P3
MUT 11
shRNA P2


Ctgf P1
Ctgf P3


Ctgf
Expression relative to shLacZ
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ P1
shDmrtc2 P3
MUT 11
shRNA P1
shRNA P3


Tnfrsf11b P1
Tnfrsf11b P2
Tnfrsf11b P3


Mmp13 P3


Tnfrsf11b
Expression relative to shLacZ
1.5
1.0
0.5
0.0
ESC
Uninfected
shLacZ P1
shDmrtc2 P3
MUT 11
shRNA P1
shRNA P2
shRNA P3

In addition to acquisition of whole well images, the Celigo instrument was used to quantify the number of Nanog-GFP+ colonies generated in each reprogramming well (Figure 5.15). These data confirmed the above observation that no shRNAs enhanced reprogramming to the level of shDmrtc2 or MUT 11. A few shRNAs induced a small increase in Nanog-GFP+ colony number compared with shLacZ including Cd109 P1, Pmaip1 P3 and Rrm2b P1, however with only an approximate 2-fold change, this is well within the realms of normal variation observed during reprogramming. These initial experiments suggest that the candidate genes analyzed are likely not the off-target underlying the enhanced reprogramming phenotype observed with shDmrtc2. There was no significant knock down observed for any of the candidate genes in the enhanced conditions (shDmrt2, MUT 11) by Q-PCR, and for several of the candidates, moderate knock down was achieved with at least one shRNA with no consequential effect on reprogramming observed. Nonetheless, more efficient shRNAs are required, which induce a more robust knock down before any confident conclusions can be drawn about these candidate genes.

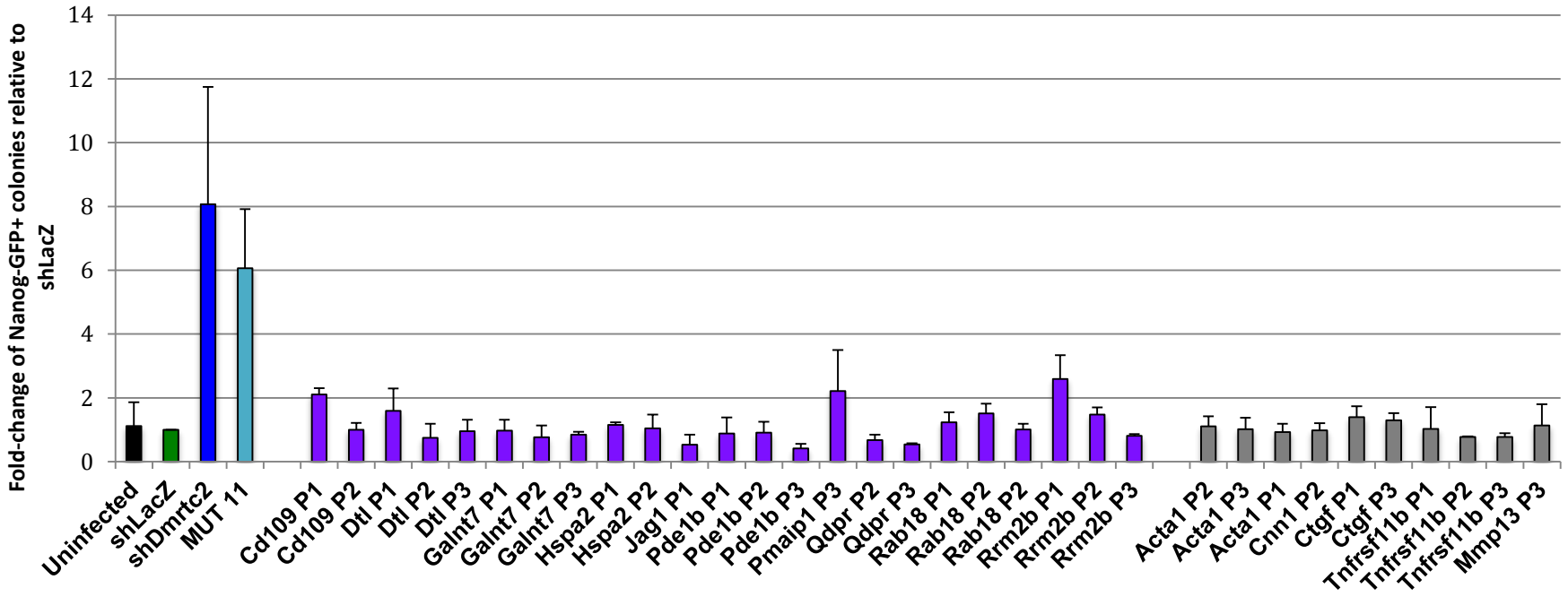


Figure 5.15. Quantification of Nanog-GFP+ colony number from Celigo whole well images. Images of reprogramming wells depicted in Figure 5.14 were quantified using Celigo software and fold-change of Nanog-GFP+ colony number relative to shLacZ was calculated for shRNAs targeting direct (purple) and indirect (grey) candidates.

At this time, the off-target effector of shDmrtc2 remains elusive, although given the striking phenotype of not only enhanced efficiency but perhaps even more importantly acceleration of reprogramming kinetics, it remains a worthy avenue for investigation. Determining the underlying cause of this phenomenon could lend some insight into the mechanism of reprogramming somatic cells to pluripotency, which is of great importance in both research and clinical contexts.

5.3 Discussion

The experiments and data discussed in this chapter give some indication that the candidate genes investigated are not likely to be the reprogramming enhancing off-target of shDmrtc2 that we sought to determine. However, no definite conclusion can be drawn due to the unclear results obtained from some of the experiments, some of which are discussed here.

Northern Blot analysis is an important validation of shRNA generation when using RNAi expression vectors as opposed to direct transfection of siRNA to cells to ensure that the exogenous expression system is working in concert with endogenous machinery to mediate RNAi production and subsequent gene silencing. In this study, a significant positive effect on reprogramming was identified with the addition of an shDmrtc2, and generation of small RNA was confirmed for this vector. However, another shRNA, MUT 11, which also promoted reprogramming was found not to produce small RNA well. The little shRNA expression detected with MUT 11 could be potentially explained in a few ways. Firstly, DNA probes with the sequence of shDmrtc2 were used for RNA hybridization in the Northern Blot analysis. Given that this already introduces a U-T mismatch, it may be possible that the additional single nucleotide changes in the mutants were enough to inhibit hybridization, and thus little or no signal could be obtained. In the case of MUT 11, it is interesting to note that there is an A to T mutation change compared with shDmrtc2, so it is plausible that this additional U-T mismatch with the probe caused robust hybridization inhibition. Another potential explanation is that the enhanced reprogramming phenotype observed with shDmrtc2 and MUT 11 is caused by different underlying mechanisms. Since a reasonable signal is detected in the Northern Blot analysis of shDmrtc2, it is likely that the phenotype is a result of expression of the shRNA. However, since little siRNA expression signal was detected with MUT 11, it is possible that a siRNA-independent mechanism might be promoting reprogramming in this context. For example, expression of the MUT 11 vector (which may or may not be producing mature siRNA) might be abolishing

some barrier to reprogramming by interfering with endogenous miRNA machinery/processing before export from the nucleus (since it seems no mature siRNA is detected from this vector), resulting in enhanced reprogramming. This may also explain the lack of candidates determined from the genome-wide expression analyses found to be significantly down-regulated in both shDmrtc2 and MUT 11 conditions. Although it is a possibility, I think it is unlikely that shRNAs with sequences that differ by only 1 nucleotide and promote reprogramming to a very similar degree would be functioning via different mechanisms. Repeating the Northern Blot analysis with a MUT 11 specific probe and using an RNA rather than a DNA probe will determine more conclusively whether MUT 11 is producing small RNA. Another possibility for the enhanced reprogramming phenotype observed with MUT 11 is that virus production with this vector may not be as robust as other vectors. If this is true, the observed increase in Nanog-GFP+ colonies could simply be a result of reduced viral/shRNA toxicity, resulting in a similar number of Nanog-GFP+ colonies being generated with MUT 11 as is observed in reprogramming with no viral infection. This could be clarified by determining viral titer for MUT 11 to confirm that virus is being produced with this vector. Taken together these experiments could give some insight as to whether the effect of this shRNA is likely to be one and the same as that observed with shDmrtc2.

Another drawback is the lack of knock down observed for many of the candidates by their reciprocal shRNAs or in the shDmrtc2 or MUT 11 samples by Q-PCR. One potential explanation for this is that retrovirus requires proliferating cells for high infection efficiency. Since MEFs have limited capacity for proliferation and the shRNAs are delivered by retroviral infection, it could be that the MEFs were not infected well enough due to slow/low proliferation rates and/or senescence. For the Q-PCR experiments, MEFs were plated in 6 wells and subsequently infected with shRNA-carrying virus and the cultures collected 3 days after infection. Since the infection efficiency was typically found to be approximately 50-75%, measured by pMXs-DsRed expression (not shown), and the cells were not sorted for shRNA incorporation prior to Q-PCR, it is possible that infection efficiency was not high

enough and/or proliferation of non-infected MEFs may have skewed the knock down results. However, although knock down of Dmrtc2 was not included as a positive control (since Dmrtc2 is not expressed in MEFs), Northern Blot analysis of MEF samples provided evidence that shDmrtc2 was robustly producing small RNA in this context, and furthermore the enhanced reprogramming phenotype was consistently maintained in every experiment suggesting that the shRNA expression vector system was working reliably. For these reasons I am confident that there should have been high enough expression of the shRNAs for knock down to be detected by Q-PCR, although sorting shRNA + MEFs samples prior to Q-PCR analysis would clarify this. Furthermore, it is unlikely that low infection efficiency was the reason for lack of knock down observed since all shRNA vectors were identical with the exception of the 21 bp shRNA sequence which varied from one vector to the next and knock down was observed for some shRNAs. Finally, the level of knock down for these candidate genes determined by RNA-seq was not exceptional with a maximum knock down of 40-50% observed. Since only a single sample of each condition was submitted for RNA-seq analysis, it is possible that the differences observed between controls and positive samples could be the result of normal variance in experimental samples. This would need to be clarified by submitting multiple replicates for each condition to determine a statistically significant difference in expression between controls and positive samples. However, it is probable that if the enhanced reprogramming phenotype is due to an off-target gene directly knocked down by shDmrtc2 then the down-regulation of this gene or genes would be fairly obvious between microarray and RNA-seq analysis. The fact that few genes have been determined through these two genome-wide analyses suggests to me that another mechanism may be in play. For example, given reports of shRNA expression vectors causing saturation of endogenous processing machinery, one possibility is that expression of shDmrtc2 is interfering with endogenous miRNAs or other regulatory non-coding RNAs (ncRNA) either through disruption of the RNAi processing pathway, or through partial complementarity resulting in off-target silencing. However, given the specificity of

shDmrtc2 expression correlating with enhanced reprogramming, I am certain that there is still some sequence specificity required for the phenotype rather than shRNA expression using this system simply being enough to overload and disrupt the endogenous process in a positive manner. Interestingly, a BLAST search of the shRNA sequence of shDmrtc2 results in several predicted ncRNAs coming up as the top hits after Dmrtc2 itself, based on sequence homology. In fact, one of these predicted genes, Gm31744, has 16 out of 21 contiguous homologous bases. In addition, a similar search for miRNA sequence homology using online tool miRBase (www.mirbase.org) identified some homology to mouse mir-221, which has been implicated in cell proliferation during angiogenesis (Nicoli et al., 2012; Santhekadur et al., 2012). Taken together, it is not unreasonable to suppose that shDmrtc2 may be exerting its effect through regulation of some factor or mechanism that is not detectable by the conventional microarray and RNA-seq techniques that we utilized.

CHAPTER 6 - Discussion

3.1 Knockdown of transient up-regulated genes during reprogramming

In the work presented in this thesis, I used retroviral based shRNA expression vectors to target candidate genes that were identified from RNA-sequencing data generated within our lab as exhibiting transient up-regulation during reprogramming. I hypothesized that the transient up-regulation could be aberrant or required for reprogramming and thus acting as either a barrier or essential function of reprogramming, respectively. In the former case, I theorized that knock down of these "UP-DOWN" genes could lead to a "short cut" during reprogramming by abolishing aberrant transient up-regulation and consequently reprogramming efficiency and/or kinetics would be enhanced. On the other hand, if transient up-regulation was essential then knock down would have a negative effect on reprogramming, and therefore overexpression could be tested to determine whether high expression of a gene resulted in enhanced reprogramming. My results concluded that knock down of several of the original candidate genes gave rise to enhanced reprogramming efficiency and/or kinetics and one of these, shDmrtc2, was further investigated to determine the role in reprogramming as it significantly, and reproducibly, enhanced reprogramming. To validate this result, I tested several alternative shRNAs targeting Dmrtc2 to try to replicate the phenotype, in addition to overexpression experiments to 'rescue' the enhancement of reprogramming. However, I could not replicate nor rescue the phenotype with these experiments and I concluded that although shDmrtc2 knocked down expression of Dmrtc2, this was not responsible for the enhanced reprogramming phenotype observed. Instead, I surmised that an off-target effect of shDmrtc2 caused the phenotype. I subsequently sought to identify the off-target causal gene by employing microarray and RNA-sequencing global gene expression analyses to determine candidate genes that had differential expression between shDmrtc2 and control samples. I carried out functional validation of candidate genes, using the same retroviral shRNA

expression vector system to knock down candidate gene expression during reprogramming but did not observe enhanced reprogramming with any candidate shRNA. Thus the causal gene responsible for the enhanced reprogramming phenotype imparted by shDmrtc2 remains elusive.

Expression of the Yamanaka factors during reprogramming causes massive global genome changes as cells transition from a somatic to a pluripotent state. In this work I investigated the role of transiently up-regulated genes in an endeavor to gain insight in to the role of such genes. One possibility is that expression could be essential for reprogramming, in which case their transient up-regulation may be a required step for progression of cells to an iPSC state and subsequently, overexpression of these genes could enhance reprogramming (if high levels are required). However, since expression is subsequently down-regulated and remains low or not detected in iPSCs and ESCs it is possible that only a short burst of expression is required and overexpression could hinder reprogramming. For example if downstream gene expression is tightly regulated, then overexpression of the primary gene could cause fatal disruption. On the other hand, transient up-regulation could represent a roadblock to reprogramming by aberrant activation of barrier genes, in which case knock down could promote reprogramming. Equally, it is possible that many of the genes within this UP-DOWN group have no function or impact on reprogramming and therefore alternative modulation would have no effect.

High reprogramming factor expression is the most obvious reason underlying massive dynamic changes in global gene expression during the first days of reprogramming. Nonetheless, other contributing factors are also notable. For example, somatic cell populations used for reprogramming are almost always highly heterogeneous. One of the most commonly used cells types for reprogramming – MEFs – are derived in a crude way involving basic and unrefined dissection of mouse embryos from which cells from various tissue types of a developing mouse are isolated. The resulting heterogeneous starting population likely gives rise to a predisposition for some cells more than others to be more

amenable to the reprogramming process depending on "starting" global gene expression status, contributed to by transiently up-regulated gene expression. Consequently, this heterogeneity is likely a contributing factor to the low reprogramming efficiency observed with most reprogramming systems. Transient up-regulation of genes is not limited to our reprogramming system and has been reported by others. Transient up-regulation of Foxd1 has been reported as a predictor of iPSC potential (Koga et al., 2014) and Bar-Nur et al. (2014) found that transiently up-regulated genes maintained high expression in response to defined media conditions, which promoted synchronous reprogramming of cells. This suggests that at least for some transiently up-regulated genes, high expression facilitates reprogramming and therefore subsequent down-regulation may contribute to a delayed and/or asynchronous reprogramming population. Takahashi et al. (2014) reported transient up-regulation of mesendodermal genes during reprogramming of human fibroblasts, suggesting that cells transitioned through a primitive streak like phase. This highlights the question of whether cells are becoming a different cell type on the route to iPSCs. These studies demonstrate that transient expression of genes can often be critical in cell processes and fate determination and there is a correlation between this pattern of gene expression and transdifferentiation. A prime example of this is reprogramming of somatic cells to iPSCs via transient expression of the four reprogramming factors in a variety of cell types. Transient up-regulation of genes, in combination with optimal culture conditions is also reported for transdifferentiation of fibroblasts or hepatocytes to neurons (Vierbuchen et al., 2010; Marro et al., 2011; Ambasudhan et al., 2011; Yoo et al., 2011). In the latter case, the authors used transient expression of exogenous Ascl1, Brn2 and Myt1l to induce neuronal fate while endogenous expression of these genes was up-regulated. A number of different cell types have been used in transdifferentiation studies, where transient up-regulation of transcription factors and/or miRNAs have been used to convert cells from one fate to another including fibroblasts to cardiomyocytes, blood progenitors or hepatocytes (Ieda et al., 2010; Szabo et al., 2010; Sekiya and Suzuki, 2011; Huang et al., 2011) and pancreatic

exocrine cells to β-cells (Zhou et al., 2008). These studies highlight the functional importance of transiently expressed genes in the context of cellular processes and fate determination, further compounding our interest in the role of the "UP-DOWN" group of transiently up-regulated genes during reprogramming in our system.

3.1 Caveats of this work

In this work, I found an shRNA vector designed against Dmrtc2 significantly accelerated kinetics and increased efficiency of reprogramming when delivered together with the Yamanaka factors. However, it turned out that the positive effect on reprogramming was due to off-target effects, rather than knockdown of Dmrtc2. Therefore, I aimed to identify the causal gene of the phenotype creating mutant shRNA vectors that maintain or abolish the reprograming enhancement activity and analyzing gene expression changes caused by the shRNAs. An important caveat of my work was that I took this approach based on a hypothesis that the enhanced reprogramming phenotype was caused by decreased mRNA levels of the causal gene via off-target effects of shDmrtc2. When shRNA is expressed, the RISC complex incorporates the processed siRNA to identify target mRNA for subsequent degradation. This is usually the case when there is perfect pairing between the siRNA and target RNA (as observed with knock down of Dmrtc2 mRNA by this shDmrtc2) usually occurring within the open reading frame sequence (Zeng et al., 2003; Agami, 2002). However, there is another major pathway by which RNAi can induce gene repression and that is by translational inhibition. This pathway is predominantly triggered when there is imperfect pairing between siRNA and the target (Agami, 2002; Saxena et al., 2003; Doench et al., 2003), and therefore is most commonly observed in a miRNA-mediated manner since miRNAs mainly act by mismatched targeting of 3'-UTRs in target genes (Zeng et al., 2003; Chu and Rana, 2006; Doench et al., 2003; Valencia-Sanchez et al., 2006). In fact, since we expect that an off-target would have imperfect complementarity to shDmrtc2, it is indeed

feasible and arguably likely that the off-target could be functioning in this manner. A way to determine whether translational inhibition is occurring is to measure the protein level of a gene. If there is a drop in the detected protein level but no difference in mRNA expression then this may suggest that translational inhibition is the mechanism in play (Gu and Kay, 2010; Chu and Rana, 2006). In other words, as opposed to siRNA mediated cleavage occurring, the mRNA persists but translation is inhibited and protein production is reduced resulting in a phenotypic change. Given that significant repression by shDmrtc2 of any gene could not be detected by the genome-wide analyses carried out, this could be an avenue for exploration going forward. However, for the RNA-sequencing analysis we used MEFs with shRNA expression without reprogramming factor expression, making the assumption that the off-target was a gene that was expressed in MEFs. It is possible that it may have not been expressed in MEFs but was rapidly up-regulated during reprogramming (as with Dmrtc2), in which case I would miss it in the RNA-sequencing results. Nonetheless, I expected that any significant differential gene expression between shDmrtc2 and controls during reprogramming would be picked up as hits in the microarray analysis. siRNA pathways have also been implicated in chromatin regulation in plants, yeast and multicellular organisms such as *C. elegans* and similar mechanisms of RNAi regulation have been shown in animals including *D. melanogaster*, though to a lesser extent (Moazed, 2009). In addition thousands of longer noncoding RNAs are well characterized in eukaryotes including Xist and HOTAIR, which modulate chromatin state. The possibility of shDmrtc2 operating by way of chromatin regulation was not probed in the scope of this study, however, non-coding RNAs should be accounted for with RNA-seq as far as those that contain a polyA tail, however, no such hits came up in the analyses. Finally, a drawback to the methods of global gene expression analyses undertaken in this work is that neither the microarray nor conventional RNA-seq would detect miRNA expression, and thus the potential that shDmrtc2 exerts its effect through modulation of miRNAs to mediate enhanced reprogramming remains a possibility.

3.2 Future directions

An alternative approach to global gene expression analysis for determining the off-target of shDmrtc2 is to use Argonaute cross-linking immunoprecipitation followed by sequencing (Ago CLIP-seq). Briefly, this technique involves *in vivo* cross-linking of protein-RNA complexes by UV light then subsequent immunoprecipitation of the protein of interest (in this case, the RISC component Ago2). A 3' radiolabelled linker is attached to RNAs allowing for visualisation of Ago2-RNA complexes following SDS-PAGE separation. After RNA extraction a 5' adaptor is ligated allowing for subsequent sequencing. This technique is widely used in studies of miRNAs and their interactions with both processing complexes and mRNAs (Clark et al., 2014; Wen et al., 2011; Chi et al., 2009). Computational analysis enables identification of putative mi/siRNA targets based on sequence match and thus is a useful tool could be a useful tool in determining the off-target of shDmrtc2 (Chou et al., 2013). A preliminary Ago CLIP-seq experiment was carried out by the Zavolan lab during the final stages of this work but no conclusions could be drawn from the data. Partly because the experiment was only performed once, and partly because no further validation was carried out due to time limitations, however, some observations could be made. It has been reported that during the CLIP protocol miRNA-target hybrids can get ligated and when such hybrids were determined for shDmrtc2, a list of putative target candidates was collated. Only reads that were supported by a significant number of independent CLIP reads (i.e. not hybridized) were included in this list and although this is not as quantitative as differential expression data, it may give some additional information or clues as to the putative off-target. Notably none of the 15 final candidates determined from the RNA-seq were among the top hits from the hybrid CLIP data, yet some genes that came up during earlier analyses (not shown), including Trp53inp1, Wisp1, Emp2 and Lpp were observed. Interestingly, the top hit in this analysis was B230219D22Rik, an uncharacterised putative protein-coding gene, highlighting the possibility that the off-target candidate could be some unknown and

uncharacterised gene. These observations are speculative at this point, but closer inspection of the results following replication of the CLIP-seq experiment in conjunction with the previous global gene expression analyses could give some further clues as to the mechanism underlying enhanced reprogramming with shDmrtc2.

Whilst shRNA can be a very useful tool in which to observe the effects of gene knock down in a relatively efficient manner, it may not the best tool now available when investigation of several genes is required, as in this study. A plausible alternative would be to employ CRISPR/Cas9 mediated gene editing, as described in Section 1.5.1. In the context of my shDmrtc2 work this approach could be employed as an alternative to shRNA in the validation of candidate genes, although with the proviso that a knock out phenotype can differ from a knock down phenotype. By using CRISPR-Cas9 technology to completely knock out rather than attempt to knock down expression of candidate genes, it could be reliably deduced whether there was an effect in the context of reprogramming that recapitulated that of shDmrtc2. Ideally, if this technology were available at the time of the original screening stage of UP-DOWN genes, it would have been the preferred method of screening. Given that there were several other shRNAs in addition to shDmrtc2 that enhanced reprogramming efficiency, it could be worth targeting these genes using a CRISPR-Cas9 system to validate the data generated using shRNA. If the phenotype observed was found to be due to specific targeting of the corresponding gene then these would be novel regulators of reprogramming not previously described.

In the context of this study whereby the targeting of unwanted genes by shDmrtc2 resulted in significant enhancement of reprogramming, positive controls are acknowledged to be of particular importance upon reflection of the work. Several so-called negative mutants were derived, that is shRNAs similar in sequence to shDmrtc2 that did not replicate the positive phenotype, and it was hypothesized that these would contribute to delineating the correct target that gave rise to increased reprogramming efficiency and kinetics upon comparison to shDmrtc2. However upon reflection it is clear that with each of these negative

mutants came additional off-targets, and thus the global gene expression for each mutant shRNA likely widened rather than narrowed the effector candidate gene list. Furthermore it is possible, if not probable, that since differential gene expression was being analyzed with this strategy that the natural variation of gene expression on a global scale in addition to the method of bulk population analysis used, obscured the resolution we sought and the answer was obviously not clear-cut.

In hindsight, a more effective strategy may have been to pursue screening of shDmrtc2 mutants, with single or double nucleotide substitutions, in order to determine several more positive control mutant shRNAs (in addition to MUT 11). By using a wider panel of positive control mutants for examination and comparison of global gene expression with one another and shDmrtc2 to determine genes that showed similar rather than differential expression between the panel of positive controls and shDmrtc2 it is likely that a smaller list of genes would be derived, and that the target gene may be more easily identified by this method. Using samples of cells for the gene expression analyses that have been sorted for the expression of reprogramming factors (mOrange) and shRNA (if fused to a reporter or tag) would further strengthen the resolution of results acquired. Importantly, this is assuming that any positive reprogramming phenotype garnered by each of the positive mutants is due to targeting of the same effector gene.

Given the striking enhanced phenotype imparted by shDmrtc2 on reprogramming there is of course a certain level of frustration with being unable to identify the cause or causes underlying this effect. The vast and complex dynamic changes which cells sustain during reprogramming have made the dissection process particularly difficult and the broad known and potentially unknown off-target effects of sh/siRNAs has only added to the complications. Technologies such as CLIP-seq and CRISPR-Cas9 could play meaningful roles in facilitating these investigations, though it could equally result in a never-ending chase. It is also possible that the phenotype caused by shDmrtc2 was due to knockdown of multiple genes, instead of one single gene. An alternative approach to understand how the

reprograming enhancement occurs would be focusing on the global gene expression changes resulting from shDmrtc2 during reprogramming in order to determine differential pathways or factors activated in this condition in comparison to conventional reprogramming with slow kinetics. In this way, investigation of indirect, rather than direct, targets of shDmrtc2 may lead to some novel insight of the reprogramming process not previously reported.

References

AGAMI, R. 2002. RNAi and related mechanisms and their potential use for therapy. *Current Opinion in Chemical Biology,* 6**,** 829-834.

AMBASUDHAN, R., TALANTOVA, M., COLEMAN, R., YUAN, X., ZHU, S., LIPTON, S. A. & DING, S. 2011. Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. *Cell Stem Cell,* 9**,** 113-8.

AMIT, M., CARPENTER, M. K., INOKUMA, M. S., CHIU, C. P., HARRIS, C. P., WAKNITZ, M. A., ITSKOVITZ-ELDOR, J. & THOMSON, J. A. 2000. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol,* 227**,** 271-8.

AMIT, M., SHARIKI, C., MARGULETS, V. & ITSKOVITZ-ELDOR, J. 2004. Feeder layer- and serum-free culture of human embryonic stem cells. *Biol Reprod,* 70**,** 837-45.

ANG, Y. S., TSAI, S. Y., LEE, D. F., MONK, J., SU, J., RATNAKUMAR, K., DING, J., GE, Y., DARR, H., CHANG, B., WANG, J., RENDL, M., BERNSTEIN, E., SCHANIEL, C. & LEMISCHKA, I. R. 2011. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell,* 145**,** 183-97.

ANOKYE-DANSO, F., TRIVEDI, C. M., JUHR, D., GUPTA, M., CUI, Z., TIAN, Y., ZHANG, Y., YANG, W., GRUBER, P. J., EPSTEIN, J. A. & MORRISEY, E. E. 2011. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell,* 8**,** 376-88.

AOI, T., YAE, K., NAKAGAWA, M., ICHISAKA, T., OKITA, K., TAKAHASHI, K., CHIBA, T. & YAMANAKA, S. 2008. Generation of Pluripotent Stem Cells from Adult Mouse Liver and Stomach Cells. *Science,* 321**,** 699-702.

BANITO, A., RASHID, S. T., ACOSTA, J. C., LI, S., PEREIRA, C. F., GETI, I., PINHO, S., SILVA, J. C., AZUARA, V., WALSH, M., VALLIER, L. & GIL, J. 2009. Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev,* 23**,** 2134-9.

BAR-NUR, O., BRUMBAUGH, J., VERHEUL, C., APOSTOLOU, E., PRUTEANU-MALINICI, I., WALSH, R. M., RAMASWAMY, S. & HOCHEDLINGER, K. 2014. Small molecules facilitate rapid and synchronous iPSC generation. *Nat Methods,* 11**,** 1170-6.

BASSIK, M. C., KAMPMANN, M., LEBBINK, R. J., WANG, S., HEIN, M. Y., POSER, I., WEIBEZAHN, J., HORLBECK, M. A., CHEN, S., MANN, M., HYMAN, A. A., LEPROUST, E. M., MCMANUS, M. T. & WEISSMAN, J. S. 2013. A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell,* 152**,** 909-22.

BATLLE, E., SANCHO, E., FRANCI, C., DOMINGUEZ, D., MONFAR, M., BAULIDA, J. & GARCIA DE HERREROS, A. 2000. The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat Cell Biol,* 2**,** 84-89.

BEN-HUR, T., IDELSON, M., KHANER, H., PERA, M., REINHARTZ, E., ITZIK, A. & REUBINOFF, B. E. 2004. Transplantation of human embryonic stem cell-derived neural progenitors improves behavioral deficit in Parkinsonian rats. *Stem Cells,* 22**,** 1246-55.

BERG, D. K., LI, C., ASHER, G., WELLS, D. N. & OBACK, B. 2007. Red Deer Cloned from Antler Stem Cells and Their Differentiated Progeny. *Biology of Reproduction,* 77**,** 384-394.

BERNS, K., HIJMANS, E. M., MULLENDERS, J., BRUMMELKAMP, T. R., VELDS, A., HEIMERIKX, M., KERKHOVEN, R. M., MADIREDJO, M., NIJKAMP, W., WEIGELT, B., AGAMI, R., GE, W., CAVET, G., LINSLEY, P. S., BEIJERSBERGEN, R. L. & BERNARDS, R. 2004. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature,* 428**,** 431-437.

BOGDANOVE, A. J. & VOYTAS, D. F. 2011. TAL Effectors: Customizable Proteins for DNA Targeting. *Science,* 333**,** 1843-1846.

BOLAND, M. J., HAZEN, J. L., NAZOR, K. L., RODRIGUEZ, A. R., GIFFORD, W., MARTIN, G., KUPRIYANOV, S. & BALDWIN, K. K. 2009. Adult mice generated from induced pluripotent stem cells. *Nature,* 461**,** 91-4.

BOLOTIN, A., QUINQUIS, B., SOROKIN, A. & EHRLICH, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology,* 151**,** 2551-61.

BOUDREAU, R. L., MONTEYS, A. M. & DAVIDSON, B. L. 2008. Minimizing variables among hairpin-based RNAi vectors reveals the potency of shRNAs. *RNA,* 14**,** 1834-44.

BOYER, L. A., LEE, T. I., COLE, M. F., JOHNSTONE, S. E., LEVINE, S. S., ZUCKER, J. P., GUENTHER, M. G., KUMAR, R. M., MURRAY, H. L., JENNER, R. G., GIFFORD, D. K., MELTON, D. A., JAENISCH, R. & YOUNG, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell,* 122**,** 947-56.

BRADLEY, A., EVANS, M., KAUFMAN, M. H. & ROBERTSON, E. 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature,* 309**,** 255-256.

BRAMBRINK, T., FOREMAN, R., WELSTEAD, G. G., LENGNER, C. J., WERNIG, M., SUH, H. & JAENISCH, R. 2008. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell,* 2**,** 151-9.

BRIGGS, R. & KING, T. J. 1952. Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proceedings of the National Academy of Sciences of the United States of America,* 38**,** 455-463.

BRONS, I. G., SMITHERS, L. E., TROTTER, M. W., RUGG-GUNN, P., SUN, B., CHUVA DE SOUSA LOPES, S. M., HOWLETT, S. K., CLARKSON, A., AHRLUND-RICHTER, L., PEDERSEN, R. A. & VALLIER, L. 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature,* 448**,** 191-5.

BUCKLEY, S. M., ARANDA-ORGILLES, B., STRIKOUDIS, A., APOSTOLOU, E., LOIZOU, E., MORAN-CRUSIO, K., FARNSWORTH, C. L., KOLLER, A. A., DASGUPTA, R., SILVA, J. C., STADTFELD, M., HOCHEDLINGER, K., CHEN, E. I. & AIFANTIS, I. 2012. Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. *Cell Stem Cell,* 11**,** 783-98.

BUECKER, C., CHEN, H. H., POLO, J. M., DAHERON, L., BU, L., BARAKAT, T. S., OKWIEKA, P., PORTER, A., GRIBNAU, J., HOCHEDLINGER, K. & GEIJSEN, N. 2010. A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. *Cell Stem Cell,* 6**,** 535-46.

BUEHR, M., MEEK, S., BLAIR, K., YANG, J., URE, J., SILVA, J., MCLAY, R., HALL, J., YING, Q. L. & SMITH, A. 2008. Capture of authentic embryonic stem cells from rat blastocysts. *Cell,* 135**,** 1287-98.

BUEHR, M. & SMITH, A. 2003. Genesis of embryonic stem cells. *Philos Trans R Soc Lond B Biol Sci,* 358**,** 1397-402; discussion 1402.

BUGANIM, Y., FADDAH, D. A., CHENG, A. W., ITSKOVICH, E., MARKOULAKI, S., GANZ, K., KLEMM, S. L., VAN OUDENAARDEN, A. & JAENISCH, R. 2012. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell,* 150**,** 1209-22.

BUGANIM, Y., FADDAH, D. A. & JAENISCH, R. 2013. Mechanisms and models of somatic cell reprogramming. *Nat Rev Genet,* 14**,** 427-39.

CANO, A., PEREZ-MORENO, M. A., RODRIGO, I., LOCASCIO, A., BLANCO, M. J., DEL BARRIO, M. G., PORTILLO, F. & NIETO, M. A. 2000. The transcription factor Snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol,* 2**,** 76-83.

CAREY, B. W., MARKOULAKI, S., HANNA, J., SAHA, K., GAO, Q., MITALIPOVA, M. & JAENISCH, R. 2009. Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proceedings of the National Academy of Sciences,* 106**,** 157-162.

CAREY, B. W., MARKOULAKI, S., HANNA, J. H., FADDAH, D. A., BUGANIM, Y., KIM, J., GANZ, K., STEINE, E. J., CASSADY, J. P., CREYGHTON, M. P., WELSTEAD, G. G., GAO, Q. & JAENISCH, R. 2011. Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell,* 9**,** 588-98.

CHANG, T. C., LIU, C. C., HSING, E. W., LIANG, S. M., CHI, Y. H., SUNG, L. Y., LIN, S. P., SHEN, T. L., KO, B. S., YEN, B. L., YET, S. F., WU, K. K. & LIOU, J. Y. 2012. 14-3-3sigma regulates beta-catenin-mediated mouse embryonic stem cell proliferation by sequestering GSK-3beta. *PLoS One,* 7**,** e40193.

CHEN, J., LIU, H., LIU, J., QI, J., WEI, B., YANG, J., LIANG, H., CHEN, Y., CHEN, J., WU, Y., GUO, L., ZHU, J., ZHAO, X., PENG, T., ZHANG, Y., CHEN, S., LI, X., LI, D.,

WANG, T. & PEI, D. 2013. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat Genet,* 45**,** 34-42.

CHEN, J., LIU, J., CHEN, Y., YANG, J., CHEN, J., LIU, H., ZHAO, X., MO, K., SONG, H., GUO, L., CHU, S., WANG, D., DING, K. & PEI, D. 2011a. Rational optimization of reprogramming culture conditions for the generation of induced pluripotent stem cells with ultra-high efficiency and fast kinetics. *Cell Res,* 21**,** 884-94.

CHEN, J., LIU, J., YANG, J., CHEN, Y., CHEN, J., NI, S., SONG, H., ZENG, L., DING, K. & PEI, D. 2011b. BMPs functionally replace Klf4 and support efficient reprogramming of mouse fibroblasts by Oct4 alone. *Cell Res,* 21**,** 205-12.

CHEN, M., DU, Q., ZHANG, H.-Y., WAHLESTEDT, C. & LIANG, Z. 2005. Vector-based siRNA delivery strategies for high-throughput screening of novel target genes. *Journal of RNAi and Gene Silencing : An International Journal of RNA and Gene Targeting Research,* 1**,** 5-11.

CHENG, J., DUTRA, A., TAKESONO, A., GARRETT-BEAL, L. & SCHWARTZBERG, P. L. 2004. Improved generation of C57BL/6J mouse embryonic stem cells in a defined serum-free media. *Genesis,* 39**,** 100-4.

CHI, S. W., ZANG, J. B., MELE, A. & DARNELL, R. B. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature,* 460**,** 479-86.

CHIA, N. Y., CHAN, Y. S., FENG, B., LU, X., ORLOV, Y. L., MOREAU, D., KUMAR, P., YANG, L., JIANG, J., LAU, M. S., HUSS, M., SOH, B. S., KRAUS, P., LI, P., LUFKIN, T., LIM, B., CLARKE, N. D., BARD, F. & NG, H. H. 2010. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature,* 468**,** 316-20.

CHO, H.-J., LEE, C.-S., KWON, Y.-W., PAEK, J. S., LEE, S.-H., HUR, J., LEE, E. J., ROH, T.-Y., CHU, I.-S., LEEM, S.-H., KIM, Y., KANG, H.-J., PARK, Y.-B. & KIM, H.-S. 2010. Induction of pluripotent stem cells from adult somatic cells by protein-based reprogramming without genetic manipulation. *Blood,* 116**,** 386-395.

CHOU, B. K., MALI, P., HUANG, X., YE, Z., DOWEY, S. N., RESAR, L. M., ZOU, C., ZHANG, Y. A., TONG, J. & CHENG, L. 2011. Efficient human iPS cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. *Cell Res,* 21**,** 518-29.

CHOU, C. H., LIN, F. M., CHOU, M. T., HSU, S. D., CHANG, T. H., WENG, S. L., SHRESTHA, S., HSIAO, C. C., HUNG, J. H. & HUANG, H. D. 2013. A computational approach for identifying microRNA-target interactions using high-throughput CLIP and PAR-CLIP sequencing. *BMC Genomics,* 14 Suppl 1**,** S2.

CHU, C. Y. & RANA, T. M. 2006. Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biol,* 4**,** e210.

CLARK, P. M., LOHER, P., QUANN, K., BRODY, J., LONDIN, E. R. & RIGOUTSOS, I. 2014. Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types. *Sci Rep,* 4**,** 5947.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science,* 339**,** 819-823.

COUZIN, J. 2009. Celebration and Concern Over U.S. Trial of Embryonic Stem Cells. *Science,* 323**,** 568-568.

COWAN, C. A., ATIENZA, J., MELTON, D. A. & EGGAN, K. 2005. Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem Cells. *Science,* 309**,** 1369-1373.

CROWLEY, T. E., NELLEN, W., GOMER, R. H. & FIRTEL, R. A. 1985. Phenocopy of discoidin I-minus mutants by antisense transformation in Dictyostelium. *Cell,* 43**,** 633-641.

CZECHANSKI, A., BYERS, C., GREENSTEIN, I., SCHRODE, N., DONAHUE, L. R., HADJANTONAKIS, A.-K. & REINHOLDT, L. G. 2014. Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. *Nat. Protocols,* 9**,** 559-574.

DING, L., PASZKOWSKI-ROGACZ, M., NITZSCHE, A., SLABICKI, M. M., HENINGER, A. K., DE VRIES, I., KITTLER, R., JUNQUEIRA, M., SHEVCHENKO, A., SCHULZ, H., HUBNER, N., DOSS, M. X., SACHINIDIS, A., HESCHELER, J., IACONE, R., ANASTASSIADIS, K., STEWART, A. F., PISABARRO, M. T., CALDARELLI, A., POSER, I., THEIS, M. & BUCHHOLZ, F. 2009. A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell,* 4**,** 403-15.

DO, J. T. & SCHÖLER, H. R. 2004. Nuclei of Embryonic Stem Cells Reprogram Somatic Cells. *STEM CELLS,* 22**,** 941-949.

DOENCH, J. G., PETERSEN, C. P. & SHARP, P. A. 2003. siRNAs can function as miRNAs. *Genes & Development,* 17**,** 438-442.

DORSETT, Y. & TUSCHL, T. 2004. siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov,* 3**,** 318-29.

DOS SANTOS, R. L., TOSTI, L., RADZISHEUSKAYA, A., CABALLERO, I. M., KAJI, K., HENDRICH, B. & SILVA, J. C. 2014. MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. *Cell Stem Cell,* 15**,** 102-10.

ECHEVERRI, C. J., BEACHY, P. A., BAUM, B., BOUTROS, M., BUCHHOLZ, F., CHANDA, S. K., DOWNWARD, J., ELLENBERG, J., FRASER, A. G., HACOHEN, N., HAHN, W. C., JACKSON, A. L., KIGER, A., LINSLEY, P. S., LUM, L., MA, Y., MATHEY-PREVOT, B., ROOT, D. E., SABATINI, D. M., TAIPALE, J., PERRIMON, N. & BERNARDS, R. 2006. Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat Meth,* 3**,** 777-779.

ECHEVERRI, C. J. & PERRIMON, N. 2006. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet,* 7**,** 373-84.

ECKER, J. R. & DAVIS, R. W. 1986. Inhibition of gene expression in plant cells by expression of antisense RNA. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 5372-5376.

EDWARDS, R. G. 2002. Personal pathways to embryonic stem cells. *Reproductive BioMedicine Online,* 4**,** 263-278.

EFE, J. A., HILCOVE, S., KIM, J., ZHOU, H., OUYANG, K., WANG, G., CHEN, J. & DING, S. 2011. Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nat Cell Biol,* 13**,** 215-22.

ELBASHIR, S. M., HARBORTH, J., LENDECKEL, W., YALCIN, A., WEBER, K. & TUSCHL, T. 2001a. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature,* 411**,** 494-498.

ELBASHIR, S. M., MARTINEZ, J., PATKANIOWSKA, A., LENDECKEL, W. & TUSCHL, T. 2001b. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *The EMBO Journal,* 20**,** 6877-6888.

ESTEBAN, M. A., WANG, T., QIN, B., YANG, J., QIN, D., CAI, J., LI, W., WENG, Z., CHEN, J., NI, S., CHEN, K., LI, Y., LIU, X., XU, J., ZHANG, S., LI, F., HE, W., LABUDA, K., SONG, Y., PETERBAUER, A., WOLBANK, S., REDL, H., ZHONG, M., CAI, D., ZENG, L. & PEI, D. 2010. Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell,* 6**,** 71-9.

EVANS, M. J. & KAUFMAN, M. H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature,* 292**,** 154-156.

FAZZIO, T. G., HUFF, J. T. & PANNING, B. 2008. An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity. *Cell,* 134**,** 162-174.

FENG, B., JIANG, J., KRAUS, P., NG, J. H., HENG, J. C., CHAN, Y. S., YAW, L. P., ZHANG, W., LOH, Y. H., HAN, J., VEGA, V. B., CACHEUX-RATABOUL, V., LIM, B., LUFKIN, T. & NG, H. H. 2009. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol,* 11**,** 197-203.

FRASER, M. J., CLSZCZON, T., ELICK, T. & BAUSER, C. 1996. Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Molecular Biology,* 5**,** 141-151.

FURUE, M., OKAMOTO, T., HAYASHI, Y., OKOCHI, H., FUJIMOTO, M., MYOISHI, Y., ABE, T., OHNUMA, K., SATO, G., ASASHIMA, M. & SATO, J. D. 2005. Leukemia inhibitory factor as an anti-apoptotic mitogen for pluripotent mouse embryonic stem cells in a serum-free medium without feeder cells. *In Vitro Cellular & Developmental Biology - Animal,* 41**,** 19-28.

FUSAKI, N., BAN, H., NISHIYAMA, A., SAEKI, K. & HASEGAWA, M. 2009. Efficient induction of transgene-free human pluripotent stem cells using a vector based on

Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences,* 85**,** 348-362.

GALLI, C., LAGUTINA, I., CROTTI, G., COLLEONI, S., TURINI, P., PONDERATO, N., DUCHI, R. & LAZZARI, G. 2003. Pregnancy: A cloned horse born to its dam twin. *Nature,* 424**,** 635-635.

GASIUNAS, G., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2012. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** E2579-E2586.

GINGOLD, JULIAN A., FIDALGO, M., GUALLAR, D., LAU, Z., SUN, Z., ZHOU, H., FAIOLA, F., HUANG, X., LEE, D.-F., WAGHRAY, A., SCHANIEL, C., FELSENFELD, DAN P., LEMISCHKA, IHOR R. & WANG, J. 2014. A Genome-wide RNAi Screen Identifies Opposing Functions of Snai1 and Snai2 on the Nanog Dependency in Reprogramming. *Molecular Cell,* 56**,** 140-152.

GINIS, I., LUO, Y., MIURA, T., THIES, S., BRANDENBERGER, R., GERECHT-NIR, S., AMIT, M., HOKE, A., CARPENTER, M. K., ITSKOVITZ-ELDOR, J. & RAO, M. S. 2004. Differences between human and mouse embryonic stem cells. *Developmental Biology,* 269**,** 360-380.

GOLIPOUR, A., DAVID, L., LIU, Y., JAYAKUMARAN, G., HIRSCH, C. L., TRCKA, D. & WRANA, J. L. 2012. A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell,* 11**,** 769-82.

GONZALEZ, F., BOUE, S. & IZPISUA BELMONTE, J. C. 2011. Methods for making induced pluripotent stem cells: reprogramming a la carte. *Nat Rev Genet,* 12**,** 231-42.

GOSSEN, M. & BUJARD, H. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences,* 89**,** 5547-5551.

GOSSEN, M., FREUNDLIEB, S., BENDER, G., MÜLLER, G., HILLEN, W. & BUJARD, H. 1995. Transcriptional Activation by Tetracyclines in Mammalian Cells. *Science,* 268**,** 1766-1769.

GRIMM, D., STREETZ, K. L., JOPLING, C. L., STORM, T. A., PANDEY, K., DAVIS, C. R., MARION, P., SALAZAR, F. & KAY, M. A. 2006. Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature,* 441**,** 537-41.

GU, S., JIN, L., ZHANG, Y., HUANG, Y., ZHANG, F., VALDMANIS, P. N. & KAY, M. A. 2012. The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell,* 151**,** 900-11.

GU, S. & KAY, M. A. 2010. How do miRNAs mediate translational repression? *Silence,* 1**,** 11.

GUMIENNY, R. & ZAVOLAN, M. 2015. Accurate transcriptome-wide prediction of microRNA targets and small interfering RNA off-targets with MIRZA-G. *Nucleic Acids Res,* 43**,** 1380-91.

GURDON, J. B. 1960. The Developmental Capacity of Nuclei Taken from Differentiating Endoderm Cells of Xenopus laevis. *Journal of Embryology and Experimental Morphology,* 8**,** 505-526.

GURDON, J. B. 1962. The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. *Journal of Embryology and Experimental Morphology,* 10**,** 622-640.

HAKELIEN, A.-M., LANDSVERK, H. B., ROBL, J. M., SKALHEGG, B. S. & COLLAS, P. 2002. Reprogramming fibroblasts to express T-cell functions using cell extracts. *Nat Biotech,* 20**,** 460-466.

HÄMMERLING, J. 1934. Entwicklungsphysiologische und genetische Grundlagen der Formbildung bei der Schirmalge Acetabularia. *Naturwissenschaften,* 22**,** 829-836.

HAN, D. W., TAPIA, N., JOO, J. Y., GREBER, B., ARAUZO-BRAVO, M. J., BERNEMANN, C., KO, K., WU, G., STEHLING, M., DO, J. T. & SCHOLER, H. R. 2010. Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. *Cell,* 143**,** 617-27.

HANNA, J., CHENG, A. W., SAHA, K., KIM, J., LENGNER, C. J., SOLDNER, F., CASSADY, J. P., MUFFAT, J., CAREY, B. W. & JAENISCH, R. 2010a. Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proceedings of the National Academy of Sciences,* 107**,** 9222-9227.

HANNA, J., MARKOULAKI, S., MITALIPOVA, M., CHENG, A. W., CASSADY, J. P., STAERK, J., CAREY, B. W., LENGNER, C. J., FOREMAN, R., LOVE, J., GAO, Q., KIM, J. & JAENISCH, R. 2009a. Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell,* 4**,** 513-24.

HANNA, J., SAHA, K., PANDO, B., VAN ZON, J., LENGNER, C. J., CREYGHTON, M. P., VAN OUDENAARDEN, A. & JAENISCH, R. 2009b. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature,* 462**,** 595-601.

HANNA, J. H., SAHA, K. & JAENISCH, R. 2010b. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell,* 143**,** 508-25.

HANSSON, J., RAFIEE, M. R., REILAND, S., POLO, J. M., GEHRING, J., OKAWA, S., HUBER, W., HOCHEDLINGER, K. & KRIJGSVELD, J. 2012. Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Rep,* 2**,** 1579-92.

HARLAND, R. & WEINTRAUB, H. 1985. Translation of mRNA injected into Xenopus oocytes is specifically inhibited by antisense RNA. *The Journal of Cell Biology,* 101**,** 1094-1099.

HENG, J. C., FENG, B., HAN, J., JIANG, J., KRAUS, P., NG, J. H., ORLOV, Y. L., HUSS, M., YANG, L., LUFKIN, T., LIM, B. & NG, H. H. 2010. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell,* 6**,** 167-74.

HOCKEMEYER, D., SOLDNER, F., COOK, E. G., GAO, Q., MITALIPOVA, M. & JAENISCH, R. 2008. A Drug-Inducible System for Direct Reprogramming of Human Somatic Cells to Pluripotency. *Cell Stem Cell,* 3**,** 346-353.

HONG, C. S., PARK, B. Y. & SAINT-JEANNET, J. P. 2007. The function of Dmrt genes in vertebrate development: it is not just about sex. *Dev Biol,* 310**,** 1-9.

HONG, H., TAKAHASHI, K., ICHISAKA, T., AOI, T., KANAGAWA, O., NAKAGAWA, M., OKITA, K. & YAMANAKA, S. 2009. Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature,* 460**,** 1132-5.

HORNUNG, V., ELLEGAST, J., KIM, S., BRZÓZKA, K., JUNG, A., KATO, H., POECK, H., AKIRA, S., CONZELMANN, K.-K., SCHLEE, M., ENDRES, S. & HARTMANN, G. 2006. 5'-Triphosphate RNA Is the Ligand for RIG-I. *Science,* 314**,** 994-997.

HOTTA, A. & ELLIS, J. 2008. Retroviral vector silencing during iPS cell induction: An epigenetic beacon that signals distinct pluripotent states. *Journal of Cellular Biochemistry,* 105**,** 940-948.

HOU, P., LI, Y., ZHANG, X., LIU, C., GUAN, J., LI, H., ZHAO, T., YE, J., YANG, W., LIU, K., GE, J., XU, J., ZHANG, Q., ZHAO, Y. & DENG, H. 2013. Pluripotent Stem Cells Induced from Mouse Somatic Cells by Small-Molecule Compounds. *Science,* 341**,** 651-654.

HU, G., KIM, J., XU, Q., LENG, Y., ORKIN, S. H. & ELLEDGE, S. J. 2009. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev,* 23**,** 837-48.

HU, K., YU, J., SUKNUNTHA, K., TIAN, S., MONTGOMERY, K., CHOI, K. D., STEWART, R., THOMSON, J. A. & SLUKVIN, II 2011. Efficient generation of transgene-free induced pluripotent stem cells from normal and neoplastic bone marrow and cord blood mononuclear cells. *Blood,* 117**,** e109-19.

HU, S., WILSON, K. D., GHOSH, Z., HAN, L., WANG, Y., LAN, F., RANSOHOFF, K. J., BURRIDGE, P. & WU, J. C. 2013. MicroRNA-302 Increases Reprogramming Efficiency via Repression of NR2F2. *STEM CELLS,* 31**,** 259-268.

HUANG, P., HE, Z., JI, S., SUN, H., XIANG, D., LIU, C., HU, Y., WANG, X. & HUI, L. 2011. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature,* 475**,** 386-9.

HUANG, Y., OSORNO, R., TSAKIRIDIS, A. & WILSON, V. 2012. In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. *Cell Rep,* 2**,** 1571-8.

HUANGFU, D., MAEHR, R., GUO, W., EIJKELENBOOM, A., SNITOW, M., CHEN, A. E. & MELTON, D. A. 2008a. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol,* 26**,** 795-7.

HUANGFU, D., OSAFUNE, K., MAEHR, R., GUO, W., EIJKELENBOOM, A., CHEN, S., MUHLESTEIN, W. & MELTON, D. A. 2008b. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat Biotechnol,* 26**,** 1269-75.

ICHIDA, J. K., BLANCHARD, J., LAM, K., SON, E. Y., CHUNG, J. E., EGLI, D., LOH, K. M., CARTER, A. C., DI GIORGIO, F. P., KOSZKA, K., HUANGFU, D., AKUTSU, H., LIU, D. R., RUBIN, L. L. & EGGAN, K. 2009. A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. *Cell Stem Cell,* 5**,** 491-503.

IEDA, M., FU, J. D., DELGADO-OLGUIN, P., VEDANTHAM, V., HAYASHI, Y., BRUNEAU, B. G. & SRIVASTAVA, D. 2010. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell,* 142**,** 375-86.

IZANT, J. G. & WEINTRAUB, H. 1984. Inhibition of thymidine kinase gene expression by anti-sense RNA: A molecular approach to genetic analysis. *Cell,* 36**,** 1007-1015.

IZANT, J. G. & WEINTRAUB, H. 1985. Constitutive and conditional suppression of exogenous and endogenous genes by anti-sense RNA. *Science,* 229**,** 345-352.

JACKSON, A. L., BARTZ, S. R., SCHELTER, J., KOBAYASHI, S. V., BURCHARD, J., MAO, M., LI, B., CAVET, G. & LINSLEY, P. S. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotech,* 21**,** 635-637.

JACKSON, A. L., BURCHARD, J., SCHELTER, J., CHAU, B. N., CLEARY, M., LIM, L. & LINSLEY, P. S. 2006. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA,* 12**,** 1179-87.

JACKSON, A. L. & LINSLEY, P. S. 2010. Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat Rev Drug Discov,* 9**,** 57-67.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science,* 337**,** 816-821.

JOSEFSEN, K. & LEE, Y. 2011. Validation of RNAi by Real Time PCR. *In:* NIELSEN, H. (ed.) *RNA.* Humana Press.

JUDSON, R. L., BABIARZ, J. E., VENERE, M. & BLELLOCH, R. 2009. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotechnol,* 27**,** 459-61.

KAJI, K., NORRBY, K., PACA, A., MILEIKOVSKY, M., MOHSENI, P. & WOLTJEN, K. 2009. Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature,* 458**,** 771-5.

KANG, L., WANG, J., ZHANG, Y., KOU, Z. & GAO, S. 2009. iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell Stem Cell,* 5**,** 135-8.

KAWAMATA, M. & NISHIMORI, K. 2006. Mice deficient in Dmrt7 show infertility with spermatogenic arrest at pachytene stage. *FEBS Letters,* 580**,** 6442-6446.

KAWAMURA, T., SUZUKI, J., WANG, Y. V., MENENDEZ, S., MORERA, L. B., RAYA, A., WAHL, G. M. & IZPISUA BELMONTE, J. C. 2009. Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature,* 460**,** 1140-4.

KAYKAS, A. & MOON, R. 2004. A plasmid-based system for expressing small interfering RNA libraries in mammalian cells. *BMC Cell Biology,* 5**,** 16.

KHORSHID, M., HAUSSER, J., ZAVOLAN, M. & VAN NIMWEGEN, E. 2013. A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. *Nat Methods,* 10**,** 253-5.

KIM, D., KIM, C. H., MOON, J. I., CHUNG, Y. G., CHANG, M. Y., HAN, B. S., KO, S., YANG, E., CHA, K. Y., LANZA, R. & KIM, K. S. 2009. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell,* 4**,** 472-6.

KIM, S., NAMEKAWA, S. H., NISWANDER, L. M., WARD, J. O., LEE, J. T., BARDWELL, V. J. & ZARKOWER, D. 2007. A mammal-specific Doublesex homolog associates with male sex chromatin and is required for male meiosis. *PLoS Genet,* 3**,** e62.

KING, T. J. & BRIGGS, R. 1956. Serial Transplantation of Embryonic Nuclei. *Cold Spring Harbor Symposia on Quantitative Biology,* 21**,** 271-290.

KITTLER, R., PELLETIER, L., MA, C., POSER, I., FISCHER, S., HYMAN, A. A. & BUCHHOLZ, F. 2005. RNA interference rescue by bacterial artificial chromosome transgenesis in mammalian tissue culture cells. *Proc Natl Acad Sci U S A,* 102**,** 2396-401.

KOCHE, R. P., SMITH, Z. D., ADLI, M., GU, H., KU, M., GNIRKE, A., BERNSTEIN, B. E. & MEISSNER, A. 2011. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell,* 8**,** 96-105.

KOGA, M., MATSUDA, M., KAWAMURA, T., SOGO, T., SHIGENO, A., NISHIDA, E. & EBISUYA, M. 2014. Foxd1 is a mediator and indicator of the cell reprogramming process. *Nat Commun,* 5**,** 3197.

KOIKE-YUSA, H., LI, Y., TAN, E. P., VELASCO-HERRERA MDEL, C. & YUSA, K. 2014. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol,* 32**,** 267-73.

KUMAR, D., GUSTAFSSON, C. & KLESSIG, D. F. 2006. Validation of RNAi silencing specificity using synthetic genes: salicylic acid-binding protein 2 is required for innate immunity in plants. *The Plant Journal,* 45**,** 863-868.

LAMBA, D. A., KARL, M. O., WARE, C. B. & REH, T. A. 2006. Efficient generation of retinal progenitor cells from human embryonic stem cells. *Proc Natl Acad Sci U S A,* 103**,** 12769-74.

LANDSVERK, H. B., HÅKELIEN, A.-M., KÜNTZIGER, T., ROBL, J. M., SKÅLHEGG, B. S. & COLLAS, P. 2002. Reprogrammed gene expression in a somatic cell-free extract. *EMBO Reports,* 3**,** 384-389.

LEE, B. C., KIM, M. K., JANG, G., OH, H. J., YUDA, F., KIM, H. J., SHAMIM, M. H., KIM, J. J., KANG, S. K., SCHATTEN, G. & HWANG, W. S. 2005. Dogs cloned from adult somatic cells. *Nature,* 436**,** 641-641.

LI, H., COLLADO, M., VILLASANTE, A., STRATI, K., ORTEGA, S., CANAMERO, M., BLASCO, M. A. & SERRANO, M. 2009a. The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature,* 460**,** 1136-9.

LI, R., LIANG, J., NI, S., ZHOU, T., QING, X., LI, H., HE, W., CHEN, J., LI, F., ZHUANG, Q., QIN, B., XU, J., LI, W., YANG, J., GAN, Y., QIN, D., FENG, S., SONG, H., YANG, D., ZHANG, B., ZENG, L., LAI, L., ESTEBAN, M. A. & PEI, D. 2010. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell,* 7**,** 51-63.

LI, S., WANG, L., BERMAN, M. A., ZHANG, Y. & DORF, M. E. 2006. RNAi screen in mouse astrocytes identifies phosphatases that regulate NF-kappaB signaling. *Mol Cell,* 24**,** 497-509.

LI, W., WEI, W., ZHU, S., ZHU, J., SHI, Y., LIN, T., HAO, E., HAYEK, A., DENG, H. & DING, S. 2009b. Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. *Cell Stem Cell,* 4**,** 16-9.

LI, Y., ZHANG, Q., YIN, X., YANG, W., DU, Y., HOU, P., GE, J., LIU, C., ZHANG, W., ZHANG, X., WU, Y., LI, H., LIU, K., WU, C., SONG, Z., ZHAO, Y., SHI, Y. & DENG, H. 2011. Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. *Cell Res,* 21**,** 196-204.

LI, Z. & RANA, T. M. 2012. A kinase inhibitor screen identifies small-molecule enhancers of reprogramming and iPS cell generation. *Nat Commun,* 3**,** 1085.

LIANG, G., HE, J. & ZHANG, Y. 2012. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nat Cell Biol,* 14**,** 457-66.

LIANG, G., TARANOVA, O., XIA, K. & ZHANG, Y. 2010. Butyrate promotes induced pluripotent stem cell generation. *J Biol Chem,* 285**,** 25516-21.

LIAO, B., BAO, X., LIU, L., FENG, S., ZOVOILIS, A., LIU, W., XUE, Y., CAI, J., GUO, X., QIN, B., ZHANG, R., WU, J., LAI, L., TENG, M., NIU, L., ZHANG, B., ESTEBAN, M. A. & PEI, D. 2011. MicroRNA Cluster 302–367 Enhances Somatic Cell Reprogramming by Accelerating a Mesenchymal-to-Epithelial Transition. *Journal of Biological Chemistry,* 286**,** 17359-17364.

LIGHT, J. & MOLIN, S. 1982. The sites of action of the two copy number control functions of plasmid R1. *Molecular and General Genetics MGG,* 187**,** 486-493.

LIGHT, J. & MOLIN, S. 1983. Post-transcriptional control of expression of the repA gene of plasmid R1 mediated by a small RNA molecule. *The EMBO Journal,* 2**,** 93-98.

LIN, T., AMBASUDHAN, R., YUAN, X., LI, W., HILCOVE, S., ABUJAROUR, R., LIN, X., HAHM, H. S., HAO, E., HAYEK, A. & DING, S. 2009. A chemical platform for improved induction of human iPSCs. *Nat Meth,* 6**,** 805-808.

LIU, L., LUO, G. Z., YANG, W., ZHAO, X., ZHENG, Q., LV, Z., LI, W., WU, H. J., WANG, L., WANG, X. J. & ZHOU, Q. 2010. Activation of the imprinted Dlk1-Dio3 region correlates with pluripotency levels of mouse stem cells. *J Biol Chem,* 285**,** 19483-90.

LIU, X., SUN, H., QI, J., WANG, L., HE, S., LIU, J., FENG, C., CHEN, C., LI, W., GUO, Y., QIN, D., PAN, G., CHEN, J., PEI, D. & ZHENG, H. 2013. Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. *Nat Cell Biol,* 15**,** 829-38.

LU, D., DAVIS, M. P., ABREU-GOODGER, C., WANG, W., CAMPOS, L. S., SIEDE, J., VIGORITO, E., SKARNES, W. C., DUNHAM, I., ENRIGHT, A. J. & LIU, P. 2012. MiR-25 regulates Wwp2 and Fbxw7 and promotes reprogramming of mouse fibroblast cells to iPSCs. *PLoS One,* 7**,** e40938.

LUCAS, D. & FRENETTE, P. S. 2014. Stem cells: Reprogramming finds its niche. *Nature,* 511**,** 301-302.

LUO, M., LING, T., XIE, W., SUN, H., ZHOU, Y., ZHU, Q., SHEN, M., ZONG, L., LYU, G., ZHAO, Y., YE, T., GU, J., TAO, W., LU, Z. & GRUMMT, I. 2013. NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. *Stem Cells,* 31**,** 1278-86.

MA, H., WU, Y., DANG, Y., CHOI, J. G., ZHANG, J. & WU, H. 2014a. Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. *Mol Ther Nucleic Acids,* 3**,** e161.

MA, H., ZHANG, J. & WU, H. 2014b. Designing Ago2-specific siRNA/shRNA to Avoid Competition with Endogenous miRNAs. *Mol Ther Nucleic Acids,* 3**,** e176.

MACFARLAN, T. S., GIFFORD, W. D., AGARWAL, S., DRISCOLL, S., LETTIERI, K., WANG, J., ANDREWS, S. E., FRANCO, L., ROSENFELD, M. G., REN, B. & PFAFF, S. L. 2011. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev,* 25**,** 594-607.

MADDALO, D., MANCHADO, E., CONCEPCION, C. P., BONETTI, C., VIDIGAL, J. A., HAN, Y. C., OGRODOWSKI, P., CRIPPA, A., REKHTMAN, N., DE STANCHINA, E., LOWE, S. W. & VENTURA, A. 2014. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature,* 516**,** 423-7.

MAHERALI, N., AHFELDT, T., RIGAMONTI, A., UTIKAL, J., COWAN, C. & HOCHEDLINGER, K. 2008. A High-Efficiency System for the Generation and Study of Human Induced Pluripotent Stem Cells. *Cell Stem Cell,* 3**,** 340-345.

MAHERALI, N. & HOCHEDLINGER, K. 2009. Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr Biol,* 19**,** 1718-23.

MAHERALI, N., SRIDHARAN, R., XIE, W., UTIKAL, J., EMINLI, S., ARNOLD, K., STADTFELD, M., YACHECHKO, R., TCHIEU, J., JAENISCH, R., PLATH, K. & HOCHEDLINGER, K. 2007. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell,* 1**,** 55-70.

MALI, P., AACH, J., STRANGES, P. B., ESVELT, K. M., MOOSBURNER, M., KOSURI, S., YANG, L. & CHURCH, G. M. 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol,* 31**,** 833-8.

MALI, P., CHOU, B. K., YEN, J., YE, Z., ZOU, J., DOWEY, S., BRODSKY, R. A., OHM, J. E., YU, W., BAYLIN, S. B., YUSA, K., BRADLEY, A., MEYERS, D. J., MUKHERJEE, C., COLE, P. A. & CHENG, L. 2010. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells,* 28**,** 713-20.

MALI, P., ESVELT, K. M. & CHURCH, G. M. 2013b. Cas9 as a versatile tool for engineering biology. *Nat Methods,* 10**,** 957-63.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013c. RNA-Guided Human Genome Engineering via Cas9. *Science,* 339**,** 823-826.

MALI, P., YE, Z., HOMMOND, H. H., YU, X., LIN, J., CHEN, G., ZOU, J. & CHENG, L. 2008. Improved Efficiency and Pace of Generating Induced Pluripotent Stem Cells from Human Adult and Fetal Fibroblasts. *STEM CELLS,* 26**,** 1998-2005.

MANSOUR, A. A., GAFNI, O., WEINBERGER, L., ZVIRAN, A., AYYASH, M., RAIS, Y., KRUPALNIK, V., ZERBIB, M., AMANN-ZALCENSTEIN, D., MAZA, I., GEULA, S., VIUKOV, S., HOLTZMAN, L., PRIBLUDA, A., CANAANI, E., HORN-SABAN, S., AMIT, I., NOVERSHTERN, N. & HANNA, J. H. 2012. The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature,* 488**,** 409-13.

MANTEI, A., RUTZ, S., JANKE, M., KIRCHHOFF, D., JUNG, U., PATZEL, V., VOGEL, U., RUDEL, T., ANDREOU, I., WEBER, M. & SCHEFFOLD, A. 2008. siRNA stabilization prolongs gene knockdown in primary T lymphocytes. *Eur J Immunol,* 38**,** 2616-25.

MARION, R. M., STRATI, K., LI, H., MURGA, M., BLANCO, R., ORTEGA, S., FERNANDEZ-CAPETILLO, O., SERRANO, M. & BLASCO, M. A. 2009. A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature,* 460**,** 1149-53.

MARRO, S., PANG, Z. P., YANG, N., TSAI, M. C., QU, K., CHANG, H. Y., SUDHOF, T. C. & WERNIG, M. 2011. Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell,* 9**,** 374-82.

MARTIN, G. R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America,* 78**,** 7634-7638.

MARTINEZ, L. A., NAGUIBNEVA, I., LEHRMANN, H., VERVISCH, A., TCHENIO, T., LOZANO, G. & HAREL-BELLAN, A. 2002. Synthetic small inhibiting RNAs: efficient tools to inactivate oncogenic mutations and restore p53 pathways. *Proc Natl Acad Sci U S A,* 99**,** 14849-54.

MATZKE, M. A. & BIRCHLER, J. A. 2005. RNAi-mediated pathways in the nucleus. *Nat Rev Genet,* 6**,** 24-35.

MEISSNER, A., WERNIG, M. & JAENISCH, R. 2007. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol,* 25**,** 1177-81.

MELTON, D. A. 1985. Injected anti-sense RNAs specifically block messenger RNA translation in vivo. *Proceedings of the National Academy of Sciences of the United States of America,* 82**,** 144-148.

MENG, X., NEISES, A., SU, R. J., PAYNE, K. J., RITTER, L., GRIDLEY, D. S., WANG, J., SHENG, M., LAU, K. H., BAYLINK, D. J. & ZHANG, X. B. 2012. Efficient reprogramming of human cord blood CD34+ cells into induced pluripotent stem cells with OCT4 and SOX2 alone. *Mol Ther,* 20**,** 408-16.

MIKKELSEN, T. S., HANNA, J., ZHANG, X., KU, M., WERNIG, M., SCHORDERET, P., BERNSTEIN, B. E., JAENISCH, R., LANDER, E. S. & MEISSNER, A. 2008. Dissecting direct reprogramming through integrative genomic analysis. *Nature,* 454**,** 49-55.

MILLER, V. M., XIA, H., MARRS, G. L., GOUVION, C. M., LEE, G., DAVIDSON, B. L. & PAULSON, H. L. 2003. Allele-specific silencing of dominant disease genes. *Proc Natl Acad Sci U S A,* 100**,** 7195-200.

MITTAL, V. 2004. Improving the efficiency of RNA interference in mammals. *Nat Rev Genet,* 5**,** 355-65.

MIYOSHI, N., ISHII, H., NAGANO, H., HARAGUCHI, N., DEWI, D. L., KANO, Y., NISHIKAWA, S., TANEMURA, M., MIMORI, K., TANAKA, F., SAITO, T., NISHIMURA, J., TAKEMASA, I., MIZUSHIMA, T., IKEDA, M., YAMAMOTO, H., SEKIMOTO, M., DOKI, Y. & MORI, M. 2011. Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell Stem Cell,* 8**,** 633-8.

MOAZED, D. 2009. Small RNAs in transcriptional gene silencing and genome defence. *Nature,* 457**,** 413-20.

MOFFAT, J. & SABATINI, D. M. 2006. Building mammalian signalling pathways with RNAi screens. *Nat Rev Mol Cell Biol,* 7**,** 177-87.

MOJICA, F. J., DIEZ-VILLASENOR, C., GARCIA-MARTINEZ, J. & SORIA, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol,* 60**,** 174-82.

MONFORT, A. & WUTZ, A. 2013. Breathing-in epigenetic change with vitamin C. *EMBO Rep,* 14**,** 337-46.

NADALUTTI, C., VIIRI, K. M., KAUKINEN, K., MAKI, M. & LINDFORS, K. 2011. Extracellular transglutaminase 2 has a role in cell adhesion, whereas intracellular transglutaminase 2 is involved in regulation of endothelial cell proliferation and apoptosis. *Cell Prolif,* 44**,** 49-58.

NAKAGAWA, M., KOYANAGI, M., TANABE, K., TAKAHASHI, K., ICHISAKA, T., AOI, T., OKITA, K., MOCHIDUKI, Y., TAKIZAWA, N. & YAMANAKA, S. 2008. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol,* 26**,** 101-6.

NALLAGATLA, S. R., TORONEY, R. & BEVILACQUA, P. C. 2008. A brilliant disguise for self RNA: 5′-end and internal modifications of primary transcripts suppress elements of innate immunity. *RNA Biology,* 5**,** 140-144.

NICHOLS, J., JONES, K., PHILLIPS, J. M., NEWLAND, S. A., ROODE, M., MANSFIELD, W., SMITH, A. & COOKE, A. 2009. Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. *Nat Med,* 15**,** 814-8.

NICHOLS, J. & SMITH, A. 2009. Naive and primed pluripotent states. *Cell Stem Cell,* 4**,** 487-92.

NICOLI, S., KNYPHAUSEN, C. P., ZHU, L. J., LAKSHMANAN, A. & LAWSON, N. D. 2012. miR-221 is required for endothelial tip cell behaviors during vascular development. *Dev Cell,* 22**,** 418-29.

NISHIKAWA, S.-I., JAKT, L. M. & ERA, T. 2007. Embryonic stem-cell culture as a tool for developmental cell biology. *Nat Rev Mol Cell Biol,* 8**,** 502-507.

NIWA, H., MIYAZAKI, J.-I. & SMITH, A. G. 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet,* 24**,** 372-376.

O'MALLEY, J., SKYLAKI, S., IWABUCHI, K. A., CHANTZOURA, E., RUETZ, T., JOHNSSON, A., TOMLINSON, S. R., LINNARSSON, S. & KAJI, K. 2013. High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature,* 499**,** 88-91.

OBOKATA, H., SASAI, Y., NIWA, H., KADOTA, M., ANDRABI, M., TAKATA, N., TOKORO, M., TERASHITA, Y., YONEMURA, S., VACANTI, C. A. & WAKAYAMA, T. 2014a. Bidirectional developmental potential in reprogrammed cells with acquired pluripotency. *Nature,* 505**,** 676-80.

OBOKATA, H., WAKAYAMA, T., SASAI, Y., KOJIMA, K., VACANTI, M. P., NIWA, H., YAMATO, M. & VACANTI, C. A. 2014b. Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature,* 505**,** 641-7.

ODORICO, J. S., KAUFMAN, D. S. & THOMSON, J. A. 2001. Multilineage Differentiation from Human Embryonic Stem Cell Lines. *STEM CELLS,* 19**,** 193-204.

OGAWA, K., MATSUI, H., OHTSUKA, S. & NIWA, H. 2004. A novel mechanism for regulating clonal propagation of mouse ES cells. *Genes to Cells,* 9**,** 471-477.

OHTSUKA, S., NISHIKAWA-TORIKAI, S. & NIWA, H. 2012. E-cadherin promotes incorporation of mouse epiblast stem cells into normal development. *PLoS One,* 7**,** e45220.

OHTSUKA, S. & NIWA, H. 2015. The differential activation of intracellular signaling pathways confers the permissiveness of embryonic stem cell derivation from different mouse strains. *Development,* 142**,** 431-7.

OKITA, K., ICHISAKA, T. & YAMANAKA, S. 2007. Generation of germline-competent induced pluripotent stem cells. *Nature,* 448**,** 313-7.

OKITA, K., MATSUMURA, Y., SATO, Y., OKADA, A., MORIZANE, A., OKAMOTO, S., HONG, H., NAKAGAWA, M., TANABE, K., TEZUKA, K.-I., SHIBATA, T., KUNISADA, T., TAKAHASHI, M., TAKAHASHI, J., SAJI, H. & YAMANAKA, S. 2011. A more efficient method to generate integration-free human iPS cells. *Nat Meth,* 8**,** 409-412.

ONDER, T. T., KARA, N., CHERRY, A., SINHA, A. U., ZHU, N., BERNT, K. M., CAHAN, P., MARCARCI, B. O., UNTERNAEHRER, J., GUPTA, P. B., LANDER, E. S., ARMSTRONG, S. A. & DALEY, G. Q. 2012. Chromatin-modifying enzymes as modulators of reprogramming. *Nature,* 483**,** 598-602.

ORKIN, S. H. & HOCHEDLINGER, K. 2011. Chromatin connections to pluripotency and cellular reprogramming. *Cell,* 145**,** 835-50.

PADDISON, P. J., CAUDY, A. A., BERNSTEIN, E., HANNON, G. J. & CONKLIN, D. S. 2002. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev,* 16**,** 948-58.

PICHLMAIR, A., SCHULZ, O., TAN, C. P., NÄSLUND, T. I., LILJESTRÖM, P., WEBER, F. & REIS E SOUSA, C. 2006. RIG-I-Mediated Antiviral Responses to Single-Stranded RNA Bearing 5'-Phosphates. *Science,* 314**,** 997-1001.

PLATH, K. & LOWRY, W. E. 2011. Progress in understanding reprogramming to the induced pluripotent state. *Nat Rev Genet,* 12**,** 253-65.

POLEJAEVA, I. A., CHEN, S.-H., VAUGHT, T. D., PAGE, R. L., MULLINS, J., BALL, S., DAI, Y., BOONE, J., WALKER, S., AYARES, D. L., COLMAN, A. & CAMPBELL, K. H. S. 2000. Cloned pigs produced by nuclear transfer from adult somatic cells. *Nature,* 407**,** 86-90.

POLO, J. M., ANDERSSEN, E., WALSH, R. M., SCHWARZ, B. A., NEFZGER, C. M., LIM, S. M., BORKENT, M., APOSTOLOU, E., ALAEI, S., CLOUTIER, J., BAR-NUR, O., CHELOUFI, S., STADTFELD, M., FIGUEROA, M. E., ROBINTON, D., NATESAN,

S., MELNICK, A., ZHU, J., RAMASWAMY, S. & HOCHEDLINGER, K. 2012. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell,* 151**,** 1617-32.

POURCEL, C., SALVIGNOL, G. & VERGNAUD, G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology,* 151**,** 653-63.

QIN, H., DIAZ, A., BLOUIN, L., LEBBINK, R. J., PATENA, W., TANBUN, P., LEPROUST, E. M., MCMANUS, M. T., SONG, J. S. & RAMALHO-SANTOS, M. 2014. Systematic identification of barriers to human iPSC generation. *Cell,* 158**,** 449-61.

RAIS, Y., ZVIRAN, A., GEULA, S., GAFNI, O., CHOMSKY, E., VIUKOV, S., MANSOUR, A. A., CASPI, I., KRUPALNIK, V., ZERBIB, M., MAZA, I., MOR, N., BARAN, D., WEINBERGER, L., JAITIN, D. A., LARA-ASTIASO, D., BLECHER-GONEN, R., SHIPONY, Z., MUKAMEL, Z., HAGAI, T., GILAD, S., AMANN-ZALCENSTEIN, D., TANAY, A., AMIT, I., NOVERSHTERN, N. & HANNA, J. H. 2013. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature,* 502**,** 65-70.

RAN, F. A., HSU, P. D., LIN, C. Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S., ZHANG, Y. & ZHANG, F. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell,* 154**,** 1380-9.

RATHBONE, A. J., LIDDELL, S. & CAMPBELL, K. H. 2013. Proteomic analysis of early reprogramming events in murine somatic cells incubated with Xenopus laevis oocyte extracts demonstrates network associations with induced pluripotency markers. *Cell Reprogram,* 15**,** 269-80.

REUBINOFF, B. E., PERA, M. F., FONG, C.-Y., TROUNSON, A. & BONGSO, A. 2000. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotech,* 18**,** 399-404.

REYNOLDS, A., LEAKE, D., BOESE, Q., SCARINGE, S., MARSHALL, W. S. & KHVOROVA, A. 2004. Rational siRNA design for RNA interference. *Nat Biotechnol,* 22**,** 326-30.

RIDEOUT, W. M., EGGAN, K. & JAENISCH, R. 2001. Nuclear Cloning and Epigenetic Reprogramming of the Genome. *Science,* 293**,** 1093-1098.

RIVAL-GERVIER, S., LO, M. Y., KHATTAK, S., PASCERI, P., LORINCZ, M. C. & ELLIS, J. 2013. Kinetics and epigenetics of retroviral silencing in mouse embryonic stem cells defined by deletion of the D4Z4 element. *Mol Ther,* 21**,** 1536-50.

ROCA, H., HERNANDEZ, J., WEIDNER, S., MCEACHIN, R. C., FULLER, D., SUD, S., SCHUMANN, T., WILKINSON, J. E., ZASLAVSKY, A., LI, H., MAHER, C. A., DAIGNAULT-NEWTON, S., HEALY, P. N. & PIENTA, K. J. 2013. Transcription factors OVOL1 and OVOL2 induce the mesenchymal to epithelial transition in human cancer. *PLoS One,* 8**,** e76773.

ROSENBERG, U. B., PREISS, A., SEIFERT, E., JACKLE, H. & KNIPPLE, D. C. 1985. Production of phenocopies by Kruppel antisense RNA injection into Drosophila embryos. *Nature,* 313**,** 703-706.

SAGE, J. 2012. The retinoblastoma tumor suppressor and stem cell biology. *Genes Dev,* 26**,** 1409-20.

SAKURAI, K., TALUKDAR, I., PATIL, V. S., DANG, J., LI, Z., CHANG, K. Y., LU, C. C., DELORME-WALKER, V., DERMARDIROSSIAN, C., ANDERSON, K., HANEIN, D., YANG, C. S., WU, D., LIU, Y. & RANA, T. M. 2014. Kinome-wide functional analysis highlights the role of cytoskeletal remodeling in somatic cell reprogramming. *Cell Stem Cell,* 14**,** 523-34.

SAMAVARCHI-TEHRANI, P., GOLIPOUR, A., DAVID, L., SUNG, H. K., BEYER, T. A., DATTI, A., WOLTJEN, K., NAGY, A. & WRANA, J. L. 2010. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell,* 7**,** 64-77.

SANTHEKADUR, P. K., DAS, S. K., GREDLER, R., CHEN, D., SRIVASTAVA, J., ROBERTSON, C., BALDWIN, A. S., FISHER, P. B. & SARKAR, D. 2012. Multifunction Protein Staphylococcal Nuclease Domain Containing 1 (SND1) Promotes Tumor Angiogenesis in Human Hepatocellular Carcinoma through Novel Pathway That Involves Nuclear Factor κB and miR-221. *The Journal of Biological Chemistry,* 287**,** 13952-13958.

SAXENA, S., JONSSON, Z. O. & DUTTA, A. 2003. Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J Biol Chem,* 278**,** 44312-9.

SCHNERCH, A., CERDAN, C. & BHATIA, M. 2010. Distinguishing between mouse and human pluripotent stem cell regulation: the best laid plans of mice and men. *Stem Cells,* 28**,** 419-30.

SCHWANK, G., KOO, B. K., SASSELLI, V., DEKKERS, J. F., HEO, I., DEMIRCAN, T., SASAKI, N., BOYMANS, S., CUPPEN, E., VAN DER ENT, C. K., NIEUWENHUIS, E. E., BEEKMAN, J. M. & CLEVERS, H. 2013. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell,* 13**,** 653-8.

SCHWARZ, D. S., HUTVÁGNER, G., DU, T., XU, Z., ARONIN, N. & ZAMORE, P. D. 2003. Asymmetry in the Assembly of the RNAi Enzyme Complex. *Cell,* 115**,** 199-208.

SEKIYA, S. & SUZUKI, A. 2011. Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature,* 475**,** 390-3.

SEMIZAROV, D., FROST, L., SARTHY, A., KROEGER, P., HALBERT, D. N. & FESIK, S. W. 2003. Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci U S A,* 100**,** 6347-52.

SHALEM, O., SANJANA, N. E., HARTENIAN, E., SHI, X., SCOTT, D. A., MIKKELSEN, T. S., HECKL, D., EBERT, B. L., ROOT, D. E., DOENCH, J. G. & ZHANG, F. 2014.

Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science,* 343**,** 84-87.

SHAO, L., FENG, W., SUN, Y., BAI, H., LIU, J., CURRIE, C., KIM, J., GAMA, R., WANG, Z., QIAN, Z., LIAW, L. & WU, W. S. 2009. Generation of iPS cells using defined factors linked via the self-cleaving 2A sequences in a single open reading frame. *Cell Res,* 19**,** 296-306.

SHI, Y., DESPONTS, C., DO, J. T., HAHM, H. S., SCHOLER, H. R. & DING, S. 2008. Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell,* 3**,** 568-74.

SHUFARO, Y. & REUBINOFF, B. E. 2004. Therapeutic applications of embryonic stem cells. *Best Pract Res Clin Obstet Gynaecol,* 18**,** 909-27.

SILVA, J., BARRANDON, O., NICHOLS, J., KAWAGUCHI, J., THEUNISSEN, T. W. & SMITH, A. 2008a. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol,* 6**,** e253.

SILVA, J., NICHOLS, J., THEUNISSEN, T. W., GUO, G., VAN OOSTEN, A. L., BARRANDON, O., WRAY, J., YAMANAKA, S., CHAMBERS, I. & SMITH, A. 2009. Nanog Is the Gateway to the Pluripotent Ground State. *Cell,* 138**,** 722-737.

SILVA, J. M., LI, M. Z., CHANG, K., GE, W., GOLDING, M. C., RICKLES, R. J., SIOLAS, D., HU, G., PADDISON, P. J., SCHLABACH, M. R., SHETH, N., BRADSHAW, J., BURCHARD, J., KULKARNI, A., CAVET, G., SACHIDANANDAM, R., MCCOMBIE, W. R., CLEARY, M. A., ELLEDGE, S. J. & HANNON, G. J. 2005. Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet,* 37**,** 1281-8.

SILVA, J. M., MARRAN, K., PARKER, J. S., SILVA, J., GOLDING, M., SCHLABACH, M. R., ELLEDGE, S. J., HANNON, G. J. & CHANG, K. 2008b. Profiling Essential Genes in Human Mammary Cells by Multiplex RNAi Screening. *Science,* 319**,** 617-620.

SINGH, S., NARANG, A. S. & MAHATO, R. I. 2011. Subcellular fate and off-target effects of siRNA, shRNA, and miRNA. *Pharm Res,* 28**,** 2996-3015.

SMITH, A. G., HEATH, J. K., DONALDSON, D. D., WONG, G. G., MOREAU, J., STAHL, M. & ROGERS, D. 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature,* 336**,** 688-690.

SMITH, Z. D., NACHMAN, I., REGEV, A. & MEISSNER, A. 2010. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat Biotechnol,* 28**,** 521-6.

SOMMER, C. A., STADTFELD, M., MURPHY, G. J., HOCHEDLINGER, K., KOTTON, D. N. & MOSTOSLAVSKY, G. 2009. Induced Pluripotent Stem Cell Generation Using a Single Lentiviral Stem Cell Cassette. *STEM CELLS,* 27**,** 543-549.

SOUFI, A., DONAHUE, G. & ZARET, K. S. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell,* 151**,** 994-1004.

SPIKE, B. T. & WAHL, G. M. 2011. p53, Stem Cells, and Reprogramming: Tumor Suppression beyond Guarding the Genome. *Genes Cancer,* 2**,** 404-19.

SRIDHARAN, R., TCHIEU, J., MASON, M. J., YACHECHKO, R., KUOY, E., HORVATH, S., ZHOU, Q. & PLATH, K. 2009. Role of the murine reprogramming factors in the induction of pluripotency. *Cell,* 136**,** 364-77.

STADTFELD, M., APOSTOLOU, E., AKUTSU, H., FUKUDA, A., FOLLETT, P., NATESAN, S., KONO, T., SHIODA, T. & HOCHEDLINGER, K. 2010. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature,* 465**,** 175-81.

STADTFELD, M., APOSTOLOU, E., FERRARI, F., CHOI, J., WALSH, R. M., CHEN, T., OOI, S. S., KIM, S. Y., BESTOR, T. H., SHIODA, T., PARK, P. J. & HOCHEDLINGER, K. 2012. Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. *Nat Genet,* 44**,** 398-405, S1-2.

STADTFELD, M., BRENNAND, K. & HOCHEDLINGER, K. 2008a. Reprogramming of pancreatic beta cells into induced pluripotent stem cells. *Curr Biol,* 18**,** 890-4.

STADTFELD, M., MAHERALI, N., BREAULT, D. T. & HOCHEDLINGER, K. 2008b. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell,* 2**,** 230-40.

STADTFELD, M., NAGAYA, M., UTIKAL, J., WEIR, G. & HOCHEDLINGER, K. 2008c. Induced Pluripotent Stem Cells Generated Without Viral Integration. *Science,* 322**,** 945-949.

STEPHENSON, E. L., MASON, C. & BRAUDE, P. R. 2009. Preimplantation genetic diagnosis as a source of human embryonic stem cells for disease research and drug discovery. *BJOG,* 116**,** 158-65.

SUBRAMANYAM, D., LAMOUILLE, S., JUDSON, R. L., LIU, J. Y., BUCAY, N., DERYNCK, R. & BLELLOCH, R. 2011. Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat Biotechnol,* 29**,** 443-8.

SWIECH, L., HEIDENREICH, M., BANERJEE, A., HABIB, N., LI, Y., TROMBETTA, J., SUR, M. & ZHANG, F. 2015. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat Biotechnol,* 33**,** 102-6.

SZABO, E., RAMPALLI, S., RISUENO, R. M., SCHNERCH, A., MITCHELL, R., FIEBIG-COMYN, A., LEVADOUX-MARTIN, M. & BHATIA, M. 2010. Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature,* 468**,** 521-6.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131**,** 861-72.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., SASAKI, A., YAMAMOTO, M., NAKAMURA, M., SUTOU, K., OSAFUNE, K. & YAMANAKA, S. 2014. Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. *Nat Commun,* 5**,** 3678.

TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126**,** 663-76.

TESAR, P. J., CHENOWETH, J. G., BROOK, F. A., DAVIES, T. J., EVANS, E. P., MACK, D. L., GARDNER, R. L. & MCKAY, R. D. 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature,* 448**,** 196-9.

THOMSON, J. A., ITSKOVITZ-ELDOR, J., SHAPIRO, S. S., WAKNITZ, M. A., SWIERGIEL, J. J., MARSHALL, V. S. & JONES, J. M. 1998. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science,* 282**,** 1145-1147.

TONG, M., LV, Z., LIU, L., ZHU, H., ZHENG, Q. Y., ZHAO, X. Y., LI, W., WU, Y. B., ZHANG, H. J., WU, H. J., LI, Z. K., ZENG, F., WANG, L., WANG, X. J., SHA, J. H. & ZHOU, Q. 2011. Mice generated from tetraploid complementation competent iPS cells show similar developmental features as those from ES cells but are prone to tumorigenesis. *Cell Res,* 21**,** 1634-7.

TZATSOS, A., PFAU, R., KAMPRANIS, S. C. & TSICHLIS, P. N. 2009. Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus. *Proc Natl Acad Sci U S A,* 106**,** 2641-6.

UTIKAL, J., POLO, J. M., STADTFELD, M., MAHERALI, N., KULALERT, W., WALSH, R. M., KHALIL, A., RHEINWALD, J. G. & HOCHEDLINGER, K. 2009. Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature,* 460**,** 1145-8.

VALENCIA-SANCHEZ, M. A., LIU, J., HANNON, G. J. & PARKER, R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev,* 20**,** 515-24.

VALLIER, L., ALEXANDER, M. & PEDERSEN, R. A. 2005. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci,* 118**,** 4495-509.

VASSILEV, L. T., VU, B. T., GRAVES, B., CARVAJAL, D., PODLASKI, F., FILIPOVIC, Z., KONG, N., KAMMLOTT, U., LUKACS, C., KLEIN, C., FOTOUHI, N. & LIU, E. A. 2004. In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2. *Science,* 303**,** 844-848.

VIERBUCHEN, T., OSTERMEIER, A., PANG, Z. P., KOKUBU, Y., SUDHOF, T. C. & WERNIG, M. 2010. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature,* 463**,** 1035-41.

VODICKA, M. 2001. Determinants for Lentiviral Infection of Non-Dividing Cells. *Somatic Cell and Molecular Genetics,* 26**,** 35-49.

WANG, H., HU, Y.-C., MARKOULAKI, S., WELSTEAD, G. G., CHENG, A. W., SHIVALILA, C. S., PYNTIKOVA, T., DADON, D. B., VOYTAS, D. F., BOGDANOVE, A. J., PAGE, D. C. & JAENISCH, R. 2013a. TALEN-mediated editing of the mouse Y chromosome. *Nat Biotech,* 31**,** 530-532.

WANG, H., YANG, H., SHIVALILA, C. S., DAWLATY, M. M., CHENG, A. W., ZHANG, F. & JAENISCH, R. 2013b. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell,* 153**,** 910-8.

WANG, T., CHEN, K., ZENG, X., YANG, J., WU, Y., SHI, X., QIN, B., ZENG, L., ESTEBAN, M. A., PAN, G. & PEI, D. 2011. The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell,* 9**,** 575-87.

WANG, T., SHI, S. B. & SHA, H. Y. 2013c. MicroRNAs in regulation of pluripotency and somatic cell reprogramming: small molecule with big impact. *RNA Biol,* 10**,** 1255-61.

WANG, T., WEI, J. J., SABATINI, D. M. & LANDER, E. S. 2014. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science,* 343**,** 80-84.

WARE, C. B., WANG, L., MECHAM, B. H., SHEN, L., NELSON, A. M., BAR, M., LAMBA, D. A., DAUPHIN, D. S., BUCKINGHAM, B., ASKARI, B., LIM, R., TEWARI, M., GARTLER, S. M., ISSA, J. P., PAVLIDIS, P., DUAN, Z. & BLAU, C. A. 2009. Histone deacetylase inhibition elicits an evolutionarily conserved self-renewal program in embryonic stem cells. *Cell Stem Cell,* 4**,** 359-69.

WARREN, L., MANOS, P. D., AHFELDT, T., LOH, Y. H., LI, H., LAU, F., EBINA, W., MANDAL, P. K., SMITH, Z. D., MEISSNER, A., DALEY, G. Q., BRACK, A. S., COLLINS, J. J., COWAN, C., SCHLAEGER, T. M. & ROSSI, D. J. 2010. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell,* 7**,** 618-30.

WARREN, L., NI, Y., WANG, J. & GUO, X. 2012. Feeder-free derivation of human induced pluripotent stem cells with messenger RNA. *Sci Rep,* 2**,** 657.

WEN, J., PARKER, B. J., JACOBSEN, A. & KROGH, A. 2011. MicroRNA transfection and AGO-bound CLIP-seq data sets reveal distinct determinants of miRNA action. *RNA,* 17**,** 820-34.

WERNIG, M., LENGNER, C. J., HANNA, J., LODATO, M. A., STEINE, E., FOREMAN, R., STAERK, J., MARKOULAKI, S. & JAENISCH, R. 2008. A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat Biotechnol,* 26**,** 916-24.

WERNIG, M., MEISSNER, A., FOREMAN, R., BRAMBRINK, T., KU, M., HOCHEDLINGER, K., BERNSTEIN, B. E. & JAENISCH, R. 2007. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature,* 448**,** 318-24.

WILLIAMS, R. L., HILTON, D. J., PEASE, S., WILLSON, T. A., STEWART, C. L., GEARING, D. P., WAGNER, E. F., METCALF, D., NICOLA, N. A. & GOUGH, N. M. 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature,* 336**,** 684-687.

WILMUT, I., SCHNIEKE, A. E., MCWHIR, J., KIND, A. J. & CAMPBELL, K. H. S. 1997. Viable offspring derived from fetal and adult mammalian cells. *Nature,* 385**,** 810-813.

WOLF, D. & GOFF, S. P. 2007. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell,* 131**,** 46-57.

WOLF, D. & GOFF, S. P. 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature,* 458**,** 1201-4.

WOLTJEN, K., MICHAEL, I. P., MOHSENI, P., DESAI, R., MILEIKOVSKY, M., HAMALAINEN, R., COWLING, R., WANG, W., LIU, P., GERTSENSTEIN, M., KAJI, K., SUNG, H. K. & NAGY, A. 2009. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature,* 458**,** 766-70.

XIE, F., YE, L., CHANG, J. C., BEYER, A. I., WANG, J., MUENCH, M. O. & KAN, Y. W. 2014. Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res,* 24**,** 1526-33.

YAKUBOV, E., RECHAVI, G., ROZENBLATT, S. & GIVOL, D. 2010. Reprogramming of human fibroblasts to pluripotent stem cells using mRNA of four transcription factors. *Biochem Biophys Res Commun,* 394**,** 189-93.

YAMANAKA, S. 2009. Elite and stochastic models for induced pluripotent stem cell generation. *Nature,* 460**,** 49-52.

YAMASAKI, S., NABESHIMA, K., SOTOMARU, Y., TAGUCHI, Y., MUKASA, H., FURUE, M. K., SATO, J. D. & OKAMOTO, T. 2013. Long-term serial cultivation of mouse induced pluripotent stem cells in serum-free and feeder-free defined medium. *Int J Dev Biol,* 57**,** 715-24.

YANG, C. S., CHANG, K. Y. & RANA, T. M. 2014. Genome-wide functional analysis reveals factors needed at the transition steps of induced reprogramming. *Cell Rep,* 8**,** 327-37.

YANG, C. S., LOPEZ, C. G. & RANA, T. M. 2011. Discovery of nonsteroidal anti-inflammatory drug and anticancer drug enhancing reprogramming and induced pluripotent stem cell generation. *Stem Cells,* 29**,** 1528-36.

YANG, C. S. & RANA, T. M. 2013. Learning the molecular mechanisms of the reprogramming factors: let's start from microRNAs. *Mol Biosyst,* 9**,** 10-7.

YANG, S. H., KALKAN, T., MORRISROE, C., SMITH, A. & SHARROCKS, A. D. 2012. A genome-wide RNAi screen reveals MAP kinase phosphatases as key ERK pathway regulators during embryonic stem cell differentiation. *PLoS Genet,* 8**,** e1003112.

YI, R., DOEHLE, B. P., QIN, Y., MACARA, I. G. & CULLEN, B. R. 2005. Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *RNA,* 11**,** 220-6.

YI, R., QIN, Y., MACARA, I. G. & CULLEN, B. R. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev,* 17**,** 3011-6.

YING, Q.-L., NICHOLS, J., CHAMBERS, I. & SMITH, A. 2003. BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. *Cell,* 115**,** 281-292.

YING, Q. L., WRAY, J., NICHOLS, J., BATLLE-MORERA, L., DOBLE, B., WOODGETT, J., COHEN, P. & SMITH, A. 2008. The ground state of embryonic stem cell self-renewal. *Nature,* 453**,** 519-23.

YOO, A. S., SUN, A. X., LI, L., SHCHEGLOVITOV, A., PORTMANN, T., LI, Y., LEE-MESSER, C., DOLMETSCH, R. E., TSIEN, R. W. & CRABTREE, G. R. 2011. MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature,* 476**,** 228-31.

YU, J., HU, K., SMUGA-OTTO, K., TIAN, S., STEWART, R., SLUKVIN, I. I. & THOMSON, J. A. 2009. Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences. *Science,* 324**,** 797-801.

YU, J., VODYANIK, M. A., SMUGA-OTTO, K., ANTOSIEWICZ-BOURGET, J., FRANE, J. L., TIAN, S., NIE, J., JONSDOTTIR, G. A., RUOTTI, V., STEWART, R., SLUKVIN, I. I. & THOMSON, J. A. 2007. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science,* 318**,** 1917-1920.

YUSA, K., RAD, R., TAKEDA, J. & BRADLEY, A. 2009. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat Meth,* 6**,** 363-369.

YUSA, K., ZHOU, L., LI, M. A., BRADLEY, A. & CRAIG, N. L. 2011. A hyperactive piggyBac transposase for mammalian applications. *Proceedings of the National Academy of Sciences,* 108**,** 1531-1536.

ZENG, Y., YI, R. & CULLEN, B. R. 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci U S A,* 100**,** 9779-84.

ZHANG, H., MA, Y., GU, J., LIAO, B., LI, J., WONG, J. & JIN, Y. 2012. Reprogramming of somatic cells via TAT-mediated protein transduction of recombinant factors. *Biomaterials,* 33**,** 5047-55.

ZHANG, Z. & WU, W. S. 2013. Sodium butyrate promotes generation of human induced pluripotent stem cells through induction of the miR302/367 cluster. *Stem Cells Dev,* 22**,** 2268-77.

ZHAO, X. Y., LI, W., LV, Z., LIU, L., TONG, M., HAI, T., HAO, J., GUO, C. L., MA, Q. W., WANG, L., ZENG, F. & ZHOU, Q. 2009. iPS cells produce viable mice through tetraploid complementation. *Nature,* 461**,** 86-90.

ZHAO, Y., YIN, X., QIN, H., ZHU, F., LIU, H., YANG, W., ZHANG, Q., XIANG, C., HOU, P., SONG, Z., LIU, Y., YONG, J., ZHANG, P., CAI, J., LIU, M., LI, H., LI, Y., QU, X., CUI, K., ZHANG, W., XIANG, T., WU, Y., ZHAO, Y., LIU, C., YU, C., YUAN, K., LOU, J., DING, M. & DENG, H. 2008. Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell,* 3**,** 475-9.

ZHENG, L., LIU, J., BATALOV, S., ZHOU, D., ORTH, A., DING, S. & SCHULTZ, P. G. 2004. An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc Natl Acad Sci U S A,* 101**,** 135-40.

ZHOU, H., WU, S., JOO, J. Y., ZHU, S., HAN, D. W., LIN, T., TRAUGER, S., BIEN, G., YAO, S., ZHU, Y., SIUZDAK, G., SCHOLER, H. R., DUAN, L. & DING, S. 2009. Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell,* 4**,** 381-4.

ZHOU, Q., BROWN, J., KANAREK, A., RAJAGOPAL, J. & MELTON, D. A. 2008. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature,* 455**,** 627-32.

ZHOU, W. & FREED, C. R. 2009. Adenoviral Gene Delivery Can Reprogram Human Fibroblasts to Induced Pluripotent Stem Cells. *STEM CELLS,* 27**,** 2667-2674.

ZHU, S., LI, W., ZHOU, H., WEI, W., AMBASUDHAN, R., LIN, T., KIM, J., ZHANG, K. & DING, S. 2010. Reprogramming of human primary somatic cells by OCT4 and chemical compounds. *Cell Stem Cell,* 7**,** 651-5.

Appendix: Relevant Publication

piggyBac transposon mediated reprogramming and flow cytometry analysis of CD44 and ICAM1 cell-surface marker changes

Springer Protocols

piggyBac transposon mediated reprogramming and flow cytometry analysis of CD44 and ICAM1 cell-surface marker changes

--Manuscript Draft--

Manuscript Number:	
Full Title:	piggyBac transposon mediated reprogramming and flow cytometry analysis of CD44 and ICAM1 cell-surface marker changes
Article Type:	Protocol Chapters
Corresponding Author:	keisuke Kaji, Ph.D. University of Edinburgh Edinburgh, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Edinburgh
Corresponding Author's Secondary Institution:	
First Author:	Sara Brightwell
First Author Secondary Information:	
Order of Authors:	Sara Brightwell
	keisuke Kaji, Ph.D.
Order of Authors Secondary Information:	
Abstract:	Generation of iPSCs is inefficient and the molecular mechanisms underlying reprogramming are not well understood. While several studies have demonstrated that reprogramming is not entirely a random process and contains predictable step-wise changes, varying degrees of cellular heterogeneity that arise in different reprogramming systems can obscure the process. Among several reprogramming systems available, delivery of polycistronic reprogramming factor expression cassettes with piggyBac transposon into mouse embryonic fibroblasts (MEFs) is one of the simplest and most robust reprogramming approaches that provides a low background of partially reprogrammed cells. Using two novel cell surface markers, ICAM1 and CD44, clear cell population changes undergoing reprogramming can be observed over a time course upon induction of the reprogramming factors. Consequently, this technique allows for easy identification of factors that enhance or delay reprogramming, and can be a useful strategy in elucidating key mechanisms for efficient generation of iPSCs.

piggyBac transposon mediated reprogramming and flow cytometry analysis of CD44 and ICAM1 cell-surface marker changes

Sara Brightwell and Keisuke Kaji

Summary

Generation of iPSCs is inefficient and the molecular mechanisms underlying reprogramming are not well understood. While several studies have demonstrated that reprogramming is not entirely a random process and contains predictable step-wise changes, varying degrees of cellular heterogeneity that arise in different reprogramming systems can obscure the process. Among several reprogramming systems available, delivery of polycistronic reprogramming factor expression cassettes with piggyBac transposon into mouse embryonic fibroblasts (MEFs) is one of the simplest and most robust reprogramming approaches that provides a low background of partially reprogrammed cells. Using two novel cell surface markers, ICAM1 and CD44, clear cell population changes undergoing reprogramming can be observed over a time course upon induction of the reprogramming factors. Consequently, this technique allows for easy identification of factors that enhance or delay reprogramming, and can be a useful strategy in elucidating key mechanisms for efficient generation of iPSCs.

Key words: Induced pluripotent stem cells (iPSCs), reprogramming, mouse embryonic fibroblasts (MEF), piggyBac transposon, flow cytometry, ICAM1 (CD54), CD44, Nanog-GFP.

1. Introduction

piggyBac (PB) transposon is the most active and widely used DNA transposon for efficient gene delivery. It can be handled as conventional plasmids without safety cabinets and shows high integration efficiency when co-transfected together with PB transferase which catalyses the integration of PB transposon[1]. Among several strategies to introduce Yamanaka reprogramming factors, piggyBac (PB) transposon is one of the easiest and safest tools to generate iPS cells (iPSCs)[2-4]. Particularly, in combination with polycistronic cassettes of reprogramming factors taking advantage of self-cleaving 2A-peptides, generation of iPSCs is robust and highly reproducible from mouse embryonic fibroblasts, which are most commonly used to study molecular mechanisms of reprogramming. Recently we reported a PB reprogramming system using cells expressing all 4 reprogramming factors, Oct4, Sox2, Klf4 and c-Myc, as visualized by ires-mOrange. We demonstrated that almost all colonies expressing mOrange can activate a pluripotency marker, Nanog-GFP[5], indicating minimal background of partially reprogrammed cells[6]. Using this system we have also demonstrated stepwise progression of reprogramming towards iPSCs with flow cytometry using ICAM1, CD44, and Nanog-GFP reporter as markers [6].

This approach allows detailed analysis of reprogramming kinetics, or the effect of factors of interest, which can be added or removed from the standard reprogramming conditions, providing a strong tool to investigate molecular mechanisms of reprogramming. In this protocol, we describe how to reprogram MEFs with PB transposon and analyse the marker expression changes during reprogramming with flow cytometry.

2. Materials

2.1 Reprogramming materials

1. **Mouse embryonic fibroblasts (MEFs)**. We use MEFs from E12.5 embryos with Nanog-GFP reporter to identify iPSC colonies live[5]. While Oct4-GFP reporter is commonly used to identify iPSC colonies, we found up-regulation of endogenous Oct4 occurs much earlier than many other pluripotency genes in the following protocol as also shown in other systems[6-8]. Reporter system and/or markers for iPSC need to be chosen with great care. Reprogramming efficiency with the following method is largely affected by passage number/proliferation rate of MEFs. We usually use MEFs less than passage 3 for reprogramming experiments. We recommend culturing MEFs in the presence of Fgf2 and heparin (MEF medium as below) for propagation to delay senescence.

2. **Basic medium** consists of 500 ml of GMEM (Sigma, G5154) supplemented with 51 ml foetal calf serum (FCS, Invitrogen, 10270, Batch 40F0240K) (*see* **Note 1**), 5.5 ml MEM

non-essential amino acids (100X, Invitrogen, 11140-036), 1140 μl 50 mM 2-mercaptoethanol (Life Technologies, 31350010), 550 μl LIF (100,000 units/ml, homemade), 2.5 ml penicillin/streptomycin (10,000 U/ml Penicillin, 10,000 μg/ml Streptomycin, Invitrogen, 15140-122), 5.5 ml 100 mM sodium pyruvate (Invitrogen 11360-039), and 5.5 ml 200 mM l-glutamine (Invitrogen, 25030-024).

3. **MEF medium** is supplemented with 10 ng/ml Fgf2 (Peprotec 100-18-B) and 1 μg/ml heparin (Sigma, H3149) to the basic medium before use. Fgf2 10,000x stock (100 μg/ml), heparin 1000x stock (1 mg/ml) are stored at -80 C for long term, and at 4C for less than 1 month. Do not repeat freeze and thaw.

4. **Transfection medium** is MEF medium without penicillin/streptomycin and heparin.

5. **Reprogramming medium** is supplemented with 500 nM Alk4/5/7 inhibitor (Alki, A83-01, TOCRIS Bioscience, 2939), 10 μg/ml vitamin C (VitC, Sigma, 1000731348) and 1 μg/ml Doxycycline (Dox, Sigma, D9891-1G) (*see* **Note 2, 3**). Alki 10,000x stock (5 mM), VitC 5,000x stock (50 mg/ml), Dox 1,000x stock (1 mg/ml) are stored at -80C for long term, and at -20C for less than 1 month.

6. **Dulbecco's phosphate buffered saline** (PBS, Sigma, D8537)

7. **Trypsin EDTA** (Invitrogen, 15090-046)

8. **Fugene HD** (Promega, E2311)

9. **Plasmids**

PB-TAP IRI attP 2LMKOSimO (piggyBac (PB) transposon for Dox-inducible expression of 2A peptide linked Myc, Klf4, Oct4, Sox2 reprogramming factors followed by ires-mOrange) Available upon request to keisuke.kaji@ed.ac.uk. PB-CAG-rtTA (PB transposon for constitutive rtTA expression vector), pCMV-hyPBase (constitutive piggyback transposase (PBase) expression vector). PB-CAG-rtTA and pCMV-hyPBase are available from Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/form/-WcLcvb-BStKQEt0xeg5MjA).

10. **Opti-MEM** (Life Technologies, 31985062)

2.2 FACS analysis materials

1. **Anti-mouse CD54 (ICAM1) biotin**, eBioscience, 13-0541-82.
2. **Anti-Human/Mouse CD44 APC**, eBioscience, 17-0441-82.
3. **Streptavidin PE-Cy7**, eBioscience, 25-431-82.
4. **FACS buffer** (2% FCS in PBS).
5. **Polystyrene round bottom FACS tubes** with or without strainer lid (BD Falcon, 352235 or 352054)

3. Methods

3.1 Reprogramming of MEF with PB transposon

1. Day -2; Seed $1.5x10^5$ MEF per well of a 6 well plate in 2.5 ml transfection medium. Penicillin/streptomycin and heparin can decrease cell viability when Fugene/DNA mix is added, therefore we use transfection medium from this point.

2. Day -1; Prepare the following DNA mix in an Eppendorf tube per well; PB-TAP IRI attP 2LMKOSimO 0.5 μg, PB-CAG-rtTA 0.5 μg, pCMV-hyPBase 0.5 μg (total 1.5 μg DNA), and add 100 μl of Opti-MEM (A). In another tube, prepare 100 μl of Opti-MEM and add 6 ul of Fugene HD (B). Mix A and B, and add to MEFs immediately. It is not necessary to incubate the DNA Fugene mix at room temperature, or change medium.

3. Day 0; Change medium to reprogramming medium. Toxicity of transfection with Fugene HD is minimal and the wells should be 70-80% confluent by this time.

4. mOrange expression should be visible 24 hours after changing medium with a fluorescence microscope. Transfection efficiency at day 2 is usually about 10% by flow cytemetry. Change medium every 2 days (*see* **Note 4**). Clusters of mOrange$^+$ ESC-like cells should appear by day 5, and we start observing Nanog-GFP reporter$^+$ colonies at around day 8. Most colonies have Nanog-GFP$^+$ cells by day 14. If MEFs with pluripotency gene reporter are not available, immunofluorescence against Nanog (eBiosciences, 14-5761-80) or Dppa4 Cosmo Bio, CAC-TMD-PB-DP4) around day 12-14 is recommended to evaluate number of fully reprogrammed iPSC colonies.

3.2 Harvesting samples for FACS time course analysis

The above Fugene transfection protocol gives 30-100 iPSC colonies depending on condition of MEFs. mOrange^{+} cell number is low at the early time points of reprogramming, and gradually increase since cells undergoing reprogramming proliferate faster. To analyse the cell surface marker changes during reprogramming taking a time course, we recommend preparation of at least the following well numbers for each time point; day 2 x10, day 4, x8, day 6 x5, day 8 x3, day 10 and onwards x1 each (*see* **Note 5**).

1. Remove media and wash cells in PBS then aspirate.
2. Lift the cells with 500 µl of trypsin EDTA per 6 well. Incubate for 1-3 minutes at 37°C then pipette to dissociate cells to single cells (*see* **Note 6**).
3. Quench with 2.5-5 ml of media into a universal tube and count cells (*see* **Note 7**).
4. Centrifuge at 1300 rpm for 3 minutes.
5. Aspirate supernatant and resuspend cell pellet in 100 µl FACS buffer per staining required, then transfer to a V-bottom 96 well plate.
6. Store plate on ice; cells are now ready for cell surface marker antibody staining.

3.3 Antibody staining of samples for FACS time course analysis

Samples are stained in 100 µl aliquots so prepare enough staining solution for 100 µl/sample + 1, e.g. if you need to stain 4 samples prepare 500 µl of **each** staining solution required. All centrifugation steps are carried out at 1300 rpm for 3 minutes.

1. Prepare ICAM1/CD44 antibody staining solution as specified in **Table 1** and foil cover/keep out of light on ice until use.
2. Centrifuge 96 well plate and remove FACS buffer with an aspirator.
3. Resuspend cells in 100 µl of ICAM1/CD44 staining solution and incubate plate on ice in the dark for 15 minutes (*see* **Note 8**).
4. Prepare streptavidin PE-Cy7 secondary stain and foil cover/keep out of light on ice until use.
5. Centrifuge plate and wash cells with 100 µl PBS. Repeat centrifugation and remove PBS.
6. Resuspend cells in 100 µl of streptavidin PE-Cy7 secondary staining solution and incubate plate on ice in the dark for 5 minutes.
7. Centrifuge and wash cells in PBS twice as above.
8. Resuspend cells in 100 µl of FACS buffer, transfer to 5 ml FACS tubes and store on ice until analysis.

3.4 Control samples required for FACS time course analysis

Appropriate controls should always be used (*see* **Note 9**).

1. **Unstained E14 ES cells**; this control should be negative for all markers.

2. **ICAM1 PE-Cy7 stained ES cells**; this is a positive control for ICAM1. All cells should be ICAM1 positive. This control also allows for compensation of leakiness into the Red 670/30 (CD44) and Yellow-Green 582/15 (mOrange) filters.

3. **CD44 stained MEFs**; this is a positive control for CD44. All cells should be CD44 positive. This control also allows for compensation of leakiness into the Yellow-Green 780/60 (ICAM1) filter.

4. **Unstained reprogramming (mOrange^{+}) sample**; this control allows for compensation of leakiness of the Yellow-Green 582/15 (mOrange) signal into the Blue 530/30 (Nanog-GFP) and Yellow-Green 780/60 (ICAM1) filters.

3.5 FACS time course analysis of reprogramming samples

This protocol is based on the use of a BD LSRFortessa cell analyser. The voltages suggested are for use with this machine and so optimization of parameters will be required for use of other flow cytometry analysers.

1. Set the cytometer up with the parameters in **Table 2** (*see* **Note 10**).

2. Run all control samples first to set up appropriate compensation and base line acquisitions for your reprogramming samples.

3. Run your reprogramming sample(s) through the analyser and gate your cells firstly to isolate the intact/live cells (this will be your P1 gate) and then gate mOrange^{+} transfected (reprogramming) cells (this will be your P2 gate) (**Figure 2**, *see* **Note 11**).

4. From the mOrange^{+} population you can observe Nanog-GFP^{+} cells when they arise (**Figure 2**).

5. A typical FACS data set for ICAM1/CD44 in our hands is shown in **Figure 3** (*see* **Note 12**).

4. Notes

1. FCS lot affects reprogramming efficiency. We have experienced that all mOrange^{+} cells died off before expressing Nanog-GFP even with an FCS lot that supported ES cell self-renewal. If the above mentioned protocol does not make any iPSC-like colonies, we recommend testing other FCS.

2. This condition gives the highest reprogramming in most of MEFs in our hands so if you are interested in observing the effects of a particular factor on reprogramming then having a lower efficiency by omitting Alki and/or vitC might be more appropriate.

3. We use doxycycline in a range from 300 ng/ml up to 1 μg/ml depending on the condition of our starting MEF and experimental requirements. In general, we find that 300 ng/ml is sufficient and preferred when starting MEF are in good condition and reprogram well. Higher

concentrations of doxycycline are used when reprogramming conditions are less optimal, in order to maintain good induction of the four factors.

4. It is important to keep the reprogramming cultures in a good condition. At later time points, if colour of media gets yellow quickly, change the media every day.

5. CD44 down-regulation usually starts to be observed between day 6 and day 8. ICAM1 up-regulation can be observed before day 10.

6. This time of trypsin EDTA treatment does not affect staining with the antibodies described here.

7. Counting cells is not strictly required, although it is useful; until you are familiar with the technique, it gives you a good idea of how many cells you can harvest from 'X' number of wells at 'X' time point during the preliminary experiments which will help you to plan and optimise future experiments. Also, if you are comparing the effect of additional factors on reprogramming, counting the cells at each time point will give some indication if your factor of interest is having any effect on proliferation of cells undergoing reprogramming (mOrange^{+}) or MEFs (mOrange^{-}). You can calculate absolute cell numbers of each gate based on the total number and % of cells in each gate.

8. If using the antibodies specified in this protocol, you can prepare the ICAM1 and CD44 antibodies in the same staining solution – there is no need to stain them separately.

9. This is important for 2 reasons: 1) over the duration of a time course experiment the power of the lasers in the FACS instrument may be reduced on any given day, even if the voltage is the same. This can produce variation in your time course data but can be identified by use of appropriate controls. 2) Sometimes some of the fluorophores we use emit light that is detected by filters other than the one we intend. This 'leakiness' can be detected by appropriate controls and compensated for.

10. These settings are only a guide of the voltages we use in our lab. Take time to adjust the voltages accordingly to your own instrument, even if it is also a BD LSRFortessa.

11. Ensure your settings allow 10,000 events from your P2 ($mOrange^+$) gate to be saved. In this example we plotted the mOrange channel against the SSC channel to set P2. The mOrange population can be plotted against any other channel. It is advisable to use the channel with the clearest separation between the positive and negative cells.

12. Variations between experiments due to MEF conditions, transfection efficiency etc. can cause the reprogramming kinetics to change and therefore the FACS data may vary by a day or two. In case you use this system to evaluate impact of factors of interest, always control experiments need to be carried out at the same time.

Acknowledgements

We thank to E. Chantzoura and L. Tosti for providing images, A. Nagy, K. Woltjen and K. Yusa for advice on the use of PB transposon for reprogramming. SB is funded by an MRC PhD Studentship. KK is supported by ERC grants ROADTOIPS (no. 261075) and the Anne Rowling Regenerative Neurology Clinic.

References

1. Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America 103 (41):15008-15013. doi:10.1073/pnas.0606979103

2. Kaji K, Norrby K, Paca A, Mileikovsky M, Mohseni P, Woltjen K (2009) Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature 458 (7239):771-775. doi:10.1038/nature07864

3. Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Hamalainen R, Cowling R, Wang W, Liu P, Gertsenstein M, Kaji K, Sung HK, Nagy A (2009) piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature 458 (7239):766-770. doi:10.1038/nature07863

4. Yusa K, Rad R, Takeda J, Bradley A (2009) Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. Nature methods 6 (5):363-369. doi:10.1038/nmeth.1323

5. Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A (2007) Nanog safeguards pluripotency and mediates germline development. Nature 450 (7173):1230-1234. doi:10.1038/nature06403

6. O'Malley J, Skylaki S, Iwabuchi KA, Chantzoura E, Ruetz T, Johnsson A, Tomlinson SR, Linnarsson S, Kaji K (2013) High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature 499 (7456):88-91. doi:10.1038/nature12243

7. Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R (2012) Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 150 (6):1209-1222. doi:10.1016/j.cell.2012.08.023

8. Di Stefano B, Sardina JL, van Oevelen C, Collombet S, Kallin EM, Vicent GP, Lu J, Thieffry D, Beato M, Graf T (2014) C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature 506 (7487):235-239. doi:10.1038/nature12885

Figure Captions

Figure 1. Typical wells at day 14 of reprogramming with PB-TAP IRI attP 2LMKOSimO vector. Top: whole well images from a 6 well plate. Bottom: high magnification images of the white square of the top images. Images were taken with Celigo S Imaging Cytometer (Nexceclom). In this system, mOrnage expression (red) is down-regulated when Nanog-GFP (green) is up-regulated without removing dox from the culture medium.

Figure 2. Typical gating strategy for cells undergoing reprogramming. Gate firstly for intact/live cells in the SSC/FSC channels (P1) and then gate the mOrange^{+} transfected/reprogramming cells by plotting mOrange against SSC (P2). This plot was from day 10 samples and Nanog-GFP^{+} cells will typically emerge around 7-10 days of reprogramming. Note Nanog-GFP^{+} cells have lower mOrnage expression.

Figure 3. ICAM1 and CD44 mark cells undergoing reprogramming. Cells initially exhibit heterogeneous expression of ICAM1 and high CD44. Decreasing ICAM1 expression is then followed by downregulation of CD44 around day 8. Finally, upregulation of ICAM1 marks cells entering the final 'iPSC' between day 8 and 14, which is concurrent with expression of Nanog-GFP (green). Figures are from reference [6] with modification. See [6] for more detail.

Table Captions

Table 1. Antibody staining dilutions for ICAM1/CD44 time course analysis.

Table 2. Guide cytometer settings for BD LSRFortessa.

Table1

Antibody	Fluorophore	Dilution in FACS buffer
CD44	APC	1/300
ICAM1	biotinylated	1/100
Streptavidin	PE-Cy7	1/1500

Table2

BD LSR Fortessa™		Excitation Line			
		488 nm	561 nm	640 nm	Voltage
Band pass (BP) filter	530±30	eGFP			300
	582±15		mOrange		455
	780±60		PECy7		470
	670±30			APC	400

Figure
Click here to download high resolution image

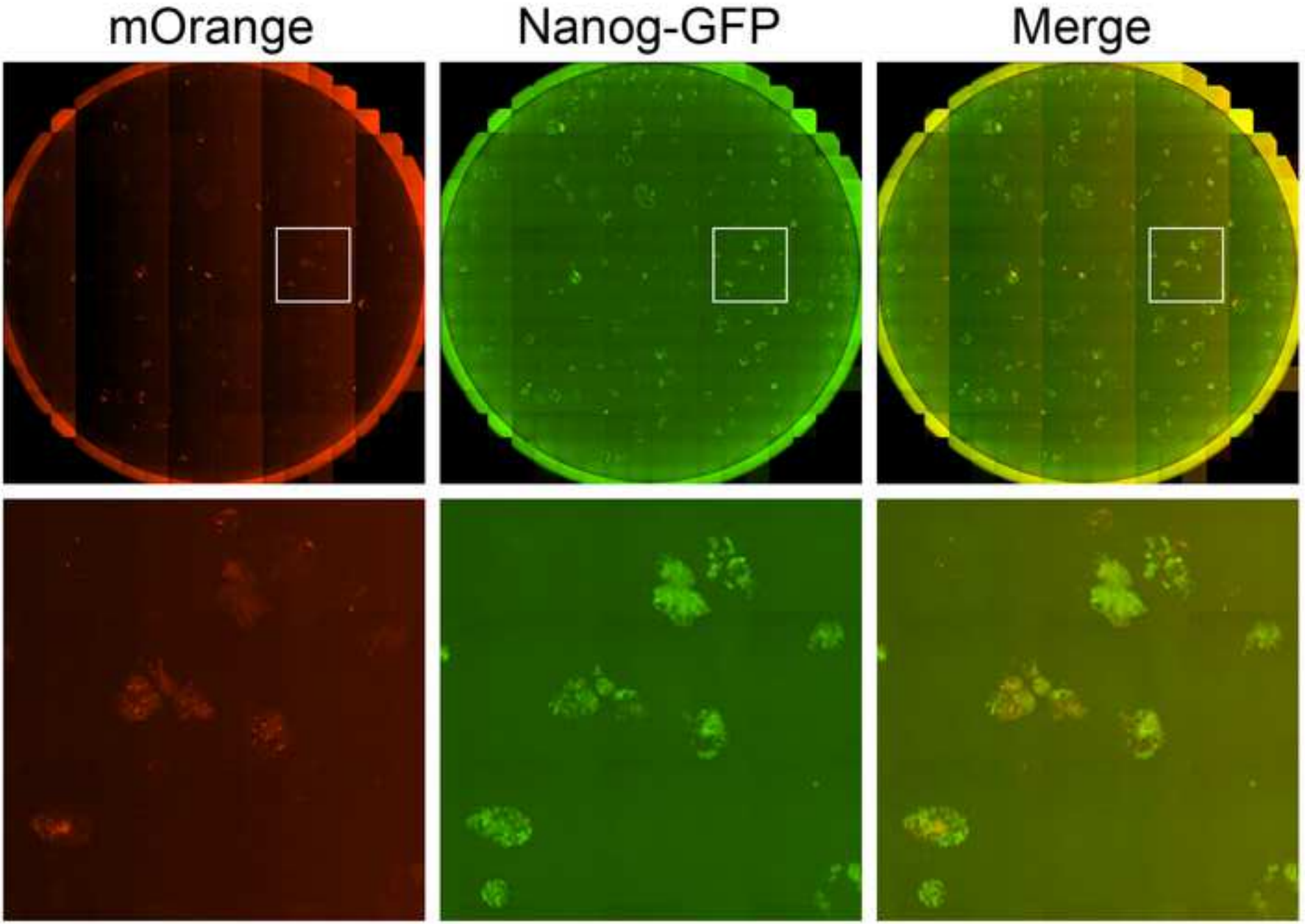


Figure
Click here to download high resolution image

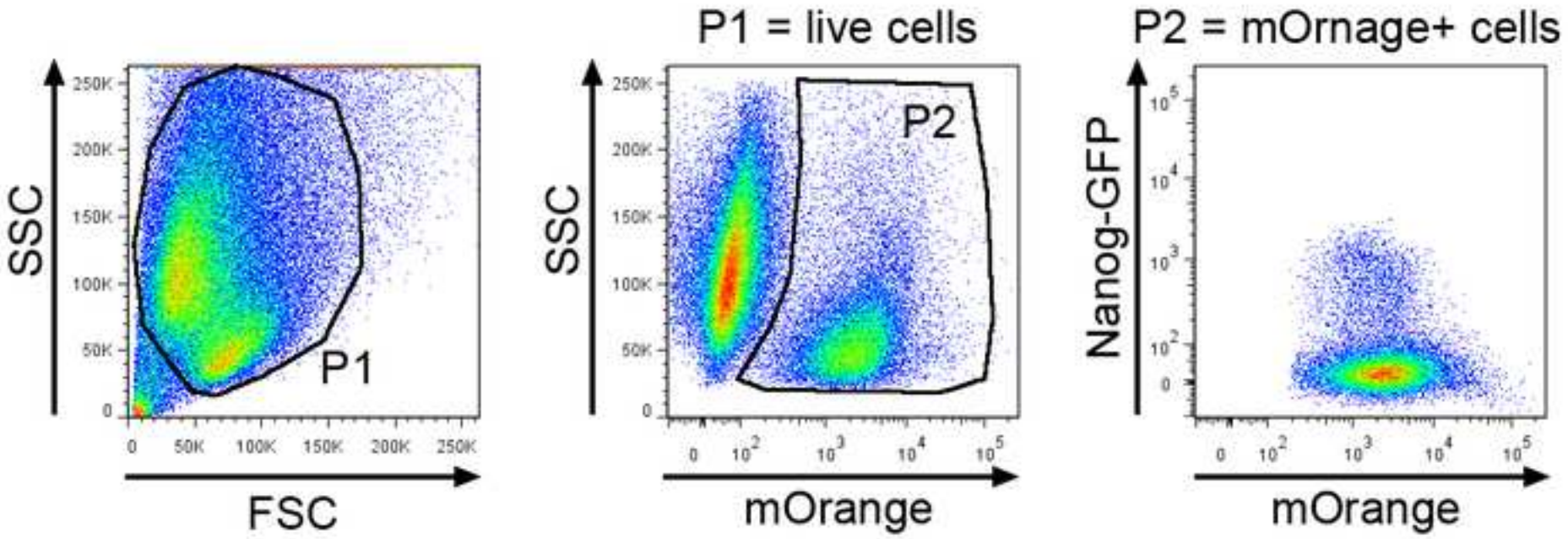


Figure
Click here to download high resolution image