Screening for nuclear reprogramming factors and analysis of DNA demethylation during *in vitro* myoblasts differentiation

Examiners: Prof. Dr. Christof Niehrs
Prof. Dr. Werner Müller

Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Master of Science (Biotechnology)
Suresh Kumar Swaminathan
born in Vellore, India

Oral-examination:

என் பெற்றோருக்கு சமர்பணம்

Dedicated to my parents

Abbreviations

Ab	Antibody
ATP	Adenosine triphosphate
BA	Beta actin
BCA	Bicinchonic Acid
BER	Base excision repair
BMP	Bone morphogenetic proteins
bp	base pairs
BSA	Bovine serum albumin
cDNA	Complementary DNA
COBRA	Combined Bisulphite Restriction Analysis
DE	Distal enhancer
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle's Medium
dn	Dominant negative
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
Dnmt	DNA methyl transferase
dNTP	Deoxynucleotides
DTT	Dithiothreitol
EC cells	Embryonic carcinoma cells
EDTA	Ethylene diamine tetraacetate
EG cells	Embryonic germ cells
EGFP	Enhanced Green Fluorescent protein
EGTA	Ethylene glycol tetraacetate
ES cells	Embryonic stem cells
FBP	FUSE binding protein
FCS	Fetal calf serum
FGF	Fibroblast growth factor
FUSE	Far Upstream Element
Gadd45	Growth arrest and DNA damage 45
GFP	Green Fluorescent protein
HEPES	Hydroxyl pipperazine ethanesulphonic acid
HRE	Hormone responsive element
HTS	High throughput screening
ICM	Inner cell mass
IL-6	Interleukin 6
JAK	Janus activated kinases
kb	Kilobase pairs
kDa	Kilodaltons
LIF	Leukemia Inhibitory Factor
LIFR	LIF receptor
Luc	Luciferase
M	Molar
MAPK	Mitogen activated protein kinase
MEF	Mouse embryonic fibroblasts
MHC	Major histocompatibility complex
MS-PCR	Methylation sensitive PCR

NER	Nucleotide excision repair
nm	Nanometer
No.	Number
PBS	Phosphate buffered Saline
PcG	Polycomb group
PE	Proximal enhancer
PEG	Poly ethylene glycol
PGC	Primordial germ cells
POU	Pit-Oct-Unc
RA	Retinoic Acid
RNA	Ribonucleic Acid
RT	Reverse transcription
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecylsulphate
siRNA	Small interfering RNA
STAT3	Signal transducer and activator of transcription
TGF	Transforming growth factor
v/v	Volume/volume

TABLE OF CONTENTS

1.1 Summary

The primary objective of this thesis was to identify nuclear reprogramming factors that would convert somatic cells back into pluripotency. Two broad screening strategies using the expression of stem cell marker *Oct4* as a molecular read out were employed.

i) Expression screening of *Xenopus* egg cDNA library: *Xenopus* eggs are totipotent as they have the capacity to give rise to whole organism. *Xenopus* egg cDNA libraries were screened by overexpression in various cell lines to isolate the genes that would upregulate the *Oct4* expression.

ii) Chemical Genomics screen: A library of about 3000 small molecules available at the DKFZ-EMBL chemical genomics core facility was screened in a HEK293 stable cell line that harbours a luciferase reporter under the control of the *Oct4* promoter and primary fibroblasts obtained from the mouse harbouring a GFP reporter under the control of the *Oct4* promoter.

I could not isolate any gene or small molecule that could specifically upregulate the *Oct4* expression by the above mentioned screens.

The second objective of this thesis concerns epigenetic changes during cellular differentiation. Several reports have indicated that DNA demethylation accompanies cellular differentiation. As a model case when C2C12 myoblasts were induced to differentiate into myotubes, the promoter of the muscle specific transcription factor *Myogenin* gene is demethylated and thereby facilitating its expression. Our group recently showed that *Gadd45alpha* mediates active DNA demethylation by a nucleotide excision repair mechanism. I tested if *Gadd45* may mediate the demethylation observed during the C2C12 differentiation by knocking down the different isoforms of *Gadd45*. *Gadd45beta* knockdown blocked the *Myogenin* promoter demethylation as analysed by MS-PCR and COBRA assays. Moreover, the *Myogenin* expression was significantly reduced. In conclusion, *Gadd45* is indeed required to relieve gene silencing during the differentiation of myoblasts into myotubes.

1.2 Zusammenfassung

Das grundlegende Ziel dieser Arbeit war die Identifizierung von nukleären, reprogrammierenden Faktoren mit der Fähigkeit, somatische Zellen in pluripotente Zellen zu konvertieren. Hierfür wurden zwei umfassende Strategien angewandt, in denen die Expression des Stammzellmarkers *Oct4* als molekularer Readout verwendet wurde.

i) Expressionsanalysen unter Verwendung von *Xenopus* cDNA Bibiliotheken: *Xenopus* Eier sind totipotent, d.h. sie können Ausgangspunkt für einen kompletten Organismus sein. Eine cDNA Bibiliothek aus *Xenopus* Eiern wurde in verschiedenen Zelllinien überexprimiert, um Gene zu isolieren, die *Oct4* Expression induzieren.

ii) Chemische genomweite Analysen: Eine Bibliothek von 3000 kleinen chemischen Verbindungen der „DKFZ-EMBL chemical genomics core facility" wurde für die Analyse in einer HEK293 stabilen Zelllinie verwendet, die ein Luziferase Reporterplasmid unter der Kontrolle des *Oct4* Promotors trägt. Die Bibiolothek wurde ebenfalls in primären Fibroblasten einer transgenen Maus getestet, die einen GFP Reporter unter der Kontrolle des *Oct4* Promotors tragen.

Ich konnte in den durchgeführten Screens kein Gen oder Molekül isolieren, welches spezifisch die *Oct4* Expression hochreguliert.

Die zweite Fragestellung dieser Arbeit befasst sich mit epigenetischen Veränderungen während der zellulären Differenzierung. Mehrere Untersuchungen belegen, dass Differenzierung mit DNA Demethylierung einhergeht. So wird zum Beispiel während der Differenzierung von C2C12 Myoblasten zu Myotuben der Promotor eines Muskel spezifischen Transkriptionsfaktors *Myogenin* demethyliert und somit dessen Transkription induziert. Kürzlich wurde in unserer Arbeitsgruppe gezeigt, dass Gadd45alpha aktive DNA Demethylierung unter Einbeziehung von Nukleotid-Entfernungsreparatur vermittelt. Ich habe getestet, ob Gadd45 ebenfalls ein Schlüsselrolle bei der Demethylierung während der C2C12 Differenzierung spielt, indem ich die Expression verschiedener Gadd45 Isoformen dezimiert habe. MS-PCR und COBRA Analysen zeigen, dass Gadd45beta Knockdown die Demethylierung des Myogenin Promotors blockiert. Darüberhinaus war die Myogenin Expression signifikant reduziert. Folglich ist Gadd45 in der Tat notwendig, Gen-Stilllegung während der Differenzierung von Myoblasten zu Myotuben aufzuheben.

2 Introduction

We begin our life from fertilization of an egg by a sperm. The resulting zygote has the capacity to develop into a complete organism and hence is referred as the totipotent stem cell. As we grow from the zygote into a complete organism following a strict developmental program, most cells, except the gametes, lose this totipotency. Is it possible to reverse this program so that somatic cells re-acquire some stem cell characteristics? Any such generated stem cells could be differentiated into tissues and organs for use in cell replacement therapies. In order to transform this wishful thinking into a medical reality we ought to discover the reagents to convert our somatic cells into a stem cell state. Above all it is important that we understand the factors that determine the stem cell state. Furthermore, in order to drive the stem cell into a tissue of our choice, we need to know about the epigenetic programs that underlie a differentiated state. These questions form the frame work of my PhD thesis.

In the first part of this thesis, I have searched for the reprogramming factors to convert a somatic cell into a stem cell state. In the second part, I have attempted to understand the mechanisms behind the epigenetic changes accompanying cellular differentiation using muscle differentiation as a model.

2.1 Biology of Stem cells

2.1.1 Characteristic properties of stem cells

Two hall-mark properties of a stem cell are:

i) Self-renewal: This refers to the ability to go through numerous rounds of cell cycles while still maintaining the undifferentiated state. To retain self-renewal, a stem cell can undergo two types of cell division (Figure 1). If a stem cell divides to yield two identical daughter cells that remain undifferentiated, the stem cell is said to have undergone a symmetric division. If one of the daughter cell remains undifferentiated and the other commits to differentiation then the stem cell is said to have undergone asymmetric division.

.

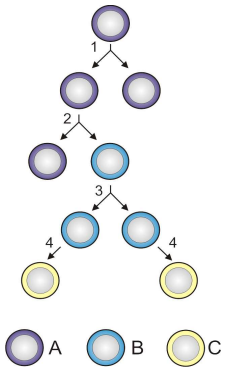


Figure 1: Stem cell renewal. A - stem cells; B - progenitor cell; C - differentiated cell; 1 - symmetric stem cell division; 2 - asymmetric stem cell division; 3 - progenitor division; 4 - terminal differentiation. Reproduced from (En.wikipedia.org/wiki/Adult_stem_cell).

Asymmetric division results in a stem cell and a progenitor cell. The progenitor cell undergoes several rounds of proliferation before terminal differentiation.

ii) Potency: Potency of a cell refers to its ability to differentiate into different cell types. Based on the potency, stem cells can be termed as

Totipotent (*Toti*: all): A totipotent stem cell can give rise to all the cell types that make up the body plus all of the cell types that make up the extra-embryonic tissues such as the placenta. For example, the zygote is totipotent as it has the capacity to give rise to a whole organism.

Pluripotent (*Pluri*: several but not all): These cells are the descendants of totipotent stem cells. They have the ability to differentiate into cells derived from the three germ layers, but cannot differentiate into extra embryonic tissues such as the chorion, amnion and other components of the placenta. For example, inner cell mass of a blastocyst, embryonic stem cells and embryonic germ cells.

Multipotent (*Multi:* many) A stem cell is considered to be multipotent if it has the ability to develop into more than one cell type of the body. Often they give rise to a closely related lineage of cells. For example, hematopoeitic stem cells differentiate into red blood cells, white blood cells and platelets.

Unipotent (*Uni*: single) cells has the ability to produce one cell type. However these cells have the capacity to self renew which distinguishes them from a differentiated cell. For example, skin stem cells are considered unipotent as they always differentiate only into skin.

All the above definitions are from http://stemcells.nih.gov/ info/glossary.asp. This section is summarised in the Figure 2.

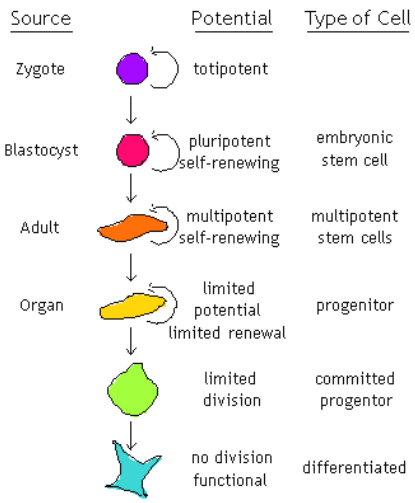


Figure 2: Potency of cells varies.
Stem cells are characterised by the properties of self-renewal and unlimited potency. Based on the potency, a cell can be termed as toti-, pluri-, multi- or uni- potent. Reproduced from (www.thebiotechclub.org/industry/emerging/stem_cells.php).

2.1.2 Derivation of stem cells

After fertilization, the mouse zygote undergoes about six cleavage divisions before implantation. At this stage, a hollow-sphere of cells termed as blastocyst is formed. A Blastocyst is made up of about 64 blastomeres and has two distinct cell populations: i) Outer trophoectoderm cells which have lost their totipotency. These cells are necessary for implantation and for formation of the placenta. ii) ICM which forms the embryo proper. ICM cells cannot give rise to extra-embryonic tissues and hence is termed pluripotent. After implantation, the ICM develops into epiblasts cells of the early egg cylinder which retain pluripotency. These cells then respond to the signals from the surrounding extra-embryonic tissues that direct differentiation and initiation of gastrulation. At the onset of gastrulation, the primordial germ cells (PGCs) are specified. At this stage, pluripotency is restricted to PGCs and thereafter to germ cells which arise from them.

Totipotent zygote and early blastomeres of an embryo undergo cleavage divisions instead of a normal cell division making it difficult to derive and establish *in vitro* cultures from them (Surani et al., 2007). ICM and PGCs however undergo a normal cell division and they respond to external signals making it possible to use them as precursors for the establishment of pluripotent cell cultures *in vitro*. Embryonic Stem (ES) cells and Embryonic Germ (EG) cells were derived from ICM (Evans and Kaufman, 1981; Thomson and Marshall, 1998) and PGCs (Matsui et al., 1992; Resnick et al., 1992; Shamblott et al., 2001) respectively. ES cell cultures have been

established from mouse (mES) (Evans and Kaufman, 1981), humans (hES) (Thomson and Marshall, 1998) and other mammalian species like bovines (Hwang, 2005), primates (Thomson et al., 1995), rabbit (Fang et al., 2006; Graves and Moreadith, 1993) etc.

Under optimal conditions, mES and hES may be cultured indefinitely. They can be induced to undergo differentiation *in vitro* as well as *in vivo* by forming teratomas upon implantation into immunosuppressed mice. To summarise, ES cells are artefacts of their *in vitro* environment because *in vivo* pluripotency of ICM is retained very transiently before it differentiates into egg cylinder and subsequently into germ layers (Surani et al., 2007).

2.2 Determinants of pluripotency

One of the key questions in the stem cell field is to understand the precise mechanism underlying self-renewal and pluripotency of stem cells. In recent years great strides were made in understanding the biology of stem cells. In the following section, I summarise various signalling pathways, transcription factors and epigenetic modifications implicated in maintaining mouse and human stem cell pluripotency.

2.2.1 Signal transduction pathways

a) LIF-STAT3 pathway

Mouse ES are routinely propagated by co-culturing on a monolayer of mitotically inactivated mouse embryonic fibroblast (MEF) feeder cells. MEFs provide the ES cell with Leukaemia Inhibitory Factor (LIF), a cytokine of IL-6 family. Indeed recombinant LIF can be used instead of MEF feeder cells highlighting the importance of LIF in maintaining mES cells (Smith, 1988; Williams et al., 1988). LIF signals via a heterodimeric receptor complex composed of a specific low-affinity LIFR beta receptor chain and gp130 chain. Cytokine induced dimerisation results in activation of janus-activated kinases (JAKs) and signal transducer and activation of transcription (STAT), particularly STAT3 as well as mitogen activated protein kinase (MAPK). However, MAPK is reported to promote differentiation (Burdon et al., 1999) and STAT3 activation is sufficient to maintain the pluripotency negating the

influence of MAPK activation (Nichols et al., 1998). The LIF/STAT3 pathway activates the targets such as growth factor Lefty 1, transcriptional regulators like c-Myc, Id1 and Id2 and groucho-like protein Aes1 (Sekkai et al., 2005). Unlike mES cells, hES can maintain pluripotency by LIF/STAT3 independent mechanisms (Humphrey, 2004). Nevertheless components of LIFR and gp130 are found in hES cells. Exogenous addition of human LIF upregulates STAT3 levels, however this alone is insufficient to maintain hES pluripotency *in vitro* (Daheron et al., 2004).

b) basic FGF pathway

Usually human ES cells are cultured in the presence of basic fibroblast growth factor (bFGF) on a fibroblast feeder cells or in a fibroblast conditioned medium (Amit et al., 2000; Xu et al., 2001). Although, bFGF is crucial for hES cell pluripotency, its exact mechanism of action is yet to be elucidated (Na Liu, 2007).

The PI3K pathway, Wnt signalling and TGFβ signalling pathways are also implicated in maintaining the pluripotency of embryonic stem cells (Boiani et al., 2002; Rao, 2004).

2.2.2 Transcription factors

a) Oct4

Oct4, also known as Oct3 and encoded by Pou5f1, is a Pit-Oct-Unc (POU) domain-containing transcription factor that binds to an octamer sequence, ATGCAAAT (Herr and Cleary, 1995; Ryan and Rosenfeld, 1997). Oct4 was first identified as a binding activity present in the extracts of undifferentiated embryonic stem cells and embryonic carcinoma Cells (Lenardo et al., 1989; Okamoto, 1990; Schoeler et al., 1989b) The presence of Oct4 in pluripotent ES and EC cells prompted the investigators to study its expression and role during early mouse embryonic development.

Expression

Expression of Oct4 follows a strict developmentally regulated pattern (Figure 3). In mouse embryos, maternal Oct4 transcripts are present and after 4-cell stage, zygotic Oct4 expression is activated in all the blastomeres until late morula. In the blastocysts

stage, Oct4 expression is restricted to ICM and is downregulated in the

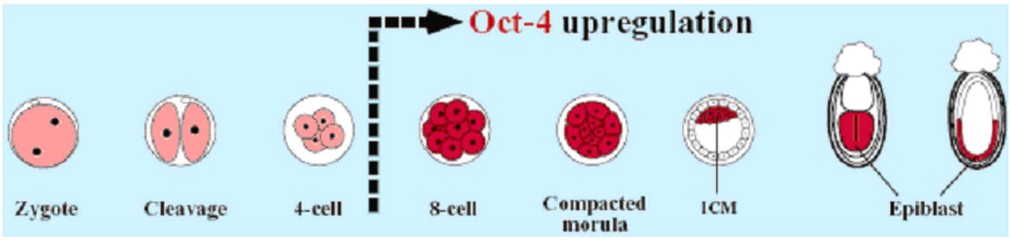


Figure 3: Oct4 expression during mouse development.
Oct4 is present as maternal transcript. Zygotic expression is activated prior to the 8 cell stage. It is expressed during morula and in the blastocyst stage its expression is restricted to ICM and later, at day 4.5, to migrating cells of differentiating primitive endoderm. Following implantation, Oct4 expression is limited to primitive ectodermal cells. Expression in primordial cells is detectable at day 8.5 (not shown). Adapted from (Pesce et al., 1998).

trophoectoderm. At implantation, Oct4 expression is maintained in the primitve ectoderm. At the end of gastrulation, Oct4 expression is significantly down regulated. Thereafter, Oct4 expression is restricted to PGCs, precursors of the germ cells (Palmieri et al., 1994; Rosner, 1990; Schoeler et al., 1989a; Yeom et al., 1996; Yeom et al., 1991) *Oct4*$^{-/-}$ embryos die at the time of implantation due to a failure to form ICM. Of note, suppression of Oct4 function does not affect the maintenance of totipotent phenotype of blastomeres prior to the formation of the blastocyst stage.

The unique Oct4 expression in the mouse embryo has lead to the hypothesis of the totipotent cycle. According to this postulate, cells losing Oct4 during embryonic development differentiate into somatic lineages whereas cells maintaining Oct4 expression retain totipotency (Yeom et al., 1996). More over, when ES or EC cells are induced to differentiate, *Oct4* mRNA expression is downregulated (Lenardo et al., 1989; Schoeler et al., 1989a). Thus Oct4 expression is associated with undifferentiated cells, *in vivo* as well as in cell culture. However studies suggesting that i) Oct4 is expressed in low levels in adult tissues (Takeda et al., 1992) and ii) neuronal differentiation of ES cells could still proceed with sustained Oct4 levels, questions the widely held initial notion that Oct4 expression is downregulated in differentiated cells (Shimozaki et al., 2003).

It is interesting to note that in mES cells, absolute levels of Oct4 is tightly regulated and is maintained within narrow limits. Any deviation from this limit promotes differentiation: An increased Oct4 levels leads to ES differentiation into endoderm

and mesoderm whereas decreased Oct4 levels promote differentiation into trophoectoderm (Niwa et al., 2000).

The *Oct4* gene has been found only in mammalian species. The human Oct4 and mouse Oct4 are 92% identical at amino acid level. Moreover human Oct4 expression pattern resembles that of mouse ortholog (Abdel-Rahman et al., 1995; Takeda et al., 1992). To date, it has been very difficult to identify closer homologs of *Oct4* in lower vertebrates. Of the several POU domain transcription factors identified in non mammalian species, zebrafish *Pou2* (*Zpou2*) and axolotl *Oct4* homolog (*Axoct4*) have a considerably very high amino acid identity (63% and 73% respectively) with mammalian *Oct4*. *Xoct-25*, *Xoct-60*, *Xoct-79* and *Xoct91* are identified as *Xenopus* POU classV homologs of mammalian *Oct4*. They have about 53-63% amino acid identity with mammalian *Oct4* (Rosemary F. Bachvarova, 2004). During *Xenopus* embryogenesis, these proteins act as suppressors of commitment during germ layer specification and thereby maintain multipotency. Moreover, *Axoct4*, *Xoct-25*, *Xoct-60* and *Xoct-91* have the ability to maintain mouse ES cells in the absence of Oct4 (Morrison and Brickman, 2006).

Regulation

The coding region of the mouse *Oct4* gene consists of five exons and its regulatory region includes a TATA less promoter which is located within 250bp of transcription initiation site. The enhancer regions, Proximal Enhancer (PE) and Distal Enhancer (DE) are located approximately 1.2kb and 2kb upstream respectively (Figure 4). Even after nearly two decades since discovery, the transcriptional regulation of the Oct4 expression is not yet completely understood. Transgenic analysis of the Oct4 expression was carried out using a LacZ reporter gene linked to various regions of the *Oct4* promoter sequences. An 18kb genomic fragment (GoF-18) was found to mimic the endogenous Oct4 expression. Deletion analysis concluded that Oct4 expression is regulated by three cis acting elements – the promoter, PE and DE (Okazawa et al., 1991; Yeom et al., 1996; Yeom et al., 1991).

The *Oct4* proximal promoter consists of a GC box to which Sp1 transcription factor proteins can bind and three repeated half sites of the consensus hormone response element (HRE). HRE resembles a canonical retinoic acid response element (RARE)

to which receptors belonging to the steroid thyroid hormone receptor family including Retinoic Acid (RAR) and retinoid X receptors (RXR) bind. Overlapping this region is a putative binding site for ELP, a nuclear orphan receptor (Ben-Shushan et al., 1995; Okazawa et al., 1991; Schoorlemmer et al., 1994). Along with Oct4, other genes implicated in early embryonic development like E-Cadherin, HCGβ have a TATA less promoter. Sp1 transcription factor mediates the expression of these TATA less genes in the undifferentiated cells and early mouse embryos (Nothias et al., 1995; Ovitt and Schoeler, 1998).

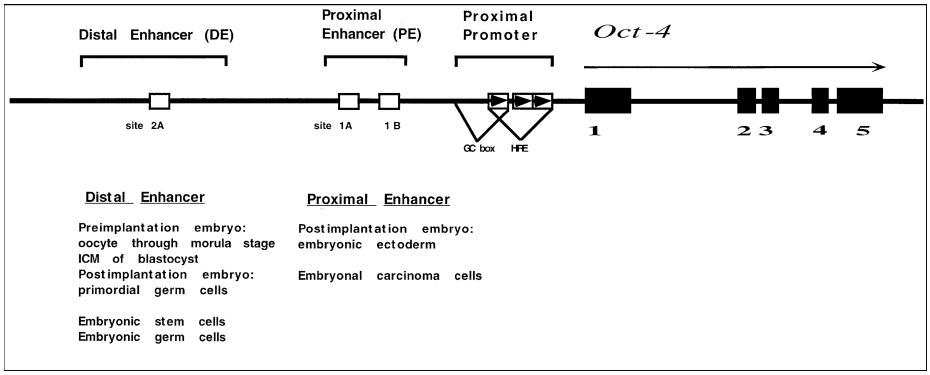


Figure 4. Genomic structure of the *Oct4* gene.
The diagram represents ~24kb of the genomic region surrounding the *Oct4* gene. The gene is divided into five exons, depicted as black boxes. The identified upstream regulatory regions include the promoter, Proximal Enhancer and Distal Enhancer and are indicated by brackets whose length is proportional to that of the region. HRE = hormone responsive element; GC box is recognized by the Sp1 family of transcription factors Adapted from (Ovitt and Schoeler, 1998).

Proximal Enhancer (PE) is a stage specific enhancer. In the developing embryo, its activity is limited to the primitive ectoderm. In contrast to its low activity in undifferentiated cells of the preimplantation embryos and ES cells, the PE is highly active in P19 EC cells. The PE has a cis-acting element which responds to retinoic acid (RA) mediated down-regulation of Oct4 expression, although no RA receptor binding site is present. This element can be subdivided into two sites (1A and 1B) and both the sites bind factors *in vitro*. However only site 1A is occupied *in vivo* in ES and EC cells (Okazawa et al., 1991; Ovitt and Schoeler, 1998; Yeom et al., 1996).

Distal Enhancer (DE) is active in undifferentiated cells of the preimplantation embryo and is also responsible for the Oct4 expression in PGCs later in the development. The DE is crucial for the Oct4 expression in ES cells and is sufficient to enhance the LacZ reporter expression driven by a TK minimal promoter. The DE has a site 2A, that is

similar to, but present in opposite orientation to that of site 1A of the PE. *In vivo*, genomic foot printing experiments has demonstrated that this site is occupied by transcription factors in undifferentiated cells. However in the cells induced to differentiate with RA treatment, this site was observed to be unoccupied (Ovitt and Schoeler, 1998; Yeom et al., 1996).

To sum up, the expression of Oct4 is temporally regulated such that the DE is active in ICM whereas the PE drives the expression in the primitive ectoderm and EC cells derived from it. The switch from DE to PE activity occurs around implantation. Following down regulation of Oct4 during gastrulation, DE regains activity in the PGCs. The selective activation of DE in the pluri/totipotent cells indicates that there are specific DNA-binding factors that may define the undifferentiated state. But these factors are not yet discovered. The aim of the thesis is to identify such upregulators of Oct4 which may be crucial in specifying pluripotency.

Down stream targets and Function:

Oct4 functions as a transcriptional activator of genes necessary for undifferentiated state of stem cells or as a suppressor of genes that are associated with the differentiated state. Target genes like *Rex1*, Osteopontin (*OPN*), and *Fgf4* are associated with pluripotency and are subjected to a positive regulation by the Oct4 transcription factor whereas expression of target genes like the alpha and beta subunits of human chrorionic gonadotropin, caudal-type homeobox transcription factor2 (Cdx2) expression are suppressed. Cdx2 triggers a pluripotent cell to differentiate into trophoectoderm. Oct4 prevents this by blocking the expression as well as the activity of Cdx2. Moreover, Cdx2 overexpression results in the downregulation of the Oct4 indicating these two proteins may constitute a negative feedback loop (Niwa et al., 2005).

b) Sox2

Sox2 is one of the key regulators of pluripotency. It is a member of the HMG-domain DNA-binding-protein family that is implicated in the regulation of transcription and chromatin architecture (Pevny and Lovell-Badge, 1997). The Oct4 and Sox2 proteins bind to the regulatory elements of the *Fgf4* gene that encodes a stem cell specific growth factor and synergistically activates *Fgf4* transcription (Yuan et al., 1995).

Interestingly regulatory regions of the *Sox2* and the *Oct4* gene have a Sox-Oct element to which Oct4/Sox2 complex binds and synergistically activate *Sox2* and *Oct4* expression respectively (Okumura-Nakanishi et al., 2004; Tomioka, 2002). Sox2 null embryos die immediately after implantation (Avilion, 2003). Knockdown of Sox2 in ES cells induces its differentiation into multiple lineages highlighting the importance of Sox2 in the maintenance of pluripotency (Ivanova, 2006).

c) Nanog

For nearly a decade, Oct4 was considered as the sole master regulator of pluripotency. The discovery of Nanog independently by the Yamanaka and Austin Smith groups (Chambers, 2003; Mitsui et al., 2003) has challenged this notion. Nanog may play an equally important role in specifying cellular pluripotency. Nanog is a homeobox containing transcription factor with an essential role in maintaining the pluripotent cells of the inner cell mass and ES cells. It is expressed in the pluripotent cells and is downregulated in differentiated cells. However using RT-PCR, *Nanog* transcripts are seen in adult cells (Hart et al., 2004). Overexpression of the *Nanog* in mES cells confers a LIF independent ability for self renewal and pluripotency. They are resistant, but not completely refractory, to the spontaneous differentiation that occurs after LIF withdrawal or by chemical induction after treatment with 3-methoxybenzamide or all-*trans* RA. *Nanog*$^{-/-}$ embryos do not develop beyond blastocyst stage due to failure to form an epiblast. *Nanog*$^{-/-}$ ES cells, derived from these blastocysts, differentiate into parietal endoderm like cells. In contrast to *Oct4* $^{-/-}$ ES cells, differentiation into trophoectoderm was not seen (Mitsui et al., 2003). Taken together, these data suggests that *Nanog* is critical for maintaining pluripotency of ICM at a stage after the initial requirement of Oct4. Overexpression of the *Nanog* promotes maintenance of pluripotent hES cells in feeder free conditions (Darr et al., 2006). Like mES cells, hES also differentiate to extra-embryonic endoderm upon *Nanog* ablation (Hyslop et al., 2005). *Nanog* may function in a similar fashion like *Oct4* by activating pluripotency specific genes like *Rex1* and by repressing genes that promote differentiation like endodermal GATA-binding protein-6, *Gata6* (Mitsui et al., 2003). Very little is known about regulation of the *Nanog* gene although Oct4, Sox2, P53 and FoxD3 have been implicated in this process (Kuroda et al., 2005; Lin, 2005; Pan et al., 2006; Rodda et al., 2005).

d) Transcriptional network in pluripotency

Oct4, Sox2 and Nanog are the main transcription factors involved in maintaining ES cell pluripotency. Oct4 and Sox2 interact to form an Oct4/Sox2 complex that positively regulates Oct4, Sox2 and Nanog expression. Oct4/Sox2/Nanog expression affects the expression of negative regulators of pluripotency like Cdx2, GCNF, Gata6, Coup-transcription factors. Cdx2 forms a reciprocal inhibitory loop with Oct4 which results in their mutually exclusive expression pattern. Oct4 is subjected to negative regulation by GCNF as well as Coup-transcription factors which are transcriptionally activated by Gata factors (Ben-Shushan et al., 1995; Fujikura et al., 2002). Moreover Nanog represses Gata6 expression thereby preventing the differentiation of ES cells into parietal endoderm. The transcriptional network is highlighted in the Figure 5.

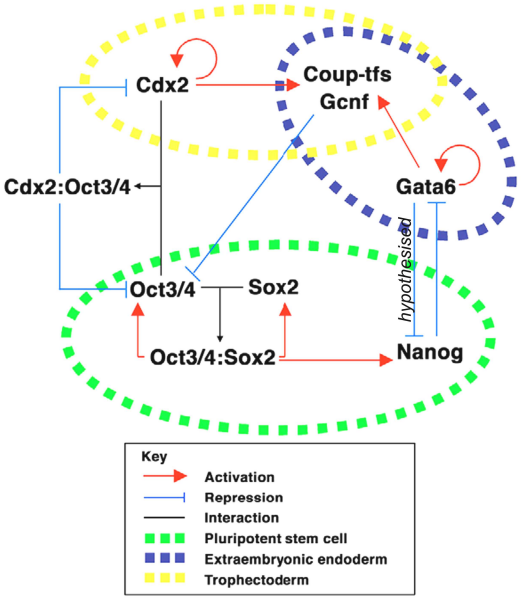


Figure 5: A transcription factor network to control ES cell pluripotency.
Oct4, Sox2 and Nanog are the core regulatory transcription factors essential for ES cell pluripotency. Oct4 is negatively regulated by differentiation inducing Cdx2, Gcnf and Coup-tfs. Nanog suppresses the differentiation promoting Gata6 expression. Repression of Nanog by Gata6 in ES cells is hypothesised. The figure is modified from (Niwa, 2007).

2.2.3 Epigenetic mechanisms

Recently many novel epigenetic features of ES cells, which may help in delineating the molecular mechanism underlying cellular pluripotency, have been reported. Some of the key discoveries covering the ES cell - chromatin state, histone modifications and DNA methylation status would be highlighted in this chapter.

a) Chromatin features

When compared to a somatic cell, ES cell chromatin has increased DNaseI hypersensitive sites, exhibits the features of acetylated histone (H4Ac) and methylated H3K4 (H3K4me3) (Meshorer, 2006), which are the hallmarks of transcriptionally active euchromatin. Due to the relaxed chromatin structure, the volume of ES cell nuclei is about double the volume of nuclei in differentiated cell (Faro-Trindade and Cook, 2006). Moreover in differentiated cells, inactive genes are positioned close to centromeric heterochromatin whereas in ES cells no such spatial compartmentalisation of inactive genes is observed (Smale, 2003). One unique feature of ES cells is the presence of so called bivalent domains in which chromatin elements that mark active state and those that mark inactive state coexist (Azuara, 2006; Bernstein, 2006).

Bivalent domains and Polycomb group of proteins:

Histone modifications are associated with different chromatin states. Histone H3 modifications like methylation of lysine 9 (H3K9me) or lysine 27 (H3K27me) are usually associated with inactive chromatin state as they recruit specific repressive proteins like HP1 and PRC1. Histone modifications like di and trimethylation of lysine 4 of H3 (H3K4me), acetylation of H3 or H4 mark transcriptionally active chromatin. Thus transcriptionally active genes are marked by H3K4me whereas inactive genes are marked by H3K9me or H3K27me. Contrary to the expectation, the inactive chromatin region of an ES cell was observed to contain H3K4me (active) and H3K9me (inactive) marks juxtaposed. Such chromatin domains are termed bivalent domains. This indicated a new model for gene regulation in pluripotent cells in which tissue specific genes are primed for expression but held in check by an opposing histone modification. Upon differentiation, these bivalent domains are resolved leading to transcriptional activation of the tissue specific genes (Spivakov and Fisher, 2007).

H3K27 methylation is catalysed by polycomb group (PcG) proteins. This indicated that the PcG proteins may play a crucial role in regulating gene expression in the ES cells. To date, four different PcG protein complexes have been characterised. Of these, Polycomb repressor complex1 (PRC1) and Polycomb repressor complex2

(PRC2) play an important role in the ES cells. PRC2 which consists of Ezh2, Eed and Suz12 catalyse methylation of H3K27. PRC1 which is made up of Ring1A and Ring1B is then recruited to H3K27me region and may repress the transcription by an unknown mechanism. In the undifferentiated ES cells, many PcG target genes carry bivalent domains. Moreover, PcG target genes are inactive in pluripotent cells but become derepressed upon differentiation. In PcG ablated ES cells, inappropriate upregulation of many tissue-specific target genes was observed. These observations have led to the conclusion that the PcG proteins play an essential role in maintaining pluripotency by preventing the premature expression of the differentiated state (Azuara, 2006; Bernstein, 2006; Boyer, 2006).

b) DNA methylation

DNA methylation at CpG islands is one of the prominent epigenetic mechanisms in mammals (Jaenisch and Bird, 2003). ES cells show unique DNA methylation profiles when compared to other cells like trophoectoderm stem cells, embryonic germ cells (Hattori et al., 2004). Moreover, mammalian genome has numerous tissue-dependent and differentially methylated regions (T-DMRs). Generally DNA methylation of T-DMRs in gene loci causes transcriptional repression whereas demethylation leads to transcriptional activation (Imamura et al., 2001). The promoter of *Oct4* is hypermethylated in differentiated cells whereas it is hypomethylated in ES cells thereby permitting its expression in the latter (Hattori et al., 2004). Similarly the T-DMR of the *Nanog* gene is hypomethylated in ES cells whereas it is heavily methylated NIH3T3 cells in which the *Nanog* gene is repressed (Hattori et al., 2004). DNA methylation is essential for proper cellular differentiation and hence normal development of the organism (Li, 2002). I will elaborate more on this aspect in the chapter 2.4

Maintanence of stem cell properties involves several mechanisms. However, we do not yet know what minimal factors are essential to establish a definitive pluripotent state. The acid test for any such authoritative pluripotency inducing factor(s) , if discovered, is their ability to convert a somatic cell to a pluripotent cell. Although we are far away from such authoritative pluripotency inducing factors, we have made a great progress in our ability to turn back the developmental clock of a somatic cell to a pluripotent state. The following section elaborates more on this theme.

2.3 Nuclear reprogramming of somatic cells

A differentiated cell like a neuron always remain as a neuron and never change, under normal circumstances, to say a liver cell. Because of such a remarkable stable state of a differentiated cell, its cellular fate is said to be 'sealed'. However, developmental biologists have shown that it is possible to reverse the cellular fate of a differentiated cell back to a pluripotent or totipotent state. Such a fate reversal is termed as nuclear reprogramming. Strategies like nuclear transfer, cell fusion, the use of pluripotent cell extracts, the use of defined reagents like small molecules or specific cocktail of genes have been exploited to demonstrate nuclear reprogramming. I discuss such strategies in the following sections.

2.3.1 Nuclear reprogramming by nuclear transfer

The procedure by which live animals are produced asexually by transferring a somatic nucleus into enucleated eggs is referred to as somatic cell nuclear transfer or animal cloning. The first animal cloning experiments in metazoans were demonstrated by hatching tadpoles that were obtained by transfering an early blastula nuclei into an enucleated oocyte of the the frog *Rana pipens* (Briggs and King, 1952). The success of nuclear transfer experiments revised the widely held notion among the nineteenth century embryologists that a cell undergoes a stable change in its genetic constitution with development and differentiation (Di Berardino, 2001). Moreover these experiments have estabished the capacity of egg cytoplasm to reprogram a somatic cell nuclei to a pluripotent state (Byrne et al., 2003).

Several years of ambhibian cloning experiments in *Rana* and *Xenopus* can be summarised as follows: i) Nuclei of early embryonic cells upon nuclear transfer yielded normal tadpoles, hence they are totipotent. ii) The success of nuclear transfer is inversely correlated with the age of the donor cell. Nucleus from early developmental stage cells can be succesfully reprogrammed than the late stage cells (Di Berardino et al., 2003). iii) Normal feeding stage larvae were obtained from a range of transplanting adult tissues like kidney, lung and skin. However they did not survive further. In no case was an adult animal obtained by nuclear transplantation from the cell of an adult frog (Gurdon, 2006; Gurdon and Byrne, 2003).

Mammalian reproductive cloning was first demonstrated in the sheep Dolly (Campbell et al., 1996) and subsequently in cows (Cibelli et al., 1998), pigs (Onishi et al., 2000), cats (Shin et al., 2002), goats (Baguisi et al., 1999), mice (Wakayama et al., 1998) etc. Moreover, mammalian clones have been obtained from the transfer of transgenic nuclei. For example, the sheep Polly was produced from a lamb cell that harbours human clotting factor IX (Wilmut, 1998).

Mammalian cloning is extremely inefficient, with most of the clones dying soon after implantation. In most of the embryos that survived, developmental and physiological abnormalities including premature death has been observed. Since many of these abnormalities are not inherited, it is thought that they arise due to an incomplete reprogramming of somatic nuclei in the egg (Gurdon and Byrne, 2003). Like amphibian cloning, the efficiency of mammalian cloning also depends on the differentiation status of the donor nuclei. For example cloning of mice from the nuclei of mouse embryonic blastomeres and mES cells are highly efficient than from the nuclei of adult cells like fibroblasts, cumulus etc (Cheong et al., 1993; Eggan et al., 2001; Rideout et al., 2000; Wakayama and Yanagimachi, 1999). Despite these progress with mammalian cloning, until now, no ES cell line has been established by the nuclear transfer of human cells (Hochedlinger and Jaenisch, 2006). Moreover this line of research is severely affected because of social and ethical concerns over oocytes donation (Wakayama and Yanagimachi, 1999). This has lead scientists to divert their attention towards developing alternative approaches for reprogramming human cells which avoid the use of human oocytes. Before discussing these alternative options, I would like to elaborate on what happens to a somatic nuclei transferred into egg cytoplam.

Transplanted nucleus undergoes morphological changes including nucleoli breakdown (Gonda et al., 2003), marked enlargement and chromatin dispersal (Gurdon, 2006). When transferred into an egg cytoplasm, DNA replication of the somatic nucleus is rapidly induced before any commencement of transcription. However DNA transcription rapidly starts if the nucleus is transferred into an *Xenopu*s oocyte (Gurdon and Byrne, 2003). The transcriptional programme of the transferred somatic nucleus is affected by several proteins which are dynamically exchanged between the egg microenvironment and the nucleus. First, somatic linker

histones are replaced with embryonic types from the recipient cytoplasm (Dimitrov and Wolffe, 1996). However the core-histones are not exchanged and changes in their post-translational modifications in the egg cytoplasm is not clearly understood (Wade and Kikyo, 2002). Second, proteins of the basal transcriptional machinery are exchanged. Upon incubation with *Xenopus* cytoplasmic extracts, proteins like TBP, TFIIB and TFIIF are eliminated from the nucleus into the egg cytoplasm. This loss is mediated by nucleosomal ATPase ISWI protein present in the egg cytoplasm (Kikyo et al., 2000). Third, methylation pattern of the somatic nucleus undergo a change. When a human thymocyte is transferred into *Xenopus* oocyte, the promoter of the *Oct4* gene undergoes active demethylation before its transcription (Byrne et al., 2003). This active demethylation could be mediated by *Gadd45alpha* (Barreto et al., 2007).

2.3.2 Nuclear reprogramming by cell fusion

Historically cell-cell fusions have been used to demonstrate the phenotypic dominance of one fusion partner over the other (Baron and Maniatis, 1986; Blau and Blakely, 1999). To date, various cell fusion combinations have been tried including the fusion of somatic cells with pluripotent cell types of ES (Tada et al., 2001), EG (Tada et al., 1997) and EC origin (Andrews and Goodfellow, 1980). In the resulting hybrids the phenotype of the pluripotent cell is dominant over that of the somatic counterpart (Flasza et al., 2003). The following evidences substantiate this:

(i) X-chromosome inactivation is essential to equalize the X-linked gene dosage between XX females and XY males. An inactive X chromosome of female PGCs is reactivated soon after they enter into the developing gonads in the mouse. This X chromosome reactivation is considered to be one aspect of the overall genomic reprogramming events that is linked to toti/pluripotency. Similarly an inactive X chromosome of the thymocyte was reactivated in the male ES cell x female thymocytes hybrids (Tada et al., 2001). X chromose reactivation was also seen in EC x thymocytes fusion (Takagi et al., 1983). However, the EC hybrids did not produce chimeras upon transplantation into blastocysts showing that the thymocyte had not undergone a complete nuclear reprogramming.

(ii) In *Oct4*-GFP transgenic mice, GFP is expressed in pluripotent cells but not in thymocytes. Fusion of such thymocytes with ES (non-GFP background) cells reactivated the GFP expression in hybrids (Tada et al., 2001).

(iii) Fusion between murine P19 and human T-lymphoma cell line resulted in inter-species hybrids in which expression of the human specific Oct4 and Sox2 were upregulated whereas a lymphocyte marker CD45 was downregulated (Flasza et al., 2003).

(iv) Recently it was shown that human fibroblasts could be reprogrammed by fusion with human ES cells (Cowan et al., 2005). A fibroblast cell that harbours a pluripotency reporter, *Rex-1* GFP was fused with GFP negative human ES cells. Although fibroblasts do not show any GFP expression, the hybrid cells express *Rex-1* GFP and exhibit other pluripotency features like expression of the Oct4 transcription factor, alkaline phosphatase activity, telomerase activity and surface antigens like SSEA4, TRA 1-61, TRA 1-81. Moreover when the expression profiles of fibroblasts and hybrids were compared, somatic gene expression was downregulated in the hybrids.

These results raise a number of questions: It is not clear whether the somatic donor nucleus has been fully reprogrammed and has regained the potential to sustain pluripotency in the absence of ES (or EC) cell genome. Since these are intraspecies fusion studies it is impossible to analyse if the endogenous *Oct4* gene expression is also induced along with the transgene reporter expression in the hybrids. Above all, one substantial technical barrier in using the tetraploid ES cell-somatic cell hybrids to generate patient specific stem cells for therapy is to remove the ES cell chromosomes after the somatic partner has undergone a complete nuclear reprogramming (Cowan et al., 2005).

A recent report (Matsumura et al., 2007) has addressed many of the above concerns. The authors have designed a universal chromosomal elimination casette (CEC) that has lox-P sites for the elimination of both the copies of chromosome 6 of a mouse ES cell that encodes the *Nanog* gene. An ES cell engineered with a CEC casette was fused with the fibroblasts to create a hybrid cell. Then cre mediated recombination led to the complete removal of chromosome 6 of the ES counterpart. In hybrids, the

Nanog gene expression was reactivated from the somatic counterpart. Pluripotency of cre-treated hybrid cells was also demonstrated by the formation of teratomas.

In regenerative medicine, major histocompatibility complex (MHC)-matched stem cells are eagerly awaited as a source for producing replacement tissues to minimise any immunological rejection. The majority of MHC class I and class II genes are clustered on human chromosome 6 and mouse chromosome 17. Selective elimination of ES cells-derived MHC gene–containing chromosomes from ES-somatic hybrid cells may provide a source of personal MHC-matched hybrid cells. The CEC approach can be used to generate MHC-personalized diploid ES cells by targeted elimination of ES cell chromosomes harboring the MHC genes and replacing both copies with somatic cell–derived MHC chromosomes using a special microcell-mediated chromosome transfer technique (Tomizuka et al., 1997) (Figure 6).

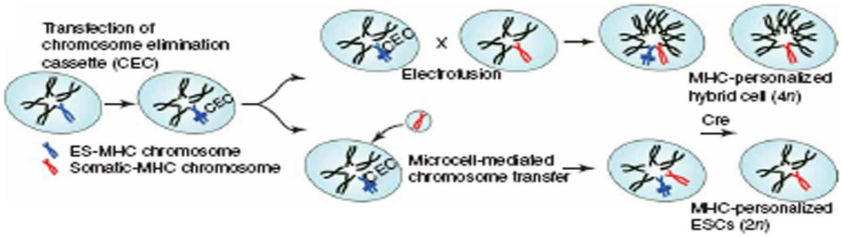


Figure 6: Scheme of generating MHC-matched stem cells.
The proposed scheme involves cell-cell fusion and chromosome elimination cassette. Adapted from (Matsumura et al., 2007). MHC refers to Major histocompatibility complex and ESCs refers to ES cells.

2.3.3 Nuclear reprogramming by cell extracts

Conventionally, Cell extracts have been widely used to investigate various nuclear processes such as DNA replication, chromosome and nuclear envelope dynamics (Martins et al., 2000; Poccia and Collas, 1996). The newfound application of cell extracts is in nuclear reprogramming. The Collas group showed for the first time that upon permeabilisation and brief incubation with T-cell extracts and neuronal cell extracts, human kidney cells HEK293T cells could be reprogrammed to induce T-cell specific genes and neuronal genes respectively (Hakelien et al., 2002). The Niehrs group developed a protocol to reprogram human cells using *Xenopus* extracts to upregulate the expression of pluripotency genes like Oct4 and GCAP (Hansis et al.,

2004). The usage of this reprogramming protocol was further extended to cells like human primary lymphocytes. In the reprogrammed lymphocytes, expression of the pluripotency markers increased. However the cells had a limited life span and did not express surface antigens characteristic of pluripotent cells indicating that reprogramming was incomplete. This study identified BRG1 (Khavari et al., 1993), a chromatin remodeling ATPase, as a necessary factor in nuclear reprogramming. To conclude, human cells were shown to be reprogrammed, although incompletely, by *Xenopus* extracts. In a conceptually related approach, pluripotent ES and EC extracts have been used to reprogram 293T cells (Taranger et al., 2005). On exposure of 293T cells to EC cell extracts: i) The promoters of pluripotent genes like *Oct4*, *Nanog* were demethylated. ii) Chromatin at the *Oct4* and *Nanog* loci underwent drastic remodeling so as to support their transcription. iii) *Oct4, Nanog* expression was drastically upregulated, as analysed by RT-PCR and immunofluorescence, with a concomitant downregulation of differentiation specific markers. iv) The capability to differentiate toward neurogenic, adipogenic, osteogenic, and endothelial lineages were enhanced.

To conclude, these studies have once again shown that cytoplasm of the pluripotent cells has the ability to reprogram a somatic cell. The major advantages with the extracts mediated reprogramming are i) They facilitate the identification of biochemical components involved in nuclear reprogramming and ii) Both the somatic cell and the reprogramming extract could be manipulated so as to obtain complete reprogramming.

2.3.4 Nuclear reprogramming by defined reagents

Several lines of evidences indicate that overexpression or loss of function of a single gene can facilitate dramatic fate changes in a somatic cell. For example, over expression of MyoD alone into fibroblasts is sufficient to induce muscle specific characters in them (Weintraub et al., 1989). In the nervous system, oligodentrocyte precursor cells could be reprogrammed into multipotential neural stem cells by sequential exposure to fetal calf serum and basic FGF (Baba et al., 2005). Ectopic expression of CCAAT/enhancer-binding protein (C/EBPα and C/EBPβ) into B cells is

sufficient to reprogram them into macrophages (Xie et al., 2004). Similarly loss of Pax5, a transcription factor necessary for B-cell development, endows the pro B-Cells with a broad developmental potential similar to that of the HSC itself (Nutt et al., 1999). Ectopic activation of β-catenin destabilizes lineage fate decisions and confers some stem cell properties on committed lymphoid or myeloid progenitors (Baba et al., 2005). In *C.elegans*, mutations in the retinoblastoma pathway components results in a soma to germline transformation (Wang et al., 2005). These findings support the hypothesis that defined factors master regulating the reprogramming of a somatic cell into a pluripotent state *may* exist and, it is possible to find them by Gain-of-function strategies. This is as well the major question of this PhD thesis.

Two different reports (Chen et al., 2004; Takahashi and Yamanaka, 2006), that have appeared during the course of this PhD thesis, contributes significantly towards delineating such defined factors. A chemical genomics screen was done to identify small molecules capable of lineage reversal (Chen et al., 2004).

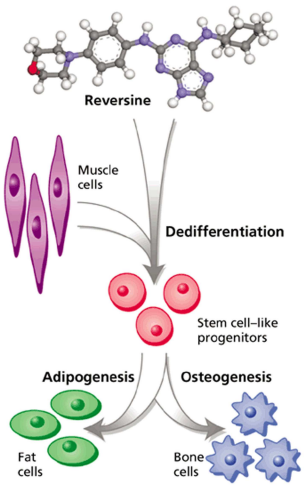


Figure7: Reprogramming by small molecules.
Reversine, a small molecule mediates the dedifferentiation of C2C12 mouse myoblasts to multipotent mesenchymal progenitor cells. These cells could then be differentiated into adipocytes or osteoblasts under appropriate differentiation conditions. This figure is reproduced from (Ding and Schultz, 2004).

This led to the discovery of myo-reversine, a 2, 6 disubstituted purine, that could dedifferentiate lineage-committed C2C12 muscle cells to become multipotent mesenchymal progenitor cells, which then can proliferate and redifferentiate into bone and fat cells (Figure 7).

In a landmark paper (Takahashi and Yamanaka, 2006), the authors have reprogrammed a differentiated mouse embryonic fibroblasts into an ES state by overexpressing a combination of four genes- *Oct4*, *Sox2*, *C-Myc* and *Klf4*. None of the candidates tested individually could reprogram the fibroblasts. Interestingly

Nanog was dispensible. The reprogrammed cells resembled but were not identical to ES cells by microarray expression analysis. Furthermore, like ES cells, the reprogrammed cells formed embryoid bodies in culture and produced teratomas when injected into mice.

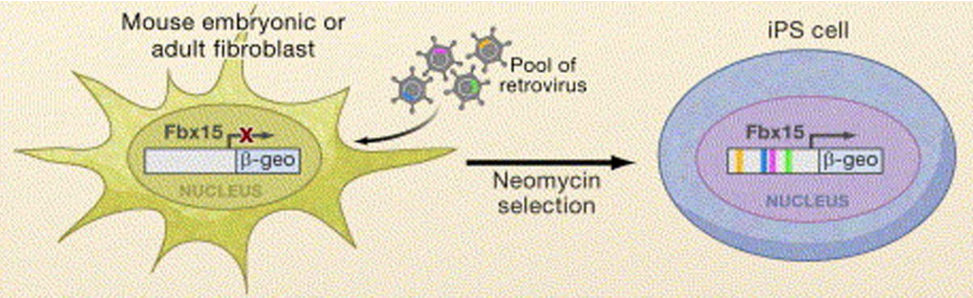


Figure 8: Reprogramming by defined factors.
Embryonic and adult mouse fibroblasts expressing a selectable marker (β-geo) driven by an ES cell-specific promoter (*Fbx15*) were transduced with retroviruses encoding *Oct4*, *Sox2*, *C-Myc* and *Klf4*. These factors reprogram the fibroblasts into an induced pluripotent (iPS) cells and have many properties similar to an ES cell. The reprogrammed cells were scored by their beta galactosidase activity and G418 resistance. This Figure is modified from (Rodolfa and Eggan, 2006).

The induced cells also differentiated to produce all three embryonic germ layers, confirming their pluripotency. However upon injection into blastocysts, the reprogrammed cells contributed to mouse chimeras only upto E13.5 and no live chimeric pup could be obtained. The *Oct4* promoter was only partially demethylated. In summary, it appears that the cells were not fully reprogrammed. Moreover it is not yet known if these four factors are sufficient to reprogram human cells. Although practical application of this approach requires more work, this seminal paper has authoritatively illustrated that nuclear reprogramming by defined factors is a reality.

To summarise this section, I have discussed the molecular determinants of pluripotency. One crucial parameter among them is the epigenetic status of a cell. Epigenetics is a broad term to describe the covalent modifications of DNA especially methylation and demethylation of its bases and histone modifications that bring about a change in the conformation of the chromatin so as to regulate the expression of a gene. Cells do not express all the genes present in their genome at a time. In fact, each cell type has its unique gene expression profile and hence its unique epigenetic signature. For example, epigenetic profile of a pluripotent cell is different from that of

a differentiated cell. The following chapter discusses epigenetic aspects focusing on DNA methylation during animal development and differentiation.

2.4 Epigenetics and cell differentiation

The course of development of organism from its pluripotent state to differentiated state involves changes in gene expression. Embryo specific genes are shut down and tissue specific genes are turned on. This change in gene expression profile depends on the methylation status of the promoter of the gene in question. Along with other standard bases, vertebrate genomes contain 5-methyl cytosine which is restricted to Cytosine followed by Guanosine (CpG) dinucleotide. In the genome of vertebrates, approximately 80% of all CpG-dinucleotides are subject to methylation. Exceptions to this rule are so-called "CpG-islands'', regions of DNA with high density of CpGs (Antequera and Bird, 1993). CpG methylation is implicated in a variety of processes, including genome organisation and stability (Niwa et al., 2000), transposon silencing, X chromosome inactivation, genomic imprinting, tissue specific gene expression and disease conditions like cancer (Jaenisch and Bird, 2003).

2.4.1 CpG methylation and gene silencing

CpG methylation is usually inversely correlated with gene expression. For example, genes, when transfected into cells, are not expresssed if their promoters have been methylated *in vitro* (Razin and Cedar, 1991). Moreover, endogenous genes could be activated by treatment of cells with a DNA demethylating drug, Aza (Mayer et al., 2000). Several models are available to explain the methylation mediated transcriptional silencing:

i) DNA methylation in the cognate transcription factor binding site could prevent the transcription factor from binding to the DNA by steric hindrance (Watt and Molloy, 1988).

ii) Attempts to digest methylated and unmethylated regions in the genome using endonucleases such as DNaseI clearly demonstrate that the former are relatively refractory to digestion, indicating a closed chromatin structure (Tamame et al., 1988). When methylated and unmethylated variants of the Herpes simplex virus thymidine

kinase (*TK*) gene were introduced into rodent cells in order to monitor *TK* activity, a time dependent repression of methylated construct was observed (Buschhausen et al., 1987; Graessmann and Graessmann, 1988). These observations indicate that the methylated construct assembles a repressive chromatin structure, in which basal-RNA polymerase II transcription machinery cannot access the promoter sequences, leading to transcriptional shutdown. Over the years, how such a repression is brought about has been understood. In short, CpG methylation sequence at the promoter region causes the binding of methylated CpG-binding proteins (MBPs) that are characterised by their methyl binding domains (MBD). These proteins, generally, then recruit Histone deacetylase complexes (HDAC) which removes the lysine acetylation of the core histones thereby shutting down the transcription (Baker and El-Osta, 2003).

iii) The methylation of mammalian genomic DNA is catalyzed by DNA methyl transferases (DNMTs). In addition to this conventional role, DNMTs have been shown to biochemically interact with histone methyl transferases and histone deacetylases (Fuks et al., 2000) thereby forming a repressive chromatin structure.

iv) In mammalian genome many CpGs are found in the intronic regions of a gene. In order to study the effect of methylation of these intragenic CpGs on transcription, a patch methylated GFP reporter (promoter is unmethylated whereas reporter is methylated) was targeted to a specific genomic loci. This study found that the methylated reporter expression was low when compared to unmethylated controls. The methylated region exclusively adopted a closed chromatin structure that depleted Polymerase II occupancy (Lorincz et al., 2004).

2.4.2 Mediators of DNA methylation and demethylation

The DNA methyl transferases (DNMTs) can be divided into *de novo* and maintenance methyl transferases (Fang et al., 2006). *De novo* DNMTs like Dnmt3a, Dnmt3b catalyses the methylation of Cytosine to 5-methylCytosine post-replicatively in an unmethylated DNA template. Maintenance DNMT, Dnmt1 essentially copies the pre-existing methylation pattern by preferentially mediating cytosine methylation of hemimethylated DNA strand during replication. Apart from these, Dnmt2 and Dnmt3L have been described. Dnmt2 has a very weak methyl transferase activity *in*

vitro and Dnmt3L has no catalytic activity by itself but physically interacts with Dnmt3a and Dnmt3b to regulate their function (Klose and Bird, 2006).

In contrast to the rich knowledge about the enzymes mediating DNA methylation, the key players of DNA demethylation are not yet fully described. The DNA demethylation could be global or local (site specific) in scale. For example, a wave of global DNA demethylation is observed during gametogenesis in mammals. Site specific demethylation is observed during tissue specific gene expression (Kress et al., 2001). For example, the promoter of the muscle specific transcription factor Myogenin is methylated in myoblasts. Upon differentiation into myocytes, the *Myogenin* promoter is demethylated concomitant with its expression (Lucarelli et al., 2001). The DNA demethylation could be occur by passive or active process, or a combination of both (Jaenisch and Bird, 2003). Passive DNA demethylation occurs by inhibition or lack of maintenance DNA methyltransferases throughout cycles of replication, whereas active DNA demethylation, being replication independent, requires specific enzymatic reactions. Such active DNA demethylation is observed *in vivo* during germline establishment and embryogenesis (see below). Despite several attempts, until now, no DNA demethylase enzyme has been isolated in mammals. In order to account for the observed active DNA demethylation, three mechanisms have been proposed (Figure 9):

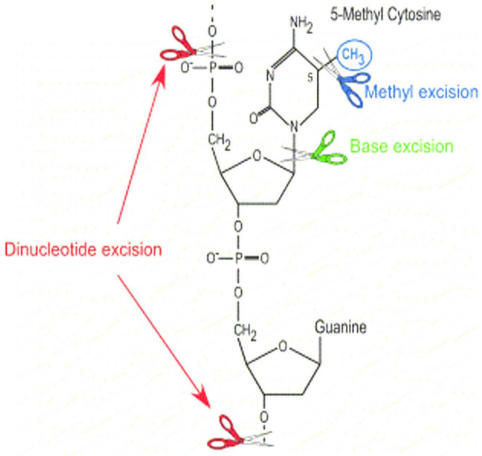


Figure 9: Possible mechanisms for DNA demethylation. Three methods via a) Direct excision of methyl group b) Base excision and c) Nucleotide excision.
This Figure is obtained from (Kress et al., 2001).

i) Direct removal of the methyl group: This involves a thermodynamically unfavorable breakage of the carbon–carbon bond that links the pyrimidine to its methyl group. The human MBD2 was reported to demethylate DNA by this

mechanism (Bhattacharya, 2004; Ramchandani et al., 1999). This claim was contested and could not be reproduced in other laboratories (Kress et al., 2001).

ii) The methylated base could be removed by base excision repair mechanism: For example, In *Arabidopsis,* DNA demethylation is mediated by DEMETER, ROS1 proteins. These are bifunctional enzyme with DNA glycosylase and DNA lyase activity. The DNA glycosylase initiates the base excision repair process by specifically excising 5-methylcytosine through cleavage of the *N*-glycosylic bond. Lyase activity of the enzyme subsequently nicks the DNA. Later other components of the base excision repair mechanism are recruited – briefly an AP endonuclease generates a 3-hydroxyl group to which a DNA repair polymerase adds an unmethylated cytosine. DNA ligase completes the repair process by sealing the nick (Gehring et al., 2006; Gong, 2006). DNA glycosylases that mediate active DNA demethylation have also been isolated from chicken extracts (Jost, 1993).

iii) The nucleotide containing the methylated base could be removed by nucleotide excision repair mechanism. Recently in this group, *Gadd45alpha* was discovered to mediate an activate DNA demethylation (Barreto et al., 2007). *Gadd45alpha* was discovered by a simple but elegant screening of *Xenopus* gastrula cDNA library for the genes that active a methylation suppressed reporter expression in HEK293T cells. Overexpression of *xGadd45alpha* resulted in a global as well as local demethylation as evident from the observed demethylation of repetitive elements and that of *Oct4* promoter respectively. Several lines of evidences presented in the paper suggest that Gadd45alpha mediates active rather than a passive demethylation.

i) Gadd45 mediated the DNA demethylation of the promoter of the injected *Oct4* gene in *Xenopus* oocytes in which no DNA replication takes place ruling out passive mechanism.

ii) Gadd45 mediated the DNA demethylation of plasmid DNA in cell culture. Since plasmid DNA as well does not replicate, the observed demethylation should be active.

iii) Gadd45 overexpression in serum starved, and hence replication arrested, cells could still induce global DNA demethylation unlike Aza that acts only on proliferating cells.

Gadd45alpha was shown to be recruited specifically to the demethylation site and mediated the observed demethylation by nucleotide excision repair mechanism. *Gadd45alpha* knockdown lead to DNA hypermethylation.Three isoforms of Gadd45, *Gadd45alpha / beta / gamma* which constitute the Gadd45 family of proteins, were earlier implicated in cell cycle regulation, cellular senescence and apoptosis. It is not clear if *Gadd45beta* and *Gadd45gamma* isoforms, as well, are implicated in demethylation. How does Gadd45 is targeted to the methylated loci is yet to be addressed.

2.4.3 DNA methylation and demethylation in development

A surge of methylation and demethylation of the genome is observed during mammalian embryogenesis as well as germline development.

Germline reprogramming: As soon as mouse primordial germ cells enter the genetic ridge, their genomes undergo a wave of active demethylation (Surani et al., 2007). However the methylation patterns are re-established by de novo methylation during early differentiation of the germ cells in the gonads. Remethylation of the male germ line seems to occur from E16 onwards whereas that of the female germline takes place only after birth during the growth of oocytes. By this cycle of erasure and regaining of methylation marks, gametes acquire sex-specific imprints. Moreover carry over of any faulty methylation imprints from adult cells into the germline is minimised (Allegrucci et al., 2005).

Reprogramming in early embryogenesis: DNA of both oocyte and sperm is heavily methylated. On fertilisation, the paternal genome undergoes genome-wide loss of methylation. Since this occurs in the absence of any DNA replication, it is referred to as active demethylation. However the maternal genome undergoes replication dependent passive demethylation. Such a genome-wide demethylation is significant as it erases all the epigenetic information originating in the highly differentiated gametes. This is essential for the establishment of a pluripotent state before specific cell lineages are determined. After implantation, a *de novo* methylation activity produces in the gastrula a methylation pattern characteristic of the adult animal. By these mechanisms, cells erase all the gamete specific epigenetic marks and acquire

embryo specific methylation marks (Reik et al., 2001). During the subsequent development, tissue-specific genes undergo specific demethylation events required for their transcriptional activation (Razin and Shemer, 1995). Several reports of tissue-specific demethylation of genes along with cellular differentiation have emerged. For example, a global demethylation is observed in differentiating murine carcinoma F9 cells and in mouse erythroleukemia (Adams et al., 1990; Bestor et al., 1984); demethylation of specific genes like *CD8alpha* and *CD8beta* during thymocytes differentiation (Fitzpatrick and Wilson, 2003), heavy and light chains of *Ig* demethylation during B cell development (Weiss and Cedar, 1997), demethylation of *IL2*, *IL4* (Fitzpatrick and Wilson, 2003), *Endothelin* (Vallender and Lahn, 2006) correlating with their expression have been reported.

Several observations have underscored the importance of DNA demethylation during muscle differentiation. 10T1/2 embryonic fibroblasts could be converted to muscles upon treatment with methyltransferase inhibitor, Aza. This study had lead to the discovery of the muscle specific transcription factor MyoD (Davis et al., 1987). Moreover, the distal control element of the MyoD promoter undergoes demethylation during somitogenesis in mice (Brunk et al., 1996) and is necessary for the MyoD expression.

A genome-wide demethylation has been observed in the mouse myoblasts induced to differentiate, which attains a maximum 2 days after the onset of differentiation (Jost et al., 2001). This global demethylation seems to be a two-step process that starts with the generation of a hemimethylated site, which in turn becomes the substrate of 5-methylcytosine DNA glycosylase. After 2 days, however, there is a partial remethylation of the genome, suggesting a broader role for this epigenetic modification during myogenesis (Jost et al., 2001). DNA remethylation is restored by the aid of a muscle specific Dnmt-1 (Aguirre-Arteta et al., 2000). This shorter isoform of the maintenance methyltransferase is not detected in myoblasts, but its expression is upregulated in differentiating myotubes. The correlation between muscle differentiation and DNA methylation is further underscored by the finding that the muscle specific transcription factor *Myogenin* promoter becomes demethylated at the onset of differentiation of the C2C12 muscle myoblasts (Lucarelli et al., 2001). Although demethylation was observed in several cases detailed above, a specific

demethylating factor was not described. It is possible that Gadd45 genes may mediate the observed demethylation.

2.5 Experimental aims and strategies

The primary aim of the thesis was to screen for the reprogramming factors that may convert a somatic cell into a pluripotent cell. I used upregulation of *Oct4*, an embryonic stem cell marker and one of the key determinants of pluripotency as the molecular readout.

Basically two broad strategies were employed: i) Expression screening of *Xenopus* egg cDNA library; ii) Chemical Genomics screen.

i) Expression screening of *Xenopus* egg cDNA library: *Xenopus* eggs are totipotent as they have the capacity to give rise to whole organism. So *Xenopus* eggs should express factors that maintain totipotency. Our own previous results shows that when differentiated HEK293T cells were incubated with *Xenopus* egg extracts, the cells aggregate into spheroids and *Oct4* expression is upregulated in these spheroids (Hansis et al., 2004). Hence I decided to screen *Xenopus* egg cDNA libraries for the factors that upregulate *Oct4* expression. Various screens were performed as will be described in detail.

ii) Chemical Genomics screen: A library of about 3000 small compounds available at DKFZ-EMBL chemical genomics core facility was used to screen a HEK293 stable cell line that harbours a luciferase reporter under the control of the Oct4 promoter and primary fibroblasts obtained from a mouse harbouring a GFP reporter under the control of the *Oct4* promoter.

As all these screens failed to identify the desired products, as a second objective of this thesis I investigated the epigenetic changes during muscle differentiation. When C2C12 myoblasts were induced to differentiate into myotubes, the promoter of a muscle specific transcription factor *Myogenin* gene is demethylated and thereby promotes its expression. However the demethylating factor was not known. Recently *Gadd45alpha* was discovered in this group to mediate active DNA demethylation by nucleotide excision repair mechanism. It is possible that *Gadd45* may mediate the

demethylation observed during the C2C12 differentiation. I addressed this question by knocking down the different isoforms of *Gadd45* and to check the consequence on the differentiation of C2C12 myoblasts into myotubes.

3 Results

The results section is divided into two main parts. Part I deals with the screening for reprogramming factors and Part II is a summary of investigation of Gadd45 as a mediator of active DNA demethylation in HCT116 Dnmt1$^{-/-}$ cells and C2C12 cells.

Part I: Screening for reprogramming factors

The main objective of this work was to screen for the nuclear reprogramming factors using Oct4 as a molecular marker. Towards this end, the following approaches were used: i) Expression screening of *Xenopus* cDNA libraries: A *Xenopus tropicalis* oocytes cDNA library and a library of cDNA prepared from the embryo stages10-30 were screened (Chapter 3.1) and ii) Chemical genomics approach in which a library of small molecules was screened (Chapter 3.2).

3.1 Expression screening of *Xenopus tropicalis* cDNA libraries

Xenopus eggs are totipotent as they have the capacity to give rise to whole organism. So *Xenopus* eggs should express factors that are essential for totipotency. Our own previous results shows that when differentiated HEK293T cells were incubated with *Xenopus* egg extracts, the cells aggregate into spheroids and the *Oct4* expression is upregulated in these spheroids (Hansis et al., 2004). Hence I decided to screen the *Xenopus tropicalis* cDNA libraries for the factors that upregulate the *Oct4* expression using two different types of readouts: the upregulation of GFP reporters that are controlled by regulatory sequences of the *Oct4* gene or upregulation of the endogenous *Oct4*.

GFP as readout

By using the upregulation of Oct4 GFP reporters as readout, the following screenings were done: i) Co-transfecting the cDNA libraries with a GFP reporter under the control of the Oct4 DE element (pDE-TK-EGFP). ii) Transfecting the libraries into Oct A cells, a HEK293 stable cell line which harbours the DE-TK-EGFP reporter. iii) Transfecting the libraries into primary fibroblasts obtained from a mouse harbouring

GFP reporter under the *Oct4* promoter and iv) Sensitising the above screens with 5–Azadeoxycytidine (Aza), a DNA demethylating agent; Trichostatin A (TSA), an inhibitor of histone deacetylases .

3.1.1 Screening by using pDE-TK-EGFP reporter

Oct4 promoter region has two important regulatory regions: Proximal Enhancer (PE) and Distal Enhancer (DE). A minimal construct containing the *Oct4* DE element was sufficient for an enhanced TK promoter driven LacZ expression in the ES cells (Yeom et al., 1996). To allow for fluorescent reporter gene readout, I subcloned the DE-TK fragment into a pEGFP vector to generate a pDE-TK-EGFP reporter construct.

Screening of Xenopus library by co-transfection of pDE-TK-EGFP reporter

HEK293T cells were cotransfected with the *Xenopus* cDNA library pools, the pDE-TK-EGFP reporter and a dsRed2 reporter. The EGFP expression was monitored every 24 hours for 3 days post transfection. The dsRed2 expression served as an internal control for the transfection efficiency. Pools which increased the EGFP reporter expression were selected for further sib-selection. Following is a list of genes that were isolated by this method. Two of them, the *growth arrest and DNA damage inducible protein alpha* (*Gadd45alpha)* and *FUSE binding protein* (*FBP*) were of potential interest.

Gadd45alpha is of interest because it was earlier identified in our lab in a separate screen designed to identify the genes that may mediate active DNA demethylation. *FBP* is interesting because of its reported expression pattern. It is expressed only in undifferentiated promonomyeloctic leukaemia cell line HL60 and the human monoblastic line U937. Upon differentiation, the expression of FBP is lost in these cells (Avigan et al., 1990; Bazar et al., 1995; Duncan et al., 1994). FBP is a sequence specific single stranded DNA binding protein. It was initially identified as a protein that bound to FUSE element present in the human *c-myc* promoter (Duncan et al., 1994). Since Gadd45 was already under intense investigation in the laboratory, I focussed on the *FBP* and performed further experiments to validate if it is indeed a regulator of the Oct4 expression.

Pool No.	Genes identified
125-4.1A	eef2-prov (elongation factor 2)
125-4.1B	actin, alpha cardiac
125-4.2D	cold-inducible RNA binding protein 2
125-4.2E	tyrosine kinase non-receptor
125-4.2G	tyrosine kinase non-receptor
126-2 1E	**growth arrest and DNA damage inducible alpha**
161-3E	MAP kinase activator XMEK3
166-3 2F	ribosomal protein L 11
166-3 2H	**FUSE binding protein-like protein**
166-3.4D	cadherin 3 type 1 preproprotein
196-3.1A	growth arrest and DNA-damage-inducible alpha
196-3.1E	gastrula zinc finger protein XLCGF57.1
196-3.12F	transmembrane 4 superfamily member 6 (Tetraspanin)
196-4.5A	ribosomal protein L10a
196-4.5D	MAP kinase activator XMEK3
196-4.11A	keratin 10,type 1, cytoskeletal

Table I: Genes isolated upon co-transfecting *Xenopus* egg library with pDE-TK-EGFP reporter.

Broadly two main approaches were taken: Overexpression of *FBP* in HEK293T cells and NIH3T3 cells to check if the endogenous *Oct4* expression is induced and overexpression of dominant negative human FBP (*dn-hFBP*) (Duncan et al., 1996) or *Engrailed–FBP* fusion constructs to study if the endogenous Oct4 expression could be downregulated in the pluripotent P19 cells.

Overexpression of the isolated *xFBP* activated the pDE-TK-EGFP reporter (Figure 10, Panel A). However *xFBP* alone or in combination with Aza and TSA could not induce the endogenous *Oct4* expression as analysed by RT-PCR in HEK293T cells (Figure 10, Panel B). It is to be noted that Aza and TSA treatment by itself could not induce the endogenous *Oct4* expression in this specific cell line.

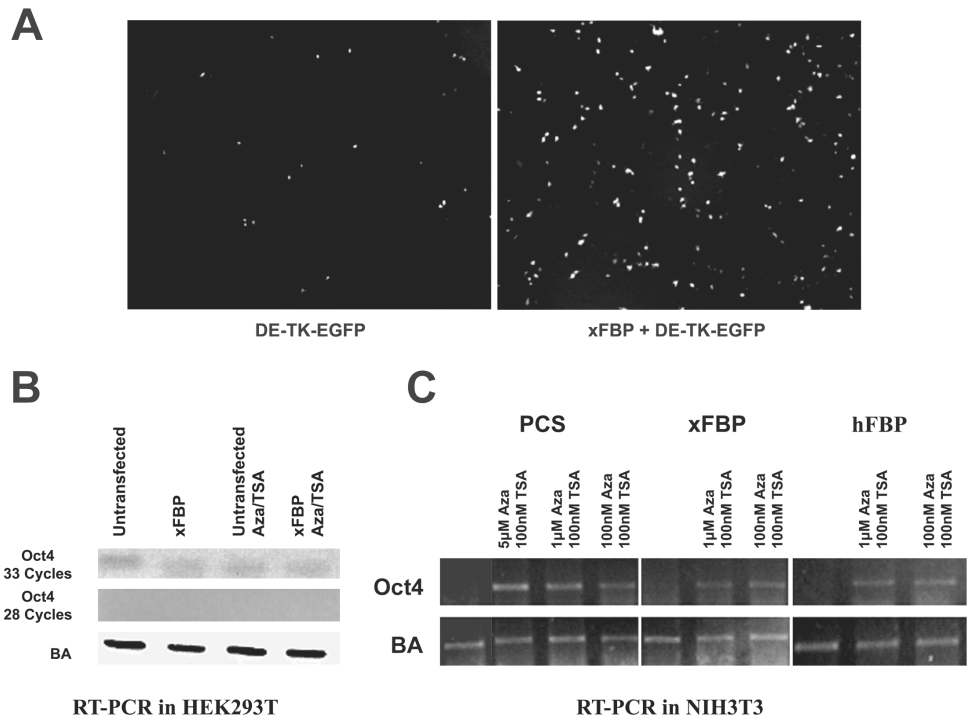


Figure 10: *FBP* activates the pDE-TK-EGFP reporter expression but does not affect the endogenous *Oct4* expression.

Panel A: HEK293T was transfected with pDE-TK-EGFP (0,3 ng) with or without FBP (20 ng-50 ng) in a 96 well plate. Total DNA transfected in each well was adjusted to 200 ng with PCS2+ plasmid and fluorescence was observed after 2 days.

Panel B: HEK293T cells were seeded in 96 well plates and transfected with the reporter and 50 ng of *xFBP* or control PCS2+. Final amount of DNA transfected in each well was adjusted to 250 ng with PCS2+. Ten hours after transfection, 200 nM Aza and 50 nM TSA were added. 24 hours later, the medium was replenished with fresh Aza. *FBP* transfection was validated by the activation of the GFP reporter. 48 hours after transfection, cells were lysed and *Oct4* expression was analysed by RT-PCR.

Panel C: NIH3T3 cells were transfected with 120 ng *xFBP* or *hFBP* along with the pDE-TK-EGFP reporter. Total amount of DNA transfected in each well was made upto 250 ng with PCS2+. *FBP* transfection was ensured by the activation of the GFP reporter (not shown). 10 hours after transfection, Aza and TSA were added. 24 hours later, the medium was replenished with fresh Aza. 48 hours after transfection, cells were lysed and *Oct4* expression was analysed by RT-PCR. BA=*Beta actin*.

An earlier report had indicated that Aza and TSA treatment could induce *Oct4* expression in NIH3T3 cells (Hattori et al., 2004). In fact, TSA and Aza strongly induced the *Oct4* expression (Figure 10, Panel C). The strength of the *Oct4* induction decreases with lower doses of Aza and this could not be compensated by *xFBP* or *hFBP* (Duncan et al., 1996) overexpression. Thus *xFBP* or *hFBP* alone or in

combination with Aza and TSA could not induce the endogenous *Oct4* expression in NIH3T3 cells.

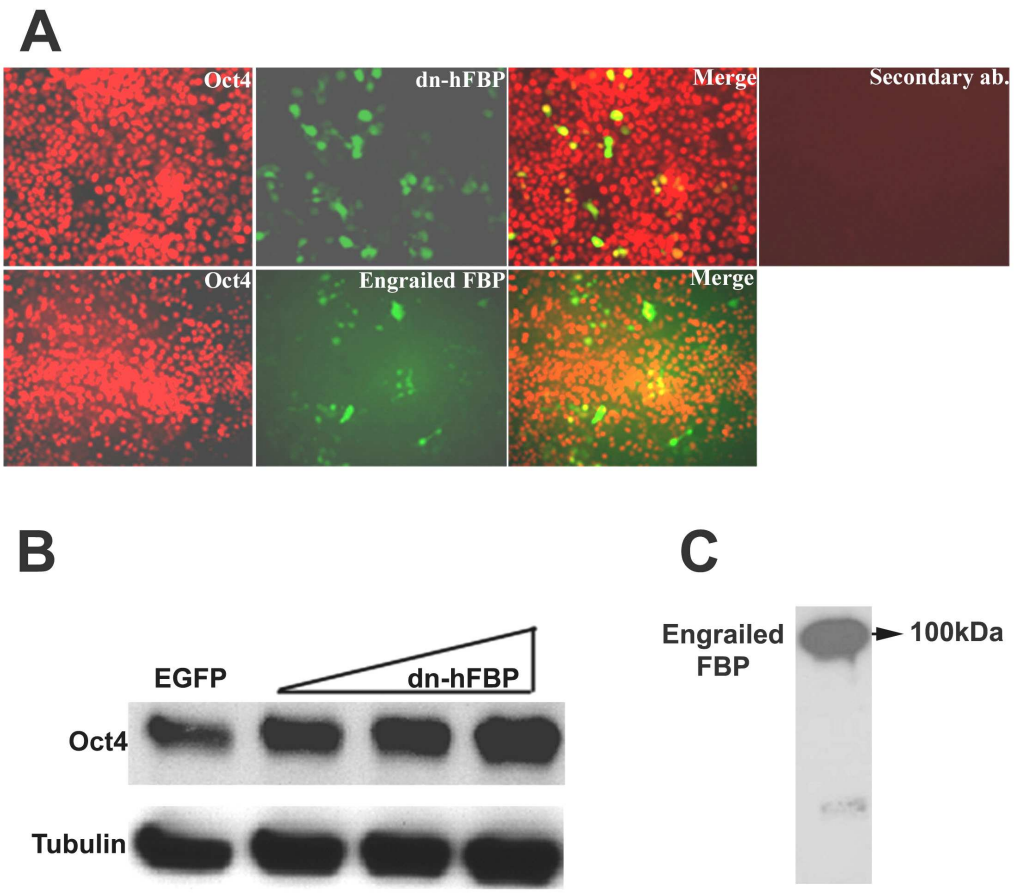


Figure 11: Dominant negative *FBP* does not reduce the Oct4 expression in P19 cells.
Panel A: P19 cells were transfected with 250 ng of dominant negative hFBP (*dn-hFBP*) or *Engrailed-FBP* and 50 ng EGFP in a 24 well plate. Cells were immunostained using anti-Oct4 primary antibody and Cy3 conjugated secondary antibody (secondary ab.). GFP expression traces the cells which are transfected.
Panel B: P19 cells were cultured in 24 well plates and transfected with EGFP or *dn-hFBP* FBP (100-400 ng). Two days later, the cells were lysed and were analysed by Western blot using anti-Oct4 antibody.Tubulin served as a loading control.
Panel C: Western blot analysis: Flag tagged Engrailed-FBP is expressed.

Concerning the second approach, I observed that *dn-hFBP* overexpression has no effect on the endogenous Oct4 expression in P19 cells as analysed by Western blot (Figure 11, Panel B). It is to be noted that only about 20% of the P19 cells are transfected by the method employed. Hence any decrease in the Oct4 expression due to the transfection of *dn-hFBP* could be masked by the endogenous Oct4 present in the rest 80% of the untransfected cells. To rule out this possibility, Oct4 expression

was analysed by immunostaining P19 cells transfected with a mixture of *dn-hFBP* and EGFP. Transfected cells were marked by the GFP expression. The transfected cells (green) retain Oct4 expression (red) which is evident from the resultant yellow colour upon merging GFP and Cy3 images (Figure 11, Panel A top).

As an alternative Loss-of-function approach, an Engrailed-FBP protein in which *hFBP* is fused to engrailed repressor domain (1-297) was used (Kessler, 1997; Smith and Jaynes, 1996). Like *dn-hFBP, Engrailed-FBP* also did not downregulate the endogenous Oct4 expression in P19 cells (Figure 11, Panel A bottom). Since overexpression of FBP did not induce the endogenous Oct4 expression in HEK293T and NIH3T3 cells and Loss-of -function approaches failed to repress the Oct4 expression in P19 cells. Based on these results, I conclude that it is unlikely that FBP could play a role in the regulation of the Oct4 expression.

To summarise, screening of the *Xenopus* library by co-transfection of the pDE-TK-EGFP reporter resulted in the isolation of false positive genes. To minimise the false positives, the next approach was to generate HEK293 cells stably harbouring the pDE-TK-EGFP construct and use this cell line for screening.

Screening of Xenopus library in HEK293/ pDE-TK-EGFP reporter cell line

In order to generate a stable pDE-TK-EGFP reporter cell line, HEK293 cells were transfected with the pDE-TK-EGFP reporter and clones were then selected by G418 resistance. Thirteen clonal pDE-TK-EGFP reporter cell lines, labelled Oct A to M, were generated. The next question was to select a suitable clone from among these thirteen cell lines in which transgene EGFP expression mimics the endogenous Oct4 expression. Towards this end, the following approaches were taken:

i) pDE-TK-EGFP cells were subjected to a reprogramming protocol (Hansis et al., 2004) using *Xenopus* egg extract treatment and then GFP as well as the endogenous *Oct4* upregulation were monitored. GFP expression was not seen in spheres and clustures obtained from any of the cell lines tested (data not shown).

ii) Aza and TSA treatment can reactivate the expression of silenced genes (Cameron et al., 1999). So I treated all the clonal cell lines with a range of doses of Aza and

TSA to select the clones in which the transgene DE-TK-EGFP could be activated along with concomitant increase in the endogenous Oct4 expression. Out of thirteen DE-TK-EGFP clonal cells tested, GFP expression could be reactivated in two cell lines namely Oct A and Oct D (Figure 12, Panel A). I analysed the expression of endogenous *Oct4* as well as the other pluripotency markers *Rex-1* and *GCAP* by RT-PCR. In comparison to the mock treated cells, Aza and TSA treatment did not upregulate *Oct4* in either of the two cell lines (Figure 12, Panel B). Other tested markers were differentially affected. *Rex-1* expression was induced strongly in Oct A cells and mildly in Oct D cells. Aza treatment alone, but not TSA or Aza and TSA combination induced the *GCAP* expression in both Oct A and Oct D cells (Figure 12, Panel B).

Since the endogenous *Oct4* was not upregulated upon Aza and TSA treatment in both the cell lines tested, it is possible to use either of them to screen the *Xenopus* cDNA library to isolate the genes that activate the transgene EGFP reporter expression. However the main caveat is any pool that could activate the EGFP expression should be further tested for its ability to upregulate the endogenous Oct4 expression. Thus Oct A cells were transfected with the *Xenopus* oocytes cDNA library pools (pool size of 300 clones). In order to sensitise the screening conditions, cells were resuspended in a medium containing 10 nM TSA 48 hours after transfection. Two pools were observed to upregulate the DE-TK-EGFP expression. Upon sib selection, 2 independent clones, *Cystatin* and *Hypothetical protein similar to human FLJ20729,* were isolated. I further checked if co-transfection of the *Cystatin* and the *FLJ20729* could result in a synergistic increase of the DE-TK-EGFP expression. However this was not the case. (Figure 13, Panel A).

The following experiments were done to validate the isolated clones. i) Transfection of *Cystatin* and *FLJ20729* singly or their cotransfection alone as well as in combination with TSA treatment failed to upregulate the endogenous *Oct4* in Oct A cells (Figure13, Panel C) or in HEK293T cells (data not shown). ii) Transfection of *Cystatin* and *FLJ20729* did not upregulate the transgene EGFP expression in Oct D cells (Figure 13, Panel B). PCS2+LacZ served as a control for transfection efficiency in Oct D cells. iii) Cotransfection of *Cystatin* or *FLJ20729* with the pDE-TK-EGFP reporter did not significantly upregulate the GFP expression in HEK293T cells (data

not shown). The results indicate that both the isolated cDNA clones were false positives.

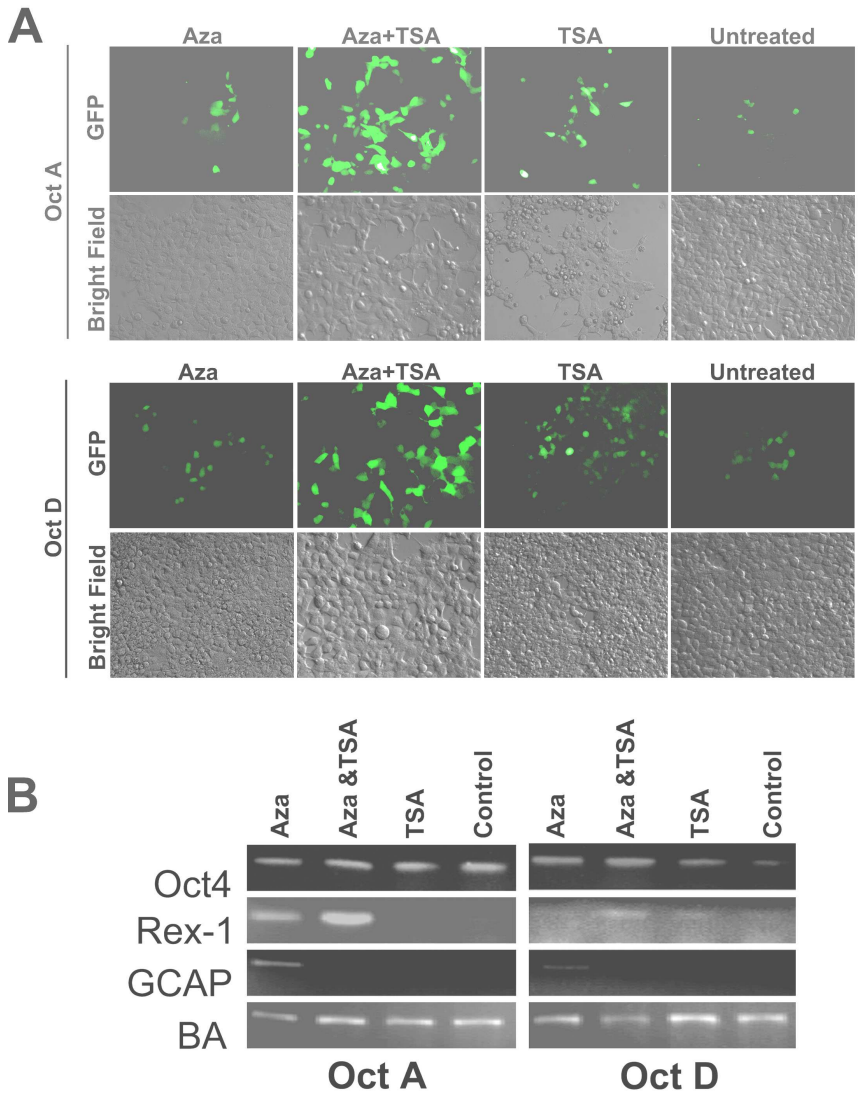


Figure 12: 5'-Azadeoxycytidine and Trichostatin A treatment reactivates pDE-TK-EGFP expression in Oct A and Oct D cells.

Panel A:

Two pDE-TK-EGFP clonal cell lines (Oct A and Oct D) were incubated in a medium containing 500 nM Aza for 4 days. Since Aza is unstable at 37˚C, fresh Aza was added every day. On the last day of incubation, 200 nM TSA was added. Aza and TSA combination significantly reactivated the EGFP expression in comparison to untreated cells.

Panel B: RT-PCR analysis in Oct A and Oct D cells treated with Aza and TSA. *Oct4* was not upregulated upon combined Aza and TSA treatment in both Oct A and Oct D cells. However *Rex-1* expression was induced. Aza treatment alone induced the pluripotency marker Germ Cell Alkaline phosphatase (*GCAP*) in both the cell lines. *Beta actin* (BA) expression was used for normalisation.

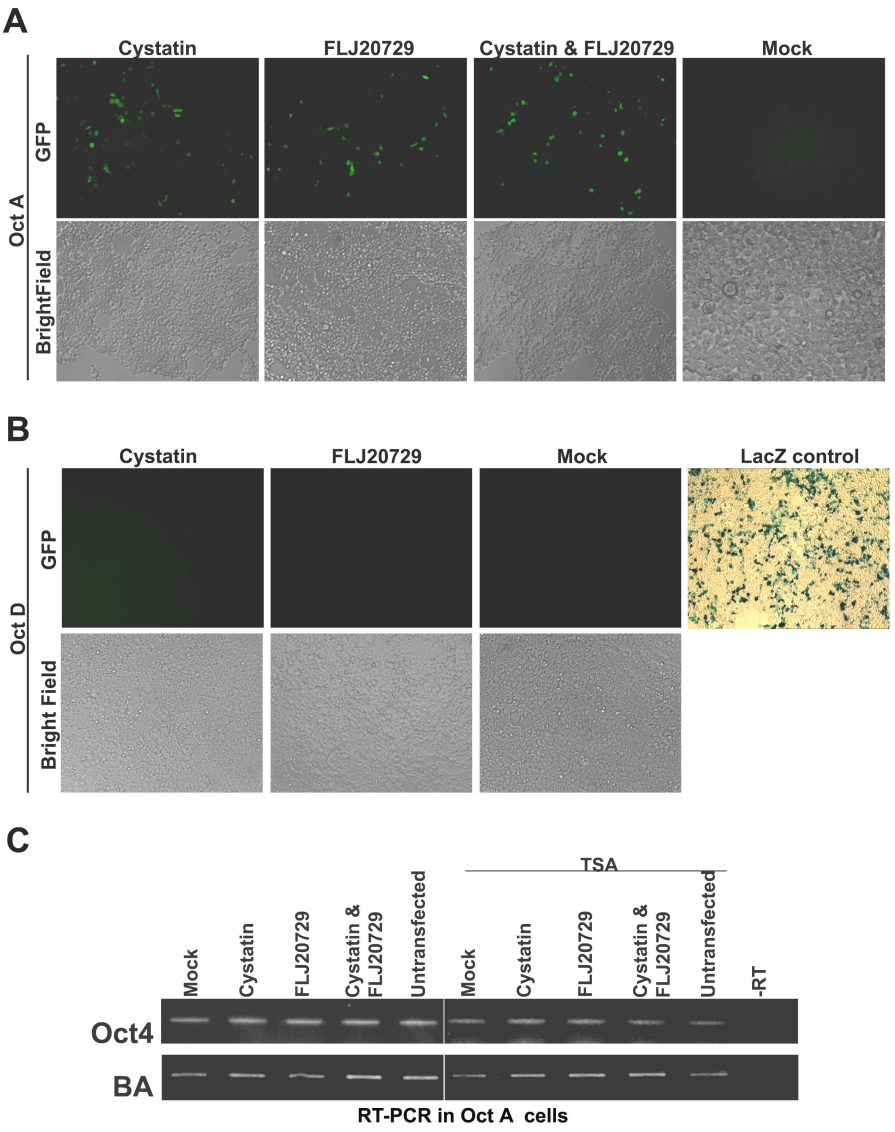


Figure 13: *Cystatin* and *FLJ20729* overexpression in Oct A and Oct D cell lines.
Panel A: Oct A cells were seeded in 24 well cell culture plates. 200 ng each of *Cystatin* or *FLJ20729* was transfected alone or in combination. Final amount of DNA transfected was adjusted to 400 ng with PCS2+ in each well. Two days after transfection, 10 nM TSA was added and incubated for further 24 hours. Overexpression of *Cystatin* as well as *FLJ20729* activated the pDE-TK-EGFP expression. Transfection of Cystatin and FLJ20729 together did not synergistically increase the reporter pDE-TK-EGFP expression.
Panel B: Oct D cells were transfected with 200 ng C*ystatin* or *FLJ20729*. Final amount of DNA transfected was adjusted to 400 ng with PCS2+ in each well. Two days after transfection, 10 nM TSA was added and incubated for further 24 hours. GFP reporter expression was not upregulated. In one of the wells, pCMV–LacZ was transfected and subsequently stained by X-Gal. This served as a transfection efficiency control for Oct D cells.
Panel C: RT-PCR analysis of *Cystatin* and *FLJ20729* transfected Oct A cells with or without TSA treatment. *Beta actin* (BA) expression was used for normalisation. *Oct4* expression was not upregulated in TSA treated as well as untreated cells.

3.1.2 Screening by using primary fibroblasts from Oct4/EGFP transgenic mice

The previous results indicated that one of the key difficulties in screening for the reprogramming factors is to obtain a reliable *Oct4* reporter cell line in which transgene expression faithfully mimicks that of the endogenous gene expression. This problem could be overcome by using primary fibroblasts isolated from a transgenic mouse, OG2, that harbours a GFP reporter inserted into an 18kb genomic fragment (GoF-18) encompassing the *Oct4* gene. In this mouse, GFP expression is restricted only to germ cells and not observed in any somatic cells (Szabo et al., 2002) thus mimicking the endogenous Oct4 expression pattern. The next approach was therefore to use the primary fibroblasts from OG2 mice for the screening of the *Xenopus* cDNA pools.

OG2 mice were a kind gift from Hans Schöler. Primary fibroblasts were isolated from the OG2 mice, at E14.5, expanded and frozen. As expected, no GFP expression was observed in these cells. The presence of the GFP transgene in fibroblasts was ensured by PCR (data not shown). A series of experiments was done to optimise the parameters to obtain better transfection efficiency. At maximum, about 40 percent of cells could be transfected.

With the optimised transfection conditions, a *Xenopus* oocytes library was transfected alone or in combination with candidate pluripotency genes like *BRG1*, *CBP, HMG-CoA*, and *Oct4* to observe for the pools that may turn on GFP expression. Overexpression of Oct4 alone in somatic cells is insufficient to induce pluripotency (Hochedlinger et al., 2005). Nevertheless I decided to cotransfect pCMV Oct4 construct with the *Xenopus* cDNA library pools expecting that this may promote a pluripotent state. The choice of the other three candidate genes for co transfection was based on the preliminary results of projects that were parallely carried out in the lab. *CBP* (Hansis, unpublished preliminary data) and *BRG1* (Hansis et al., 2004) were implicated in nuclear reprogramming of somatic cells. *HMG-CoA* was one among the initial candidates obtained from another screen carried out to isolate proteins that may play an active role in DNA demethylation (Barreto, unpublished). On this basis, the following set of screens was done.

i) Pools of *Xenopus* oocytes library were co-transfected with pDsRed2 that encodes a red fluorescent protein and served as a control for transfection efficiency.

ii) Pools of *Xenopus* oocytes were co-transfected with 3 ng each of *BRG1*, *CBP,* pDsRed2

iii) Pools of *Xenopus* oocytes were co-transfected with 3 ng each of *HMG-CoA*, *BRG1*, *CBP*, pDsRed2

iv) *Xenopus* oocytes were co-transfected with cDNA pools along with 8 ng pCMV Oct4, pDsRed2

v) Xenopus oocytes cDNA pools were co-transfected with 8 ng pCMV Oct4, 3 ng each of *HMG-CoA*, *BRG1*, *CBP*, pDsRed2.

In all the above cases 48 hours after transfection, 25 nM TSA was added to further sensitise the screening. Cells were observed by UV fluorescence for the GFP expression. However I could not isolate any *Xenopus* oocytes cDNA pool which could induce the GFP expression in OG2 cells by the above mentioned screening parameters (data not shown).

Endogenous Oct4 readout

3.1.3 Screening by using 293T cells with endogenous Oct4 readout

The pDE-TK-EGFP reporter based approaches were not fruitful to isolate the regulators of Oct4 expression. As an alternative, I wished to use the endogenous Oct4 instead of EGFP as readout in the screens. This would minimise the chances for isolating any false positive clones arising as an artefact of the reporter cell lines.

However the main disadvantage of using this method is it is laborious; moreover only those pools which strongly induce the endogenous Oct4 could be isolated whereas weakly active pools could be missed out. *Xenopus* egg cDNA library pools were transfected into HEK293T cells. Two days later, the cells were analysed by Western blotting using anti-Oct4 antibody. Lysates from the P19 cells that expressed

endogenous Oct4 were used as a control for blotting. Tubulin expression was used as loading control. One representative blot is shown in the Figure 14. None of the cDNA pools could induce any endogenous Oct4 expression as analysed by Western blot.

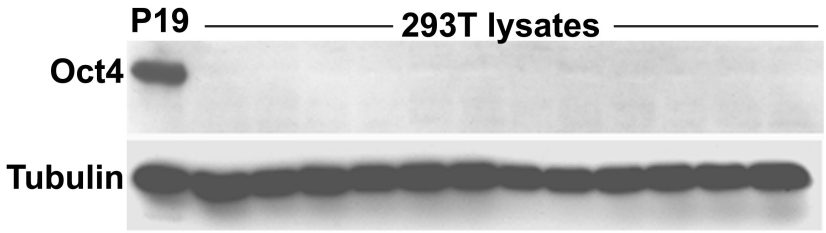


Figure 14: *Xenopus* cDNA library pools did not upregulate the endogenous Oct4 expression in HEK293T cells.
HEK293T cells were transfected with *Xenopus* st10-30 cDNA library or oocytes cDNA library. Two days later, the cells were lysed in 30 µl NOP buffer and 40% of the lysate was resolved by a 12% SDS PAGE gel, blotted to a nitrocellulose membrane and immunodetected with anti Oct4 antibody. Lysates from P19 cells (lane 1) were included in each blot as a positive control for the immunodetection. The blots were stripped and reprobed using anti tubulin antibody which served as a loading control.

3.1.4 Overview of screenings

Cells	Library used	Read out	Result
HEK293T	Oocytes and *X.tropicalis* egg library cotransfected with pDE-TK-EGFP1	Increase in GFP Signal	False positives were isolated
pDE-TK-EGFP1 stable 293 cell line	Oocytes library and TSA treatment	GFP Signal	False positives
HEK293T	*X.tropicalis* library	Endogenous Oct4 expression by Western blot	No clones were isolated
HEK293T	Oocytes library and TSA treatment	Endogenous Oct4 expression by Western blot	No clones were isolated
OG2 fibroblasts (maintained in ES Medium)	Oocytes library and TSA treatment	GFP Signal	No clones were isolated
OG2 fibroblasts (ES Medium)	Oocytes library + BRG1, CBP, TSA	GFP Signal	No clones were isolated
OG2 fibroblasts (ES Medium)	Oocytes library + BRG1,HMG CoA, CBP,TSA	GFP Signal	No clones were isolated
OG2 fibroblasts (ES Medium)	Oocytes library + pCMV Oct4	GFP Signal	No clones were isolated
OG2 fibroblasts (ES Medium)	Oocytes library + pCMV Oct4, HMG CoA, BRG1,CBP,TSA	GFP Signal	No clones were isolated

Table 2: Summary of different approaches used in the screening of *Xenopus* cDNA libraries for factors that upregulate the Oct4 expression.

The above table (Table 2) is a summary of different ways by which *Xenopus* cDNA libraries were screened for the factors that upregulate the Oct4 expression. However I did not isolate any clones that upregulated endogenous Oct4 by any of these methods.

3.2 Screening of small molecules library

Nuclear reprogramming of somatic cells has been achieved by nuclear transplantation, egg extracts treatment and cell fusion. Recently a small molecule named myo-reversine was found to dedifferentiate lineage-committed muscle cells to become multipotent mesenchymal progenitor cells, which can proliferate and redifferentiate into bone and fat cells (Chen et al., 2004). The discovery of myo-reversine implies that chemical genomics may be used as a valuable approach to dissect the complex biological process of nuclear reprogramming (Ding and Schultz, 2004).

A chemical genomics screen was set up with the help of the DKFZ-EMBL core facility. A pilot screening of a library of ~3000 small molecules was performed in OG2 fibroblasts (chapter 3.2.1) and GoF-9 Luc cells (chapter 3.2.2) so as to find out any small molecules that could induce the Oct4 expression.

3.2.1 Screening of the small molecules library using OG2 fibroblasts

Fibroblasts obtained from E14.5 embryos of Oct4 GFP transgenic mice (OG2) line were used for screening the small molecule library. Induction of GFP fluorescence is used as the readout. Prior to the pilot screen, many parameters were considered for the optimisation of screening conditions. This includes i) Optimisation of Cell Culture conditions. ii) Optimisation of different methods for the GFP readout.

Opitimisation of Cell Culture conditions

The ability of the small molecules to induce the nuclear reprogramming of OG2 cells was monitored until 5 days post introduction of the chemical library. So it was necessary to optimise the initial plating density of cells so as to ensure a linear growth during the period of investigation. As the small molecule library was dissolved in

DMSO, I tested the tolerance of OG2 cells for the presence of DMSO in the growth medium. Moreover, serum albumin in the growth medium could interfere with the permeability of the small molecules into somatic cells. Therefore it is desirable to perform the screening in a medium with a reduced serum concentration. Thus I titrated the serum concentration to determine the minimal concentration that supports the normal growth of OG2 cells.

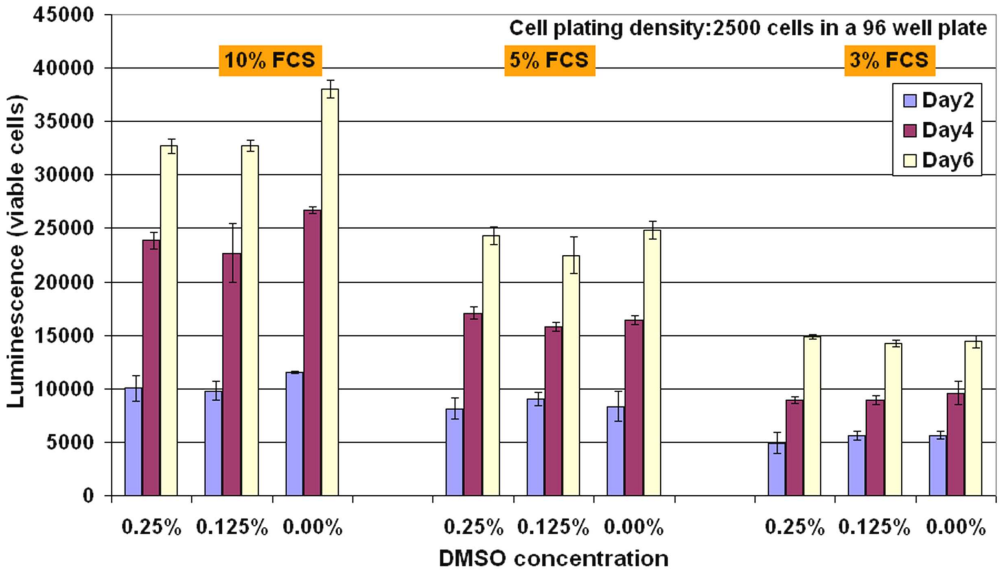


Figure 15: Optimisation of cell culture conditions for the screening of small molecules library in OG2 fibroblasts.
The parameters to be optimised were cell plating density, DMSO and FCS levels. OG2 fibroblasts were plated at different densities in 96 well plates. Plating density of 2500 cells was found to support a linear growth of cells as shown in the panel. The cells were grown in a medium consisting of varying FCS and DMSO concentration as indicated. Cell proliferation was measured at the end of 2, 4, 6 days after the initial plating by luminometry using Vialight Cell proliferation/Cytotoxicity kit (Cambrex). Luminescence measured is directly proportional to the number of viable cells in the culture dish. 10% FCS supported the best growth rate whereas 3% FCS poorly supported the cell growth. DMSO, in the concentration range tested, did not exhibit any cytotoxicity.

OG2 cells were plated at different plating densities. A range of DMSO concentration (1% to 0%, final concentration v/v) as well as serum (3%-10%, final concentration v/v) were tested that could support the linear growth for about 6 days in culture. The Growth kinetics was measured using Vialight Cell proliferation/Cytotoxicity kit. The luminescence obtained by this method is directly proportional to the healthy viable cells in the medium. From the data (Figure15),

i) DMSO had little effect on the growth kinetics of OG2 fibroblasts.

ii) Initial plating density of 2500 cells ensured a linear growth curve for ensuing 6 days of culture (data from other plating density not shown).

iii) 3% FCS retarded the growth of OG2 cells whereas 10% FCS supported a maximal growth.

Based on these data, 2500 cells were plated initially and the cells were grown in a medium containing 7.5% FCS for the screening. An empirical 3 μM of the small molecule library was added to the cells. The final DMSO concentration was 0.075% v/v.

Optimisation of different methods for the GFP read out:

The effect of small molecules on the OG2 cells was read out by the GFP fluorescence. Several options of detecting the GFP signal were considered i) by FACS; ii) by a plate reader; and iii) by microscopy. An ideal readout method must be able to measure even a very low GFP signal intensity over the background noise with a very high confidence. Any such readout method that has a very high limit of sensitivity would facilitate the identification of small molecules that may reprogram only very few OG2 fibroblasts.

In order to test and select the best option for reading a GFP signal, I made use of NIH3T3/pDE-TK- EGFP stable cells for the preliminary optimisation experiments. This cell line was generated by transfecting pDE-TK-EGFP into NIH3T3 cells and then a selection with G418. As NIH3T3 cells are differentiated, pDE-TK-EGFP was not expressed in many clones obtained. However a leaky expression of the GFP was observed in few clones which were further expanded and propagated. The resultant stable cell line was heterogeneous and exhibited a mosaic, from strong to very weak, fluorescence expression pattern.

In order to measure the limit of sensitivity of various GFP readout options, GFP expressing NIH3T3/pDE-TK-EGFP cells were serially diluted in GFP negative NIH3T3 cells. After 24 hours of incubation under standard culture conditions, GFP

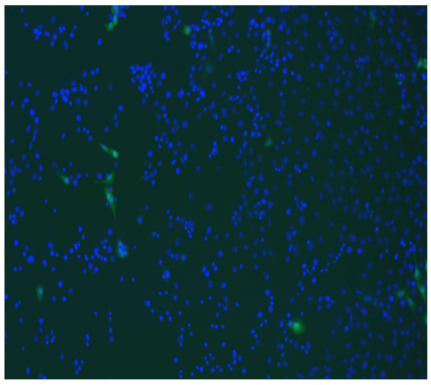
signal was read by i) FACS; ii) Using a plate reader; iii) Using CellWoRx automated image analysis system

FACS analysis was the best method. As few as 0.06 % GFP expressing cells amidst a population of non GFP negative background cells could be identified (Figure16, Panel A). Unfortunately, I could not employ FACS readout because of major technical difficulties in adapting the FACS protocol to a high throughput screening (HTS) format.

A

GFP negative cells	GFP positive cells	% GFP cells as gated by FACS
0	40000	61,0
20000	20000	29,4
30000	10000	14,0
35000	5000	6,8
37500	2500	3,7
39,000	1250	1,6
39,500	625	0,8
39,700	313	0,31
39,900	156	0,35
39,900	78	0,15
40,000	39	0,06
40000	0	0,00

B



C

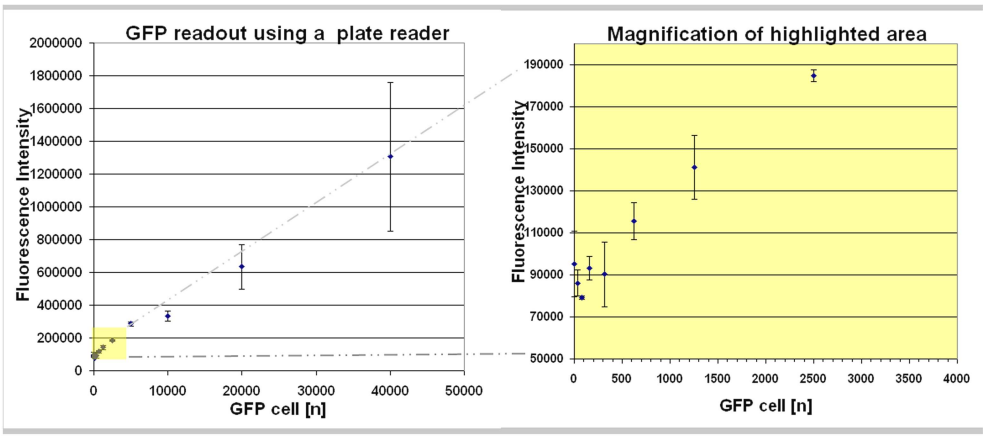


Figure 16: Limit of sensitivity of different GFP read outs.
GFP expressing NIH3T3/pDE-TK-EGFP cells were serially diluted in GFP negative NIH3T3 cells as tabulated in the Panel A. After 24 hours, GFP was read by FACS **(Panel A)** or by using a plate reader **(Panel C)**. By FACS, as little as 0.06% GFP expressing cells in a GFP negative background could be gated. However by using a plate reader, a background noise of about 100,000 fluorescent units was observed. In order to measure any significant signal with confidence over this back ground noise, a plating density of 1250 GFP expressing cells in a total of 40,000 cells (3.1%) is necessary.
Panel B shows a representative GFP image obtained using automated fluorescence microscopy. Cells were stained by Hoechst dye (blue) and then images were taken using CellWoRx imaging system. Very rare GFP expressing cells in a background of non-GFP cells could be easily spotted with a high resolution.

When the fluorescence was measured using a plate reader, there was a significant background noise of about 100000 fluorescence units (Figure16, Panel C). To measure a significant signal over this noise with a high confidence, a plating density of 1250 GFP expressing cells in a total of 40,000 cells is necessary. On the assumption that both GFP expressing and non expressing cells have similar growth kinetics it could be inferred that a small molecule that induces nuclear reprogramming in as little as 3.1 % of OG2 cells could be detected using the plate reader. One major concern is very high error rates observed in the measured fluorescence intensity among replicates. When these plates were observed under a UV microscope (data not shown), the GFP cells were observed to be clustered. It is likely that this uneven distribution of GFP cells has resulted in high error rates. To test this possibility, I lysed the cells using a lysis buffer so that the GFP molecules are more uniformly distributed in the lysate and then measured the fluorescence using the plate reader. Four different lysis buffers were tested. i) Lysis buffer (Cambrex); ii) Passive lysis buffer (Promega); iii) NaCl lysis buffer (Cruciat et al., 2006) and iv) NOP buffer. The NaCl lysis buffer helped in obtaining a maximal fluorescence signal with very minimal standard deviations (11%) compared to 35% standard deviation observed when the GFP signal was read without lysing the cells (data not shown).

With the initial optimised conditions, the small molecule screening was carried out. Briefly, OG2 fibroblasts cells were plated in a 96 well plate and 3 μM of each of the small molecule (n=3140) from the pilot library was added. After 5 days of incubation, the cells were washed with PBS and lysed using the NaCl lysis buffer. The fluorescence emission at 535 nm was read by a plate reader. However none of the small molecules tested could induce any GFP signal in the OG2 cells (data not shown).

As an alternative approach, the GFP expression was observed by fluorescence microscopy using an automated CellWoRx imaging system. OG2 cells were plated in a 384 well plates and incubated with 3 μM chemical library for 5 days. The cells were stained by Hoechst dye and then fluorescence images at 488 nm (GFP) and 350 nm (Hoechst) channels were captured. NIH3T3/pDE-TK-EGFP cells were used as an experimental control (Figure16, Panel B). Once again, none of the small molecules in the pilot library could induce a GFP signal.

3.2.2 Screening of the small molecules library using a luciferase reporter

In a HTS scenario, GFP reporter assays suffer from certain disadvantages: i) Quantitative measurements are difficult because of the background fluorescence from the cells and the reagents; ii) The lack of enzymatic amplification of GFP signal make it less sensitive and thus limiting its use to reporter constructs in which GFP is driven by a strong promoter (Naylor, 1999). These drawbacks could be avoided by using a cell line in which a luciferase reporter is driven by the *Oct4* promoter as a tool to screen the small molecule library. Towards this purpose, GoF-9 Luc construct was generated in which the Luciferase (Luc) reporter is driven by a 9kb length complete mouse *Oct4* promoter subcloned from the GoF-18 GFP (Yoshimizu et al., 1999).

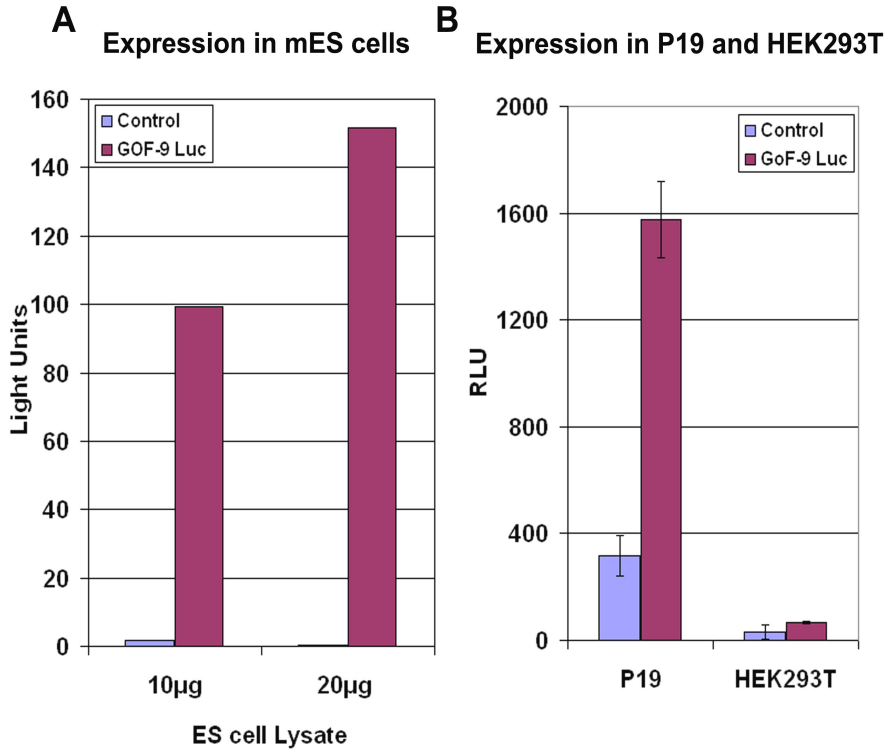


Figure 17: GoF-9 Luc is expressed in pluripotent cells but not in differentiated cells.
Panel A: Mouse ES (mES) cells were electroporated with 30 µg of linearised GoF-9 Luc or GFP Luc (Control) and were subjected to a G418 selection. The survived clones were lysed and different amounts of lysate (10 or 20 µg) were taken for quantification of the luciferase activity. GoF-9 Luc is expressed in mES cells.
Panel B: 50 ng of GoF-9 Luc or GFP Luc (Control) was contransfected with 1 ng Renilla Luc in a 96 well plate. Total amount of DNA transfected was adjusted to 100 ng with PCS2+. Fire fly luciferase reporter assays were performed 48 hours after transfection using the Dual-Luciferase Reporter Assay System and normalized after *Renilla* luciferase activity. GoF-9 Luc is active in P19 cells but not in HEK293T.

GoF-9 Luciferase was expressed in pluripotent cells such as mouse ES cells and mouse embryonic P19 cells and it had a minimal basal expression in HEK293T immortalized cells (Figure 17). Clonal HEK293/GoF-9 Luc cells were generated by transfecting HEK293 with the GoF-9 Luc construct and further selection by G418 resistance. Sixteen such clonal cells were obtained. In order to select a clonal cell line in which the luc reporter expression mimics the endogenous Oct4 expression, all HEK293/GoF-9 Luc cell line were subjected to nuclear reprogramming by *Xenopus* egg extracts (Hansis et al., 2004). After 5 days of incubation, the luciferase expression was measured (Figure 18, Panel A).

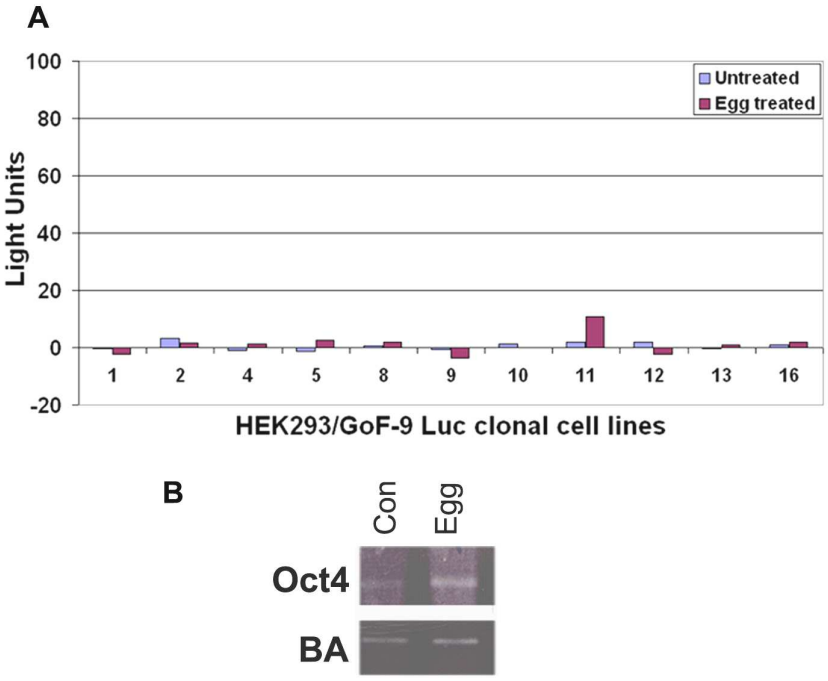


Figure 18: *Xenopus* egg extracts did not upregulate the luciferase expression in HEK293/GoF-9 Luc clonal cells.
Panel A: Each of the thirteen HEK293/GoF-9 Luc clonal cell lines were digitonin permeabilised and treated with *Xenopus* egg extracts. After 5 days of culturing, spheroids and clustures of cells were observed in eleven cell lines. They were lysed and luciferase expression was measured. Equal amounts of untreated and egg treated lysates from each cell line was analysed.
Panel B: HEK293 cells were reprogrammed in parallel to analyse the efficiency of egg extracts used. RT-PCR analysis of spheres isolated from such cells is shown here. *Oct4* is upregulated in the extract treated cells (Egg) than untreated cells (Con) and *Beta actin* (BA) is used as a loading control.

No significant upregulation of the luciferase expression was observed with any of the tested reporter cell lines. A slight upregulation of the luciferase expression was noted in clone 11. However the absolute value of luciferase expression measured was very low and fell within the range of background noise measured by the luminometer. So no specific clonal cell line could be selected by this approach. *Oct4* upregulation as analysed by RT-PCR of HEK293 cells served as a positive experimental control (Figure 18, Panel B).

As an alternative approach to select the best clonal cell line, I used reprogramming by cell fusion strategy. It was earlier reported that somatic cells could be reprogrammed into pluripotent state by a PEG mediated fusion with mouse P19 cells (Flasza et al., 2003) I intended to select a HEK293/GoF-9 Luc stable cell line in which the luciferase upregulation mimics the endogenous Oct4 expression by this approach. Towards this end, I generated P19 cells that carry a puromycin resistance cassette (referred to as P19/puro). These cells were fused with each of the sixteen HEK293/GoF-9 Luc clonal cell lines that carry a G418 resistance cassette and then subjected to simultaneous selection with G418 and puromycin for 3 weeks. Unfused parental cells died whereas fused cells alone survived. Fusion colonies obtained from six of the sixteen HEK293/GoF-9 Luc clonal cell lines analysed namely # 2, 11, 7, 13, 15, 16 alone survived the drug selection (Figure 19, Panel A). Each of these colonies was split into two and subjected to a luciferase expression analysis and RT-PCR analysis.

None of the fusion colonies from cell lines # 7,13,15,16 show any upregulation in the luciferase expression (data not shown). However many fusion colonies obtained from cell lines # 2 and 11 upregulated the GoF-9 Luc expression (Figure 19, Panel B). Three colonies obtained with the cell line 2 and eight colonies obtained with the cell line 11 showed a significant upregulation of the luc expression**.** I intended to check if the endogenous human *Oct4* is upregulated in these fusion colonies by RT-PCR (Figure 19, Panel C). Upon using SK 104/106 primer pair that specifically amplified human *Beta actin* (*BA*) ortholog for normalising the amounts of cDNAs to be used, I did not observe any specific *BA* expression. This prevented me from proceeding further to measure the expression of the human Oct4 ortholog in these fusion colonies. In contrast, *BA*

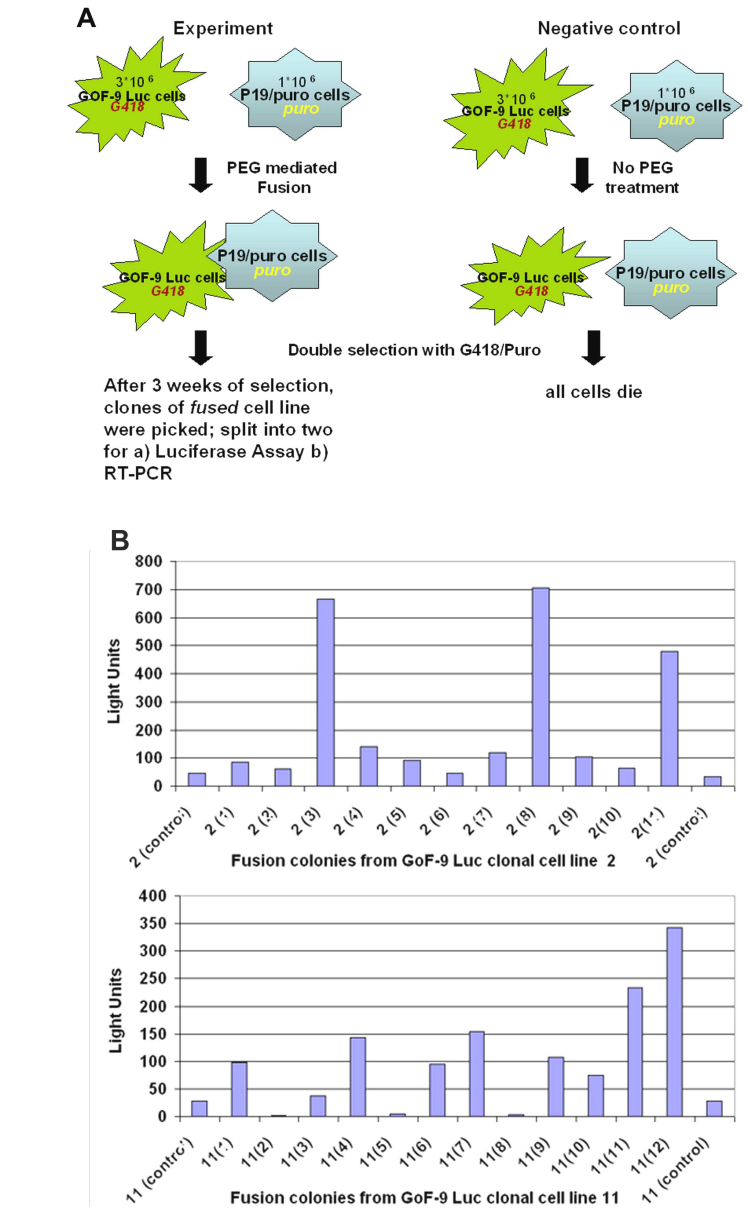


Figure 19: Poly Ethylene Glycol mediated fusion of HEK293/GoF-9 Luc cells and P19/puro cells.

Panel A: The experimental Scheme is described. Each of the HEK293/GoF-9 Luc clonal cell line (containing G418 resistance cassette) was fused with P19/puro (containing puromycin cassette) cells by using poly ethylene glycol (PEG). The fused cells were selected under combined treatment of G418 and puromycin. Only successfully fused clones survived whereas unfused control cells died. Sixteen HEK293/GoF-9 Luc clonal cell lines were fused with P19/puro cells

Panel B: Luciferase expression was significantly upregulated in fusion colonies obtained from the cell line 2 and the Cell line 11. Out of eleven fusion clones obtained with cell line 2, colonies 2(3), 2(8), 2(11) and out of twelve fusion clones obtained with the cell line 11, colonies 11(1), 11(4), 11 (6), 11(7), 11(9), 11(10), 11(11), 11(12) showed a significant upregulation of the Luc activity over that of control unfused cells. Equal amounts of protein were taken for analysis in each lane.

Panel C: RT-PCR analysis of fusion colonies. The objective was to check if endogenous human *Oct4* is upregulated in those fusion colonies in which the luc transgene expression is upregulated. As a representative case, results of clones 11(10), 11(11), 11(12) are shown here. The primer pair SK104/106 which specifically amplifies human (*Beta actin*) BA ortholog was used for normalisation. With this primer pair, BA expression was seen in control unfused cells but not in any of the fusion colonies. So I could not proceed to verify if the endogenous *Oct4* (human ortholog) is upregulated. E270/E21 primer recognises both mouse and human BA orthologs. With this primer pair, BA expression was seen in control as well as fusion colonies. –RT = Reverse Transcriptase negative samples.

expression was observed upon using a different primer pair (E270/271) that could amplify both mouse as well as human *BA* orthologs. The results indicate that in the colonies obtained by fusing mouse cell line P19 with human HEK293/GoF-9 Luc cells, mouse gene expression remains unaffected whereas expression of human genes is somehow affected. This result is in agreement with a old literature reporting the specific elimination of human chromosomes in human-mouse hybrid cells upon serial culture (Weiss and Green, 1967). When the experiments were designed in such a way that human counterpart is necessary for surviving a drug selection, it was observed that hybrid cells retained only a human chromosome or portion thereof that bears the essential gene that conferred drug resistance (Littlefield, 1964; Matsuya et al., 1968; Migeon and Miller, 1968). So it is highly likely that fusion colonies from HEK293/GoF-9 Luc and P19/puro cells retained only the human chromosome in which GoF-9 Luc cassette, and hence the G418 resistance gene, is integrated. However for want of time, I did not do any experiments to verify this possibility. I used cell line 11 for the pilot screen of the small molecule library.

A small molecule screen was done in the following manner: 5000 GoF-9 Luc cells were plated in 96 well plates. Small molecule library was added at three different concentrations: 4, 12 or 32 μM. After incubation for three days, luciferase expression was measured.

Out of 3140 small molecules tested, the majority of them had no effect (Figure 20, Panel B). Fifty eight small molecules upregulated the GoF-9 luciferase reporter (Figure 20, Panel A). However none of these small molecules dramatically increased the luciferase expression and at the maximum, a two fold upregulation of the reporter activity was observed.

Only 35 out of the 58 lead small molecules could be successfully procured for further analysis. A dose response assay was performed in order to validate the results obtained in the pilot screen. Only 8 out of those tested small molecules could reproducibly upregulate the luc reporter. However, none of them could upregulate the endogenous *Oct4* as analysed by RT-PCR (Figure 21).

To conclude this part, I could not isolate any gene or small molecule that could upregulate the Oct4 expression in a somatic cell.

A

I. Decreasing activity: Small molecule is maximally active at 4 µM

S.No	Hit ID	4 µM %	12 µM %	32 µM %
1	1525-03195	301	118	115
2	1532-07572	262	142	58
3	1542-00405	241	166	125
4	1527-02960	236	154	82
5	1503-01410	235	209	139
6	1527-34453	232	153	93
7	1547-00090	230	78	72
8	1527-30435	217	78	70
9	1532-08042	207	54	27
10	1542-01665	200	193	89
11	1532-04568	188	78	49
12	1511-00680	188	128	94
13	1556-01283	188	99	97
14	1544-00205	182	130	89
15	1542-04266	180	122	53
16	1519-01353	180	128	100
17	1545-07771	177	83	60
18	1544-00257	176	119	88
19	1526-26205	174	169	163
20	1532-07995	171	133	62
21	1536-00312	167	77	46
22	1542-01615	166	102	66

II. Increasing activity: Small molecule is maximally active at 32 µM

S.No	Hit ID	4 µM %	12 µM %	32 µM %
23	1535-01263	133	231	382
24	1511-01249	138	199	310
25	1542-03533	124	125	307
26	1527-30569	151	201	260
27	1519-02718	116	205	241
28	1547-01766	138	146	230
29	1511-00921	112	192	215
30	1520-05267	94	99	210
31	1526-25309	130	138	204
32	1518-14792	104	131	203
33	1535-03232	98	115	197
34	1518-15710	153	170	194
35	1520-03859	99	113	192
36	1530-00778	101	106	190
37	1535-00595	108	109	189
38	1545-01993	88	123	186
39	1503-07490	110	130	186
40	1511-01269	108	151	184
41	1503-07516	85	132	184
42	1520-00666	149	180	184
43	1520-03086	75	123	181

III. Other small molecules that showed one value > 200

S.No	Hit ID	4 µM %	12 µM %	32 µM %
44	1506-09770	99	222	178
45	1539-03834	90	201	173
46	1530-01561	74	204	169
47	1519-01626	110	244	162
48	1503-05161	111	223	146
49	1520-02103	104	216	146
50	1536-00588	162	249	201
51	1526-22945	160	212	201
52	1511-01778	160	236	145
53	1559-05832	129	85	235
54	1544-00290	123	241	214
55	1536-00722	132	103	209
56	1547-01333	116	88	207
57	1503-00269	79	78	204
58	1511-01160	102	241	186

B

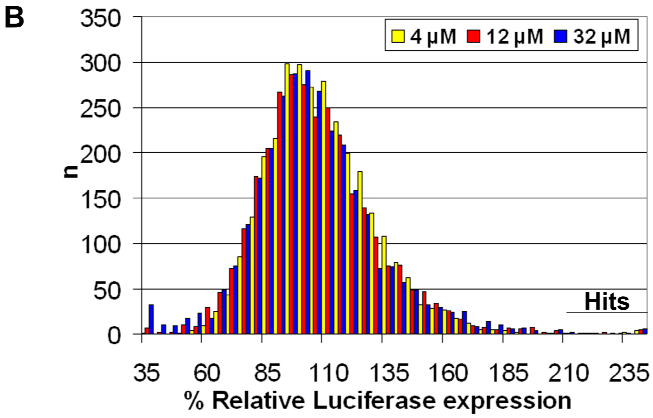


Figure 20: Small molecule screening in HEK293/ GoF-9 Luc Cells:
Panel A: About 5000 HEK293/GoF-9 Luc 11 cells were seeded in 96 well plates and incubated with 4, 12, or 32 µM of the small molecule library. A total of 3140 small molecules were screened. After 3 days of incubation, the luciferase expression was measured. The activity of the small molecule is estimated as % Relative Luciferase expression (Luciferase expression upon small molecule treatment/ Luciferase expression of control cells * 100) Fifty eight small molecules that upregulated the luciferase expression are listed. Based on their activity profile, these compounds were grouped under three headings. I. Decreasing activity: Those small molecules which are maximally active at 4 µM but toxic at higher concentrations; II. Increasing activity: Those small molecules which are active at higher concentrations; III. other small molecules which does not fit the category I or II. For easy reading, the % Relative Luciferase expression values are colour coded. Red denotes the maximum activity and white denotes the least activity. The intermediate activity profiles are denoted by yellow and saffron colours.
Panel B: Histogram of luciferase expression profile obtained from the pilot screening of small molecules library in HEK293/GoF-9 Luc 11 cells. X axis: n, number of small molecules. Y axis: % Relative Luciferase expression.

A

S.No	Hit ID	PilotScreen at 4 µM %	PilotScreen at 12 µM %	PilotScreen at 32 µM %	Validation of Hits Maximal Activity %	Validation of Hits Maximal Activity at µM
1	1518-14792	104	131	203	205	50
2	1520-00666	149	180	184	173	17
3	1525-03195	301	118	115	185	17
4	1526-25309	130	138	204	202	50
5	1527-02960	236	154	82	185	17
6	1527-30435	217	78	70	175	2
7	1527-34453	232	153	93	165	0.6
8	1536-00722	132	103	209	163	50

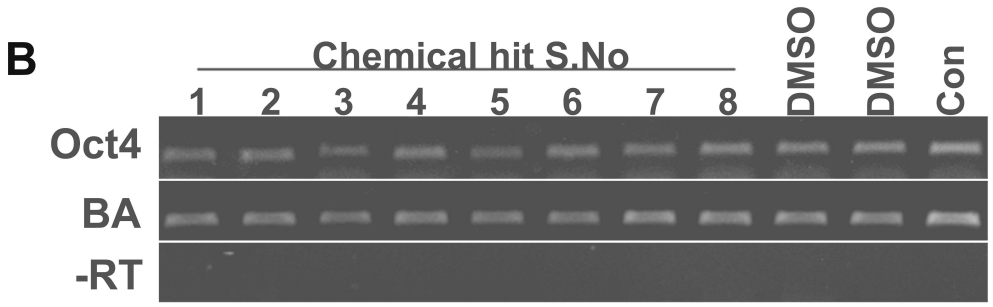
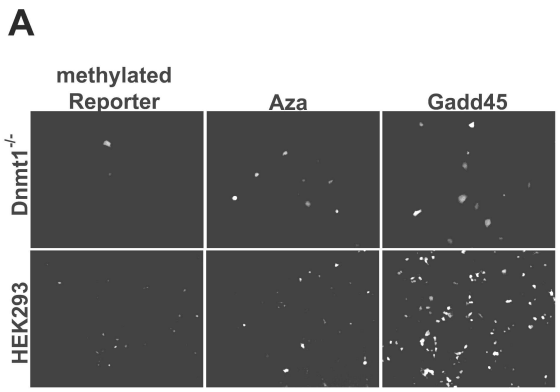


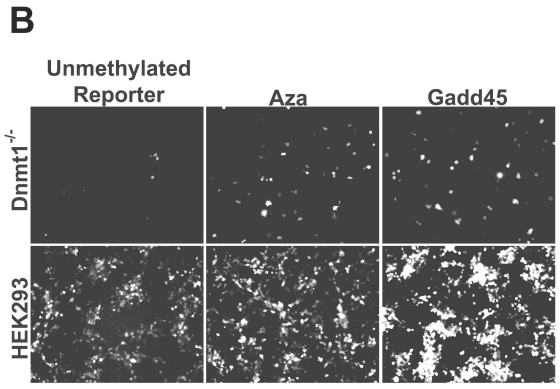
Figure 21: Validation of Hits obtained in the primary screen of small molecules library in GoF-9/Luc cells.
Panel A: HEK293/GoF-9 Luc 11 cells were incubated with the lead small molecules shortlisted from the primary screen. A serial range of dose (from 50 µM to 0,02 µM) was tested for each of the candidates. Luciferase upregulation could be validated for only 8 small molecules. % Relative Luciferase expression obtained from the pilot screen as well as that of validation experiment is tabulated.
Panel B: RT-PCR analysis for checking the endogenous *Oct4* expression in HEK293/GoF-9 Luc 11 cells after incubation with 8 of the validated small molecules (IDs of Small molecules # 1-8 as in the Panel A; Dose used: as in the Panel A under the maximal activity list). DMSO control, in duplicates and untreated cells (con) served as negative controls. None of the small molecules tested could upregulate endogenous *Oct4* expression. BA, human *Beta actin* was used for normalisation. –RT refers to the Reverse Transcriptase negative samples.

Part II: Gadd45 mediated DNA demethylation

3.3 Gadd45 as a mediator of active DNA demethylation in Dnmt1$^{-/-}$ cells

Gadd45alpha was isolated in this laboratory by screening a *Xenopus* expression library for genes capable of promoting active DNA demethylation. *Gadd45alpha* overexpression activates methylation-silenced reporter plasmids and promotes global DNA demethylation. *Gadd45alpha* knockdown leads to DNA hypermethylation (Barreto et al., 2007).





C

Experiment	Gadd45 + Reporter	Aza + Reporter	Reporter only
10 ng mEGFP	64 +/- 7	56 +/- 6	30 +/- 2
25 ng mEGFP	79 +/- 3	67 +/- 1	37 +/- 3
5ng EGFP	235 +/- 6	259 +/- 3	138+/-3

Out of approx. 7000 cells counted in each case

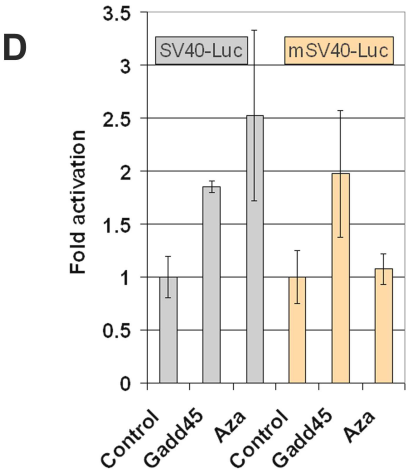


Figure 22: *Gadd45* activates reporter plasmids in Dnmt1$^{-/-}$ cells.
Transcriptional activation of methylated (**Panel A**) and unmethylated (**Panel B**) PCS2+ EGFP reporter in HEK293T cells and HCT116 Dnmt1$^{-/-}$ cells. Transient transfections were carried out in 96 well plates. A total amount of 100 ng DNA per well containing 25 ng PCS2+EGFP reporter *in vitro* methylated using the CpG DNA methylase from Spiroplasma sp. strain MQ1(M.SssI)) or 5 ng PCS2+ EGFP and 50 ng *xGadd45alpha* was transfected. Three hours later, 500 nM Aza was added to the indicated wells. 48 hours post transfection, photographs of GFP expressing cells were taken with an inverted fluorescent microscope.
Panel C: Statistics of Gadd45/Aza activation of methylated as well as non-methylated PCS2+ EGFP reporter transfected in a HCT116 Dnmt1$^{-/-}$ cells as observed by fluorescence microscopy.
Panel D: Luciferase reporter assay of HCT116 Dnmt1$^{-/-}$ cells transiently transfected with a total of 100 ng DNA containing 20 ng fire fly luciferase reporter (PGL3) plasmid, 50 ng *xGadd45alpha*, 2,5 ng *Renilla* luciferase reporter plasmid in a 96 well plate. After three hours of transfection, 500 nM Aza was added in the indicated lanes. Reporter plasmids were produced in dam- and dcm- bacteria strain SCS110 and *in vitro* methylated using HpaII- and HhaI-methylase. Fire fly luciferase reporter assays were performed 48 hours after transfection using the Dual-Luciferase Reporter Assay System and normalized after *Renilla* luciferase activity. Error bars s.e.m, (n=3)

One of the key questions I wanted to analyse was to check if *Gadd45alpha* mediates active rather than passive DNA demethylation. I overexpressed Gadd45 along with a GFP or luciferase reporter in HCT116 Dnmt1$^{-/-}$ cell line in which the maintenance

methyltransferase Dnmt1 is mutated (Rhee et al., 2000). *Gadd45alpha* transfection activated the expression of methylated as well as unmethylated pCS2+ EGFP (Figure 22). Similarly *Gadd45alpha* induced the activation of a methylated as well as an unmethylated luciferase reporter, SV40-Luc (Figure 22, Panel D).

Aza is an inhibitor of Dnmt. In a cell line in which Dnmt is mutated, I reasoned that Aza would not activate the expression of a GFP reporter or SV40-Luc reporter and included it to serve as a negative control in the above experiment. Contrary to my expectations, Aza activated the unmethylated as well as methylated PCS2+ EGFP expression in the Dnmt1$^{-/-}$ cells. Moreover, Aza activated the unmethylated SV40-Luc expression but not methylated SV-40 Luc (Figure 22).

These results in HCT116 Dnmt1$^{-/-}$ cells indicate that i) *Gadd45alpha* mediated activation is promoter- and reporter independent; ii) Gadd45 does not act by somehow affecting Dnmt1 maintenance methylation, thereby diluting methylation levels passively with each round of replication and iii) Aza could mediate demethylation via Dnmt1 independent mechanisms.

3.4 Role of Gadd45 in muscle differentiation

A Genome-wide loss of DNA methylation is usually associated with cellular differentiation. This is observed in early development of mouse embryos (Mayer et al., 2000; Oswald et al., 2000), differentiating myoblasts (Jost et al., 2001; Jost and Jost, 1994; Scarpa et al., 1996) Friend erythroleukemia cells (Adams et al., 1990; Bestor et al., 1984; Jost and Jost, 1995; Razin et al., 1988) and teratocarcinoma cells (Bestor et al., 1984; Razin et al., 1984).

Given that *Gadd45alpha* has been implicated in active DNA demethylation (Barreto et al., 2007), I wanted to check if Gadd45 mediated DNA demethylation plays a significant role in cellular differentiation by using the mouse myoblasts differentiation as a model system. When the mouse myoblasts differentiates into myotubes, a genome wide demethylation occurs (Szyf et al., 1992), expression of Myogenin, a muscle specific transcription factor, is induced (Lucarelli et al., 2001; Scarpa et al., 1996). Demethylation of the 5'-flanking region of the promoter of the *Myogenin* gene precedes its expression (Lucarelli et al., 2001).

The Specific aims of this work are to establish the C2C12 muscle differentiation assay and to study the effect of siRNA mediated knockdown of *Gadd45* on the promoter methylation status and expression of the *Myogenin* gene.

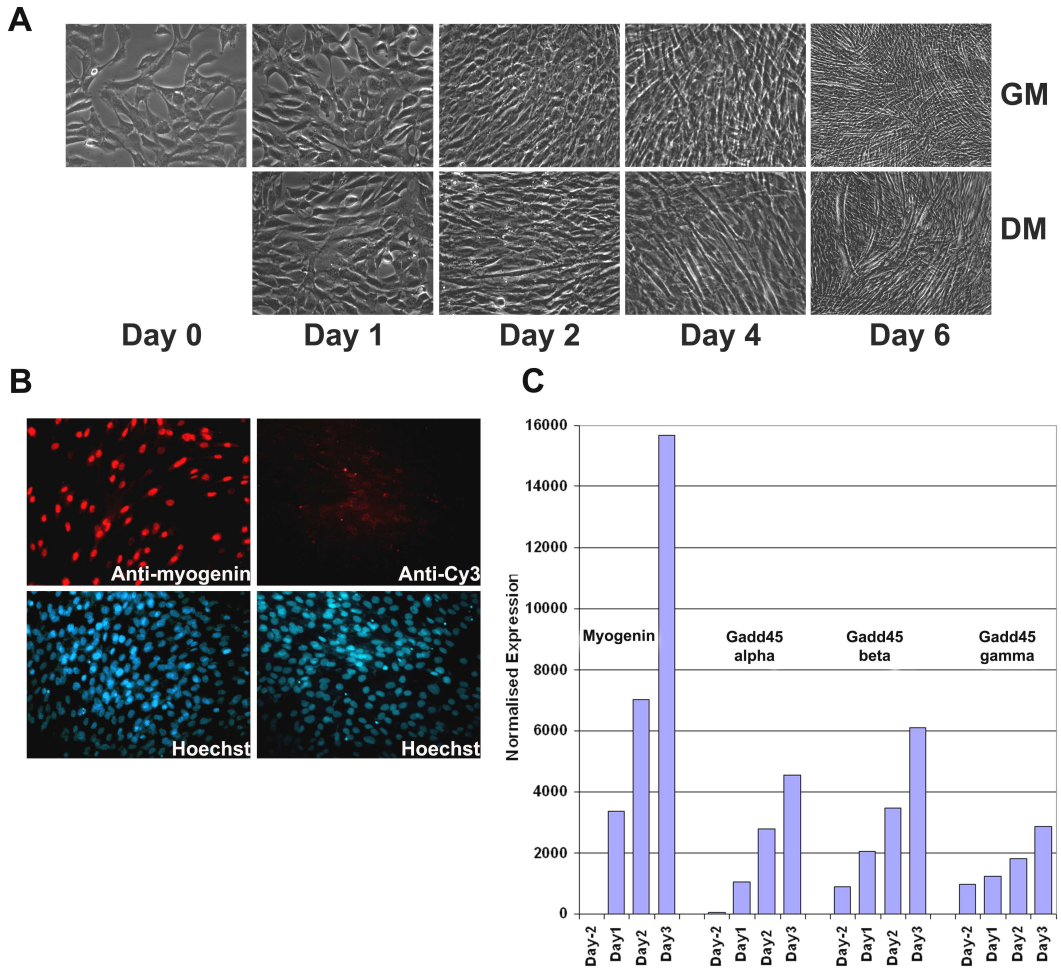


Figure 23: *Myogenin* is induced upon differentiation of C2C12 myoblasts.
Panel A: C2C12 myoblasts were plated in a 6 well plate in a growth medium (GM) containing DMEM and 10% FCS (Day -2). Two days later (Day 0), GM is replaced with differentiation medium (DM) consisting of DMEM and 2% Horse Serum. By Day 2, myotubes start appearing in the DM treated cells. They can be distinguished by their long tapering morphology. By Day 4, more C2C12 cells differentiate into myotubes.
Panel B: Two days after the induction of differentiation, C2C12 cells were immunostained using anti-myogenin antibody and then with Cy3 conjugated secondary antibody. The nuclei were Hoechst stained.
Panel C: Gene expression analysis of differentiating C2C12 myoblasts. Cells were treated as indicated in the Panel A. RNA was isolated from cells on Days -2, 1, 2 and 3. Expression of *Myogenin*, *Gadd45alpha*, *Gadd45beta* and *Gadd45gamma* were analysed by real time RT-PCR. *Myogenin* is induced only upon addition of DM. Similarly, expressions of different Gadd45 genes are upregulated with the progression of the differentiation.

C2C12 myoblasts cells are normally grown in a growth medium (GM) consisting of complete DMEM medium supplemented with 10% FCS. Differentiation is induced by suspending the cells in a differentiation medium (DM) consisting of DMEM and 2% Horse Serum (Figure 23, Panel A). The efficiency of differentiation achieved by this protocol is about 30% as analysed from immunostaining of Myogenin (Figure 23, Panel B). From Real time PCR analysis, *Myogenin* expression is observed to be induced and expression of all the three isoforms of *Gadd45* is upregulated during the differentiation of C2C12 cells (Figure 23, Panel C).

To analyse if Gadd45 is necessary for the differentiation of C2C12 cells, I knocked down *Gadd45 alpha, beta, gamma* using specific siRNAs against them. RNA and genomic DNA were harvested to study the gene expression and methylation status of the *Myogenin* promoter respectively. Knockdown efficiency of *Gadd45gamma* is poor (data not shown). *Gadd45beta* knockdown is efficient as observed from a 50% reduction in its transcript levels (Figure 24). With *Gadd45beta* knockdown, *Myogenin* transcript levels are significantly reduced. Expression of other isoforms of *Gadd45* is not perturbed. With *Gadd45alpha* siRNA, a humble 30% reduction in its transcript levels was observed. It has no significant effect on *Myogenin* expression.

Using genomic DNA extracted from the above samples, methylation status of the *Myogenin* promoter was analysed by a HpaII/Methylation sensitive multiplex PCR (HpaII/MS multiplex PCR) assay in which methyl sensitive enzyme (HpaII) restriction is coupled to a PCR amplification. The assay yields an amplified product only if the DNA fragments that is to be amplified by specific pairs of primers fails to be cut by a restriction endonuclease. The analysis of the amplified products obtained from the HpaII-treated samples allows the determination of the methylation status of the single CCGG site in the 5'-flanking region of the *Myogenin* promoter. If a product is obtained, then the template is methylated. If there is no product, CCGG site in question is demethylated. Figure 25 shows the results of methylation patterns of the *Myogenin* gene as analysed by an HpaII/MS multiplex PCR.

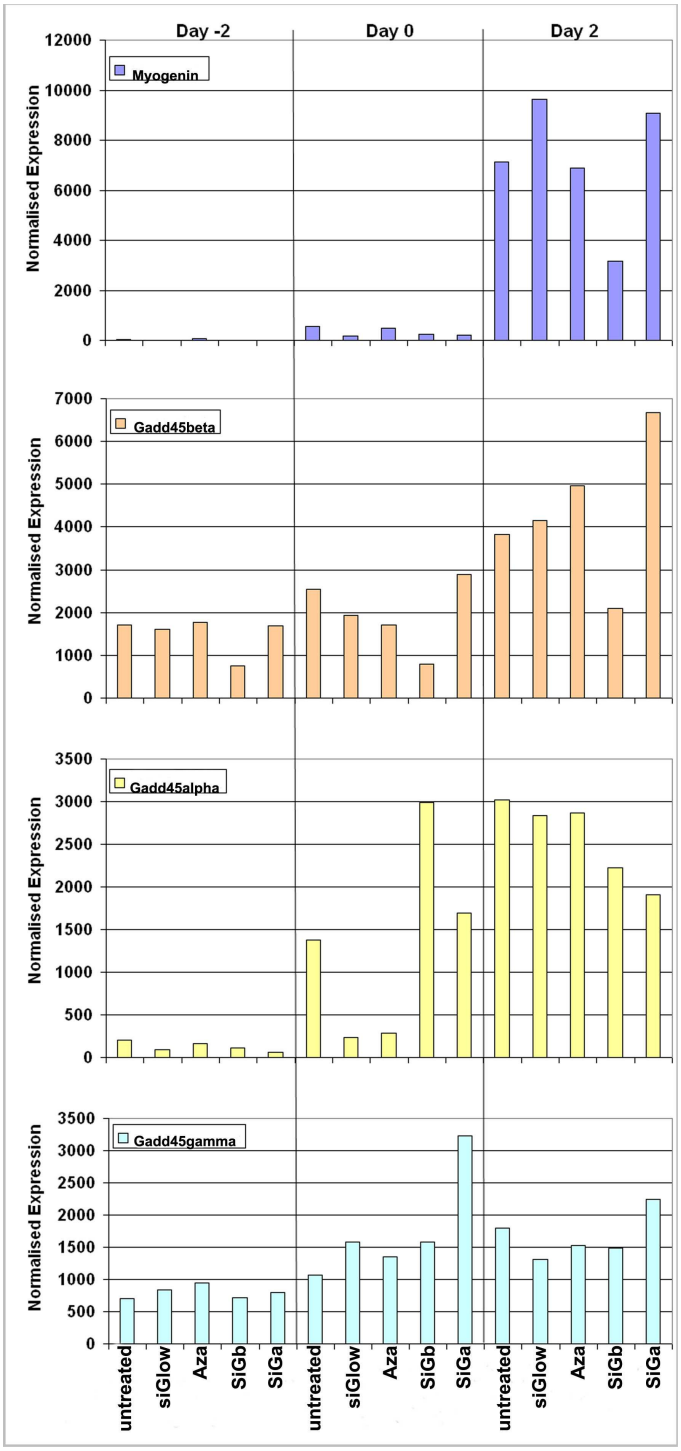


Figure 24: siRNA mediated knockdown of *Gadd45beta* decreases the *Myogenin* expression.
C2C12 cells were treated as in the Figure 23 except that they were transfected twice, on Days -2 and -1, with siRNA smart pools against *Gadd 45alpha* (siGa), *Gadd45beta* (siGb) or control (siGlow). 5 μM of fresh Aza was added to specific wells on each day. DM was added on day 0. Cells were harvested on days -1, 0, 2. Expression of *Myogenin*, *Gadd45alpha*, *Gadd45beta*, and *Gadd45gamma* were analysed by real time RT-PCR and normalised to the levels of *Beta actin* expression. Note that i) *Myogenin* expression was significantly upregulated with the addition of DM. Aza treatment did not affect the *Myogenin* expression.
ii) *Gadd45beta* knockdown downregulated the *Myogenin* expression.
iii) *Gadd45alpha* knockdown efficiency was poor and *Myogenin* expression did not change.

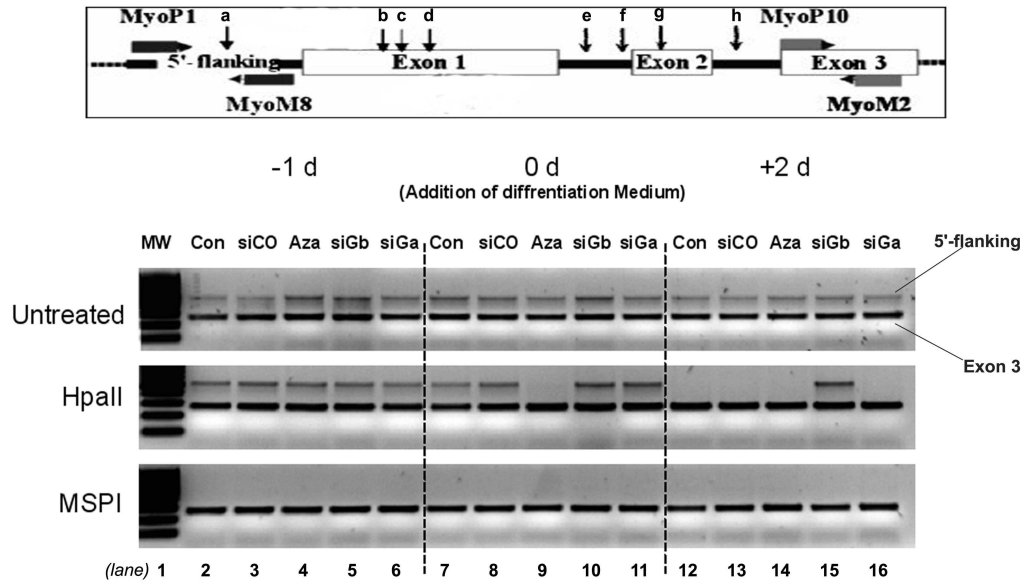


Figure 25: *Gadd45beta* knockdown blocks the demethylation of 5'-flanking region of the *Myogenin* gene as analysed by MS PCR.
(top) schematic drawing of the *Myogenin* gene showing the relative positions of annealing of various primers used in the study. Arrows (a-h) indicate HpaII (CCGG) site.
C2C12 cells were treated as in the Figure 23 except that they were transfected twice, on Days -2 and -1, with siRNA smart pools against *Gadd 45alpha* (siGa), *Gadd45beta* (siGb) or control siGlow (siCO). As an additional control, untransfected control (Con) cells were also included. 5 µM of fresh Aza was added to specific wells on each day that served as a positive control. DM was added on Day 0. Cells were harvested on Days -1(d-1), 0 (0d), 2 (+2d) and genomic DNA was isolated. The undigested DNA (Untreated) and samples digested with HpaII or MspI served as a template in multiplex PCR using P1M8 primer pair that amplifies 5'-flanking region and P10M2 pair that amplifies exon 3 of the *Myogenin* gene that serves as a loading control. PCR products corresponding to sizes 499bp (5'-flanking) and 266bp (Exon 3) were seen in all the lanes in undigested DNA samples. However 2 days after addition of the DM, demethylation is complete. This is blocked by si*Gadd45beta.* MW= 100bp ladder

Uncut control DNA samples (Figure 25 Untreated, lanes 2-16) produced two bands as expected: Upper one corresponds to the 5'-flanking region of the *Myogenin* promoter. The lower band is obtained by the amplification of exon 3 of the *Myogenin* gene that serves as a loading control for DNA being analysed. The 5'-flanking region is amplified in all the HpaII digested day -1 samples (Figure 25 HpaII, lanes 2-6). This shows that the 5'-flanking CCGG site of the *Myogenin* promoter of C2C12 grown in GM is methylated. 48 hours of incubation with 5 µM Aza induces a demethylation of CCGG site as evident from the absence of PCR amplification (HpaII, lane 9). Other day 0 samples grown in GM (HpaII, lane 7, 8, 10, 11) still remain methylated. 2 days after the addition of DM, the 5'-flanking site of the *Myogenin* promoter is

demethylated (HpaII, lanes 12, 13). *Gadd45beta* knockdown prevents the demethylation of this CCGG site (HpaII, lane 15) whereas *Gadd45alpha* knockdown has no effect (HpaII, lane 16). When HpaII is replaced by its methylation insensitive isoschizomer MspI digestion and then coupled with multiplex PCR, the formation of the upper band corresponding to the 5'-flanking region is inhibited (MspI, lanes 2-16).

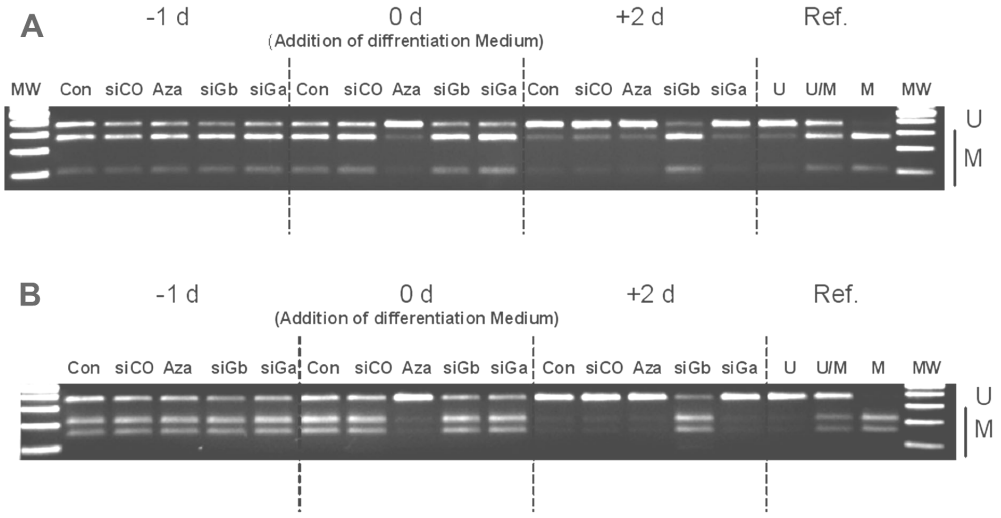


Figure 26: Gadd45beta knockdown blocks the demethylation of 5'-flanking region of the *Myogenin* gene as analysed by COBRA.
The region between nucleotides 1191- 1549 of the *Myogenin* gene is being analysed. The 359bp region is PCR amplified and cloned into a TOPO2.1 vector to obtain pMyogref construct and was then subjected to MssI invitro methylation. Unmethylated (U), fully methylated (M) and 1:1 ratio of unmethylated and methylated construct (U:M) were subjected to bisulphite conversion and then subjected to HpyCH4IV digestion (Panel A) and Hpy99I digestion (Panel B). The pattern of digestion products obtained with pMyogref served as a reference (Ref.).
Genomic DNA was obtained from C2C12 cells in the experimental set up described in Figure 25, bisulphite converted, PCR amplified and subjected to HpyCH4IV digestion (Panel A) and Hpy99I digestion (Panel B). Initially (on d-1 and d0), the investigated DNA is a mixture of methylated and unmethylated alleles. Two days after the addition of DM (+2d), a complete demethylation is observed. This demethylation is prevented by *siGadd45beta* (siGb) but not by *siGadd45alpha* (siGa). Con, untransfected control cells; siCo, siGlow transfected.

Further conformation was done by a COBRA assay (see materials and methods for details). Genomic DNA from the siGadd45 transfected C2C12 cells were bisulphite converted; nt 1191- 1549 of the *Myogenin* gene (M95800.1) which comprises 9 CpGs was amplified and then subjected to HpyCH4IV digestion (Figure 26, Panel A) and Hpy99I digestion (Figure 26, Panel B). The obtained results clearly show that Gadd45beta but not Gadd45alpha is required for the *Myogenin* promoter demethylation.

4 Discussion

Part I: Screening for reprogramming factors

The primary objective of this thesis was to identify nuclear reprogramming factors that would upregulate the pluripotency marker Oct4 expression in somatic cells. Towards this end, two broad screening strategies were employed.

4.1 Screening of the *Xenopus* cDNA libraries

Xenopus egg cDNA libraries were screened by overexpression in various cell lines to isolate the genes that would upregulate Oct4 expression The main readouts were i) GFP upregulation and ii) Endogenous Oct4 expression by Western blot.

As a first screen, *Xenopus* egg cDNA libraries were co-transfected with the Oct4 reporter plasmid pDE-TK-EGFP in HEK293T cells. Many of the isolated genes that upregulated the reporter activity were false positives (Table 1; Figures 10 and 11). It is possible that the immortalising t-antigen of HEK293T cells could bind to the SV40 Ori present in the reporter construct resulting in an episomal replication. This along with the observed background expression of the reporter plasmid may have contributed to the isolation of many false positive candidates.

The next screen was done in HEK293 stable cell line harbouring pDE-TK-EGFP reporter construct. Two clones, *Cystatin* and *FLJ20729*, were found to activate EGFP activation. However these clones did not activate the endogenous *Oct4* and failed in other validating experiments leading me to conclude that the isolated clones were false positives (Figure 13). The choice of the reporter cell line is a major concern. This cell line was selected on the basis of EGFP activation by Aza and TSA treatment (Figure 12). However no concomitant upregulation of the endogenous *Oct4* expression was observed. It is possible that in this cell line, the pDE-TK-EGFP transgene had integrated into a locus that is amenable to Aza treatment and hence the observed reporter activation is more reflective of the geography of the transgene location.

The next set of screens was done using the primary OG2 fibroblasts. *Xenopus* oocytes library was transfected alone or in combination with probable pluripotency facilitators like Oct4 (a stem cell marker), BRG1 (necessary factor for nuclear reprogramming), HMG-CoA and TSA (chromatin modifiers). I did not isolate any genes that upregulated the GFP expression.

The next approach was to move away from using GFP as readout and to monitor expression changes directly at the level of endogenous Oct4 by Western blot (Figure 14). I did not isolate any clones by this approach.

4.2 Screening of the small molecules library

The greatest advantage with this chemical genomics approach is the availability of a huge repertoire of synthetic molecules. This increases the probability of finding a candidate molecule mediating the desired cellular function.

A library of about 3000 small molecules available at the DKFZ-EMBL chemical genomics core facility was screened in OG2 cells to isolate the small molecules that may upregulate the GFP fluorescence. I did not find any such small molecule. As an alternative, the screen was repeated in HEK293/GoF-9 Luc clonal Cell line 11 to screen the small molecule library. Eight small molecules did show very modest upregulation of the luciferase activity but none of them did upregulate endogenous *Oct4* expression by RT-PCR (Figures 20 and 21). It is possible that the reporter upregulation observed with the small molecules could be cell line specific artefact arising out of the genomic loci of integration of the transgene.

One major concern with the latter screen is the choice of the reporter cell line. These cells were selected on the basis of observed upregulation of the luc reporter in several but not all colonies obtained after chemical fusion with pluripotent P19 cells and subsequent drug selection. The RT-PCR analysis indicated that the endogenous genes of human counterpart were suppressed (Figure 19). It is already reported that in mouse-human cell fusions, the human chromosomes were progressively lost (Matsuya et al., 1968; Migeon and Miller, 1968). So it is possible that in P19/HEK293 GoF-9 Luc fusions all the human chromosomes except the region conferring G418 resistance which is necessary for the survival were lost. Had I used

human N-Tera2 cells as the pluripotency partner, instead of mouse P19 cells, I could have overcome this problem. However even in that scenario, it would be difficult to distinguish between the upregulation of human Oct4 of the HEK293T cells from the background human Oct4 contributed by NTera2 cells. As an alternative, mouse somatic cells could be used to generate transgenic reporter cell lines and then could be subjected to fusion with NTera2 cells.

4.3 'SWOT' analysis of the screening for reprogramming factors

SWOT Analysis is a strategic planning tool used to evaluate the ***S***trengths, ***W***eaknesses, ***O***pportunities, and ***T***hreats involved in a project or in a business venture or in any other situation of an organization or individual requiring a decision in pursuit of an objective (http://en.wikipedia.org/wiki/SWOT). Here, I have attempted to do a SWOT on this screening for elusive reprogramming factors hoping that it would be useful for designing fresh screening strategies *tabula rasa*.

1.Strengths: As against the general tendency of avoiding high risk projects, this lab encourages to use off-the-beaten track approaches to find answers to some of the challenging problems in the developmental biology. The high quality of *Xenopus* expression library, exemplified from the discoveries made using it (Barreto et al., 2007; Davidson et al., 2005) is a definitive strength. DKFZ-EMBL chemical biology core facility nurtures any chemical genomics projects *ab initio*.

2.Weakness: The fulcrum of a screening project resides on the library being screened; the cells being used; the methodology of screen and readout being employed. As the quality of the libraries employed is assured, I would discuss other two points largely from the point of view of the results obtained with this thesis.

Cell lines: i) I principally used HEK293 cells and murine embryonic fibroblasts. The problem with both of them is they are far highly differentiated. Even *Xenopus* egg extract treatment could not completely reprogram HEK293T cells (Hansis et al., 2004). Moreover forced expression of factors like Yamanaka genes were not highly efficient in reprogramming mouse embryonic fibroblasts into pluripotent cells. Given this, extremely differentiated cells may not be very ideal for screening for

reprogramming factors. Adult progenitor cells or differentiated cell lines like C3H10T1/2 which has shown multipotential capability could be considered as suitable alternatives.

ii) Most of the screens were done in a Oct4 reporter cell lines in which reporter activity did not mimic the endogenous Oct4 expression. This has resulted in the isolation of genes or small molecules which turned out to be cell-line specific artefacts.

iii) A readout that favours single cell analysis should be used. Most of the work relied on GFP by microscopy in which such analysis was possible. However in the screen done to read endogenous Oct4, Western blot analysis was used. As immunostaining allows the gene expression studies at a single-cell level, would have been more insightful. However, it should be noted that the local concentration of protein should be significantly high to be analysed by immunostaining.

iv) One more key weakness is I relied exclusively on Oct4 as the read out. Instead, upregulation of a combination of pluripotency markers as readout could have been used.

3.Opportunities: Reprogramming factors would definitely help us in widening our academic knowledge of the mammalian developmental biology. They would be of high use in therapeutics and probably may help us usher into an era of personalised medicine.

4. Threats:

a. Threat of hypothesis: Are there any master regulators for puripotency? If they exists they should be able to convert a differentiated cell into a pluripotent cell. However no such genes have been described yet and unless identified they would remain *ipse dixit*. At least, the identification of Yamanaka genes is comforting that it is possible to reprogram somatic cells into stem cell state to an extent by a defined cocktail of genes.

b. Other Threats: This is materialistic: It is always possible that hit or miss projects like this could fail to translate into a scientific publication.

The fundamental objective of doing this SWOT is to minimise the weakness associated with the project and maximise the strengths. Following ideas may be considered in this pursuit.

i) The priority should be in developing a reliable tool for reprogramming assay rather than the screen itself. One could try to reprogram adult progenitor cells as well as multipotential cells like C3H10T1/2, C2C12 or hematopoeitic cells using *Xenopus* egg extracts; forced expression of Yamanaka genes and cell fusion with pluripotent cells. It is highly possible that these cells are more efficiently reprogrammable than 293T cells. One can investigate if the efficiency of reprogramming could be increased by subjecting the cells to more than one round of *Xenopus* egg extracts treatment.

ii) Once a best reprogrammable cell line is selected, then reporter cell lines could be generated. Effort should be to generate multi-reporter cell line. For example a cell line harbouring *Oct4-geo*/*Sox2-RFP*/*Nanog–GFP* could be generated. Upon reprogramming, such a cell line should express beta galactosidase, survive G418 selection, and upregulate GFP and RFP expression. Moreover, the endogenous Oct4, Nanog, Sox2 should be upregulated.

iii) Once the above reprogrammable tool is available, then isolating the reprogramming factors should be achievable. Several screening strategies can be considered. For example, a) Reprogramming by *Xenopus* extracts would confer G418 resistance in the reporter cell line. The unrestrictive availability of *Xenopus* eggs makes them the easily available totipotent proteome source. Reprogramming factors could be isolated by the biochemical fractionation approaches. b) Inorder to cover the complete genome, atleast 50,000 genes of the *Xenopus* expression cDNA libraries (that are available in the lab) should be screened. 50 pools of 1000 cDNA clones could be transfected alone or in sequential combination with others (pool 1 with pool 2; pool 1 with pool 3 and so on) and select those pools that confers resistance to G418 selection. Moreover in such cells GFP and RFP reporters should as well get activated. c) Another modification that could be considered is to prime the reporter cell line by reprogramming the cells with a sub-optimal dose of *Xenopus* egg extracts and then use such 'primed cells' for library screening. d) Other two peculiar features of stem cells that could be exploited are – short G1 phase in stem cells and the enormous tolerance of ES cells for a 3-fold reduction in levels of genomic m^5C. Are there any

somatic cell lines which have unusually shorter (than usual) G1 phase? If so they could be considered for reprogramming assay as well as a tool for screening. *Xenopus* cDNA library transfection could be combined with siRNA knockdown of DNA methyl transferases or knockdown of Rb components as well.

Part II: Gadd45 mediated DNA demethylation

4.4 Gadd45alpha in DNA demethylation of Dnmt1$^{-/-}$ cells

Gadd45alpha was discovered in this laboratory to have a DNA demethylation activity. Several experiments reported in this study (Barreto et al., 2007) indicate that Gadd45 acts by active demethylation. Apart from these experiments, I overexpressed Gadd45 in HCT116 Dnmt1$^{-/-}$ cell line in which the maintenance methyltransferase Dnmt1 is mutated (Figure 22). Gadd45 induced the GFP as well as luciferase reporter activation in these cells. This indicates that Gadd45 does not act by affecting Dnmt1 maintenance methylation, thereby diluting the methylation levels with each round of replication. Obviously it would have been more informative if I were able to show Gadd45 mediated demethylation in a compound knock out cells in which Dnmt1 as well as Dnmt3 have been depleted. However these cells were poorly transfectable (data not shown) making such analysis difficult.

One interesting out come of this set of experiments was obtained with Aza treatment. Reasoning that Aza should not activate a reporter construct in a methyltransferase negative background, I included it to serve as a negative control. Surprisingly Aza did activate unmethylated GFP and luciferase reporters as well as methylated GFP reporter but not methylated SV40-Luc. This observation is puzzling and probably this discrepancy could arise due to difference in the absolute levels of methylation of these two constructs. It is possible that SV40-Luc construct is heavily methylated and probably requires higher dose of Aza to relieve of its methylation. More experiments are necessary to validate this finding that Aza could mediate demethylation by Dnmt1 independent mechanisms. Indeed, Aza treatment has been shown to activate the Gadd45 expression (Schneider-Stock et al., 2005). It is possible that in Dnmt1$^{-/-}$ cells, such a Gadd45 induction by Aza could mediate reporter activation. This claim could be verified by analyzing the reporter activity after knocking down Gadd45 in Dnmt1$^{-/-}$

cells treated with Aza. However those experiments are beyond the scope of this thesis study. Recently a truncated Dnmt1 protein which retains a hemimethylase activity has been reported in HCT116 Dnmt1$^{-/-}$ cells (Egger et al., 2006; Spada et al., 2007). Dnmt1$^{-/-}$ cells being hypomorphs for Dnmt1, this may mask the real phenotype as one would expect from a complete Dnmt 1 knockout in these cells.

4.5 Role of Gadd45 in muscle differentiation

Demethylation of individual genes has been observed during cell type specific differentiation, a necessary step in the process of tissue specific transcriptional activation. I tested if Gadd45 plays a role in development using *in vitro* differentiation of C2C12 myoblasts as a model system.

All the three isoforms of Gadd45 were expressed in C2C12 myoblasts. When induced to differentiation, upon serum starvation, their expression levels were significantly upregulated (Figure 23). This result is in confirmation with an earlier publication that reported an increased *Gadd45alpha* during muscle differentiation (Cam et al., 2006). It is possible that Gadd45 is upregulated because of the serum starvation induced stress response and does not have any direct ramifications on muscle differentiation. This probability is discounted by the fact that knockdown of *Gadd45beta* reduced the Myogenin induction (Figure 24). This is because of inhibition of the demethylation of the myogenin promoter as analysed by MS-PCR and COBRA assays (Figures 25 and 26). There are 23 CpG dinucleotides present in the 5'-flanking region of the *Myogenin* promoter. Out of these, the one at nt position 1233 is a part of HpaII site hence easy to analyse its methylation status by MS-PCR. It was reported earlier this site undergoes demethylation during C2C12 differentiation (Lucarelli et al., 2001). I have found that *Gadd45beta* knockdown prevented the observed demethylation at this site (Figure 25). To analyse the methylation status of other CpGs, one needs to employ techniques like bisulphite sequencing or COBRA. However both these methods suffer from inherent disadvantages: The bisulphite sequencing method is a labour intensive and time consuming whereas the scope of COBRA analysis is limited by the availability of appropriate restriction enzymes. I was able to design COBRA assay for analysing three other CpGs (1339-CGACG-1443 by Hpy99I and

1448-ACGT-1551 by HpyCH4IV) and found that these sites as well undergo *Gadd45beta* mediated demethylation during C2C12 differentiation (Figure 26).

From the above discussion, I could conclude that *Gadd45beta* is necessary for the demethylation of the *Myogenin* promoter during C2C12 differentiation which leads to the induction of *Myogenin* expression.

Open Questions:

This observation leads to many open questions to be analysed in order to conclude the role of Gadd45 in muscle differentiation:

i) Out of siRNA reagents against three different *Gadd45* isoforms, *Gadd45beta* siRNA alone is very efficient as evident from its ability to significantly decrease the *Gadd45beta* transcript levels. However the knockdown efficiencies obtained with siRNA reagents, under the conditions used, against *Gadd45alpha* (Figure 24) and *gamma* (data not shown) are poor. It is necessary to use different siRNA concentrations as well as reagents from different suppliers to obtain a high efficient knockdown of alpha as well as gamma isoforms in order to rule in or rule out their role during C2C12 differentiation.

ii) Does Gadd45beta affect the entire myogenesis program or is it *Myogenin* promoter specific? As a result of myogenesis, muscle specific creatinine kinase (CK) activity is upregulated in the myotubes (Lucarelli et al., 2001). One could address if the *Gadd45beta* knockdown reduces such a CK activity. Earlier reports have indicated that global DNA demethylation occurs during muscle differentiation. What would happen to such a demethylation when Gadd45beta is inhibited could be analysed.

iii) Gadd45beta is present in C2C12 before the induction of differentiation, however the *Myogenin* promoter remains methylated in these cells. Probably, Gadd45beta protein is not targeted to the methylated loci at this stage. With the induction of differentiation, the targeting complex could be triggered to load the Gadd45 to an appropriate methylated locus. Two questions that need to be addressed are: a) To analyse by Chromatin IP if Gadd45beta is recruited to the methylated loci in the

Myogenin promoter; b) What factors contribute to such a differential targeting during normal growth and upon induction of differentiation could be addressed.

Mechanism:

If Gadd45 is implicated in muscle demethylation what would be the mechanism of action?

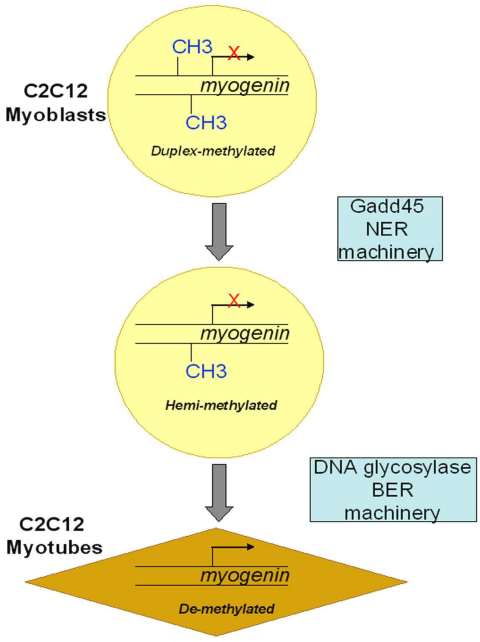


Figure 27: Model for DNA demethylation during *in vitro* muscle differentiation.
Both nucleotide excision machinery (NER) and base excision machinery (BER) may be involved in demethylation.

Gadd45beta knockdown in human RKO cells affects MLH1 promoter demethylation and hence its transcriptional activation. Moreover, *Gadd45beta* siRNA knockdown in HeLa and RKO cells induces global DNA hypermethylation. Similar results were observed with *Gadd45alpha* knockdown (Barreto et al., 2007). Hence it is highly possible that Gadd45beta mediates DNA demethylation like Gadd45alpha via nucleotide excision mechanism (NER). Moreover in G8 mouse myoblasts cells, maximal loss of DNA methylation is attained after 2 days of differentiation, followed by a gradual remethylation. During this period, other changes were also noted: DNA methyl transferase activity and DNA synthesis decreased. Moreover a number of hemimethylated DNA strands increased along with increased methylcytosine glycosylase activity (Jost et al., 2001; Jost and Jost, 1994). Taken together, it is possible that C2C12 genome demethylation may involve NER via Gadd45 as well as base excision repair (BER) via methyl cytosine glycosylase. Such a model accounts

for the rapid active demethylation observed in the *Myogenin* promoter immediately after the induction of differentiation as well as high DNA glycosylase activity observed later during differentiation in G8 myoblasts. The observed reduction in DNA methyltransferase activity may prevent the remethylation of actively demethylated DNA strands.

The proposed model could be further validated by analysing if other components of NER and BER machinery affect C2C12 differentiation.

Role of Gadd45 in development *in vivo*

What is the role of Gadd45 in development *in vivo*? In zebrafish, *Gadd45beta* is implicated in the somite segmentation (Kawahara et al., 2005). In mouse, *Gadd45beta* has been shown to be necessary for chondrogenesis (Ijiri et al., 2005). *Xenopus Gadd45gamma* is shown to be necessary for neurogenesis (de la Calle-Mustienes et al., 2002). In mammals, *Gadd45alpha*$^{-/-}$ (Hollander et al., 1999), *Gadd45beta*$^{-/-}$ (Lu et al., 2004), *Gadd45gamma*$^{-/-}$ (Hoffmeyer et al., 2001) mice grew to adulthood without showing severe phenotypic differences from their wild-type littermates. This indicates that Gadd45 deficiency has no effect on normal mouse development. The main lacuna in analysing these mouse lines is that each of them is deficient only in one Gadd45 isoform. It is possible that this deficiency is compensated by the presence of the other two isoforms so that the required Gadd45 threshold that supports a normal development could be maintained. This necessitates the analysis of development in mice in which all three isoforms have been knocked out. As another suggestion if a dominant negative Gadd45 is available, it could be used to generate conditional mouse lines so as to study the role of Gadd45 in development.

This raises another question: Is any such Gadd45 dosage compensatory mechanism observed i*n vitro* in cell culture? *Gadd45alpha* knockdown or *Gadd45beta* knockdown alone is sufficient to cause genome hypermethylation in RKO and HeLa (Barreto et al., 2007) and promoter hypermethylation in C2C12 cells (Figures 25 and 26). When *Gadd45beta* is knocked down in C2C12 cells, *Gadd45alpha* and *gamma* transcript levels were similar to that of the control transfected. These observations indicate that such compensatory mechanisms may not exist *in vitro*.

To conclude, the primary objective of this thesis to isolate genes which may reprogram somatic cells into stem cell state remained unfulfilled and the second objective to investigate the role of Gadd45 in the *in vitro* differentiation of myoblasts shows some promising trends.

5 Materials and Methods

5.1 Equipments and materials

5.1.1 Equipments

PCR thermocyclers (MJ Research), heating blocks (Eppendorf), centrifuges and microcentrifuges(Heraeus, Sorval, Beckman), fluroescent microscope (Nikon), digital cameras (Sony, Visitron Systems), spectrophotometer (Beckman), luminometer (luminoskan Ascent, labsystems), power supplies (Biorad, Consort), gel UV photo documentation (Biostar and Intas) , electroporator (BioRad), balances (Sartorius, Kern), photolaboratory (Agfa), micropipettes (Gilson), multi-channel pipettes, vortexes, agarose gel and PAGE minigel chambers, microwave oven , shakers, rotars, homogenisors, Neubauer cell counter, humidified incubators for cell culture, light Cycler 480 (Roche).

5.1.2 Materials

a) Reagents and consumables

Reagents and consumables were bought from Biochrom AG, Bio-Rad, BioZym, CalBiochem, Eppendorf, Falcon, Fluka, Fuji, Gibco, Invitrogen, Kodac, Labstar, Nunc, Merck, Sartesdt, Serva, Sigma, Roche, Roth, Whatmann and others:

Acetic Acid; Agar; Agarose; Ampicillin; Aprotinin; ATP; Blocking reagent; BSA; chloroform; Copper sulphate; Complete,mini protease inhibitor cocktail; Creatine Phosphate; Creatine Kinase; Cytochalasin B; DEPC; Digitonin; DMEM; DMSO; dNTPs; EDTA; EGTA; Ethanol; Ethidiumbromide; FCS; Fugene6; G418; Glycerol; Glycogen; HEPES; Hoechst reagent; Horse Serum; Hydrochloric acid; Isopropanol; Kanamycin; Leupeptin; LIF; Magnesium acetate;Magnesium chloride; PBS (cell culture grade); Pepstain A; Phenol; PEG 1500; Potassium acetate; Potassium chloride; Potassium hydroxide; Puromycin; rRNA; RNAguard; SDS; Sodium acetate; Sodium glutamate (cell culture grade); Sodium hydroxide; β-mercaptoethanol; Triton X-100; Trypsin-EDTA; Tween20; X-gal; X-ray film etc.

b) Enzymes and special kits

Enzymes - DNase I; Proteinase K; RNase A; Restriction enzymes; T4 DNA ligase, Taq DNA polymerase , *in vitro* methylating enzymes etc were brought from New England Biolabs, Promega, Fermantas, Invitrogen and Roche.

Special Kits – Accuprime Taq polymerase system (Invitrogen); DNeasy blood & tissue kit (Qiagen); Epitect Bisulphite kit (Qiagen); Expand High fidelity PCR system (Roche); Qiaprep spin mini-prep kit (Qiagen); Qiaprep midi -prep kit (Qiagen); Qiaquick PCR purification kit (Qiagen); TOPO cloning kit (Invitrogen), ViaLight cell proliferation/cytotoxicity kit (Cambrex) etc.

c) Cell lines

HEK293T; HEK293; NIH3T3; P19; C2C12; HCT116 Dnmt1$^{-/-}$. Apart from these cells, following reporter cell lines were generated and used in this study (Table III).

Name	Parental Cell line	Transgene	Comments
HEK293/GoF-9 Luc (clone11)	HEK293	GoF-9 Luc	Clonal
Oct A and Oct D	HEK293	pDE-TK-EGFP	Clonal
NIH3T3/pDE-TK-EGFP	NIH3T3	pDE-TK-EGFP	Non-clonal and selected for GFP expression.
OG2 fibroblasts obtained from OG2 mice (Szabo et al., 2002).	Fibroblasts	GoF-18 GFP	Primary cells. No GFP expression is seen in these cells.
P19/pur	P19	pPur	Non-clonal

Table III: Reporter cell lines used in this study.

d) Antibodies

Mouse anti tubulin cl.512 (Sigma); Oct4 antibody SC-5279 (Santa Cruz); Mouse anti myogenin monoclonal antibody, clone F5D (Abcam); Anti mouse Cy3 (Dianova)

e) Libraries

i) ***Xenopus* oocytes cDNA expression library:** This is a complementary DNA library prepared with *Xenopus tropicalis* oocytes mRNA.

ii) X**tSt10-30 cDNA expression library:** This is a *Xenopus tropicalis* complementary DNA library prepared with mRNA from embryos at stage 10, 20 and 30. Individual colonies (100,000) were arrayed in 384-well plates and, from each plate, four sets of 96 colonies were combined to prepare plasmid DNA pools. Both these *Xenopus* cDNA libraries were generated by W. Wu

iii) **Small molecule library:** This library is available at the DKFZ-EMBL chemical genomics core facility. The complete library has 55,000 compounds drawn from Tripos Leadscreen collection and form the laboratory of Prof. A. Giannis, University of Leipzig. A minimal library consisting of about 3140 compounds representing the spectrum of the complete library is used for the pilot screening.

f) Constructs

pDE-TK-EGFP: pDE-TK-LacZ (Yeom et al., 1996) was used as the parental construct. DE-TK region was amplified by a high fidelity PCR kit ;digested with Asp718/SmaI and ligated into pEGFP-1 (Clonetech).

GoF-9 Luc: GoF-18 EGFP (Yoshimizu et al., 1999) construct is digested with NotI to release 18 kb Oct4 elements including the regulatory regions, EGFP transgene and the coding sequence. This 18kb fragment is digested with Asp718 to obtain a 9kb fragment (GoF-9) that exclusively consists of *Oct4* regulatory regions which was then inserted into a modified pEGFP1 construct from which EGFP casette was removed and luciferase coding sequences from PGL3 was inserted to yield a GoF-9 Luc.

Engrailed-FBP: Engrailed was obtained from Eng-Sia construct (Kessler, 1997). FBP was PCR amplified from hFBP expression construct (Duncan et al., 1996) and then inserted in frame into a Flag-PCS2+ construct.

Other constructs that were used are:

pPUR (Clonetech)

pDSRed2-N1 (BD biosciences)
pCMV Oct4 (Tomilin et al., 2000)
pCMVmCBP (Chrivia et al., 1993)
pBS(KS+)CeBRG1 (de La Serna et al., 2000);
pCR 2.1-TOPO (Invitrogen)
SV-40 Luc and pCMV-EGFP -unmethylated as well as *in vitro* methylated constructs (Barreto et al., 2007).

g) siRNA

siRNA smart pools were designed using http://www.dharmacon.com/ sigenome/default.aspx. Smart pools of siGadd45alpha (M-042505-00), siGadd45 beta (M-040541-00), siGadd45 gamma (M-040263-00) and siGlow control (D-001600-01) were obtained from Dharmacon.

5.2 General molecular methods

Molecular biological techniques such as preparation of electrocompetent E.Coli (XL-1 blue or SCS 110 strains) cells, transformation, plasmid preparation, restriction digestion, agarose gel electrophoresis of DNA, purification of DNA from gels have been performed according to standard methods (Joseph Sambrook and Russell, 2001) or according to the instructions of the manufacturers. DNA oligonucleotides for sequencing and PCR were synthesised by Qiagen . Sequencing of DNA samples were carried out by Dr. Hunziker of the central sequencing facility of the DKFZ.

5.3 Screening of the libraries

5.3.1 Screening of the *Xenopus* oocytes cDNA library

1. Preparation of plasmid DNA pools of three hundred clones.

i) 30 µl of the initial complete oocytes library is diluted 1:1000 by dissolving in 30 ml LB containing 6 g glycerol, mixed well and aliquoted into eppendorf tubes and stored at -80˚C.

ii) The titre of the 1:1000 diluted library was estimated. Based on the titre, 1:1000 diluted library was diluted further and spread on LB Amp plates so as to obtain 300 colonies upon incubation for atleast 20 hours at 37˚C.

iii) The bacterial colonies were collected by adding 2.4 ml of LB and using a bacterial spreader the colonies were dissolved to an even suspension. 200 µl of this suspension was mixed with glycerol and was frozen. From the rest, plasmid DNA was prepared. Quality of each preparation was ensured by $A_{260/280}$ >1.8. 330 such pools were prepared.

2. Transfection: Generally, the cDNA pools were transfected into cells and observed after 48 hours. The choice of cells, modifications involving cotransfection of other constructs and reagents like Aza/TSA have been explained in the results section.

3. Sib-selection

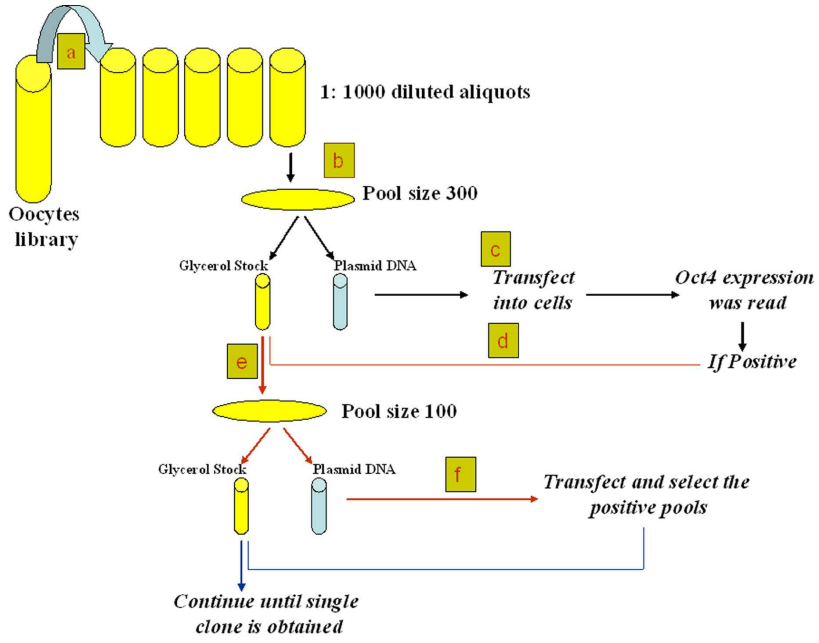


Figure 28: Overview of screening of *Xenopus* oocytes cDNA library.
The oocytes library is not arrayed. Hence subpools had to be made for each round of screening. a) The library is diluted 1:1000 and stored at -80˚C. The titre of the diluted library was estimated; b) Plasmid DNA from 300 pooled bacterial clones was prepared; c) DNA pools were transfected into cells and *Oct4* expression was read out; d) Glycerol stocks of all the positive pools were spread onto LB plates so as to get 100 bacterial clones; e) Plasmid DNA from 100 pooled bacterial clones was prepared; f) They were transfected and *Oct4* expression was read. This procedure could be repeated until a single clone is isolated or a similar strategy explained in the 5.3.2 could be used.

The pools that activated Oct4 expression were scored as positive. The glycerol stocks of these pools were diluted; their titre was calculated and spread onto LB Amp plates so as to get 100 colonies after incubation. These colonies were collected by spreading

with LB. An aliquot was mixed with glycerol and frozen. The rest was used to obtain plasmid DNA. This was then transfected into cells and once again those pools (of size 100) that activated Oct4 were selected. This procedure was repeated until a single clone was obtained.

5.3.2 Screening of the XtSt10-30 cDNA expression library

This library is made up of 11*96 well master plates and each well is a pool of plasmid DNAs from 96 bacterial clones. Glycerol stocks of these bacterial clones is also arrayed separately in a 96 well format. The library is transfected and positive pools were scored. The bacterial stock plate of the selected positive pools were replica-inoculated into two 96 well plates filled with LBAmp medium and grown at 37 ˚C for 48 hours. From one of them, the bacterias were collected by pooling along the row so that 8 such pools - row pools-(made up of 12 clones each) were obtained. In the another plate , the bacterias were pooled along the column so that 12 such pools -column pools-(made up of 8 clones each) were obtained. Plasmid DNAs were prepared from these 20 pools and transfected into cells. One would expect the clone of interest to be present in one of the 8 row pools as well as in one of the 12 column pools. The individual clone is spotted by the co-ordinates of the row-pool and that of the colum pool that activates the Oct4 expression after transfection into cells.

5.3.3 Screening of the small molecules library

The optimisation and methodology is explained in the results section.

5.4 Reprogramming by using *Xenopus laevis* egg extracts

5.4.1 Preparation of *Xenopus laevis* egg extracts

Eggs were obtained by priming females with 600 IU human chorionic gonadotropin and were dejellied with 2% cysteine HCl. Eggs were washed two times with ice-cold extraction buffer (50 mM HEPES/KOH [pH 7.4], 50 mM KCl, 5 mM $MgCl_2$, 2 mM β-mercaptoethanol, 5 mM EGTA, second wash including 10 µg/ml each cytochalasin B [Sigma, Munich, Germany], leupeptin, aprotinin, pepstatin A [all Calbiochem

Biosciences, La Jolla, CA]). Eggs or embryos were centrifuged at 1000 rpm for 1 min and transferred to 2 ml Eppendorf tubes after excessive buffer was carefully removed. Eggs and embryos were crushed by centrifugation at 10,000 × g for 15 min at 4°C. The middle layer was collected and recentrifuged at 16,000 × g to clear the extract. The extract was substituted with 2% glycerol, and aliquots were snap frozen in liquid nitrogen and stored at −80°C.

5.4.2 Reprogramming experiment

293T cells (1.5×10^6 per reprogramming experiment) were harvested by flushing with PBS and transferred to 15 ml Falcon tubes. Cells were pelleted and resuspended in 0.5ml ice-cold transport buffer (20 mM HEPES [pH 7.3], 110 mM KAc, 5 mM NaAc, 2 mM MgAc, 1 mM EGTA, 2 mM DTT, 1 μg/ml each of aprotinin, pepstatin A, and leupeptin[protease inhibitors from Sigma]). For reprogramming, cells were pelleted and resuspended in 0.5 ml transport buffer ± 17.5 μg/ml digitonin (Calbiochem) and incubated for 5 min on ice. Cells were washed in transport buffer, resuspended in 20 μl reprogramming mix (1 μl 20 mg/ml BSA [Roche], 1 μl 20 mM ATP, 1 μl 100 mM phosphocreatine, 1 μl 400 U/ml creatine kinase [all Sigma], 1 μl RNAsin [40 U/μl], 5 μl transport buffer, 10 μl egg extract) and incubated for 30 min at 37°C. Incubation was terminated by addition of 0.5 ml ice-cold transport buffer. 293T cells were centrifuged, resuspended in cell culture medium, and cultivated in 4 cm petri dishes at 37°C with 10% CO_2. For the standard assay, spheres were harvested at day 5 (day 1 = experiment)

5.5 Cell based assays

5.5.1 Cell fusion

HEK293 GoF-9 Luc cells and P19/pur cells were trypsinised and counted. They were mixed in the ratio 3:1 (HEK293: P19). The mixed cells were pelleted and then washed with plain DMEM medium. To the pellet, PEG 1500 solution (Roche) was added drop-by-drop very gently and then incubated for 3-4 minutes at 37˚C. The cells were washed in plain DMEM medium and resuspended in complete medium and then distributed in 2*20 cm plates. P19 and HEK293 cells mixed together but not fused by PEG was used as a negative control. After 48 hours, puromycin and G418 (Final

concentration Puro: 0.6 µg/ml and G418: 2 mg/ml) were added. The media was changed every alternate day until no cells survived in the negative control. Fusion Colonies obtained from experimental plates were divided into two for i)RT-PCR analysis and ii) luciferase assay.

5.5.2 DNA transfections

DNA transfections using Fugene6 (Roche) were done according to the manufacturer's instructions. For maximal transfection efficiency, the volume of the transfection reagent to be used was optimised for each of the cell lines (Table IV).

Cell lines	Reagent / Method	DNA (µg): transfection reagent (µl)
HCT116Dnmt1$^{-/-}$ cells	Fugene	1:6
HEK293	Fugene	1:3
HEK293T	Fugene	1:3
NIH3T3	Fugene	1:6
OctA/OctD cells	Fugene	1:5
OG2 fibroblasts	Fugene	1:9
P19	Fugene	1:5
mES cells	Electroporation	

Table IV: DNA transfection conditions used for different cell lines.

5.5.3 Establishment of OG2 fibroblasts culture

i) On day 14.5 of pregnancy, the OG2 mouse was sacrificed and rinsed with 70% ethanol.

ii) The abdominal skin was opened and pinned to the side. The body was opened with new operating tools.

iii) The whole uterus was taken out and put into a sterile 10 cm dish containing PBS. Under the cell culture hood, the embryos were removed from the uterus and placed into a new 10 cm-dish containing PBS.

iv) Each embryo was cleaned off extra-embryonic tissue and debris. The liver was removed with a bowed pipette and the embryos were directly put into a clean 10 cm dish without any PBS.

v) The embryo was sheared into small fragments until a viscous solution was got. 5 ml of trypsin EDTA was added and transferred with a cut-off blue tip to a conical flask with a magnetic stirrer. The contents were stirred for 10 minutes.

vi) The supernatant was transferred to a falcon tube and the trypsin was inactivated with complete DMEM medium. This mixture was pipetted up and down and left aside for 10 minutes and then pelleted. The cells were resuspended with complete DMEM. Typically 10 ml of the medium was used for each of the embryos obtained.

vii) Day 3: cells were transferred from 10-cm dish to one maxi dish

viii) Day5: Each maxi-dish was split 1:3

ix) Day 7: The cells were trypsin–EDTA digested and frozen in three cryovials per maxi plate.

5.5.4 Differentiation of C2C12 cells

C2C12 cells were plated in 6 well dishes. Typically $1*10^6$ cells are plated. 24 hours after culturing complete growth medium, the cells were washed with plain DMEM medium and then suspended in DMEM consisting of 2% HCS. The medium was then changed every alternative day. Typically after 4 days of induction, myotubes were seen.

5.5.5 siRNA transfection in C2C12 cells

Day -5: C2C12 cells were freshly thawed .

Day -4: Thawed cells were expanded in 20 cm dishes .

Day -2: These Passage 1 cells were trypsnised, counted and $1*10^5$ cells were suspended in complete DMEM without antibiotics in each well of a 6 well plate. After about 5-6 hours of plating, cells were transfected with siRNA. 120 nM siGlow Control or 40 nM si*Gadd45alpha* or 40 nM si*Gadd45beta* smart pools were suspended in antibiotic free 250 µl optiMEM. 8 µl Dharmafect3 is suspended in 250 µl optimem. These siRNA and transfection reagents were mixed and incubated for 20 minutes and then added to cells.

Day -1: After 24 hours of initial transfection, the medium is sucked away and the transfection of siRNAs is once again repeated .

Day 0: Growth medium was removed and differentiation medium was added.

Day +2: End of experiment

To One well, 5 µM Aza was added and seved as a control. As the Aza is unstable at cell culture conditions, fresh Aza was added everyday during the course of the experiment.

5.5.6 HpaII/MS multiplex PCR

C2C12 cells were induced to differentiate as indicated above. Genomic DNA was isolated on day-2, day 0, day 2 of differentiation using the DNeasy blood & tissue kit (Qiagen). 500 ng genomic DNA was digested with HpaII or MspI overnight in a 20 µl enzymatic reaction. Undigested DNA served as a control. 2 µl of each of them was used as a template in a multiplex PCR using Accuprime Taq polymerase system (Invitrogen) according to the program

95˚C, 3 minutes
94˚C, 45 seconds
62˚C, 45 seconds
68˚C, 45 seconds
-31 cycles-
68˚C, 2 minutes

The products were then resolved by 1.5% Agarose gel.

Primers used in MS-PCR of m*Myogenin* is listed in the table V.

Purpose	Primer ID and Sequence
5'-flanking region	**MyoP1** TGGAGTGGTCCTGATGTGGTAGTGG **MyoM8** ACCCAGAGATAAATATAGCCAACGC
Exon3, loading Control	**MyoP10** TCCATCGTGGACAGCATCACG **MyoM2** TAAGGAGTCAGCTAAATTCCCTCGC

Table V: Primers used in MS-PCR of m*Myogenin.*

5.5.7 Combined Bisulphite Restriction Analysis (COBRA)

Background:

The bisulphite treatment based methods for determining the methylation status of cytosine residues in a DNA molecule depend on the reaction of bisulphite with cytosines in single stranded DNA. Cytosines are converted to uracils (see Figure 29) whereas 5-methylcytosines (5-mC) are unreactive.

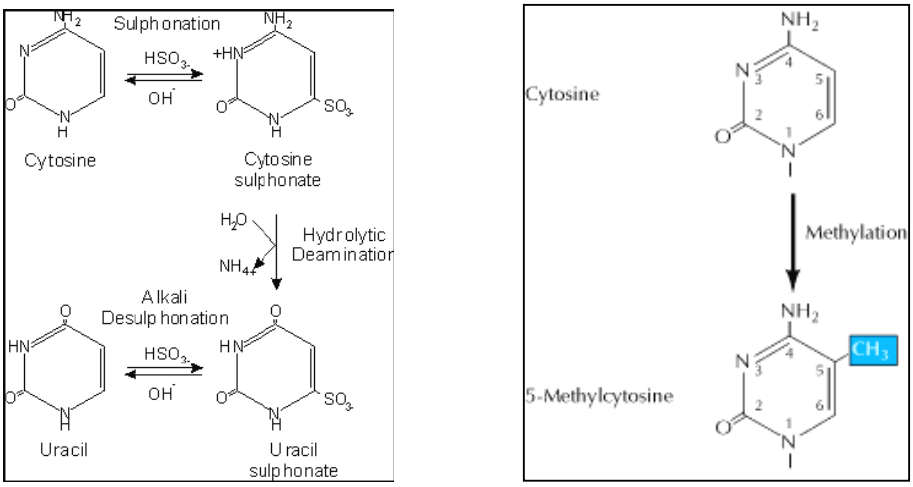


Figure 29: Bisulphite reaction.
Left panel: Basis of bisulphite modification; Right panel: a cartoon of cytosine and methyl cytosine

The modified DNA strands can be amplified using Polymerase Chain Reaction (PCR). It is important that the primers used in the PCR reaction do not contain CpG dinucleotides so that the amplification step does not discriminate between templates according to their original methylation status. Because the *Taq* polymerase is permissive for uracil, it will incorporate thymine in the new synthesized strand in those positions in which a uracil was present in the template DNA molecule. This combination of bisulphite treatment and PCR amplification results in the conversion of unmethylated cytosine residues to thymine and methylated cytosine residues to cytosine. The PCR-product can be on one hand sequenced directly, or cloned and sequenced to give methylation data from single DNA molecules or in the other hand subjected to a restriction enzyme digestion as in COBRA (Combined Bisulphite Restriction Analysis). The methylation-dependent sequence conversion induced after bisulphite treatment and PCR amplification can lead to a loss of a pre-existing site

(thereby creating in some cases a new restriction enzyme site) or it can lead to a methylation-dependent retention of pre-existing sites.

Basis of COBRA of 5'-flanking region of the Myogenin gene

Genomic sequence M95800.1 myogenin nt 1191 - 1549

accagctgccttggaccatggaggagagagtaggcaggaggcccggtaggagtaattgaaaggagcagatgagacg
ggggaatgcacccacccccaccttccctgccccacaggntgtggagaaatgaaaactaatcaaattacagccgacggcc
tcccgacccgtgcacaggagccgcctgggccaggggcaggcctgcagggtggggtgggggcaaaaggagagggaa
ggggaatcacatgtaatccactggaaacgtcttgatgtgcagcaacagcttagagggggctcaggtttctgtggcgttgg
ctatatttatctctgggttcatgccagcagggagggtttaaatggcac

Keys:
HpaII (this site was reported to be demethylated during differentiation)
HpyCH4IV; Hpy99I

The above 358bp region of the *Myogenin* gene is bisulphite converted and PCR amplified. CpGs in the *Myogenin* gene of C2C12 cells are expected to be methylated when cultured in the growth medium and undergo demethylation when the cells are induced to differentiate. If this is the case, bisulphite treatment should not modify CpGs in the undifferentiated state and hence the 358 bp region would yield two bands upon digestion with HpyCHIV or Hpy99I. Upon differentiation, CpGs would be demethylated and hence would be modified by bisulphite treatment thereby HpyCHIV or Hpy99I sites would be lost.

Thus if a 358 bp product alone is obtained with HpyCHIV or Hpy99I digestion of bisulphite converted DNA, it indicates, *Myogenin* remains demethylated. If shorter digestion products are obtained, it indicates the methylation status of the myogenin gene.

Brief protocol:

i) siRNA was transfected into C2C12 cells as explained in chapter 5.5.5 and genomic DNA was isolated on day -2, day 0, day +2 using Dneasy blood & tissue kit.

ii) Genomic DNA was bisulphite converted using Epitect Bisulphite kit (Qiagen) according to the manufacturer's instructions.

iii) Efficiency of bisulphite conversion was ensured by absence of any PCR products using primers SK173 and SK174 that specifically recognises only unconverted DNA template.

iv) Bisulphite converted DNA was used as a template and using primers SK171 and SK172, a 358 bp region of the myogenin gene (region 1191 to 1549 of the genbank accession number M95800.1) was amplified using Accumprime Taq polymerase system (Invitrogen) using 2.25mM Magnesium chloride and annealing at 56˚C.

V). The obtained PCR product was then digested by using HpyCH4IV or Hpy99I and resolved by agarose gel electrophoresis.

Primers used in COBRA of *mMyogenin* is listed in table VI

Purpose	Primers
For Bisulphite converted DNA	**SK171** ATTAGTTGTTTTGGATTATGGAGGA **SK172** TACCATTTAAACCCTCCCTACTAAC
For unconverted DNA	**SK173** ACCAGCTGCCTTGGACCATGGAGGA **SK174** TGCCATTTAAACCCTCCCTGCTGGC

Table VI: Primers used in COBRA of m*Myogenin.*

5.5.8 Reverse Transcription-Polymerase Chain Reaction

RNA isolation

i) From Spheroids obtained after nuclear reprogramming with *Xenopus* extracts:

RNA can be obtained by CsCl method or by modified trizol protocol.

CsCl method: Spheres, clusters, and cells were picked by observing under the microscope and transferred to 100 µl denaturing solution (4 M GnSCN, 0.5% sarcosinate, 1% β-mercaptoethanol, 20 µg rRNA), laid over 100 µl 5.7 M CsCl, and spun for 3.5hr at 174,000 × g in a TL-100 Tabletop Ultracentrifuge (Beckman Coulter, Fullerton, CA) using a TLA 100.2 rotor. The supernatant was discarded in several steps with new pipette tips to carefully avoid DNA contamination. The RNA pellet was washed in 70% EtOH, dried, resuspended in 20 µl DEPC-H_2O and

reprecipitated with sodium acetate. The washed and dried pellet was resuspended in DEPC-H_2O.

Trizol protocol: This method provides an easy option to isolate RNA from Spheres, clustures and cells. The protocol involves cell lysis using 1 ml of trizol and subsequent steps according to the manufacturer's instructions and the following modifications: It is important to add RNase free glycogen as a carrier to precipitate the RNA. Isopropanol precipitation was done overnight at -20˚C.

ii) From cell cultures: RNA was isolated by lysing by using Trizol method according to the manufacturers instructions.

DNase digestion

1µg of RNA	` x µl
10X DNAse buffer (Fermentas)	1.2 µl
RNA guard (40 units/µl) (Fermentas)	0.5 µl
DNase (Fermentas)	1 unit
Final Volume adjusted with DEPC treated water to	12 µl

Digestion was done at 37˚C for 40 minutes. 1 µl of 25 mM EDTA was added and incubated at 65˚C for 15 minutes.

First Strand Synthesis:

6 µl of DNase treated RNA was used for first strand synthesis. The rest was used for – RT control. First strand synthesis was essentially done using Superscript II cDNA synthesis kit (Invitrogen).

A 20 µl reaction volume can be used for 1 ng–5 µg of total RNA or 1–500 ng of mRNA.

i) Following components were added to a nuclease-free microcentrifuge tube:

Total RNA	6 µl
Random primers (50 ng/µl)	2 µl

dNTP Mix (5 mM each)	2 µl
Sterile, distilled water to	12 µl

ii) Incubated the mixture to 65°C for 5 min and quick chilled on ice. The contents of the tube were collected by brief centrifugation and added:

5X First-Strand Buffer	4 µl
0.1 M DTT	2 µl
RNA guard (40 units/µl)	1 µl

iii). Mixed the contents of the tube gently and incubated at 25°C for 2 min.

IV) Added 1 µl (200 units) of SuperScript II RT and mix by pipetting gently up and down. The cDNA was synthesized by incubating the tubes at 25°C for 10 min; then at 42°C for 90 min. The reaction was inactivated by heating at 70°C for 15 min. The cDNA can now be used as a template for amplification in PCR or Real time PCR.

RT-PCR

i) The PCR reactions were mixed on ice in 0.5 µl PCR tubes (master mix was prepared):

10X PCR Buffer (Perkin-Elmer)	5.0 µl
2 mM dNTP Mix	2.5 µl
Forward primer (10 µM)	1.0 µl
Reverse primer (10 µM)	1.0 µl
Taq DNA polymerase (home made)	0.5 µl
cDNA from first-strand reaction	1.0 µl
autoclaved, distilled water to	25.0 µl

ii) The RT-PCR settings were
95˚C, 1 minute
55˚C-62˚C (varies for each primer pair), 1 minute
72˚C, 1 minute

18-45 cycles depending on the target

72˚C, 10 minutes

20˚C, 10 minutes

iii) After the PCR, 5 µl of 6X loading dye was added and then 15 µl of the mixture was loaded onto 1.5-2% agarose gel, separated at 6 V/cm for 30 min and visualized with ethidium bromide staining. Normalisation of the cDNA samples was performed according to *Beta actin* levels. PCR cycle numbers were optimized to be in the linear range of amplification. The primers used for RT-PCR are listed in table VII:

Gene	Primer ID and Sequence
***Beta actin*(Human and Mouse orthologs)**	**E270** GACCCAGATCATGTTTGAGACC **E271** AGGTCCAGACGCAGGATG
hBeta actin	**Mao20** GCTATCCCTGTACGCCTCTG **Mao21** CCATCTCTTGCTCGAAGTCC **SK104** CATTCCAAATATGAGATGCATTG **SK106** CACGAAGGCTCATCATTCAA
mBeta actin	**mBA1** GTGGGCCGCTCTAGGCACCAA **mBA2** CTCTTTGATGTCACGCACGATTTC
hOct4	**SK14** GACAACAATGAAAATCTTCAGGAGA **SK15** TTCTGGCGCCGGTTACAGAACCA
mOct4	**SK12** TGCTGGAGAAGTGGGTGGAGG **SK13** GTAGAGTGTGGTGAAGTGGGGG
hRex-1	**SK18** GCGTACGCAAATTAAAGTCCAGA **SK19** CAGATCCTAAACAGCTCGCAGAAT
hGCAP	**GCAP-1** CCATATCCTGAGGTGGATCAG **GCAP-2** CAGGGAGGGAAGGTAATGAGT

Table VII: RT-PCR primers.

Real Time PCR using UPL Probes:

For Real-Time PCR employing UPL probes, the Roche LightCycler480 Probes Master was used. This Master mix is 2x concentrated and contains a Hot Start Taq Polymerase and dNTPs. Typically 20 µl reaction was performed in triplicates.

Setting up of the reaction (1x):

2x **PROBES** Master	10.0 µl
Primer Mix	1.0 µl
(forward + reverse, 10µM each)	
UPL probe	0.2 µl
cDNA template	x µl
ddH_20 add	20 µl

Precautions:

• The UPL probe is light sensitive! Do not expose to light for longer periods!!

• Real-Time PCR is generally very sensitive! To avoid deviations because of variations in the pipetting volume, dilute your DNA template and use higher volume (optimally 2-5 µl)

PCR plates: Special 96 Well plates are used for Real-Time PCR as the conventional plates often show an intrinsic fluorescence which affects the fluorescence acquisition. After pipetting the plate was centrifuged for 3 min at 2000 rpm in a swing out rotor to remove air bubbles.

PCR-Program

Step I: denaturation (HotStart Taq) 95°C, 5-10 minutes

Step II: amplification, 40-50cyc 95°C, 10 seconds

Annealing Temperature, 15-30 seconds

72°C, 1 second Acquisition*

3rd cooling, 40°C 10 seconds

* After every round of elongation the fluorescence is measured by the LightCycler.

Amount of fluorescence is monitored in Real-Time so that you can follow up the amplification. Melting curve has not to be performed using this format. Values can direcly be used for analysis. Typically a standard curve was generated for each primer

pair being used using a template that the primer pair could successfully amplify. Primers used in the Real time PCR are described in the table VIII

Gene	Primer ID and Sequence	UPL Probe ID
mBeta actin	**SK157** AAGGCCAACCGTGAAAAGAT **SK158** GTGGTACGACCAGAGGCATAC	56
mGADD45alpha	**SK159** GCTGCCAAGCTGCTCAAC **SK160** TCGTCGTCTTCGTCAGCA	40
mGADD45beta	**SK161** CGGCCAAACTGATGAATGT **SK162** ATCTGCAGAGCGATATCATCC	79
mGADD45gamma	**SK163** CCGTGGCCAGGATACAGTT **SK164** AAGTTCGTGCAGTGCTTTCC	76
mMyogenin	**SK165** CCTTGCTCAGCTCCCTCA **SK166** TGGGAGTTGCATTCACTGG	63

Table VIII: Real time PCR primers.

5.5.9 Luciferase assay

Luciferase assays reporter assays in P19 and HCT116Dnmt1$^{-/-}$ cells were carried out in 96 well plates in triplicates with a volume of 50 µl per well. Firefly and Renlla luciferase activities were measured using the Dual Luciferase Assay Kit (Promega) on the Fluroskan Ascent FL (Labsystems). The Firefly luciferase activity was normalized against Renilla luciferase activity. In case of mES cells and fusion colonies, the cells were lysed in passive lysis buffer and equal amounts of protein in control and experimental samples were taken for analysis. For small molecule screening in HEK293/GoF-9 Luc clonal cell line 11, ATPLITE (PerkinElmer) was used to measure the luciferase readout.

5.5.10 Beta-galactosidase staining

OctD cells were rinsed with 0.5 ml of PBS and fixed with 0.5% glutaraldehyde in PBS at room temperature for 5 minutes. The cells were then rinsed two times with PBS and incubated for 15 minutes in 0.1% Triton X 100 diluted in PBS. 0.5 ml of staining solution per well (5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$, 0.25

mg/ml X-Gal in PBS). The cells were incubated until blue staining was visible and then rinsed several times with PBS.

5.6 Protein methods

5.6.1 Cell lysis

Tissue culture cells were washed with PBS on ice and the cells were lysed with NOP buffer [2% NP-40,150 mM NaCl, 20 mM Tris, 2 mM PMSF and 1X protease inhibitory cocktail (Roche)] and the spinned at maximum speed at 4˚C to remove the debris and stores at -20˚C.

5.6.2 Protein concentration estimation (BCA Method)

The protein to be analyzed reacts with Cu^{2+} in an alkaline solution to make Cu^{1+} ions. These Cu^{1+} ions are then chelated by the BCA which converts the apple-green color of the free BCA to the purple color of the copper-BCA complex. This can be quantitatively measured at 562 nm. The complex is quite linear with increasing protein concentration over a working range of 20 mg/ml to 2000 mg/ml.

BSA was used to gernerate a standard curve. Different dilutions of BSA and protein to be measured were incubated with 1 ml of 1:50 copper sulphate (4% v/v): BCA solution and incubated at 37˚C by vortexing for about 15 minutes. The samples were measured using spectrophotometer.

5.6.3 SDS-PAGE and Western Blotting

SDS-PAGE and western Blotting were carried out according to standard protocols (Joseph Sambrook and Russell, 2001)and signals were detected by enhanced chemiluminescence (Pierce).

5.6.4 Immunostaining

Cells were grown in 24 well plates. They were washed twice with PBS and fixed with paraformaldehyde. The cells were rinsed with PBS and incubated in a blocking solution made up of 1% BSA in PBS for 1 hour. The cells were then incubated with primary antibody overnight at 4˚C. After extensive washing the cells were incubated

for 1 hour in secondary antibody. The cells were washed and observed under microscope.

5.7 Bioinformatic tools

Genomic sequences were extracted from the public databases NCBI (www.ncbi.nih.gov) and the Sanger centre (www.ensembl.org). Sequence searches and comparisons were performed with BLAST (www.ncbi.nlm.nih.gov/BLAST). Primer design and sequence analyses were carried out with the integrated bioinformatic package HUSAR (http://genome.dkfz-heidelberg.de). Literature searches were done with PubMed (www.pubmed.gov). Real time PCRs primers and their UPL probes were obtained from https://www.roche-applied-science.com/sis/rtpcr/upl/adc.jsp

6 Literature

Abdel-Rahman, B., Fiddler, M., Rappolee, D. and Pergament, E. (1995). Expression of transcription regulating genes in human preimplantation embryos. *Hum Reprod* **10**, 2787-92.

Adams, R. L. P., Hanley, A. and Rinaldi, A. (1990). DNA demethylation in erythroleukaemia cells. *FEBS Letters* **269**, 29-31.

Aguirre-Arteta, A. M., Grunewald, I., Cardoso, M. C. and Leonhardt, H. (2000). Expression of an Alternative Dnmt1 Isoform during Muscle Differentiation. *Cell Growth Differ* **11**, 551-559.

Allegrucci, C., Thurston, A., Lucas, E. and Young, L. (2005). Epigenetics and the germline. *Reproduction* **129**, 137-49.

Amit, M., Carpenter, M. K., Inokuma, M. S., Chiu, C. P., Harris, C. P., Waknitz, M. A., Itskovitz-Eldor, J. and 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.

Andrews, P. W. and Goodfellow, P. N. (1980). Antigen expression by somatic cell hybrids of a murine embryonal carcinoma cell with thymocytes and L cells. *Somatic Cell Genet* **6**, 271-84.

Antequera, F. and Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* **90**, 11995-9.

Avigan, M. I., Strober, B. and Levens, D. (1990). A far upstream element stimulates c-myc expression in undifferentiated leukemia cells. *J Biol Chem* **265**, 18538-45.

Avilion, A. A. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126-140.

Azuara, V. (2006). Chromatin signatures of pluripotent cell lines. *Nature Cell Biol.* **8**, 532-538.

Baba, Y., Garrett, K. P. and Kincade, P. W. (2005). Constitutively Active [beta]-Catenin Confers Multilineage Differentiation Potential on Lymphoid and Myeloid Progenitors. *Immunity* **23**, 599-609.

Baguisi, A., Behboodi, E., Melican, D. T., Pollock, J. S., Destrempes, M. M., Cammuso, C., Williams, J. L., Nims, S. D., Porter, C. A., Midura, P. et al. (1999). Production of goats by somatic cell nuclear transfer. *Nat Biotechnol* **17**, 456-61.

Baker, E. K. and El-Osta, A. (2003). The rise of DNA methylation and the importance of chromatin on multidrug resistance in cancer. *Experimental Cell Research* **290**, 177-194.

Baron, M. H. and Maniatis, T. (1986). Rapid reprogramming of globin gene expression in transient heterokaryons. *Cell* **46**, 591-602.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S. K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F. et al. (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445**, 671-5.

Bazar, L., Harris, V., Sunitha, I., Hartmann, D. and Avigan, M. (1995). A transactivator of c-myc is coordinately regulated with the proto-oncogene during cellular growth. *Oncogene* **10**, 2229-38.

Ben-Shushan, E., Sharir, H., Pikarsky, E. and Bergman, Y. (1995). A dynamic balance between ARP-1/COUP-TFII, EAR-3/COUP-TFI, and retinoic acid receptor:retinoid X receptor heterodimers regulates Oct-3/4 expression in embryonal carcinoma cells. *Mol Cell Biol* **15**, 1034-48.

Bernstein, B. E. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326.
Bestor, T. H., Hellewell, S. B. and Ingram, V. M. (1984). Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. *Mol. Cell. Biol.* **4**, 1800-1806.
Bhattacharya, B. (2004). Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood* **103**, 2956-2964.
Blau, H. M. and Blakely, B. T. (1999). Plasticity of cell fate: Insights from heterokaryons. *Seminars in Cell & Developmental Biology* **10**, 267-272.
Boiani, M., Eckardt, S., Schoeler, H. R. and McLaughlin, K. J. (2002). Oct4 distribution and level in mouse clones: consequences for pluripotency. *Genes Dev* **16**, 1209-1219.
Boyer, L. A. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353.
Briggs, R. and King, T. J. (1952). Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proc Natl Acad Sci U S A* **38**, 455-63.
Brunk, B. P., Goldhamer, D. J. and Emerson, J. C. P. (1996). Regulated Demethylation of themyoDDistal Enhancer during Skeletal Myogenesis. *Developmental Biology* **177**, 490-503.
Burdon, T., Stracey, C., Chambers, I., Nichols, J. and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev. Biol.* **210**, 30-43.
Buschhausen, G., Wittig, B., Graessmann, M. and Graessmann, A. (1987). Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. *Proc Natl Acad Sci U S A* **84**, 1177-81.
Byrne, J. A., Simonsson, S., Western, P. S. and Gurdon, J. B. (2003). Nuclei of adult mammalian somatic cells are directly reprogrammed to oct-4 stem cell gene expression by amphibian oocytes. *Curr Biol* **13**, 1206-1213.
Cam, H., Griesmann, H., Beitzinger, M., Hofmann, L., Beinoraviciute-Kellner, R., Sauer, M., Huttinger-Kirchhof, N., Oswald, C., Friedl, P., Gattenlohner, S. et al. (2006). p53 family members in myogenic differentiation and rhabdomyosarcoma development. *Cancer Cell* **10**, 281-293.
Cameron, E. E., Bachman, K. E., Myohanen, S., Herman, J. G. and Baylin, S. B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat Genet* **21**, 103-7.
Campbell, K. H. S., McWhir, J., Ritchie, W. A. and Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature* **380**, 64-66.
Chambers, I. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-655.
Chen, S., Zhang, Q., Wu, X., Schultz, P. G. and Ding, S. (2004). Dedifferentiation of Lineage-Committed Cells by a Small Molecule. *J. Am. Chem. Soc.* **126**, 410-411.
Cheong, H. T., Takahashi, Y. and Kanagawa, H. (1993). Birth of mice after transplantation of early cell-cycle-stage embryonic nuclei into enucleated oocytes. *Biol Reprod* **48**, 958-63.
Chrivia, J. C., Kwok, R. P., Lamb, N., Hagiwara, M., Montminy, M. R. and Goodman, R. H. (1993). Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature* **365**, 855-9.
Cibelli, J. B., Stice, S. L., Golueke, P. J., Kane, J. J., Jerry, J., Blackwell, C., Ponce de Leon, F. A. and Robl, J. M. (1998). Cloned transgenic calves produced from nonquiescent fetal fibroblasts. *Science* **280**, 1256-8.

Cowan, C. A., Atienza, J., Melton, D. A. and Eggan, K. (2005). Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem Cells. *Science* **309**, 1369-1373.

Cruciat, C.-M., Hassler, C. and Niehrs, C. (2006). The MRH Protein Erlectin Is a Member of the Endoplasmic Reticulum Synexpression Group and Functions in N-Glycan Recognition. *J. Biol. Chem.* **281**, 12986-12993.

Daheron, L., Opitz, S. L. and Zaehres, H. (2004). LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* **22**, 770-778.

Darr, H., Mayshar, Y. and Benvenisty, N. (2006). Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development* **133**, 1193-1201.

Davidson, G., Wu, W., Shen, J., Bilic, J., Fenger, U., Stannek, P., Glinka, A. and Niehrs, C. (2005). Casein kinase 1 gamma couples Wnt receptor activation to cytoplasmic signal transduction. *Nature* **438**, 867-72.

Davis, R. L., Weintraub, H. and Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000.

de la Calle-Mustienes, E., Glavic, A., Modolell, J. and Gomez-Skarmeta, J. L. (2002). Xiro homeoproteins coordinate cell cycle exit and primary neuron formation by upregulating neuronal-fate repressors and downregulating the cell-cycle inhibitor XGadd45-gamma. *Mech Dev* **119**, 69-80.

de La Serna, I. L., Carlson, K. A., Hill, D. A., Guidi, C. J., Stephenson, R. O., Sif, S., Kingston, R. E. and Imbalzano, A. N. (2000). Mammalian SWI-SNF complexes contribute to activation of the hsp70 gene. *Mol Cell Biol* **20**, 2839-51.

Di Berardino, M. A. (2001). Animal Cloning - the route to new genomics in agriculture and medicine. *Differentiation* **68**, 67-83.

Di Berardino, M. A., McKinnell, R. G. and Wolf, D. P. (2003). The golden anniversary of cloning: a celebratory essay. *Differentiation* **71**, 398-401.

Dimitrov, S. and Wolffe, A. P. (1996). Remodeling somatic nuclei in Xenopus laevis egg extracts: molecular mechanisms for the selective release of histones H1 and H1(0) from chromatin and the acquisition of transcriptional competence. *Embo J* **15**, 5897-906.

Ding, S. and Schultz, P. G. (2004). A role for chemistry in stem cell biology. *Nat Biotech* **22**, 833-840.

Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Avigan, M. and Levens, D. (1994). A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. *Genes Dev* **8**, 465-80.

Duncan, R., Collins, I., Tomonaga, T., Zhang, T. and Levens, D. (1996). A unique transactivation sequence motif is found in the carboxyl-terminal domain of the single-strand-binding protein FBP. *Mol Cell Biol* **16**, 2274-82.

Eggan, K., Akutsu, H., Loring, J., Jackson-Grusby, L., Klemm, M., Rideout, W. M., 3rd, Yanagimachi, R. and Jaenisch, R. (2001). Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc Natl Acad Sci U S A* **98**, 6209-14.

Egger, G., Jeong, S., Escobar, S. G., Cortez, C. C., Li, T. W., Saito, Y., Yoo, C. B., Jones, P. A. and Liang, G. (2006). Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. *Proc Natl Acad Sci U S A* **103**, 14080-5.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156.

Fang, Z. F., Gai, H., Huang, Y. Z., Li, S. G., Chen, X. J., Shi, J. J., Wu, L., Liu, A., Xu, P. and Sheng, H. Z. (2006). Rabbit embryonic stem cell lines derived from fertilized, parthenogenetic or somatic cell nuclear transfer embryos. *Experimental Cell Research* **312**, 3669-3682.

Faro-Trindade, I. and Cook, P. R. (2006). A conserved organization of transcription during embryonic stem cell differentiation and in cells with high C value. *Mol Biol Cell* **17**, 2910-20.

Fitzpatrick, D. R. and Wilson, C. B. (2003). Methylation and demethylation in the regulation of genes, cells, and responses in the immune system. *Clin Immunol* **109**, 37-45.

Flasza, M., Shering, A. F., Smith, K., Andrews, P. W., Talle, P. and Johnson, P. (2003). Reprogramming in inter-species embryonal carcinoma-somatic cell hybrids induces expression of pluripotency and differentiation markers. *Cloning Stem Cells* **5**, 339-354.

Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki Ji, J. and Niwa, H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev* **16**, 784-9.

Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L. and Kouzarides, T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat Genet* **24**, 88-91.

Gehring, M., Huh, J. H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J. J., Goldberg, R. B. and Fischer, R. L. (2006). DEMETER DNA Glycosylase Establishes MEDEA Polycomb Gene Self-Imprinting by Allele-Specific Demethylation. *Cell* **124**, 495-506.

Gonda, K., Fowler, J., Katoku-Kikyo, N., Haroldson, J., Wudel, J. and Kikyo, N. (2003). Reversible disassembly of somatic nucleoli by the germ cell proteins FRGY2a and FRGY2b. *Nat Cell Biol* **5**, 205-10.

Gong, Y. (2006). NSPc1 is a cell growth regulator that acts as a transcriptional repressor of p21Waf1/Cip1 via the RARE element. *Nucleic Acids Res.* **34**, 6158-6169.

Graessmann, A. and Graessmann, M. (1988). DNA methylation, chromatin structure and regulation of Herpes simplex virus tk gene expression. *Gene* **74**, 135-7.

Graves, K. H. and Moreadith, R. W. (1993). Derivation and characterization of putative pluripotential embryonic stem cells from preimplantation rabbit embryos. *Mol Reprod Dev* **36**, 424-33.

Gurdon, J. B. (2006). From nuclear transfer to nuclear reprogramming: the reversal of cell differentiation. *Annu Rev Cell Dev Biol* **22**, 1-22.

Gurdon, J. B. and Byrne, J. A. (2003). The first half-century of nuclear transplantation. *Proc Natl Acad Sci U S A* **100**, 8048-52.

Hakelien, A. M., Landsverk, H. B., Robl, J. M., Skalhegg, B. S. and Collas, P. (2002). Reprogramming fibroblasts to express T-cell functions using cell extracts. *Nat Biotechnol* **20**, 460-6.

Hansis, C., Barreto, G., Maltry, N. and Niehrs, C. (2004). Nuclear reprogramming of human somatic cells by xenopus egg extract requires BRG1. *Curr Biol* **14**, 1475-80.

Hart, A. H., Hartley, L., Ibrahim, M. and Robb, L. (2004). Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. *Dev. Dyn.* **230**, 187-198.

Hattori, N., Nishino, K., Ko, Y. G., Hattori, N., Ohgane, J., Tanaka, S. and Shiota, K. (2004). Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. *J Biol Chem* **279**, 17063-9.

Herr, W. and Cleary, M. A. (1995). The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev* **9**, 1679-93.
Hochedlinger, K. and Jaenisch, R. (2006). Nuclear reprogramming and pluripotency. *Nature* **441**, 1061-7.
Hochedlinger, K., Yamada, Y., Beard, C. and Jaenisch, R. (2005). Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **21**, 465-477.
Hoffmeyer, A., Piekorz, R., Moriggl, R. and Ihle, J. N. (2001). Gadd45gamma is dispensable for normal mouse development and T-cell proliferation. *Mol Cell Biol* **21**, 3137-43.
Hollander, M. C., Sheikh, M. S., Bulavin, D. V., Lundgren, K., Augeri-Henmueller, L., Shehee, R., Molinaro, T. A., Kim, K. E., Tolosa, E., Ashwell, J. D. et al. (1999). Genomic instability in Gadd45a-deficient mice. *Nat Genet* **23**, 176-184.
Humphrey, R. K. (2004). Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. *Stem Cells* **22**, 522-530.
Hwang, W. S. (2005). Patient-specific embryonic stem cells derived from human SCNT blastocysts. *Science* **308**, 1777-1783.
Hyslop, L., Stojkovic, M. and Armstrong, L. (2005). Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. *Stem Cells* **23**, 1035-1043.
Ijiri, K., Zerbini, L. F., Peng, H., Correa, R. G., Lu, B., Walsh, N., Zhao, Y., Taniguchi, N., Huang, X. L., Otu, H. et al. (2005). A novel role for GADD45beta as a mediator of MMP-13 gene expression during chondrocyte terminal differentiation. *J Biol Chem* **280**, 38544-55.
Imamura, T., Ohgane, J., Ito, S., Ogawa, T., Hattori, N., Tanaka, S. and Shiota, K. (2001). CpG island of rat sphingosine kinase-1 gene: tissue-dependent DNA methylation status and multiple alternative first exons. *Genomics* **76**, 117-25.
Ivanova, N. (2006). Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-538.
Jaenisch, R. and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33**, S245-S254.
Joseph Sambrook and Russell, D. W. (2001). Molecular Cloning: A laboratory manual. New York: CSHL Press.
Jost, J.-P. and Jost, Y.-C. (1995). Mechanism of active DNA demethylation during embryonic development and cellular differentiation in vertebrates. *Gene* **157**, 265-266.
Jost, J.-P., Oakeley, E. J., Zhu, B., Benjamin, D., Thiry, S., Siegmann, M. and Jost, Y.-C. (2001). 5-Methylcytosine DNA glycosylase participates in the genome-wide loss of DNA methylation occurring during mouse myoblast differentiation. *Nucl. Acids Res.* **29**, 4452-4461.
Jost, J. (1993). Nuclear Extracts of Chicken Embryos Promote an Active Demethylation of DNA by Excision Repair of 5-Methyldeoxycytidine. *PNAS* **90**, 4684-4688.
Jost, J. P. and Jost, Y. C. (1994). Transient DNA demethylation in differentiating mouse myoblasts correlates with higher activity of 5-methyldeoxycytidine excision repair. *J. Biol. Chem.* **269**, 10040-10043.
Kawahara, A., Che, Y. S., Hanaoka, R., Takeda, H. and Dawid, I. B. (2005). Zebrafish GADD45beta genes are involved in somite segmentation. *Proc Natl Acad Sci U S A* **102**, 361-6.

Kessler, D. S. (1997). Siamois is required for formation of Spemann's organizer. *Proc Natl Acad Sci U S A* **94**, 13017-22.

Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B. and Crabtree, G. R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature* **366**, 170-174.

Kikyo, N., Wade, P. A., Guschin, D., Ge, H. and Wolffe, A. P. (2000). Active Remodeling of Somatic Nuclei in Egg Cytoplasm by the Nucleosomal ATPase ISWI. *Science* **289**, 2360-2362.

Klose, R. J. and Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**, 89-97.

Kress, C., Thomassin, H. and Grange, T. (2001). Local DNA demethylation in vertebrates: how could it be performed and targeted? *FEBS Lett* **494**, 135-40.

Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S. Y. and Suemori, H. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol* **25**, 2475-2485.

Lenardo, M. J., Staudt, L., Robbins, P., Kuang, A., Mulligan, R. C. and Baltimore, D. (1989). Repression of the IgH enhancer in teratocarcinoma cells associated with a novel octamer factor. *Science* **243**, 544-6.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* **3**, 662-73.

Lin, T. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nature Cell Biol.* **7**, 165-171.

Littlefield, J. W. (1964). Selection of Hybrids from Matings of Fibroblasts in Vitro and Their Presumed Recombinants. *Science* **145**, 709-10.

Lorincz, M. C., Dickerson, D. R., Schmitt, M. and Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol* **11**, 1068-75.

Lu, B., Ferrandino, A. F. and Flavell, R. A. (2004). Gadd45beta is important for perpetuating cognate and inflammatory signals in T cells. *Nat Immunol* **5**, 38-44.

Lucarelli, M., Fuso, A., Strom, R. and Scarpa, S. (2001). The dynamics of myogenin site-specific demethylation is strongly correlated with its expression and with muscle differentiation. *J Biol Chem* **276**, 7500-6.

Martins, S. B., Eide, T., Steen, R. L., Jahnsen, T., Skalhegg, B. S. and Collas, P. (2000). HA95 is a protein of the chromatin and nuclear matrix regulating nuclear envelope dynamics. *J Cell Sci* **113 Pt 21**, 3703-13.

Matsui, Y., Zsebo, K. and Hogan, B. L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841.

Matsumura, H., Tada, M., Otsuji, T., Yasuchika, K., Nakatsuji, N., Surani, A. and Tada, T. (2007). Targeted chromosome elimination from ES-somatic hybrid cells. *Nat Meth* **4**, 23-25.

Matsuya, Y., Green, H. and Basilico, C. (1968). Properties and uses of human-mouse hybrid cell lines. *Nature* **220**, 1199-202.

Mayer, W., Niveleau, A., Walter, J., Fundele, R. and Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature* **403**, 501-2.

Meshorer, E. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105-116.

Migeon, B. R. and Miller, C. S. (1968). Human-mouse somatic cell hybrids with single human chromosome (group E): link with thymidine kinase activity. *Science* **162**, 1005-6.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M. and Takahashi, K. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642.

Morrison, G. M. and Brickman, J. M. (2006). Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. *Development* **133**, 2011-2022.

Na Liu, M. L. X. T. Z. H. (2007). Molecular mechanisms involved in self-renewal and pluripotency of embryonic stem cells. *Journal of Cellular Physiology* **211**, 279-286.

Naylor, L. H. (1999). Reporter gene technology: the future looks bright. *Biochem Pharmacol* **58**, 749-57.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D. and Chambers, I. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-391.

Niwa, H. (2007). How is pluripotency determined and maintained? *Development* **134**, 635-46.

Niwa, H., Miyazaki, J. and Smith, A. G. (2000). Quantitative expression of Oct-3//4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-376.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917-29.

Nothias, J. Y., Majumder, S., Kaneko, K. J. and DePamphilis, M. L. (1995). Regulation of gene expression at the beginning of mammalian development. *J Biol Chem* **270**, 22077-80.

Nutt, S. L., Heavey, B., Rolink, A. G. and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**, 556-562.

Okamoto, K. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60**, 461-472.

Okazawa, H., Okamoto, K., Ishino, F., Ishino-Kaneko, T., Takeda, S., Toyoda, Y., Muramatsu, M. and Hamada, H. (1991). The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. *Embo J* **10**, 2997-3005.

Okumura-Nakanishi, S., Saito, M., Niwa, H. and Ishikawa, F. (2004). Oct-3//4 and Sox2 regulate Oct-3//4 gene in embryonic stem cells. *J Biol Chem* **280**, 5307-5317.

Onishi, A., Iwamoto, M., Akita, T., Mikawa, S., Takeda, K., Awata, T., Hanada, H. and Perry, A. C. (2000). Pig cloning by microinjection of fetal fibroblast nuclei. *Science* **289**, 1188-90.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W. and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Current Biology* **10**, 475-478.

Ovitt, C. E. and Schoeler, H. R. (1998). The molecular biology of Oct-4 in the early mouse embryo. *Mol Hum Reprod* **4**, 1021-1031.

Palmieri, S. L., Peter, W., Hess, H. and Schoeler, H. R. (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev Biol* **166**, 259-67.

Pan, G., Li, J. and Zhou, Y. (2006). A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *FASEB J* **20**, 1730-1732.
Pesce, M., Gross, M. K. and Schoeler, H. R. (1998). In line with our ancestors: Oct-4 and the mammalian germ. *BioEssays* **20**, 722-32.
Pevny, L. H. and Lovell-Badge, R. (1997). Sox genes find their feet. *Current Opinion in Genetics & Development* **7**, 338-344.
Poccia, D. and Collas, P. (1996). Transforming sperm nuclei into male pronuclei in vivo and in vitro. *Curr Top Dev Biol* **34**, 25-88.
Ramchandani, S., Bhattacharya, S. K., Cervoni, N. and Szyf, M. (1999). DNA methylation is a reversible biological signal. *PNAS* **96**, 6107-6112.
Rao, M. (2004). Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Developmental Biology* **275**, 269-286.
Razin, A. and Cedar, H. (1991). DNA methylation and gene expression. *Microbiol Rev* **55**, 451-8.
Razin, A., Levine, A., Kafri, T., Agostini, S., Gomi, T. and Cantoni, G. L. (1988). Relationship between Transient DNA Hypomethylation and Erythroid Differentiation of Murine Erythroleukemia Cells. *PNAS* **85**, 9003-9006.
Razin, A. and Shemer, R. (1995). DNA methylation in early development. *Hum Mol Genet* **4 Spec No**, 1751-5.
Razin, A., Webb, C., Szyf, M., Yisraeli, J., Rosenthal, A., Naveh-Many, T., Sciaky-Gallili, N. and Cedar, H. (1984). Variations in DNA Methylation during Mouse Cell Differentiation in vivo and in vitro. *PNAS* **81**, 2275-2279.
Reik, W., Dean, W. and Walter, J. (2001). Epigenetic reprogramming in mammalian development. *Science* **293**, 1089-1093.
Resnick, J. L., Bixler, L. S., Cheng, L. and Donovan, P. J. (1992). Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550.
Rhee, I., Jair, K. W., Yen, R. W., Lengauer, C., Herman, J. G., Kinzler, K. W., Vogelstein, B., Baylin, S. B. and Schuebel, K. E. (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. *Nature* **404**, 1003-7.
Rideout, W. M., 3rd, Wakayama, T., Wutz, A., Eggan, K., Jackson-Grusby, L., Dausman, J., Yanagimachi, R. and Jaenisch, R. (2000). Generation of mice from wild-type and targeted ES cells by nuclear cloning. *Nat Genet* **24**, 109-10.
Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B. and Ng, H. H. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280**, 24731-24737.
Rodolfa, K. T. and Eggan, K. (2006). A Transcriptional Logic for Nuclear Reprogramming. *Cell* **126**, 652-655.
Rosemary F. Bachvarova, T. M. M. D. N. P. K. M. R. P. A. D. J. (2004). Gene expression in the axolotl germ line: Axdazl, Axvh, Axoct-4, and Axkit. *Developmental Dynamics* **231**, 871-880.
Rosner, M. H. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**, 686-692.
Ryan, A. K. and Rosenfeld, M. G. (1997). POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev* **11**, 1207-25.
Scarpa, S., Lucarelli, M., Palitti, F., Carotti, D. and Strom, R. (1996). Simultaneous myogenin expression and overall DNA hypomethylation promote in vitro myoblast differentiation. *Cell Growth Differ* **7**, 1051-8.
Schneider-Stock, R., Diab-Assef, M., Rohrbeck, A., Foltzer-Jourdainne, C., Boltze, C., Hartig, R., Schonfeld, P., Roessner, A. and Gali-Muhtasib, H. (2005). 5-aza-Cytidine Is a Potent Inhibitor of DNA Methyltransferase 3a and Induces

Apoptosis in HCT-116 Colon Cancer Cells via Gadd45- and p53-Dependent Mechanisms. *J Pharmacol Exp Ther* **312**, 525-536.

Schoeler, H. R., Balling, R., Hatzopoulos, A. K., Suzuki, N. and Gruss, P. (1989a). Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *EMBO J.* **8**, 2551-2557.

Schoeler, H. R., Hatzopoulos, A. K., Balling, R., Suzuki, N. and Gruss, P. (1989b). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J.* **8**, 2543-2550.

Schoorlemmer, J., van Puijenbroek, A., van Den Eijnden, M., Jonk, L., Pals, C. and Kruijer, W. (1994). Characterization of a negative retinoic acid response element in the murine Oct4 promoter. *Mol Cell Biol* **14**, 1122-36.

Sekkai, D., Gruel, G., Herry, M., Moucadel, V., Constantinescu, S. N., Albagli, O., Tronik-Le Roux, D., Vainchenker, W. and Bennaceur-Griscelli, A. (2005). Microarray analysis of LIF/Stat3 transcriptional targets in embryonic stem cells. *Stem Cells* **23**, 1634-42.

Shamblott, M. J., Axelman, J., Littlefield, J. W., Blumenthal, P. D., Huggins, G. R., Cui, Y., Cheng, L. and Gearhart, J. D. (2001). Human embryonic germ cell derivatives express a broad range of developmentally distinct markers and proliferate extensively in vitro. *Proc Natl Acad Sci U S A* **98**, 113-8.

Shimozaki, K., Nakashima, K., Niwa, H. and Taga, T. (2003). Involvement of Oct3/4 in the enhancement of neuronal differentiation of ES cells in neurogenesis-inducing cultures. *Development* **130**, 2505-2512.

Shin, T., Kraemer, D., Pryor, J., Liu, L., Rugila, J., Howe, L., Buck, S., Murphy, K., Lyons, L. and Westhusin, M. (2002). A cat cloned by nuclear transplantation. *Nature* **415**, 859.

Smale, S. T. (2003). The establishment and maintenance of lymphocyte identity through gene silencing. *Nature Immunol.* **4**, 607-615.

Smith, A. G. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-690.

Smith, S. T. and Jaynes, J. B. (1996). A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo. *Development* **122**, 3141-50.

Spada, F., Haemmer, A., Kuch, D., Rothbauer, U., Schermelleh, L., Kremmer, E., Carell, T., Langst, G. and Leonhardt, H. (2007). DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J. Cell Biol.* **176**, 565-571.

Spivakov, M. and Fisher, A. G. (2007). Epigenetic signatures of stem-cell identity. *Nat Rev Genet* **8**, 263-271.

Surani, M. A., Hayashi, K. and Hajkova, P. (2007). Genetic and Epigenetic Regulators of Pluripotency. *Cell* **128**, 747-762.

Szabo, P. E., Hubner, K., Schoeler, H. and Mann, J. R. (2002). Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mech Dev* **115**, 157-60.

Szyf, M., Rouleau, J., Theberge, J. and Bozovic, V. (1992). Induction of myogenic differentiation by an expression vector encoding the DNA methyltransferase cDNA sequence in the antisense orientation. *J. Biol. Chem.* **267**, 12831-12836.

Tada, M., Tada, T., Lefebvre, L., Barton, S. C. and Surani, M. A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *Embo J* **16**, 6510-6520.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr. Biol.* **11**, 1553-1558.

Takagi, N., Yoshida, M. A., Sugawara, O. and Sasaki, M. (1983). Reversal of X-inactivation in female mouse somatic cells hybridized with murine teratocarcinoma stem cells in vitro. *Cell* **34**, 1053-1062.

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676.

Takeda, J., Seino, S. and Bell, G. I. (1992). Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues. *Nucleic Acids Res* **20**, 4613-20.

Tamame, M., Antequera, F. and Santos, E. (1988). Developmental characterization and chromosomal mapping of the 5-azacytidine-sensitive fluF locus of Aspergillus nidulans. *Mol Cell Biol* **8**, 3043-50.

Taranger, C. K., Noer, A., Sorensen, A. L., Hakelien, A.-M., Boquest, A. C. and Collas, P. (2005). Induction of Dedifferentiation, Genomewide Transcriptional Programming, and Epigenetic Reprogramming by Extracts of Carcinoma and Embryonic Stem Cells. *Mol. Biol. Cell* **16**, 5719-5735.

Thomson, J. A., Kalishman, J., Golos, T. G., Durning, M., Harris, C. P., Becker, R. A. and Hearn, J. P. (1995). Isolation of a Primate Embryonic Stem Cell Line. *PNAS* **92**, 7844-7848.

Thomson, J. A. and Marshall, V. S. (1998). Primate embryonic stem cells. *Curr. Top. Dev. Biol.* **38**, 133-165.

Tomilin, A., Remenyi, A., Lins, K., Bak, H., Leidel, S., Vriend, G., Wilmanns, M. and Schoeler, H. R. (2000). Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. *Cell* **103**, 853-64.

Tomioka, M. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4[ndash]Sox-2 complex. *Nucleic Acids Res.* **30**, 3202-3213.

Tomizuka, K., Yoshida, H., Uejima, H., Kugoh, H., Sato, K., Ohguma, A., Hayasaka, M., Hanaoka, K., Oshimura, M. and Ishida, I. (1997). Functional expression and germline transmission of a human chromosome fragment in chimaeric mice. *Nat Genet* **16**, 133-43.

Vallender, T. W. and Lahn, B. T. (2006). Localized methylation in the key regulator gene endothelin-1 is associated with cell type-specific transcriptional silencing. *FEBS Lett* **580**, 4560-6.

Wade, P. A. and Kikyo, N. (2002). Chromatin remodeling in nuclear cloning. *European Journal of Biochemistry* **269**, 2284-2287.

Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369-74.

Wakayama, T. and Yanagimachi, R. (1999). Cloning of male mice from adult tail-tip cells. *Nat Genet* **22**, 127-8.

Wang, D., Kennedy, S., Conte, D., Kim, J. K., Gabel, H. W., Kamath, R. S., Mello, C. C. and Ruvkun, G. (2005). Somatic misexpression of germline P granules and enhanced RNA interference in retinoblastoma pathway mutants. *Nature* **436**, 593-597.

Watt, F. and Molloy, P. L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev* **2**, 1136-43.

Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B. and Miller, A. D. (1989). Activation of Muscle-Specific Genes in Pigment, Nerve, Fat, Liver, and Fibroblast Cell Lines by Forced Expression of MyoD. *PNAS* **86**, 5434-5438.

Weiss, A. and Cedar, H. (1997). The role of DNA demethylation during development. *Genes Cells* **2**, 481-6.

Weiss, M. C. and Green, H. (1967). Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc Natl Acad Sci U S A* **58**, 1104-11.

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. and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-7.

Wilmut, I. (1998). Cloning for medicine. *Sci Am* **279**, 58-63.

Xie, H., Ye, M., Feng, R. and Graf, T. (2004). Stepwise Reprogramming of B Cells into Macrophages. *Cell* **117**, 663-676.

Xu, C., Inokuma, M. S., Denham, J., Golds, K., Kundu, P., Gold, J. D. and Carpenter, M. K. (2001). Feeder-free growth of undifferentiated human embryonic stem cells. *Nat Biotechnol* **19**, 971-4.

Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., Hubner, K. and Schoeler, H. R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-94.

Yeom, Y. I., Ha, H. S., Balling, R., Scholer, H. R. and Artzt, K. (1991). Structure, expression and chromosomal location of the Oct-4 gene. *Mech Dev* **35**, 171-9.

Yoshimizu, T., Sugiyama, N., De Felice, M., Yeom, Y. I., Ohbo, K., Masuko, K., Obinata, M., Abe, K., Schoeler, H. R. and Matsui, Y. (1999). Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Development, Growth & Differentiation* **41**, 675-684.

Yuan, H., Corbi, N., Basilico, C. and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9**, 2635-45.

7 Publication

A part of this thesis has been in published:

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S. K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F. et al. (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445**, 671-5.

Acknowledgments

The nice thing about this section is one need not cite any references! As I myself have done with the theses of many others, I am fully aware of the fact that this section would receive the maximum ‘page views’. Jokes apart, I would now like to express my sincere thanks to all those who helped me navigate my PhD studies.

Christof, a bouquet of thanks for teaching me - to think big, to take some risks in life, to make not-so-bad power points, for encouraging me when I came up with some new ideas, for correcting me when I blundered and for having tolerated my idiosyncrasies all these days.

Guillermo = Impeccable mentor + Cheerful friend + a support I can always rely on. Thanks for everything you have been to me.

Rather than the availability of state of the art scientific facilities, the true asset of our group is a fantastic set of co-workers it has been bestowed with. Thank you all, and reprogramming subgroup in particular, for your cordiality and for providing the perfect niche for doing some exciting science.

I would like to Thank Joe Lewis and Peter Sehr of EMBL for helping with chemical genomics, Frank of DKFZ for mouse stem cell experiments, Guillermo, Gary and Gokul for critical reading of this thesis and Andrea for German translation of Summary. I wish to acknowledge the funding of the GK484 to support the initial three years of my studies.

It is difficult to imagine my PhD days at Heidelberg without the help and support of my close friends and buddies as well as the dynamic local Indian diaspora. Thank you all for making my stay here very enchanting and eventful.

அம்மா, அப்பா, அண்ணா, அக்கா- என்னை செதுக்கிய உங்களின் கடல்கொள்ளா அன்பிற்கும், அரவணைப்பிற்கும் முன்னால் சிறுமைப்பட்டுப்போய் நெகிழ்ச்சியுடன் நிற்கின்றேன்.