Characterisation of the pluripotency determinant Nanog

Adam Yates

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

2007

I declare that the work described in this thesis is my own, except where otherwise stated.

Adam Yates

Acknowledgments.

I would like to thank my supervisors Ian and Austin for the opportunity to study for my PhD in their labs, along with their constant guidance and support throughout the last three years. Big thanks go to Morag and Bianca who taught me all the molecular biology I needed to know, and ensured things ran smoothly in the lab. Dougie, with his unique sense of humour, made the long hours in the tissue culture labs enjoyable, and was always on hand to dish out advice whenever it was needed (and that was often). I also thank Nick, for calling a spade a spade, and generally keeping me on the right track with all things biochemical. There is a long list of people who have made my life in and around the lab enjoyable, but I particularly thank Tilo, Steve, José, and Jenny for advice and encouragement throughout my time in the lab. Outside of the lab, many people including Alistair, Craig, Chris, Jason, Maurice, John and Ollie have ensured I have always been entertained, either on the green baize, at the oche, or in the pubs of Edinburgh. I also thank my friends from back home in Manchester for their frequent trips north of the border.

I also owe a big debt of gratitude to my family, particularly Mum and Dad for their constant support and encouragement, and trips to Edinburgh throughout my studies-thank you!

Finally, I thank Laura, for her endless love and support during the last three years.

Abstract

Mouse embryonic stem (ES) cells are continuous undifferentiated cell lines derived directly from the inner cell mass of blastocysts. These cells have two defining characteristics: self-renewal and pluripotency. Self-renewal is the capacity to produce at least one identical daughter cell at each cell division, while pluripotency is the potential to differentiate into cellular derivatives of all three primary germ layers.

Nanog is a divergent homeodomain protein with the capacity to direct constitutive self-renewal in absence of otherwise obligatory cytokine stimulation. *Nanog* is expressed in the early mouse embryo and is essential for the specification of pluripotent cells. However the mechanism by which Nanog governs pluripotency is incompletely understood. In this thesis experiments are presented that further the functional characterisation of Nanog.

In the mouse embryo, *Nanog* is normally down regulated in cells prior to delamination and ingression through the primitive streak. To address the consequence of *Nanog* over-expression *in vivo*, a revertible *Nanog* over-expressing cell line has been generated which can be tracked in the embryo. Results show that the modest 2-3 fold increase in *Nanog* expression does not cause any overt phenotype at this stage and *Nanog* over-expressing cells can be detected in the mesoderm of mouse embryos.

Nanog is shown to exist in multimers in ES cells. The domain mediating multimerisation is identified as a tryptophan repeat motif and the functional consequence of deletion of this domain is investigated.

To identify Nanog partner proteins, a biotinylation tagging system in ES cells has been designed, constructed, and implemented. This led to the identification of putative Nanog partner proteins via mass-spectrometry. Three Nanog partner proteins, Esrrb, HDAC2, and Wdr5 have been confirmed by co-immunoprecipitation. In addition, the SLQQ motif within the Nanog homeodomain is shown to be the site of interaction between Nanog and Sall4. This SLQQ motif is found at a similar location in only one other homeodomain protein, Oct4. Consistent with these observations Sall4 is also shown to bind Oct4.

Abbreviations

a.a.	amino acids	Jak	Janus kinas
β-gal	β-galactosidase	kb	kilobase
β-geo	β-gal/*neo* fusion	*lacZ*	beta-galatosidase
BIO	biotin tag	LIF	leukaemia inhibitory
bps	base pairs		factor
cDNA	complimetary DNA	LIF-R	LIF receptor
ChIP	chromatin	MAPK	mitogen activated protein
	immunoprecipitation		kinase
CMV	cytomegalovirus	MOPS	3-N-morpholino
DNA	deoxyribonucleic acid		propanesulfonic acid
DsRed2	*Discomia sp.* Red2	mRNA	messenger RNA
	fluorescent protein	*neo*	*neomycin*
EC	embryonal carcinoma		*phosphotransferase*
EDTA	ethylenediamine	NLS	nuclear localisation signal
	tetraacetic acid	pac	puromycin acetyl
EG	embryonic germ		transferase
EMSA	electromobility shift	PAGE	polyacrylamide gel
	assay		electrophoresis
ERK	extracellular receptor	PBS	phosphate buffered saline
	kinase	PCR	polymerase chain reaction
ES	embryonic stem	PGCs	primordial germ cells
FCS	Foetal calf serum	pgk	phosphoglycerate kinase
g	gravitational force	RNA	ribonucleic acid
GCNF	germ cell nuclear factor	R	resistance
GFP	green fluorescent protein	rpm	revolutions per minute
ICM	inner cell mass	SDS	sodium dodecyl sulphate
IL-6	interleukin 6	SH2	src homology 2
IRES	internal ribosome entry	sIL6-R	soluble IL6 receptor
	site	TEV	tobacco etch virus

Contents.

Chapter 2: Materials and Methods.

Chapter 3: Biochemical Characterisation of Nanog.

Chapter 4: Investigation of the *in vivo* consequences of *Nanog* over-expression during mouse development.

Chapter 5: A biotin tagging strategy to identify Nanog interacting proteins.

Chapter 6: Concluding remarks.

References.

Appendices

List of Figures

List of Tables

Chapter 1.

Introduction

1.1 Pluripotent cell lines.

1.1.1 Discovery of embryonic stem cells.

Embryonic stem cells are characterised by the cardinal attributes of self-renewal and pluripotency. Self-renewal describes the capacity, at each cell division, to produce at least one identical daughter cell. Pluripotency is the ability to generate tissues of all three primary germ layers, as shown by teratoma formation (Evans and Kaufman, 1981; Martin, 1981), and in aggregation culture (Doetschman et al., 1985). That ES cells are truly pluripotent is most remarkably demonstrated following re-introduction into a host embryo to create a chimaera, and subsequent transmission of the ES cell genome through the mouse germline to form an entire mouse (Bradley et al., 1984).

Mouse ES cells are derived from pre-implantation embryos (Martin, 1981; Evans and Kaufman, 1981) with the embryonic source of ES cells being the epiblast (Brook and Gardner, 1997). The experiments leading to the derivation of ES cells were informed by studies in teratoma biology. Teratomas are disorganised solid tumours containing many different tissues representative of the three primary germ layers, and can be benign or malignant with the latter referred to as teratocarcinomas. In 1954, Leroy Stevens described the identification of an inbred mouse line that had a high frequency (1%) of testicular teratoma and these tumours were of "pleiomorphic character," that is to say, contained many different tissue types (Stevens and Little,

1954). At low frequency, these tumours could be transplanted, indicating the presence of undifferentiated cells. These cells are now known as EC (embryonal carcinoma) cells and represent the stem cell population sustaining the teratocarcinoma (reviewed by Andrews, 2002). Subsequent studies showed that transplanting genital ridges or pre-gastrulation embryos to ectopic sites could generate teratocarcinomas (Stevens, 1968; Stevens, 1964; Solter *et al.*, 1970). In the latter case the embryological origin of teratocarcinoma was shown to be the epiblast (Diwan and Stevens, 1976). Important in functionally defining a stem cell is the concept of clonality, that is to say that a single stem cell is able to produce each of the potency restricted cell types in the lineage it supports. Indeed, EC cells were shown to possess this property, with a single EC cell being able to generate complex teratomas (Kleinsmith and Pierce, 1964). Embryonal carcinoma cells can be cultured indefinitely whilst retaining this "histogenetic potentiality", and this caused excitement as these cells were hailed as a tool suitable for the study of differentiation pathways *in vitro* (Finch and Ephrussi, 1967). Furthermore, in 1974, it was demonstrated that EC cells, when reintroduced into the mouse embryo, are able to contribute to the developing foetus (Brinster, 1974). However, the ability of EC cells to efficiently differentiate *in vitro* and *in vivo* was ultimately found to be low. EC cells are often aneuploid, and as such are unable to generate functional gametes during meiosis, a trait that precludes passage through the germ-line. A great step forward was made in 1981 when embryonic stem (ES) cell lines were derived directly from the inner cell mass of mouse embryos without having passed through a teratocarcinoma (Martin, 1981; Evans and Kaufman, 1981). ES cells grow indefinitely in culture whilst been able to differentiate both *in vitro* and into complex

teratomas, that is, possess the key attributes of self-renewal and pluripotency. Unlike EC cells, embryonic stem cells are stably diploid (Evans and Kaufman, 1981) and are able to transmit their genome through the germline to generate an entire mouse (Bradley et al., 1984).

1.1.2   Other pluripotent cell types.

Mouse ES cells are not the only pluripotent cell type that that can be propagated *in vitro*. Explanted primordial germ cells (PGCs) can give rise to embryonic germ (EG) cell lines that can be cultured indefinitely *in vitro* (Resnick et al., 1992; Matsui et al., 1992). EG cells are able to contribute to mouse embryogenesis and pass through the germline (Labosky et al., 1994; Stewart et al., 1994). However, EG cells cannot be considered as ES cell equivalents, as EG cells undergo global erasure of imprints at the latter stages of PGC development, which leads to diminished developmental potential (Tada et al., 1998). More recently, there have been reports of the generation of pluripotent stem cells lines derived directly from neonatal testes, albeit with a very low frequency (Kanatsu-Shinohara et al., 2004). Hybrids formed by fusing somatic cells with EC cells retain pluripotentiality (Miller and Ruddle, 1977). Subsequent experiments have shown that pluripotency can be induced by fusing somatic cells with undifferentiated mouse ES cells (Tada et al., 2001) and human ES cells (Cowan et al., 2005), and that this ability can be increased by a key regulator of ES cells (Silva et al., 2006). More excitingly, direct ectopic expression of four ES cell specific transcription factors is sufficient to induce pluripotency, or re-program, adult somatic cells (Takahashi and Yamanaka, 2006).

1.1.3 Why are ES cells interesting to study?

1.1.3.1 ES cells: a tool to study developmental biology.

ES cells provide a useful tool to study mouse development. Firstly, ES cells represent an *in vitro* cell-line sharing many characteristics with the pluripotent cells of the early embryo. Whether ES cells represent a direct counterpart of an embryonic cell type or a cell culture artefact remains unclear (Smith, 2001; Buehr and Smith, 2003). Secondly, they can be used as an *in vitro* model for a variety of differentiation pathways (reviewed by Smith, 2001), and to produce a homogenous source of tissue specific stem cells (Conti *et al.*, 2005). Furthermore, mouse ES cells are amenable to genetic manipulation allowing functional analysis of gene mis-expression both in ES cells and in the mouse. More specifically, loss of gene function experiments can be studied using the technique of homologous recombination through gene targeting in ES cells (Thomas and Capecchi, 1987). Thus ES cells can act as a cellular vector to transmit the altered genome through the mouse germline (Thompson *et al.*, 1989) allowing gene function to be analysed both during development and in adult mice. An additional gene ablation strategy namely RNAi (RNA interference) has been shown to be efficient in ES cells (Kunath *et al.*, 2003). Coupled with the fact that ES cells can be expanded in culture indefinitely, this makes ES cells suitable for performing loss-of-function phenotypic screens, for example in the identification of key regulators of ES cell self-renewal (Ivanova *et al.*, 2006). Gene function can also be investigated in ES cells by over-expressing genes either via standard stable additive transgenesis or via high efficiency extra-chromosomal expression of episomal DNA (Gassmann *et al.*, 1995). The latter has proved useful in a number of experimental setups including the identification of single transcription factors that

can drive a particular differentiation program (Fujikura *et al.*, 2002), molecular dissection of the functional domains of key proteins in ES cells (Niwa *et al.*, 2002), and a gain of function screen for key ES cell regulators (Chambers *et al.*, 2003).

1.1.3.2 ES cells and their promise in regenerative medicine.

Mouse ES cells have been shown to differentiate into myriad tissues in *in vitro* aggregation cultures (Doetschman *et al.*, 1985; Martin, 1981), in teratomas (Martin, 1981; Evans and Kaufman, 1981), and *in vivo* (Bradley *et al.*, 1984; Beddington and Robertson, 1989). Embryonic stem cells have also now been derived from human blastocysts (Thomson *et al.*, 1998). The resultant cell lines generate complex teratomas when transplanted to ectopic sites in the mouse (Thomson *et al.*, 1998) and are able to differentiate into somatic cell types in culture (Reubinoff *et al.*, 2000). Mouse ES cell studies show that terminally differentiated cell types can be generated *in vitro*, which are functional upon reintroduction into mice (reviewed by Smith, 2001; reviewed by Prelle *et al.*, 2002). One striking example is that ES cell derived cultures of oligodendrocytes are able to re-myelinate axons in myelin deficient mice (Liu *et al.*, 2000). Clearly such a regenerative approach to medicine would be advantageous in the treatment of human degenerative diseases such as multiple sclerosis (a de-myelination disease). However, to achieve these goals more efficient protocols of derivation, propagation and differentiation of human ES cells are required, along with technical advances to either derive patient specific ES cell lines from somatic cell nuclear transfer embryos, or to entirely bypass the requirement for an embryo phase. In addition to usage in regenerative medicine, ES cells harbouring a disease causing mutation will be useful in the *in vitro* study of these diseases.

These considerations are however beyond the scope of this thesis. Key to the efficient and routine handling of human ES cells will be the identification of the molecular mechanisms underpinning self-renewal, which permit the propagation of uniformly undifferentiated cells whilst retaining the ability to efficiently and specifically differentiate on cue. Mouse ES cells can be readily propagated, and it is possible, indeed probable, that lessons learnt from studies into mouse ES cell self-renewal circuitry may be transferable to the human system.

1.2   Early mouse embryogenesis

The development of an adult mouse from a fertilised egg involves many rounds of cell-division and cell movements that are governed by dynamic, tightly controlled gene expression patterns, and perception and response to environmental cues. The development of the early mouse embryo has been well reviewed (Beddington and Robertson, 1999; Tam and Behringer, 1997) and the salient points with relevance to this thesis are outlined here, and depicted in Figure 1.1. Fertilisation followed by multiple cleavage rounds results in a ball shaped embryo known as a morula (E3.0) containing outer polar cells and inner apolar cells. At E3.5 the major morphological event of blastulation occurs whereby the distinct tissues of the ICM and the trophectoderm (TE) form. These subpopulations are non-interchangeable, with TE generating all the placental tissues and the ICM generating the embryo proper along with the supporting extra-embryonic endoderm and mesoderm (Gardner, 1983). One day later (E4.5), the primitive endoderm forms on the blastocoelic surface of the ICM, which will later generate the parietal and visceral endoderm. The remaining undifferentiated ICM tissue is now referred to as the epiblast and generates the entire
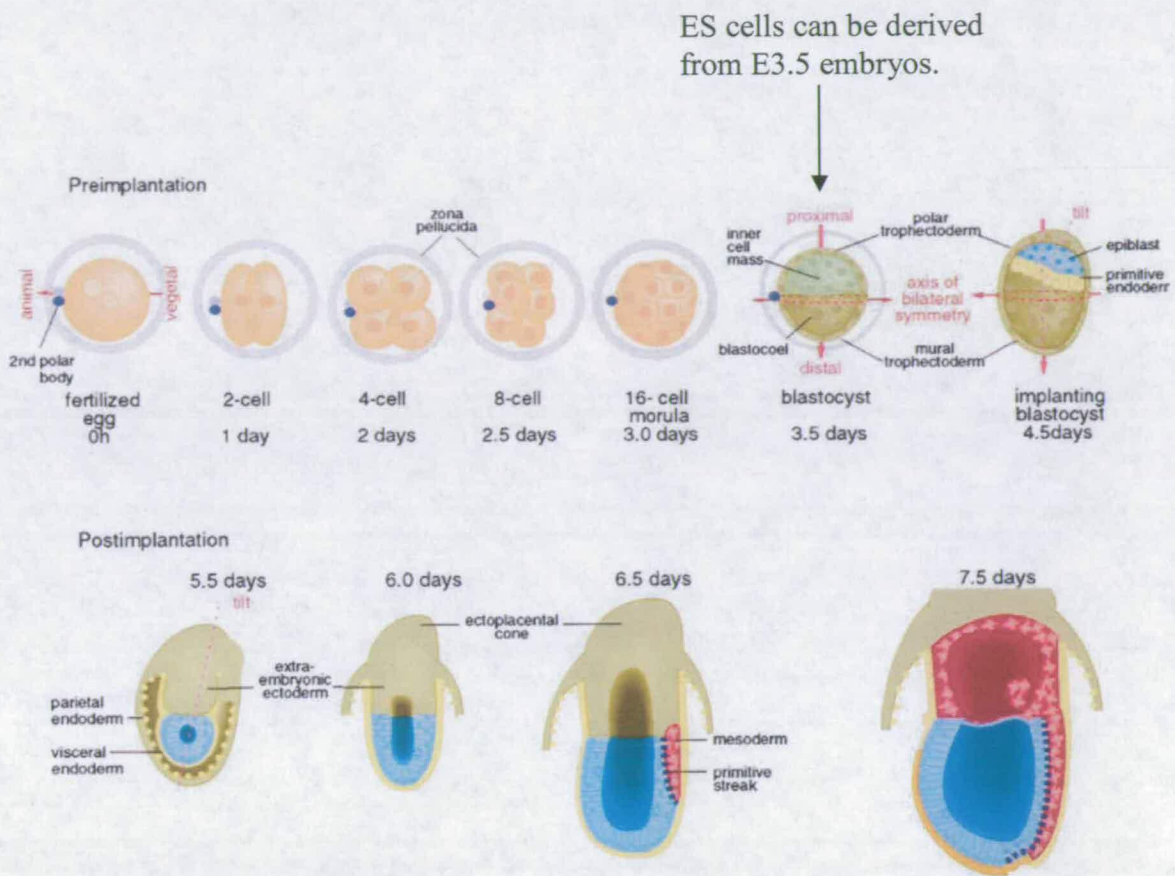
Figure 1.1- Schematic representation of early mouse development. Modified from Beddington and Robertson (1999).

After fertilisation, the embryo undergoes a number of cleavage divisions to generate a morula (ball shaped embryo) at E3.0. At E3.5 the major event of blastulation occurs. At E4.5 the primitive endoderm forms on the surface of the fluid filled blastocoelic cavity, with the remaining ICM tissue now being referred to as the epiblast. ES cells can be derived from the ICM of blastocysts. Over the next two days the embryo implants and the epiblast undergoes epithelialisation. At E6.5 the embryo begins gastrulation- a complex set of cellular movements during which cells of the epiblast generate mesoderm. These cellular movements begin at the proximal end of the epiblast and extend distally generating the primitive streak.

embryo proper and extra-embryonic mesoderm. Over the next two days the embryo implants into the uterine wall, the epiblast becomes epithelialised and adopts a cup shape and gastrulation begins at E6.5. Gastrulation describes a complex set of cellular movements whereby epiblast cells undergo an epithelial to mesenchymal transition to form mesoderm, with these movements starting at the proximal end of the epiblast and extending to the distal tip. This process demarcates the embryonic anterior-posterior axis with the proximal tip of the primitive streak marking the posterior end of the embryo.

1.3   Factors governing ES cell self-renewal.

1.3.1   Extrinsic factors governing ES cell self-renewal.

To both developmental biologist and clinician, it is important to define the minimal chemical conditions required for efficient ES cell propagation, in order to understand the true nature of the pluripotent cells of the early embryo and to be able to produce xeno-free human ES cell culture conditions for clinical use, respectively.

1.3.1.1   From feeder cells to the LIF/Stat3 pathway.

Initial derivation of mouse ES cells relied on co-culture with a feeder layer of STO fibroblasts to support propagation (Evans and Kaufman, 1981; Martin, 1981). This reliance on fibroblast support cells suggested feeders provided a signal to ES cells which favoured maintenance of the undifferentiated state during culture (Smith and Hooper, 1983). Indeed, medium conditioned by Buffalo rat liver cells was shown to inhibit mouse ES cell differentiation (Smith and Hooper, 1987). The effective component of the so called 'differentiation inhibiting activity' was identified as

leukaemia inhibitory factor (LIF) via fractionation of conditioned media (Smith *et al.*, 1988; Williams *et al.*, 1988). Furthermore, *lif*[/-] feeder cells have a reduced capacity to support ES cell self-renewal (Stewart *et al.*, 1992). LIF is a member of the IL6 (Interleukin 6) family of cytokines and indeed provision of IL6 together with soluble IL6 receptor can substitute for LIF both in ES cell cultures (Yoshida *et al.*, 1994) and during ES cell derivation (Nichols *et al.*, 1994). LIF acts by first binding to the low affinity LIF receptor, LIFR. This binary complex then interacts with gp130 to form a high affinity LIF receptor (Zhang *et al.*, 1997). The formation of this complex induces phosphorylation of Janus kinases which then phosphorylate specific tyrosine residues on the cytoplasmic tail of gp130 (Narazaki *et al.*, 1994; Stahl *et al.*, 1994). The phosphorylated gp130 acts as docking site for SH2 (Src homology 2) domain containing proteins which are themselves then phosphorylated by JAKs (Lutticken *et al.*, 1994; Stahl *et al.*, 1994). The upshot of this intricate membrane associated assembly is the initiation of signalling cascades including phosphorylation and activation of Stats (Signal transducer and activator of transcription) (Lutticken *et al.*, 1994; Stahl *et al.*, 1995) and activation of the Ras-MAPK (mitogen activated protein kinase) pathway (Boulton *et al.*, 1994; Yin and Yang, 1994; Sheng *et al.*, 1997). When phosphorylated Stats translocate to the nucleus, they exert their function as activators of transcription (Ihle, 1996). In mouse ES cells, the key positive event downstream of the LIF receptor is activation of Stat3 which is required for efficient self-renewal (Niwa *et al.*, 1998). In addition, using a conditionally activatable form of Stat3, Stat3 was shown to be sufficient for mouse ES cell self-renewal (Matsuda *et al.*, 1999) although this not strictly true, due to the continued requirement for foetal calf serum in the culture. LIF also stimulates the Ras-MAPK signalling cascade,

although this is not required for, and indeed antagonises ES cell self-renewal (Burdon *et al.*, 1999; reviewed by Burdon *et al.*, 2002). It has also been suggested that LIF dependent ERK suppression occurs in mouse ES cells and is transduced through the PI3K (phosphoinositide-3 kinase) signalling pathway (Paling *et al.*, 2004). Although *lif, lifr, gp130* are co-expressed in the early mouse embryo in a reciprocal pattern suggestive of paracrine signalling (Nichols *et al.*, 1996), targeted disruption of either *lif, lifr,* or *gp130*, does not impair normal blastocyst development (Yoshida *et al.*, 1996; Stewart *et al.*, 1992; Ware *et al.*, 1995). The gp130 signalling pathway is only required for blastocyst development during diapause, with this adaptive physiological mechanism likely providing the basis for ES cell responsiveness to gp130 signalling (Nichols *et al.*, 2001). LIF independent paracrine signalling is able to support self-renewal in mouse ES cells to a certain degree (Dani *et al.*, 1998) and it is possible that the same mechanism maintains an undifferentiated ICM population in non-delayed blastocysts. Interestingly, human ES cells, although expressing the required components of the LIF/Stat3 pathway, do not require its activation for efficient self-renewal (Humphrey *et al.*, 2004; Daheron *et al.*, 2004). This may be due to either additional effective signalling pathways or intrinsic factors being present at higher effective concentrations.

1.3.1.2   From FCS to BMP.

Defining the contribution of FCS to ES cell culture would permit self-renewal in a completely chemically defined media. Recently, BMP2/4 was identified as the critical component of FCS which, together with LIF allows efficient self-renewal without the neural differentiation that normally ensues during culture with LIF alone

(Ying *et al.*, 2003). BMP signals are transduced by the Smad proteins (reviewed by Massague and Wotton, 2000), and their effective targets for mouse ES cell self-renewal are the Id (Inhibitor of differentiation) proteins. Indeed, forced expression of Id proteins can circumvent the requirement for continued BMP stimulation (Ying *et al.*, 2003). The gene targets of Id proteins in ES cells are unclear although it is thought they may prevent premature neural differentiation via repression of pro-neural factors such as Mash1 (Ying *et al.*, 2003). In apparent contradiction, an alternative explanation for the mechanism of maintaining self-renewal via BMP signalling in ES cells has been offered which is via inhibition of the ERK and p38 MAPK pathways (Qi *et al.*, 2004). Forced expression of Smad1/4 led to non-neural differentiation even in the presence of LIF, which suggests the balance of the LIF/Stat3 and BMP/Smad signals is critical for efficient self-renewal (Ying *et al.*, 2003). Stat3 and active Smad1 can physically interact in neuroepithelial cells and can co-operate in transcriptional regulation (Nakashima *et al.*, 1999). This physical association is also detected in mouse ES cells (Ying *et al.*, 2003), although whether it is functionally significant remains unclear.

1.3.1.3   Wnt signalling and ES cell self-renewal

Wnt signalling has previously been shown to be important for self-renewal of haematopoeitc stem cells (Reya *et al.*, 2003), and recently has also been implicated in ES cell self-renewal (Sato *et al.*, 2004). This study used a pharmacological inhibitor of GSK3 (glycogen synthase kinase 3), which effectively activates the canonical wnt pathway to 'maintain' pluripotency. However, GSK3 is known to be involved in other pathways and therefore the observed phenotype may not be directly

attributable to wnt signalling. Furthermore, these experiments are carried out in high density culture on feeder cells and are not passaged throughout the duration of the experiment. Further experiments in feeder-free, chemically defined media, that involve clonal expansion and serial passaging, will be required to unequivocally add wnts to the ES cell self-renewal signalling repertoire.

1.3.2 Intrinsic factors governing ES cell self-renewal.

Recently, four proteins have been proposed to play central roles in directing transcriptional networks that define both mouse and human ES cell self-renewal and pluripotency (Boyer *et al.*, 2005; Loh *et al.*, 2006; Wu *et al.*, 2006). These are the divergent homeodomain protein Nanog, the POU (Pit/Oct/Unc) domain Oct4, the HMG (high mobility group) containing protein Sox2, and the spalt family protein Sall4. These proteins will be discussed here along with other proteins important for ES cell renewal that are studied in this thesis. Particular attention will be given to Nanog- the major focus in this thesis.

1.3.2.1 Oct4

Oct4 is homeodomain protein of the POU class of transcription factors, which can regulate a wide range of target genes (Saijoh *et al.*, 1996; Matoba *et al.*, 2006). A subset of Oct4 target genes in ES cells such as *fgf4* are co-regulated by Sox2 (Yuan *et al.*, 1996). Together, Oct4 and Sox2 with Nanog have been localised to many common gene targets in mouse and human ES cells (Loh *et al.*, 2006; Boyer *et al.*, 2005). Oct4 is expressed in the oocyte and expression is then limited to the blastomeres, pluripotent cells of the early embryo and germ cells (Rosner *et al.*,

1990; Scholer *et al.*, 1990b; Yeom *et al.*, 1996; Pesce *et al.*, 1998). In addition, some nascent mesodermal cells express *Oct4* mRNA transiently after ingression through the primitive streak (Yeom *et al.*, 1996). Oct4 is a key regulator of the pluripotent cell type as evidenced by the trophoblastic differentiation of $Oct4^{-/-}$ blastocysts (Nichols *et al.*, 1998) and upon conditional repression in mouse ES cells (Niwa *et al.*, 2000). Recently, experiments have shown that the mechanism of Oct4 mediated suppression of trophectoderm differentiation relies on physical and genetic interaction with a second transcription factor, Cdx2 (Niwa *et al.*, 2005). Oct4 and Cdx2 exist in a state of reciprocal inhibition whereby Cdx2 is prevented from driving trophectoderm differentiation by both Oct4 suppression of *Cdx2* expression along with blockade of Cdx2 activity in an Oct4-Cdx2 repressive complex, whilst Cdx2 inhibits Oct4 auto-regulation (Niwa *et al.*, 2005; Smith, 2005). It has been noted that constitutive *Oct4* expression cannot negate the requirement for gp130 signalling. On the contrary, elevated *Oct4* expression initiates a developmental process that mirrors that generated by LIF withdrawal in which cells expressing markers of primitive endoderm and mesoderm are generated (Niwa *et al.*, 2000). This may reflect the *in vivo* situation whereby a transient burst of Oct4 expression is observed in the nascent primitive endoderm (Palmieri *et al.*, 1994). Based on these data the existence of another factor which may be active in the absence of LIF, yet is maximally effective in the presence of LIF has been suggested (Niwa, 2001; Chambers, 2004). This unidentified factor is thought to be important for limiting Oct4 activity and for the maintenance of the pluripotent state.

1.3.2.2   Sox2

Sox2 is a HMG transcription factor which acts together with Oct4 by binding a composite Oct/Sox binding site to regulate transcription of target genes (e.g. *fgf4*) both in ES cells and in pluripotent cells *in vivo* (Yuan *et al.*, 1996; Ambrosetti *et al.*, 1997). Functionally important Oct/Sox sites have also been found at other genes with ES cell specific expression including *Utf1* (Nishimoto *et al.*, 1999), as well as Oct4 (Okumura-Nakanishi *et al.*, 2005; Chew *et al.*, 2005) and Sox2 (Chew *et al.*, 2005; Tomioka *et al.*, 2002) themselves. Composite Oct/Sox sites are non-palindromic and occur with conserved comparative directionality to permit side chain interactions between the HMG domain of Sox and the POU-specific domain of Oct which stabilise the Oct-Sox-DNA ternary complex (Williams *et al.*, 2004; reviewed by Chambers, 2005). *Sox2*$^{-/-}$ inner cell masses cannot give rise to ES cells when explanted *in vitro* (Avilion *et al.*, 2003). *Sox2*$^{-/-}$ embryos develop until E6.5, considerably later than defects in *Oct4*$^{-/-}$ embryos became apparent. It is currently unclear whether this difference reflects the fact that Sox2 may be activating only a subset of Oct4 target genes or whether an earlier requirement for Sox2 is satisfied by long-lived maternal protein (Avilion *et al.*, 2003). Whether Sox2 in necessary for ES self-renewal remains unclear, and experiments to acutely remove *Sox2* during ES cell culture could address this question. Certainly forced *Sox2* expression does not effect increased self-renewal, and rather increases the efficiency of neural differentiation in appropriate culture conditions (Zhao *et al.*, 2004).

1.3.2.3   Sall4

Sall4 is a zinc finger transcription factor of the spalt family that was first identified in *Drosophila* as region specific homeotic gene (Kuhnlein *et al.*, 1994). Sall4 has recently been identified as an interacting partner protein of Nanog in mouse ES cells (Wu *et al.*, 2006). Nanog and Sall4 can reciprocally transcriptionally activate each other suggesting the existence of a transcriptional feed-forward loop between *Nanog* and *Sall4* (Wu *et al.*, 2006). Nanog and Sall4 co-occupy many genomic loci including *Nanog, Oct4, Sox2,* and *Esrrb* suggesting Nanog and Sall4 may act together to regulate some target genes (Wu *et al.*, 2006). Genetic disruption of both *Sall4* alleles reveals a cell-autonomous requirement for Sall4 in the epiblast, and no *Sall4^{-/-}* ES cell lines could be established from *Sall4^{+/-}* intercrosses (Elling *et al.*, 2006). However sequential targeting of *Sall4* alleles in ES cells revealed that proliferation compromised *Sall4^{-/-}* clones can be obtained, albeit only at very low frequency (2%), due to preferential recombination at the already mutated allele (Sakaki-Yumoto *et al.*, 2006). This suggests that the homozygous *Sall4* mutation is detrimental to ES cells. This proliferation defect however is not due to modulation of expression of two genes previously implicated in controlling ES cell proliferation, *Utf-1* (Nishimoto *et al.*, 2005) and *Eras* (Takahashi *et al.*, 2003), as northern analysis shows *Utf-1* and *Eras* expression levels are unchanged in *Sall4^{-/-}* ES cells (Sakaki-Yumoto *et al.*, 2006). Sall4 also activates *Oct4* in ES cells (Zhang *et al.*, 2006). *In vivo* siRNA reduction of *Sall4* expression in the one cell embryo results in decreased *Oct4* expression and expansion of the *Cdx2* expression domain into the ICM (Zhang *et al.*, 2006). It therefore seems Sall4 may, through Oct4, modulate *Cdx2* expression and may be important in the first lineage determination decision *in vivo*. This data

perhaps explains why *Sall4*[-/-] trophoblast stem (TS) cell lines (Tanaka *et al.*, 1998), yet not ES or extraembryonic endoderm stem (XEN) cell lines (Kunath *et al.*, 2005) can be obtained from *Sall4*[-/-] blastocysts (Elling *et al.*, 2006).

1.3.2.4   Nanog.

1.3.2.4.1   Identification of Nanog.

Nanog was identified in two concurrent screens for functional proteins important in mouse ES cell self-renewal. The first involved an *in silico* subtraction method which identified Nanog as one of several transcripts specifically expressed in ES cells (Mitsui *et al.*, 2003). The second strategy involved directly selecting cDNAs that were capable of directing self-renewal in the absence of otherwise obligatory LIF stimulation (Chambers *et al.*, 2003). It was subsequently shown that *Nanog* over-expression can also circumvent the requirement for BMP/Smad signalling as well as Stat3 activation, allowing efficient self-renewal in a completely defined minimal media (N2B27) (Ying *et al.*, 2003) as well as conferring resistance to pro-differentiation stimuli (Chambers *et al.*, 2003).

1.3.2.4.2   Nanog protein

Nanog is a 305 amino acid polypeptide which, in simple terms can be considered a three domain protein containing a 96 residue serine rich N-terminal region, a divergent homeodomain, and a 150 residue C-terminal region (Mitsui *et al.*, 2003; Chambers *et al.*, 2003) (Figure 1.2). Homeodomain proteins are particularly interesting to the developmental biologist as they have numerous roles development and evolution (Gehring, 1987). The Nanog homeodomain is most closely related to

```
           10              20              30    ·        40
MS VGL PGP HS LPS S E EAS NS GN AS S MP AVF HP EN YS CLQG
                     N-terminal domain
           50              60    ·        70          80
S ATE ML CT E AAS PRPS S EDL PL QGS PDS S TS PKQKLS S PE

           90     ·        100            110           120
ADKGP EEEE NKVLAR KQ KMR T VF S QAQL CAL KD R F QKQKY
                     Homeodomain
           130            140            150           160
L S LQQMQELS S I LNLS YKQVKTWF QN QRMKC KRWQ KNQWL

           170            180            190           200
KTS NGLI QKGS AP VE YP S I HCS YP QGY LVN AS GS L S MWGS
                     C-N
           210            220            230           240  ·
QT WT N P T WS S QT WT N P T WN N QT WT N P T WS S QAWT A QS WN G
                  Tryptophan repeat
           250            260            270           280
QP WN A APL HN E GE DE L QP YV QL QQN E S AS DL E VN L EATRE
                     C-C
           290            300     305
S HAHES TP QAL EL F LN YS V TP P GE I
```

Figure 1.2- Amino acid sequence of mouse Nanog protein.

The coloured boxes show the boundaries of putative domains within Nanog. The C-terminal domain begins at K156 and is further divided due to the presence of the tryptophan repeat into C-N (K156-M197), tryptophan repeat (W198-W243), and C-C (N244-I305).

the NK2 family yet lies outwith this family due to the lack of characterisitic sequence motifs (Lints *et al.*, 1993). Nanog protein localises to the nuclear compartment (Chambers, 2005), and the human Nanog homeodomain has been shown to contain a nuclear localisation signal (Do *et al.*, 2006). The C-terminal half of Nanog contains an unusual sequence repeated ten times, characterised by a tryptophan residue repeated every 5th amino acid. Orthologues of Nanog are recognisable in many organisms including human (Chambers *et al.*, 2003), rat (Chambers *et al.*, 2003), pig (Yang *et al.*, 2004), and chicken (Canon *et al.*, 2006). Furthermore, human Nanog protein is functionally conserved, having a reduced but detectable ability to direct cytokine independent self-renewal in mouse ES cells (Chambers *et al.*, 2003). In Gal4 DNA binding domain fusion experiments, both the N-terminal domain and C-terminal domain have been shown to possess transactivation potential with the latter shown to be more potent (Pan and Pei, 2003). Subsequently, the C-terminal domain was found to contain two distinct transactivation domains; one being the 50 amino acid tryptophan repeat in which at least some of the tryptophan residues are functionally important, and the second being the C-terminally adjacent 58 residues (Pan and Pei, 2005). These studies were performed using both Gal4 Nanog fusions reported by a Gal4 responsive promoter and also a reporter driven by a multimerised Nanog binding site identified by SELEX (systematic evolution of ligands through exponential enrichment) (Pan and Pei, 2005). It will be interesting to examine whether the transactivation domains are functional at endogenous target genes in ES cells. Interestingly, the C-terminal (but not the N-terminal) transactivation potential is conserved in human Nanog (Oh *et al.*, 2005). Although not yet reported, Nanog may also possess repressive function, possibly dependent on specific interacting

partners as has been shown for distinct multi-subunit complexes of Gata-1 (Rodriguez *et al.*, 2005).

1.3.2.4.3   *Nanog* expression *in vivo*

Nanog is expressed, and has a functional role in the early mouse embryo (Mitsui *et al.*, 2003; Chambers *et al.*, 2003). Nanog mRNA can first be visualised at the compacted morula stage with expression being restricted to the apolar inner cells (Chambers *et al.*, 2003) which, one day later will form the ICM (inner cell mass) (Johnson and Ziomek, 1981). The expression of *Nanog* at this stage is complementary to that of *Cdx2*, which is required for the specification and differentiation of the trophectoderm (Strumpf *et al.*, 2005). *Nanog* is expressed in the ICM but is down-regulated immediately prior to implantation (Chambers *et al.*, 2003). The *in situ* hybridisation data in Figure 1.3 shows *Nanog* expression in the early mouse embryo. In the early post implantation egg-cylinder stage embryo, *Nanog* mRNA is detectable as a gradient with the highest mRNA levels in the proximal posterior region of the embryo (Hart *et al.*, 2004). *Nanog* expression is rapidly down-regulated as cells delaminate and ingress through the primitive streak forming mesoderm (Hart *et al.*, 2004; Morkel *et al.*, 2003). Whether *Nanog* down-regulation is required for mesoderm formation is a question that is addressed in this thesis. Later in development, *Nanog* is co-expressed with *Oct4* in the pluripotent PGCs (primordial germ cells) (Chambers *et al.*, 2003; Yamaguchi *et al.*, 2005) which go on to generate functional gametes. *Nanog* expression has been reported in some somatic cell types (Yan *et al.*, 2005; Carlin *et al.*, 2006) However, as these data

anterior ← → posterior

anterior ← → posterior
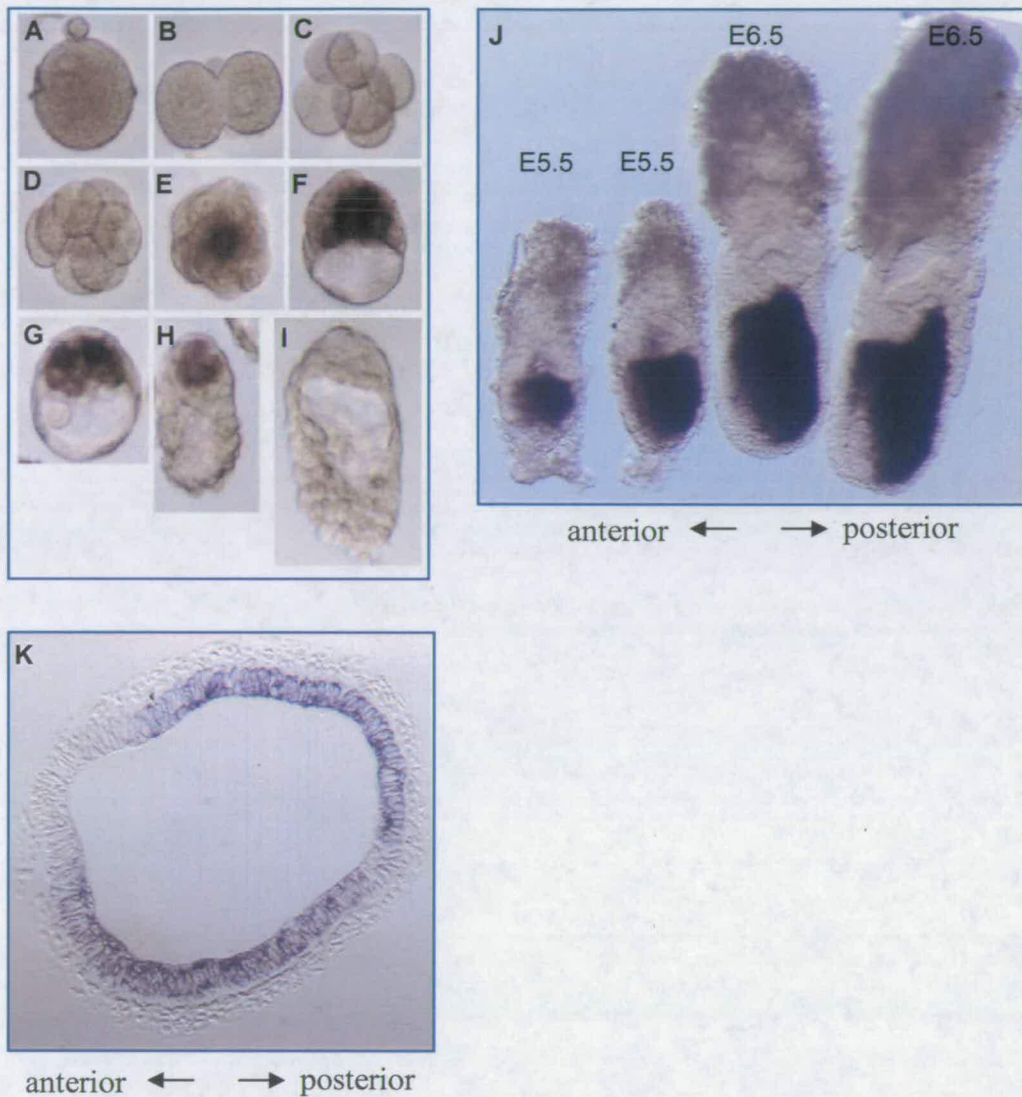
Figure 1.3- *In situ* **hybridisation showing** *Nanog* **mRNA expression during mouse embryogenesis. Taken directly from Chambers** *et al* **(2003)(A-I), and Chambers' unpublished data (J+K).**

In situ hybridisation was used to detect Nanog mRNA which is visulaised as a purple signal. Panels show one cell (A), two cell (B), 6 cell (C), 8 cell (D), late morula (E), early blastocyst (F), expanded blastocyst (G), hatched blastocyst (H), implanting blastocyst (I), E5.5 and 6.5 embryos (J), and a transverse section through an E7.5 embryo (K).

Nanog expression commences in the compacted morula and in maintained in the ICM prior to down-regulation at implantation. *Nanog* is re-expressed in the post-implantation embryo with highest expression in the proximal posterior region of the embryo (J), and expression is extinguished upon de-lamination and ingression through the primitive streak, (K).

generally rely on RT-PCR, the existence of *Nanog* retrogenes may mean this detection is artefactual (Robertson *et al.*, 2006; Booth and Holland, 2004).

1.3.2.4.4   Consequences of *Nanog* mis-expression.

Nanog was identified as a molecule able to direct cytokine independent mouse ES cell self-renewal, and elevated expression of *Nanog* endows mouse ES cells with the properties of increased self-renewal efficiency and resistance to pro-differentiation stimuli (Chambers *et al.*, 2003). These properties are exemplified by the cell line EF4, (*Nanog* over-expressing) which is BMP4 and LIF independent, and when reverted to the wild-type via Cre recombinase excision of the *loxP* flanked additive Nanog transgene, also reverts to LIF and BMP dependency (Chambers *et al.*, 2003; Ying *et al.*, 2003). Quantitative western blotting shows the extent of Nanog over-expression necessary to release mouse ES cells from LIF dependency is 5-6 times the wild-type expression level (Yates and Chambers, 2005). Nanog is therefore a key mediator of mouse ES cell self-renewal that is normally expressed at limiting concentrations but, when over-expressed, can 'lock-in' ES cell identity. A paradigm for the importance of transcription factors overcoming an expression level threshold has been shown in the haematopoeitic stem cell system in which the homeodomain protein HoxB4 shows maximal self-renewal at high expression levels (reviewed by Klump *et al.*, 2005). *NANOG* over-expression has also been suggested to bypass the requirement for feeders and conditioned media in human ES cells, although this may be due the irreversible gene expression profile changes that occur in these experiments (Darr *et al.*, 2006).

Classical gene targeting of *Nanog,* performed in ES cells grown on a fibroblast feeder layer resulted in loss of pluripotency and differentiation into extra-embryonic endoderm cells which expressed *Gata6* (Mitsui *et al.*, 2003). In addition, Nanog$^{-/-}$ embryos do not support an epiblast compartment with extraembryonic endoderm differentiation ensuing (Mitsui *et al.*, 2003). *Gata6* is normally expressed in (Morrisey *et al.*, 1996), and required for (Koutsourakis *et al.*, 1999) the developing primitive endoderm, and indeed forced expression of *Gata6* in mouse ES cells drives extra-embryonic endoderm differentiation (Fujikura *et al.*, 2002). It has therefore been suggested that Nanog may act to repress *Gata6* in ES cells (Mitsui *et al.*, 2003; Chambers and Smith, 2004; Ralston and Rossant, 2005). Coupled with the knowledge that *Oct-4* expression is required to prevent trophectoderm differentiation, this may seem a neat mechanism in which two proteins, namely Oct4 and Nanog maintain the identity of pluripotent cells both *in vitro* and *in vivo* by preventing differentiation into the two alternative cell fates of trophectoderm and primitive endoderm, respectively. However, recent experiments involving the acute deletion of *Nanog* reveal it is non-essential for the maintenance of pluripotency in healthy mouse ES cell cultures (Chambers unpublished). Rather, *Nanog* expression appears to oscillate in ES cells such that transiently *Nanog* negative cells are provided with a "window of indecision" during which they may perceive environmental signalling cues and differentiate accordingly (Chambers unpublished). In the absence of pro-differentiation stimuli the transiently *Nanog* negative cells can re-express *Nanog* and are thereby shielded from differentiation (Chambers unpublished). This explains how *Nanog*$^{-/-}$ ES cells may be maintained in an undifferentiated state when adhering to a stringent pro-self-renewal culture regimen. Nanog is however required *in vivo* both

for the establishment of pluripotent cells in the pre-implantation epiblast (Mitsui *et al.*, 2003) and for the completion of germ cell development following entry of PGCs to the genital ridge (Chambers unpublished). This study also suggests the re-expression of *Nanog* in the post-implantation embryo may serve to protect the pluripotent post-implantation epiblast cells from premature differentiation. A complementary experimental approach shows that *Nanog* expression in ES cells plays a significant role in re-establishment of the pluripotent state when fused to tissue restricted non-pluripotent stem cells or terminally differentiated somatic cells (Silva *et al.*, 2006). Chambers *et al* note that the common theme between Nanog enhancing reprogramming of non-pluripotent cells via cell fusion, and requirement of Nanog for PGC maturation is epigenetic erasure, which suggests that Nanog may be key to this process (Chambers unpublished).

1.3.2.4.5   Regulation of *Nanog*

It appears that an appropriate level of *Nanog* expression is required to balance prevention of precocious differentiation, with the need to permit differentiation upon receipt by the cell of appropriate environmental signals. An Oct/Sox site has been identified in the *Nanog* gene (Chambers and Smith, 2004), within 180 base pairs of the transcription initiation site (Chambers, 2005; Wu da and Yao, 2005). Given the functional importance of Oct/Sox sites in the promoters of other ES cell specific genes (Ambrosetti *et al.*, 1997; Nishimoto *et al.*, 1999; Tokuzawa *et al.*, 2003) including *Oct4* and *Sox2* themselves (Okumura-Nakanishi *et al.*, 2005; Chew *et al.*, 2005; Tomioka *et al.*, 2002) these sites were further investigated. Indeed, endogenous Oct4 and Sox2 are bound to the *Nanog* promoter in ES and EC cells

(Rodda et al., 2005; Kuroda et al., 2005). Moreover, transient transfection assays in ES and EC cells suggest that both Oct4 and Sox2 positively regulate *Nanog* expression (Rodda et al., 2005; Kuroda et al., 2005). Oct4 is not however required for *Nanog* transcription in vivo, as evidenced by *Nanog* expression in all blastocysts of an Oct4$^{+/-}$ intercross (Chambers et al., 2003). It remains unclear whether the apparent contradiction of in vitro data (Rodda et al., 2005; Kuroda et al., 2005) and in vivo data (Chambers et al., 2003) reflects a difference in Oct/Sox regulation of *Nanog* during establishment as opposed to maintenance of pluripotency.

In addition to Oct4 and Sox2, the forkhead transcription factor FoxD3 can also activate the *Nanog* promoter in transient transfection assays (Pan et al., 2006). Chromatin immunoprecipitation experiments suggest this effect could be direct as epitope tagged FoxD3 can bind the *Nanog* promoter (Pan et al., 2006). *FoxD3$^{-/-}$* embryos are indistinguishable from the wildtype at the blastocyst stage, although later in development at E6.5, *FoxD3$^{-/-}$* embryos have a defect in the epiblast (Hanna et al., 2002). This suggests that FoxD3 may be dispensable in cells of the ICM yet required for maintenance of pluripotent epiblast cells after implantation. In ES cells, reciprocal positive regulation of *Nanog* and *FoxD3* expression (Loh et al., 2006; Pan et al., 2006) may be important for lineage priming as FoxD3 can activate the endodermal genes *FoxA1* and *FoxA2*. However, precocious differentiation is prevented via co-repression of FoxD3 activity by a physical interaction with Oct4 (Guo et al., 2002).

Clearly however, some mechanism of *Nanog* repression is required for epiblast cells and ES cells to be able to differentiate. The transcription factors p53 (Lin *et al.*, 2005), GCNF (Gu *et al.*, 2005), and Tcf3 (Pereira *et al.*, 2006) may be able to repress *Nanog*. However, whether these transcription factors can repress *Nanog* in self-renewing wildtype ES cells is unclear. $p53^{-/-}$, $GCNF^{-/-}$, and $Tcf3^{-/-}$ ES cells are all capable of differentiation suggesting that either additional factors, or a combination of these transcription factors is required for complete *Nanog* repression (Lin *et al.*, 2005; Gu *et al.*, 2005; Pereira *et al.*, 2006). In ES cells, LIF stimulation does not result in a global increase in Nanog levels (Chambers *et al.*, 2003). Recent work suggests that Stat3 in combination with Brachyury (*T*) can directly bind and activate the *Nanog* promoter when ES cells are cultured in reduced LIF conditions in order to prevent mesoderm differentiation (Suzuki *et al.*, 2006a; Suzuki *et al.*, 2006b). However, under normal self-renewing conditions Brachyury and Stat3 do not regulate *Nanog* expression in ES cells (Suzuki *et al.*, 2006b)

Constitutive activation of canonical wnt signalling enhanced ES cell self-renewal in short term, high density culture (Sato *et al.*, 2004) possibly via β-catenin mediated up-regulation of *Nanog* expression (Takao *et al.*, 2006). Genetic disruption of the wnt signalling components *β-catenin* and *wnt3a* result in the *in vivo* down-regulation of *Nanog* at E6.5 and the mesoderm in these embryos does not form (Morkel *et al.*, 2003). Whether the wnt/ β-catenin pathway impinges directly on *Nanog* in the mouse embryo is not yet established.

The Grb2-Mek-Erk pathway is also thought to repress *Nanog* expression (Hamazaki *et al.*, 2006) and this may reflect the increased *Nanog* expression and lack of primitive endoderm differentiation of *Grb2*$^{-/-}$ blastocysts (Chazaud *et al.*, 2006). Such an Erk mediated *Nanog* repression may explain the observation that ES cells grown in the presence of Mek inhibitors exhibit enhanced self-renewal (Burdon *et al.*, 1999). The various regulatory factors governing *Nanog* expression are depicted schematically in Figure 1.4.

1.3.2.4.6 Regulation by Nanog

The transcriptional networks of Nanog together with Oct4 and Sox2 have begun to be elucidated in mouse and human ES cells, with these three proteins co-occupying many genomic sites (Loh *et al.*, 2006; Boyer *et al.*, 2005). In fact Nanog has been proposed to be bound to nearly 1500 genes in mouse ES cells. Exactly how many of these targets are truly dependent on Nanog binding for expression is unclear at present. One example of a gene co-regulated by Nanog, Oct4 and Sox2 is *Esrrb*, the orphan nuclear receptor. Esrrb appears functionally important for ES cells as specific RNAi mediated knock-down of *Esrrb* results in morphological flattening of ES cells concomitant with a loss of expression of alkaline phosphatase, a marker of the undifferentiated state (Loh *et al.*, 2006). In addition, *Nanog* over-expression induces increased *Esrrb* expression (Loh *et al.*, 2006).

The consensus DNA sequence recognised by Nanog was identified by systematic evolution of ligands by exponential enrichment (SELEX) and is characterised by an ATTA typical of homeodomain recognition sequences (Mitsui *et al.*, 2003). The
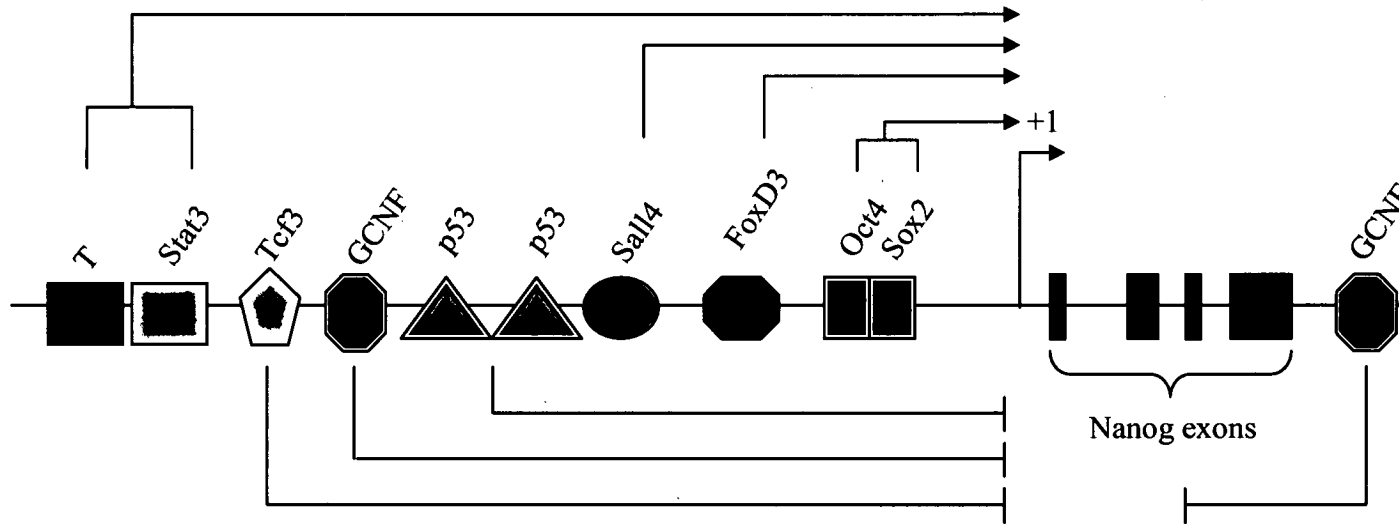
Figure 1.4- Schematic representation of all known transcription factors that regulate expression of the *Nanog* promoter.

Modified from Pan and Pei (2006). Coloured shapes indicate binding sites of regulatory factors. Black rectangles represent *Nanog* exons. The binding sites are not spaced to scale. Direct binding has been demonstrated by either ChIP data or EMSA or both. Primary references are provided in the main text (Section 1.3.2.4.5).

Nanog binding site has been identified in the enhancers of *Gata6* and *Rex1* (Mitsui *et al.*, 2003). Furthermore, ChIP data shows that Nanog can bind the promoter region of the pluripotency asscociated gene *Rex1* (Shi *et al.*, 2006) and *Gata6*, a gene expressed in primitive endoderm (Wang *et al.*, 2006). It has not yet been shown that Nanog binding to *Gata6* results in *Gata6* repression. However, such a mechanism could explain the primitive endoderm cell types generated upon *Nanog* deletion in the mouse embryo or by classical gene targeting in ES cells. It is possible that Nanog is acting both as an activator and repressor; repressing pro-differentiation genes such as *Gata6* whilst activating pluripotency associated genes such as *Rex1*.

Nanog and Oct4 have been proposed to exist in a positive feedback loop with the spalt family transcription factor *Sall4* (Wu *et al.*, 2006; Zhang *et al.*, 2006). Genetic disruption of *Sall4* is not compatible with efficient mouse ES cell self-renewal (Elling *et al.*, 2006; Sakaki-Yumoto *et al.*, 2006), suggesting the feed-forward loop involving Oct4, Nanog and Sall4 may be important for maintaining mouse ES cell self-renewal. Although LIF stimulation does not globally increase Nanog levels in ES cells, growth in reduced LIF concentrations may lead to transient up-regulation of *Nanog* by Brachyury and Stat3 to prevent premature mesoderm differentiation (Suzuki *et al.*, 2006a). It has been proposed that this involves Nanog negatively and indirectly regulating *Brachyury* by physically sequestering the active Smad1 required for *Brachyury* expression (Suzuki *et al.*, 2006b).

1.4 Protein-Protein interaction technology.

1.4.1 Background

A detailed understanding of the dynamic nature of the proteome is required to dissect cellular functions. Since 1979 when the first protein-protein interactions were identified (Lane and Crawford, 1979; Linzer and Levine, 1979), proteins have been identified in large multi-protein complexes in many systems. To characterise a protein of interest, detailed knowledge of its interacting partners is useful as this offers insight into the biological function of the protein. Many techniques have been devised to screen for interacting partners and these will be summarised here. These systems generally fall into one of two categories; (i) affinity based approaches in which proteins of interest are isolated from complex mixtures based on their binding to particular ligands. Identification of the purified protein and associated factors is then attained by using antibodies in a candidate based approach or, more recently, by mass-spectrometry. (ii) Functional complementation of a genetic system whereby activation of a transcription unit is dependent on protein-protein interaction between a bait (protein of interest), and a prey protein (unknown partner protein) e.g. yeast 2 hybrid.

1.4.2 Affinity based approaches

All these approaches involve purification of a protein of interest based on particular affinity properties. For identification of co-purifying proteins for which there is no prior knowledge, mass-spectrometry is the method of choice (depicted schematically in Figure 1.5).
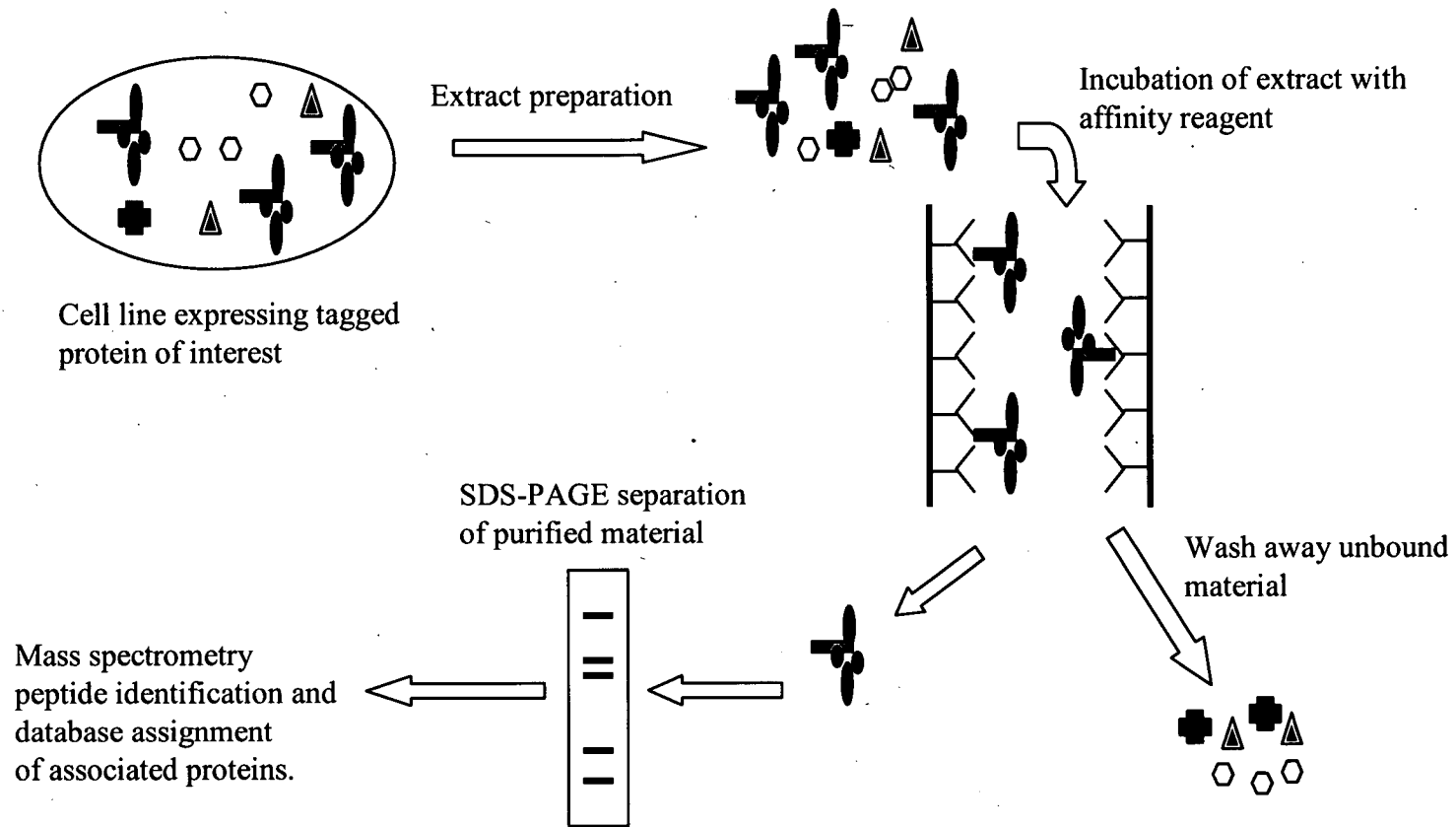
Figure 1.5- Schematic depiction of affinity based purification methods.

This schematic diagram gives an overview of the one step affinity purification procedure used in this thesis- BIO tagging. Nuclear extract is prepared from cells co-expressing a BIO tagged transcription factor of interest and a biotin ligase. The nuclear extract is incubated with immobilised streptavidin for 2h at 4oC and subsequently washed extensively. The specifically bound proteins are then eluted from the solid phase via boiling in Laemmli buffer or incubation with TEV protease. Eluted proteins are then resolved via SDS-PAGE before being subjected to in gel trypsinisation and mass-spectrometric identification of purified peptides. This generic scheme also applies to co-immunoprecipitation or TAP tagging methodologies.

1.4.2.1   Co-immunoprecipitation

Co-immunoprecipitation involves purification of a protein of interest using a specific antibody under conditions that permit maintenance of higher order complex integrity. Although this can be performed using antibodies against the endogenous protein, this is often insufficient as these antibodies may purify different protein isoforms, homologues, or even non-related proteins leading to the mis-identification of partners. In addition, some antibodies will not be of sufficiently high affinity to identify all interactions. Furthermore, it has been suggested that using an antibody raised directly against the protein of interest in unwise due to the possibility that it will compete with partner proteins for interaction at the immune epitope (Monti *et al.*, 2005a). To bypass some of the problems with antibodies against endogenous proteins, fusions between the protein of interest and short epitope tags with highly specific affinity properties can be employed (Terpe, 2003). Tagged proteins have been used successfully to isolate many proteins from complex mixtures and subsequently identify interacting partner proteins in a number of cell types. The use of epitope tags introduces additional concerns. For example, epitope tags may alter the three dimensional structure of the protein thus affecting function and/or interactions with partner proteins. This problem is illustrated by the abolition of TGF-β induced Smad3 phosphorylation upon fusion of an epitope tag to the C-terminus of Smad3 (Liu *et al.*, 1997). Furthermore, unless targeted to the endogenous genes, the addition of epitope tags also necessitates introduction of a transgene which will alter the stoichiometry of the protein with respect to its partners which could affect the interactions observed (Monti *et al.*, 2005b). One major caveat with all co-immunoprecipitation studies is that it is difficult to unequivocally conclude that the

biochemical interaction observed is present *in vivo* as it has been suggested that many non-specific interactions may occur after cell lysis *in vitro* on the derivatised beads (Monti *et al.*, 2005a).

1.4.2.2 Tandem affinity purification (TAP) tagging.

TAP tagging provides a generic method for purification of proteins and associated partner proteins at (or close to) endogenous expression levels (Rigaut *et al.*, 1999). The first generation of TAP tag was utilised in yeast, and identified a novel U1 snRNP subunit along with all previously known U1 snRNP subunits (Rigaut *et al.*, 1999). Furthermore, eight times less starting material was required than for standard antibody mediated affinity procedures and the non-specific background proteins were much reduced compared to standard antibody mediated purifications (Rigaut *et al.*, 1999). The TAP tag is bipartite in nature, consisting of a protein A domain separated from a calmodulin binding protein (CBP) domain by a tobacco etch virus (TEV) protease site (Puig *et al.*, 2001). Cells expressing the TAP tagged protein of interest are lysed under native conditions and subjected to the first step of purification by binding to IgG beads. The bound material is cleaved from the IgG beads with TEV protease and bound to calmodulin coated beads prior to elution with EGTA and subsequent mass spectrometry analysis. A benefit of this system is the decreased background afforded by the two step purification as well as the fact that TEV protease release from the solid phase will leave any non-specific proteins *in situ*. Recently, the TAP tag system has been successfully used in a mammalian system to identify the protein interaction networks in a signal transduction pathway (Bouwmeester *et al.*, 2004). TAP tags have also been modified to increase their

efficiency in mammalian systems by for example, introducing a Flag tag (Knuesel *et al.*, 2003) or a biotin tag (Drakas *et al.*, 2005) instead of the calmodulin binding protein, the latter of which is estimated to increase the yield of purified protein by about three fold compared to the yeast TAP tag. Widespread use of the TAP tag in mammalian cells lags behind yeast as there is often a low yield of the fusion protein, there is difficulty obtaining the required cell mass, and there is competition between the endogenous and tagged protein for partner proteins (Drakas *et al.*, 2005).

1.4.2.3   Biotinylation (BIO) tagging.

Biotin is a naturally occurring co-factor for many metabolic enzymes, yet is only active when covalently added to enzymes via the action of biotin ligase (Chapman-Smith and Cronan, 1999). Biotin is able to bind streptavidin in the strongest non-covalent interaction in nature ($K_d \sim 10^{-15}$). This property led to the generation of peptide tags which can be biotinylated in several cell types, but these tags were large and inefficiently biotinylated in mammalian cells (Parrott and Barry, 2000; Parrott and Barry, 2001). Recently, however, shorter tags of approximately 23 amino acids have been developed that can be biotinylated as efficiently as natural biotin acceptor proteins (Beckett *et al.*, 1999). Furthermore, it has been demonstrated that upon fusion of a BIO tag to the transcription factor Gata1 and co-expression in mouse erythroleukaemic cells with the *E.coli* biotin ligase BirA, biotinylated Gata1 can be efficiently purified on immobilised streptavidin in a single step protocol (de Boer *et al.*, 2003; Rodriguez *et al.*, 2006). Performing these experiments under non-denaturing conditions permitted identification by mass-spectrometry of functionally important partner proteins present in distinct activator and repressor complexes

(Rodriguez *et al.*, 2005). This approach appears to be particularly useful as it can be performed in a single step, with no requirement for antibodies or for tag cleavage. Moreover, there are very few naturally biotinylated contaminating proteins (de Boer *et al.*, 2003). Biotin tagging has subsequently been employed to purify proteins from ES cells and transgenic mice (Driegen *et al.*, 2005).

1.4.2.4   Mass spectrometry

Each of the affinity purification protocols described requires protein identification using mass-spectrometry. Peptide mass fingerprinting is a powerful mass spectrometry technique which allows identification of proteins based on the mass: charge ratio of the ionised peptides derived from proteolytic digestion of a protein mixture (reviewed by Yates, 2000). Peptide mass fingerprints together with knowledge of the protease used to digest the protein mixture are used to search databases containing predicted peptide masses from all known proteins. Due to the exquisite resolution of mass-spectrometry it is possible to identify many of the proteins in a mixture. However, the signals produced by single step mass-spectrometer e.g. MALDI-TOF (Matrix assisted laser desorption ionisation- time of flight) are often too complex for the complete identification of all proteins in a mixture. This is because large numbers of peptides from several proteins are co-detected, and it is impossible to correctly identify which protein a particular peptide represents. In these cases innovations such as LC/MS/MS (Liquid Chromatography tandem mass spectrometry) are useful. LC/MS/MS first separates the protein mixture via liquid chromatography, before introduction to the mass spectrometer. The technique works by fragmenting peptide ions and obtaining sequence data from mass:

charge values. Individual peptides can be selected out of complex mixtures of other ions and analysed by fragmentation. The fragmentation chemistry is such that the introduced peptides collide with a gas, and the peptide is broken at peptide bonds producing an ion series from which sequence can be determined (Peng and Gygi, 2001; Yates, 2000; Monti *et al.*, 2005a). It has been suggested that mass-spectrometry should be used only to ascertain that a product of a particular gene is present rather than the unequivocal identification of a particular protein (Rappsilber and Mann, 2002) because mass spectrometry data is only partially experimental.

1.4.3   Yeast two hybrid systems.

Two hybrid systems involve functional reconstitution of a transcription factor. Some transcription factors can be physically separated into independent domains, for example the activation domain (AD) and DNA binding domain (DBD) of Gal4p (Keegan *et al.*, 1986). This knowledge was exploited in yeast to show that if one member of an interacting partner pair is fused to Gal4p-AD and the other to Gal4p-DBD, that Gal4p activity is reconstituted by virtue of their close physical proximity (Fields and Song, 1989). Subsequently, this technology has been further developed and is now routinely used to screen for novel interactors using a 'bait' protein of interest fused to the Gal4p-AD and a "prey" cDNA library from a cell-type/ tissue of choice fused to Gal4p-DBD (Vidal and Legrain, 1999). Vidal and Legrain (1999) outline the advantages and disadvantages of yeast two hybrid technology and they are summarised here. The yeast 2-hybrid system has the advantage of being performed *in vivo* so issues of *ex vivo* non-physiological interactions are diminished. Furthermore, large libraries containing potential interactors can be screened in a

single experiment. Yeast two hybrid screens also have no reliance on expensive antibodies or mass-spectrometry. However, there are a number of drawbacks including the high false positive rate, as the screen is not only for proteins that *do* interact *in vivo* but also for those that merely *can*. Moreover, as the screen reads-out based on transcriptional activation, some proteins in the introduced library can have 'self-activator' activity. Additionally, yeast two hybrid systems predominantly identify only direct interactions rather than the whole complexes identified via affinity purification procedures. Finally, and of particular importance when analysing mammalian protein interaction, the proteins may not be correctly folded or post-translationally modified in yeast cells. Mammalian two hybrid systems are available (reviewed by Lee and Lee, 2004) but are more suited to use for confirming putative protein-protein interaction rather than for performing screens as with the yeast system.

1.5   Aims

The broad aims of this thesis are to perform a preliminary biochemical characterisation of Nanog protein and to assess the effects of *Nanog* over-expression *in vivo*. Experiments will be performed to assess the multimerisation capacity of Nanog as well as attempting to identify Nanog interacting partner proteins using a candidate approach. To extend the search for Nanog partner proteins, an unbiased biotinylation tagging strategy will be designed, constructed, and implemented, and preliminary confirmation of putative partner proteins performed via co-immunoprecipitation. Additionally, the phenotypic consequences of *Nanog* over-expression during development will be investigated via the design and generation of

a *Nanog* over-expressing cell line reagent that can be tracked in the mouse embryo. Together, these experiments should allow progress to be made in understanding the molecular mechanisms underlying Nanog function.

Chapter 2.

Materials and Methods

Unless otherwise stated, chemicals were obtained from Fisher, and oligonucleotides synthesised by Sigma. The water used for all procedures was milliQ water (Millipore) which was monitored for electrical resistance during purification, and used at $18.2m\Omega$.

2.1   Culture and manipulation of ES cells.

2.1.1   Routine culture of mouse ES cells.

ES Media

500ml	GMEM (Sigma).
11ml	Sodium pyruvate/ L- Glutamine (Invitrogen)
51ml	Foetal Calf Serum (Invitrogen).
5.5ml	100x Non-essential amino acids (Gibco).
555µl	0.1M 2-mercaptoethanol. (BDH)
555µl	LIF (prepared by ISCR tissue culture staff)

1X Trypsin solution

0.186g EDTA was dissolved in 500ml PBS and filter sterilised. 5ml Chick serum (Sigma) and 5ml trypsin (2.5%) (Invitrogen) was added and mixed. Trypsin was stored at $-20^{o}C$ and the final concentration was 0.025%.

Sodium pyruvate/ L- Glutamine solution

The 11ml aliquot was made by mixing 5.5ml Sodium pyruvate (100mM) with 5.5ml L-Glutamine (200mM).

Protocol   ES cells were cultured according to (Smith, 1991).

- Routinely, the media was changed at least every 2 days with pre-warmed ES cell media.

- Cells were routinely passaged every 3-4 days when cells were 80-90% confluent.

- Tissue culture coated flasks/ plates (Iwaki) were coated with 0.1% gelatin (Sigma)/ Dulbecco PBS 20 min. before passaging cells.

- ES cell media was removed from the cells via aspiration.

- Cells were washed twice with pre-warmed Dulbecco PBS (Sigma).

- Trypsin solution was added to the cells so that the monolayer was entirely covered.

- The flask/ plate was placed in 37°C/ 7% CO_2 incubator for ~1minute.

- The flask/ plate was tapped to dislodge the cells.

- Trypsin was neutralised by addition of 4x volume of ES cell media.

- The cells were centrifuged at 1200rpm (250g) (ALC PK120; Annita) for 3 min. in universal tubes.

- The cell pellet was resuspended in 10ml ES media.

- Routinely cells were split 1:5 at each passage into pre-warmed ES cell media.

- Cells were gassed with 5% CO_2/ air and returned to the 37°C incubator.

2.1.2   Transfection of DNA into mouse ES cells.

2.1.2.1   Stable transfection.

- ES cell media was changed at least 2h prior to transfection.

- 9.5ml ES cell media was added to gelatinised 100mm plates and placed in 37°C/7% CO_2 incubator to equilibrate.

- The cells were trypsinised and neutralised with ES cell media.

- The cells were pelleted at 1200rpm (250g) for 3min. (ALC PK120; Annita).

- The cell pellet was washed in pre-warmed PBS.

- The cells were counted using a haemocytometer.

- The cells were pelleted at 1200rpm (250g) for 3min..

- The cells were resuspended such that 10^7 cells were in a volume of 0.7ml PBS.

- 100μg of linearised (usually *Sca*I) DNA in a volume of 0.1ml 1X PBS was placed in an electroporation cuvette (Biorad).

- 0.7ml (10^7 cells) were added to the cuvette containing the DNA and mixed gently.

- The cuvettes were left 3 min. at room temperature.

- Cells were electroporated at 0.8kV and 3μF using a GenePulser (Biorad).

- The cells were removed with a plugged Pasteur pipette and added to 9.2ml of prewarmed ES cell media.

- 0.5ml (5×10^5 cells) cell suspension was added to the plate and swirled to distribute.

- Selection was started 24h post-transfection (see Table 2.1).

2.1.2.2 Transient transfection.

- 30 minutes prior to transfection, 10^6 mouse ES cells were plated into a well of gelatinised 6 well plate in a volume of 2ml ES cell media.

- 3µl Lipofectamine 2000 (Invitrogen) was diluted into 250µl ES cell medium without FCS and incubated at room temperature for 5 minutes.

- 3µg plasmid DNA was diluted into 250µl ES cell medium without FCS and incubated at room temperature for 5 minutes.

- The diluted DNA and Lipofectamine 2000 were mixed and incubated at room temperature for 20 minutes.

- The Lipofectamine/DNA mixture was added drop-wise to the plated ES cells and swirled to distribute.

- Plates were returned to the 37°C/ 7%CO_2 incubator.

- Cells are replated and selection started (see Table 2.1) 24-36h post transfection.

Table 2.1. Antibiotic concentrations used for drug selection in mammalian cells

ANTIBIOTIC	SUPPLIER	STOCK CONC.	WORKING CONC.
G418	PAA	200mg/ml	200µg/ml
Puromycin	Sigma	5mg/ml	1-2µg/ml
Blasticidin S HCl	Invitrogen	5mg/ml	5-15µg/ml
Hygromycin B	Roche	50mg/ml	100-200µg/ml

2.1.2.3   Picking mouse ES cell colonies.

The media was removed and the cells were washed twice with warm PBS. 5µl trypsin was taken up in a yellow tip and expelled over a colony. The colony was then scraped and sucked up with the yellow tip and placed into a gelatinised well of a 96 well plate. 180µl ES cell media was added and tituration was used to break up the colony. The plate was then placed in the 37%/ 7%CO_2 incubator.

2.1.3   LIF independence assay.

For LIF independence assays of stable ES cell lines, 600 cells were plated/ well of a 6 well plate (64 cells/cm^2) in the presence or absence of LIF, and left for 6 days prior to inspection and staining. For LIF independence assays following transient transfection (lipofection), cells were routinely plated at $5x10^4$ cells per 100mm diameter plate (640 cells/ cm^2) and selected in appropriate antibiotics in the presence or absence of LIF for 10-12 days prior to alkaline phosphatase staining.

2.1.4   Freezing mouse ES cells.

- ~$1x10^6$ ES cells were pelleted

- Cells were washed in freezing mix (ES cell media containing 10% DMSO)

- The pellet was then resuspended in ~1ml freezing mix, transferred to a cryotube (Nunc) and placed immediately in the -80°C freezer.

- The next day the cells were transferred to the liquid N_2 storage tank (-170°C).

2.1.5   Thawing mouse ES cells.

- The vial was removed from the liquid N_2 and placed immediately in the $37^{\circ}C$ waterbath to thaw.

- The thawed cells were transferred to a universal tube containing 9ml warmed ES cell media.

- The cells were pelleted via centrifugation at 1000rpm (200g) (ALC PK120; Annita).

- The pellet of ES cells was carefully resuspended in ES cell media, and transferred to a gelatinised flask, and placed in a $37^{\circ}C/7\%CO_2$ incubator.

2.1.6   Staining of mouse ES cells

2.1.6.1   Alkaline phosphatase staining.

An alkaline phosphatase staining kit was used (Sigma)

FIXATIVE SOLUTION (Store at $4^{\circ}C$)

25ml Citrate solution (18mM Citric acid; 9mM Sodium citrate; 12mM NaCl)
8ml Formaldehyde
65ml Acetone

STAIN

400µl FRV alkaline solution (in kit) and 400µl Sodium Nitrite solution (in kit) were mixed and incubated for 2 minutes at room temperature. The Alkaline/Nitrite was added to 18ml dH$_2$O. Finally, 400µl Napthol solution (in kit) was added.

Protocol

- The media was aspirated.

- The cells were washed with warm PBS.

- 2ml fixative solution was added for 45s

- The fixative solution was removed and the cells were washed in dH_2O

- 2ml stain was added and incubated for 25min. at room temperature in the dark.

- The stain was removed and wells were washed with dH_2O.

- Plates were allowed to air dry before microscope analysis.

2.1.6.2   X-Gal staining of ES cells and embryos.

<u>PO_4 Buffer (0.1M pH 7.33)</u>

75ml 1M disodium hydrogen orthophosphate+ 25ml 1M Sodium dihydrogen

orthophosphate- make up to 1l with dH_2O.

<u>Fixative</u>	<u>Wash Buffer</u>
0.2% glutaraldehyde	2mM $MgCl_2$
2mM $MgCl_2$	0.1% sodium deoxycholate
5mM EGTA	0.02% NP-40
	0.05% BSA

Both the fixative and wash were made up with 0.1M PO_4 buffer pH7.33.

<u>Stain</u>

(1) Dissolve 50mg X-Gal in 1ml dimethyl formamide.

(2) 50ml wash + 12mg spermidine

82mg $K_3Fe(CN)_6$

105mg $K_4Fe(CN)_6$

15mM NaCl (0.25ml 3M stock)

Mix (1) and (2), filter, and freeze down in Eppendorf tubes.

- The stain was thawed and spun down before use.

- The adherent cells or the dissected embryos were rinsed in PBS, fixed for 5min (cells) or 60min (embryos), washed 3x 10min., before addition of the stain and incubation at 37°C in the dark. Cells were checked for the appearance of a blue colour every hour (or left overnight).

2.1.7   Metaphase spreads of mouse ES cells.

Mouse ES cells were plated the day before preparing the spreads (20% confluent T_{25} tissue culture flask).

- The next day cells were trypsinised and neutralised with ES cell media. The detached cells were pelleted by centrifugation.

- The supernatant was removed and the cells were resuspended in ~2ml 0.56% KCl for 6 minutes. 50µl fixative (3:1, methanol: acetic acid) was added in a drop-wise fashion, prior to re-pelleting the cells for a further 5 min.

- The supernatant was removed and 1ml fixative was added.

- The fixed cells were then refrigerated for at least 30min.

- The fixed cells were re-pelleted and resuspended in 200µl fixative.

- 20µl of fixed cells were dropped on a clean microscope slide, allowed to air dry, and stained with 10% Giemsa stain (BDH).

2.1.8　　　Embryological techniques

2.1.8.1　　　Morula aggregation and embryo transfer.

The protocols were based on those from (Nagy, 2002; Nagy, 2001)

Aggregations and transfers were performed by ISCR staff.

PB1 media		G2 media	
136mM	NaCl	90.08mM	NaCl
2mM	KCl	5.5mM	KCl
1mM	$CaCl_2$	0.5mM	$NaH_2PO_4.2H_2O$
1mM	KH_2PO_4	1mM	$MgSO_4.7H_2O$
0.5mM	$MgCl_2.6H_2O$	3.15mM	Glucose
8mM	Na_2HPO_4	11.74mM	Na Lactate
0.3mM	Na pyruvate	0.1mM	Na pyruvate
5.5mM	Glucose	25mM	$NaHCO_3$
1.5g/500ml	BSA	1mM	Glutamine
		1.8mM	$CaCl_2.2H_2O$
		1X Essential amino acids and 1X non-essential amino acids (Gibco)	

Morula Aggregation

- ES cells were trypsinised and neutralised as usual, plated in ES cell medium in bacterial plates, and placed in a $37^{\circ}C/7\%CO_2$ incubator for ~2h prior to aggregation. This allows cells to adhere to one another loosely forming strings of cells. For morula aggregation experiments ES cell lines of passage number <15 were used.

- Aggregation plates: Drops of G2 complete media were placed in primaria dishes, and covered with mineral oil (Sigma). An indentation was made in the dish at the centre of each drop using a Hungarian darning needle (ND-09; from BLS Ltd).

- The aggregation plates were placed at $37^{\circ}C/7\%CO_2$ to equilibrate.

- The oviducts from pregnant (E2.5) F_1 female mice were cut out, the fat was trimmed off and the oviducts were flushed with PB1 media using a flattened needle into a 60mm dish of PB1.

- The zona pellucida was removed from the embryos by washing in Acid Tyrode's solution.

- Each morula was picked up using a drawn Pasteur mouth pipette and placed into an indent in the aggregation plate.

- ES cells clumps were then added to each indent containing a morula (~8cells/morula). Before adding the ES cells they were washed in a large drop of G2.

- Aggregation plates were then incubated at $37^{o}C$/ 7% CO_2 overnight.

Transfer of embryos

- The next day, embryos were transferred to 2.5dpc CBA/ BL/6 (F_1 hybrid) pseudopregnant females.

- The anaesthetic Avertin (2.5%) was given via intra-peritoneal injection with the dose being dependent on mouse size (0.015-0.017 ml/g body weight)

- A dorsal incision was made in the abdominal cavity and the ovary was exposed. Ovary stimulation was monitored by the presence of red dots which report the vascularisation of the corpus luteum.

- The uterine horn was pierced with a small needle.

- A fine glass needle was used to introduce the embryos into the uterine horn.

- The skin was sealed with wound clips.

- After the procedure the animal was given a painkilling injection of Caprofen ($4\mu g$/g body weight).

- At the desired time point the animal was sacrificed via cervical dislocation and embryos were photographed and stained as appropriate.

2.1.8.2   Sectioning of embryos

- Embryos were embedded in 100% paraffin (BDH) and left to set.

- Serial sections were taken through the embryo at 6 micron intervals using a microtome (Anglia Scientific 0325)

- These sections were floated on a 37°C water bath containing sterile dH_2O.

- The sections were lifted onto microscope slides and allowed to dry before microscopic examination.

2.1.9   FACS analysis

Cells were trypsinised as described (section 2.1.1) and resuspended in ice cold PBS/ 5% FCS at a density of ~$1x10^6$ cells/ml. Cells were analysed using the FACScalibur apparatus (Becton Dickinson).

2.2 Biochemical Techniques.

2.2.1 Preparation of nuclear extract.

Buffer A

10mM HEPES/ KOH pH7.9
1.5mM MgCl$_2$
10mM KCl
+ EDTA free complete protease inhibitor cocktail (Roche)

Buffer C
(100mM HENG)

20mM HEPES/ KOH pH7.9
1.5mM MgCl$_2$
20% Glycerol
100mM KCl
0.2mM EDTA
+ EDTA free complete protease inhibitor cocktail (Roche)

2.2M HENG

20mM HEPES/ KOH
1.5mM MgCl$_2$
20% Glycerol
2.2M KCl
0.2mM EDTA
+ EDTA free complete protease inhibitor cocktail (Roche)

For 25 150cm^2 plates of 80% confluent mouse ES cells;

- Media was supplemented with 0.1µg/ml biotin (Sigma) 24h before lysis.

- Media was aspirated and cells were washed with 5ml room temperature PBS.

- 2ml trypsin was added and the plates were incubated at 37°C for ~1min.

- Trypsin was neutralised with 8ml ES cell media.

- Three plates worth of cells were pooled in 1 universal tube.

- The cells were pelleted via centrifugation at 1200rpm (250g) in a benchtop centrifuge (ALC PK120; Annita)

- The pellets were washed in PBS and pooled.

- The pooled cells were pelleted at 1200rpm (250g) (ALC PK120; Annita)

- The cells were carefully resuspended in 50ml ice cold Buffer A in a 50ml Corning tube.

- The cells were lysed by incubation on ice for 20 min.

- The cells were vortexed for 90s-120s.

- 10µl of cell suspension was mixed with 10µl Unna Stain/ methyl pyronin (a gift from J. Strouboulis), pipetted onto a microscope slide and covered with a cover slip. The cells were examined under the light microscope (the free nuclei stain a pale blue colour). When >90% of the nuclei were free from cytoplasm the protocol was continued. If <90% of cells were lysed, vortexing was repeated.

- The nuclei were pelleted via centrifugation at 4°C at 340g (Sigma 4K15 centrifuge).

- The supernatant (cytoplasmic fraction) was removed and discarded.

- The pellet (nuclei) was resuspended in 6ml ice cold buffer C.

- 2.2M HENG was added drop-wise until the DNA precipitated.

- The Corning tube was rotated at 4°C for 30min.

- The liquid was decanted into a SW40.1 tube (Beckman), ultracentrifuged for 60min at 4°C at 40000 rpm (198,000g) using a pre-chilled SW41T-4 rotor in a Beckman L7-65 ultracentrifuge.

- The supernatant is the nuclear extract which was split into 1ml aliquots in non-stick microcentrifuge tubes (Alpha Laboratories), snap frozen and stored at -80°C.

- The salt concentration is calculated by comparing the sample nuclear extract to a dilution series of KCl in the concentration range 100-500mM using an AKTA FPLC system (Amersham).

- Protein concentration was calculated using a ND 1000 spectrophotometer (Nanodrop).

2.2.2 Binding biotinylated material to streptavidin beads and preparation for mass-spectrometry.

Solutions required;

1M HENG

10mM HEPES/ KOH pH7.9
1.5mM $MgCl_2$
0.25mM EDTA
20% Glycerol
1M KCl

0M HENG

10mM HEPES/ KOH pH7.9
1.5mM $MgCl_2$
0.25mM EDTA
20% Glycerol

1M HENG/ 0% Glycerol

10mM HEPES/ KOH pH7.9
1.5mM $MgCl_2$
0.25mM EDTA
1M KCl

0M HENG/ 0% Glycerol

10mM HEPES/ KOH pH7.9
1.5mM $MgCl_2$
0.25mM EDTA

Using these solutions it was possible to easily make HENG buffers of desired KCl and glycerol concentration without affecting the concentration of the other components.

Bead Wash Buffer

250mM HENG Buffer
0.3% NP-40

2x Laemmli Buffer

125mM Tris pH6.8
25% Glycerol
4% SDS
0.01% bromophenol blue
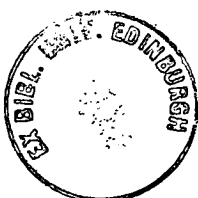
<u>50X Protease inhibitor</u>

Dissolve 1 EDTA free Complete protease inhibitor cocktail tablet (Roche) in 1ml PBS.

<u>Stock solutions:</u>

100mM $MgCl_2$,

RNAseA (100mg/ml; Roche),

10% NP-40 (Sigma)

Protocol

- Nuclear extract aliquots were thawed on ice

- Approximately 15mg of nuclear extract was added to a 50 ml Corning tube and made up to a final volume of 30ml with HENG buffer. The final concentration should be 100mM KCl/ 7-12% Glycerol. NP-40 was added to a final concentration of 0.3%, $MgCl_2$ to a final concentration of 10mM and protease inhibitors at 1X.

- A 100µl aliquot was removed and stored at -20°C (for nuclease efficiency assay).

- 80µl RNAseA or 100µl Benzonase (Novagen; 125U/ mg protein) was added.

- The diluted nuclear extract was incubated on ice for 2h.

- 200µl Dynabeads M-280 (Invitrogen) were blocked with 200ng/µl chicken egg extract (Sigma) in PBS for 1h at room temperature in a rotary incubator.

- Block buffer was removed using a magnetic rack (Dynal-MPCS).

- The beads were added to the diluted nuclear extract and incubated at 4°C with tumbling for 2h.

- The beads were pelleted via centrifugation at 4°C for 5min at 340g (Sigma 4K15 centrifuge).

- The supernatant was removed and a 1ml aliquot stored at -20°C (for western analysis, and a 100µl aliquot stored at -20°C (for nuclease efficiency assay).

- The beads were transferred to a fresh microcentrifuge tube and washed 5x 5min. with 1ml bead wash buffer.

- **Either;** The beads were boiled in 30µl 2x Laemmli Sample buffer. 1µl was retained for western analysis.

  Or; The bound material was trypsinised on the beads (see later).

- Bound material was subjected to SDS-PAGE on a 4-12% Bis-Tris NuPage gel in MOPS running buffer (Invitrogen) for 70min at 200V.

- The gel was stained with colloidal blue stain kit (Invitrogen).

- Each lane was cut into 20 slices and each slice cut into 3 pieces, and submitted for in gel tryptic digests and mass-spectrometry at Erasmus MC (Rotterdam).

On bead trypsinisation (based on Rybak *et al.*, 2005)

- Beads were washed once in 50mM Ammonium bicarbonate then resuspended in 250µl 50mM Ammonium bicarbonate.

- Sequencing grade trypsin (Roche) was reconstituted in 50mM Ammonium bicarbonate and added to the bead suspension at 60ng/mg input protein.

- The beads were incubated overnight at 37°C in a rotary incubator.

- The beads were immobilised using the magnetic rack, the supernatant (trypsinised material) was transferred to a fresh microcentrifuge tube and stored at -20°C prior to mass-spectrometry.

- Boiling the M-280 beads (post trypsinisation) in Laemmli buffer followed by western analysis can confirm that bound material has been efficiently digested.

Routinely 1% of the input, bound, and unbound material was retained for western analysis. Immmunoblotting with both α-Nanog and streptavidin-HRP can provide an estimate of purification efficiency.

Nuclease activity was monitored by treating the 100μl aliquots of the binding mixture (pre and post nuclease) with proteinase K, phenol chloroform extracting the nucleic acid, and adding DNA loading dye prior to electrophoresis on a fresh 1% TBE agarose gel. Degradation of nucleic acid was assessed by viewing the gel under UV transillumination.

2.2.3   Mass spectrometry

2.2.3.1 Preparation of samples for Mass-Spectrometry.

Performed by Jeroen Demmers at the Erasmus MC (Rotterdam) as part of a collaboration with J. Strouboulis and colleagues. Protocol taken from (Rodriguez *et al.*, 2006).

- Following electrophoresis by SDS-PAGE the gel was stained overnight with Colloidal Blue, according to the manufacturer's instructions.
- The gel was destained in several changes of ddH2O until the background (i.e. the non-protein containing part of the gel) was completely destained. This usually takes several hours (i.e. more than 12 hours).

- The destained gel was photographed to provide a record of the purification experiment.

- 20-25 microfuge tubes were rinsed in 60% acetonitrile.

- The entire lane were cut out lengthwise and divided into at least 20 gel slices. Each gel slice was cut into 3 pieces and set of three pieces was placed in a separate tube.

- Each gel slice was destained in 100μl of destaining solution (25mM ammonium bicarbonate in 50% acetonitrile) for 20-30min. This step was repeated until the gel slice became completely destained (usually 3-4 times). Alternatively, gel slices can be destained overnight at 4oC.

- Each gel slice was dehydrated in 100μl of 100% acetonitrile for 5-10min at room

    temperature. The plug became hard and white at this step.

- The gel slices were reduced with freshly prepared 6.5mM DTT solution for 45-60min at 37°C.

- The solution was discarded and proteins in the gel slices were alkylated by adding 100μl of 54 mM iodoacetamide solution and incubating for 60min at room temperature in the dark.

- The solution was discarded and the gel slices were washed in 100μl of gel slice destaining solution for 15 min at room temperature. This step was repeated once more.

- The washing solution was discarded and the gel slices were dried in 100μl of 100% acetonitrile for 10min. The solution was again discarded and the gel slices were dried at room temperature.

- Proteins were in-gel digested in 15μl of 10 ng/μl modified trypsin at (diluted from the 100x stock in 50 mM ammonium bicarbonate) for 30 min on ice. 15 μl of 50 mM ammonium bicarbonate were added to the samples followed by overnight incubation at 37°C.

- Samples were equilibrated to room temperature. 30μl of 2% acetonitrile in 0.1% formic acid were added to the samples and incubated at room temperature for 15 min. The samples were then vortexed briefly and sonicated for 1 minute.

- The supernatants were collected in separate tubes and the remaining gel slices were treated with 30μl of 50% acetonitrile in 0.1% formic acid and incubated as above. Samples were again vortexed and sonicated as above and the supernatants were collected and pooled with the corresponding supernatants the previous step.

- The samples were vacuum dried in a vacuum centrifuge for 45-60 minutes.

- The eluted peptides were now ready for analysis by mass spectrometry.

2.2.3.2   Mass-Spectrometry analysis.

Performed by Jeroen Demmers at the Erasmus MC (Rotterdam) as part of a collaboration with John Strouboulis and colleagues.

Nanoflow LC-MS/MS was performed on an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ linear ion trap mass spectrometer (Thermo) operating in positive mode and equipped with a nanospray source. Peptide mixtures were trapped on a ReproSil C18 reversed phase column (Dr Maisch GmbH;

column dimensions 1.5 cm × 100 μm, packed in-house) at a flow rate of 8 μl/min. Peptide separation was performed on ReproSil C18 reversed phase column (Dr Maisch GmbH; column dimensions 15 cm × 50 μm, packed in-house) using a linear gradient from 0 to 80% B (A = 0.1 M acetic acid; B = 80% (v/v) acetonitrile, 0.1 M acetic acid) in 70 min and at a constant flow rate of 200 nl/min using a splitter. The column eluent was directly sprayed into the ESI source of the mass spectrometer. Mass spectra were acquired in continuum mode; fragmentation of the peptides was performed in data-dependent mode. Peak lists were automatically created from raw data files using the Mascot Distiller software (version 2.1; MatrixScience). The Mascot search algorithm (version 2.1, MatrixScience) was used for searching against the NCBInr database (release NCBInr_20061209.fasta; taxonomy: *Mus musculus*). The peptide tolerance was typically set to 2 Da and the fragment ion tolerance to 0.8 Da. A maximum number of 2 missed cleavages by trypsin were allowed and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively. The Mascot score cut-off value for a positive protein hit was set to 60. Individual peptide MS/MS spectra with Mowse scores below 40 were checked manually and either interpreted as valid identifications or discarded.

2.2.4 Phenol:Chloroform extraction from nuclear extracts.

- 100μl extract was diluted in 400μl PBS and 5μl proteinase K (10mg/ml; Sigma), and 15μl 20% SDS were added. This was incubated overnight at 55°C.

- 500μl Phenol: Chloroform: Isoamyl alcohol (25:24:1; Fluka) was added to the overnight proteinase K digest and shaken vigorously.

- The material was spun for 3 min. at full speed in a microcentrifuge (13,000rpm in a Biofuge pico centrifuge, Heraeus).

- The aqueous phase was transferred to a fresh microcentrifuge tube and 0.7 volumes of propan-2-ol and 50μl 5M NaCl were added prior to vortexing.

- The material was spun down for 10 min. at full speed at 4°C in a microcentrifuge.

- The pellet was washed in ice cold 70% Ethanol and allowed to air dry.

- The pellet was resuspended in 30 μl 1X DNA loading buffer.

2.2.5   Size exclusion chromatography. (Rodriguez *et al.*, 2006)

<u>Column running buffer</u>   <u>2x Laemmli Sample Buffer</u>

20mM HEPES (KOH pH7.9)   4% SDS
0.5mM EGTA   125mM Tris pH6.8
1mM $MgCl_2$ 25% Glycerol
200mM KCl   0.01% Bromophenol blue.
10% Glycerol   5% 2- mercaptoethanol

- Nuclear extract was thawed on ice and microcentrifuged at full speed for 5 minutes at 4°C.

- The pump (AKTA FPLC; Amersham) and the Superose 6 column (Amersham) were washed with column running buffer. Running program was set as follows: Flow rate, 100μl/min; Sample volume loop, 200μl; Fraction volume, 500μl; Elution length, 1 column volume of column running buffer; Alarm pressure, 0.5MPa.

- The calibration standards (Amersham) were run through the column to establish the elution volume of protein complexes.

- The sample was injected onto the column (200µl =~600µg nuclear extract). NB. Injected volume must not exceed 1% of the total column volume (24ml).

- 500µl fractions were collected.

- The protein fractions were precipitated with Tricloroacetic acid (Sigma). 125µl 100% TCA was added to each 500µl fraction, incubated on ice for 1h. Precipitated protein was spun in a microcentrifuge at full speed for 15 minutes at 4°C, washed in 1% TCA, re-pelleted, washed in ice cold acetone, re-pelleted, and finally boiled in 50µl Laemmli sample buffer.

- 10µl of all even numbered fractions were subjected to SDS-PAGE, immunoblotting, and probing with α-Nanog antibody.

2.2.6 Histone tail binding protocols

- Streptavidin M-280 beads (Invitrogen) (50µl/reaction) were blocked with 200ng/ul chicken egg extract (Sigma) in PBS for 1h at room temperature.

- 2µg of either unmodified or K4 dimethylated, biotinylated H3 peptide tails (NEB) were bound to 50µl blocked beads in non-stick microcentrifuge tubes (Alpha Labs) in 100mM HENG buffer (see 2.2.1), 0.3% NP-40, 10% Glycerol, mini-complete protese inhibitors (Roche), 1mM PMSF, in a volume of 500µl. A bead only negative control was also performed to control for non-specific background binding.

- The reaction was incubated at 4°C for 3h with rotation.

- After the biotinylated peptide had pre-incubated with the beads, the unbound peptide was washed away via 3 washes with 300mM HENG/ complete EDTA free protease inhibitor (Roche)/ 1mM PMSF, using the magnetic rack to immobilise the beads at each wash.

- The immobilised peptides (and bead only control) were added 3 tubes of diluted ES cell nuclear extract (300µg protein, in 100mM HENG buffer, 0.3% NP-40, 10% glycerol, EDTA free complete protease inhibitor (Roche), 1mM PMSF) in a volume of 500µl and incubated for 2h at 4°C with tumbling.

- The beads were washed 8 times with 300mM HENG containing EDTA free complete protease inhibitor (Roche), 1mM PMSF.

- The beads were boiled in 50µl 2x Laemmli buffer for 10min. and either stored at -20°C or subjected to SDS-PAGE.

2.2.7   SDS-PAGE Electrophoresis and Western Blotting.

Running Buffer: MOPS Running Buffer (Invitrogen)
Transfer Buffer: 25mM Tris, 0.21M Glycine, 20% Methanol.
TBS: 10mM Tris pH7.6-8.0, 150mM NaCl.
Western wash Buffer: 0.5M NaCl, 0.3% Triton X-100- made up in TBS.
Western stripping Buffer: 62.5mM Tris pH6.8, 2%SDS.
2X Laemmli Buffer: 125mM Tris pH6.8, 4% SDS, 25% Glycerol, 0.01% bromophenol blue, 5% 2-mercaptoethanol.

- All samples are prepared in Laemmli buffer prior to loading. To prepare whole cell lysates, 1×10^6 cells were lysed in 200µl 1X Laemmli buffer.

- Both the upper and lower chamber of the XCell*Surelock* module (Invitrogen) were filled with running buffer, and 500μl antioxidant reagent (Invitrogen) was added to the upper chamber.

- Routinely, 10μl of samples were loaded onto 4-12% Bis-Tris gels (Invitrogen). 10μl of 'See Blue Plus 2' protein size ladder (Invitrogen) was loaded. Electrophoretic separation was performed at 200V for ~70 minutes. This provides resolution of proteins in the 20-200kDa range.

- The gel was removed from the housing and submerged in ice cold transfer buffer and the gel foot was cut off.

- The transfer sandwich was assembled (as shown) below the surface of ice cold transfer buffer to minimise the risk of trapping air bubbles.

WESTERN TRANSFER SANDWICH
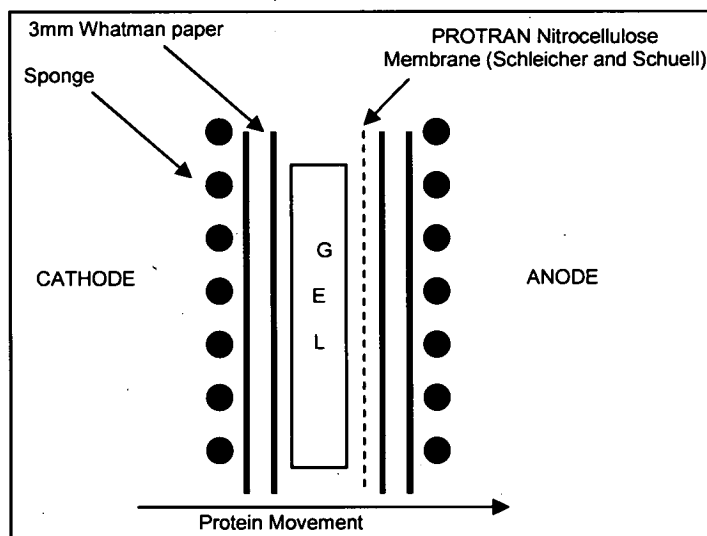
- Transfer was performed at 4°C at 395mA constant for 70 minutes.

- The transfer sandwich was disassembled and transfer efficiency monitored by visually checking that the size markers were on the membrane and not remaining in the gel.

- The membrane was washed briefly in TBS/ 0.05% NP-40.

- The membrane was blocked overnight in 10% Non- fat dry Milk in TBS/ 0.05% NP-40.

- The membrane was incubated for 2h at room temperature with the primary antibody diluted (Table 2.2) in 5% Non- fat dry milk in TBS/ 0.15% NP-40.
- The membrane was washed 3x 15min. in western wash buffer.
- The membrane was tumbled for 1h at room temperature with the secondary antibody diluted (Table 2.2) in 5% Non-fat dry milk in TBS/ 0.15% NP-40.
- The membrane was washed 3x 15min. in western wash buffer.
- The membrane incubated with Super-Signal West Pico reagent (Pierce) for 5min at room temperature.
- The membrane was wrapped in cling film and exposed to Hyperfilm (Amersham) for 10s- 30min depending on the signal strength.
- Film was developed using a SRX-101A developer (Konica-Minolta).

Stripping Western Blot Membranes

The western blot nitrocellulose membrane was incubated with stripping buffer in a sealed container at 70°C for 40 min. The stripping buffer was thouroughly washed away before re-applying the blocking solution.

Table 2.2 Antibodies used for western blotting

ANTIBODY	SUPPLIER	CAT NUMBER	WESTERN DILUTION	SECONDARY REAGENT
α75x36 (Nanog)	Produced in house	N/A	1:3000	Rabbit-HRP
α Oct-4-c-10	Santa Cruz	Sc5279	1:1000	Mouse-HRP
α HA	Covance	MMS-101P	1:2000	Mouse-HRP
α Flag-M2	Sigma	F3165	1:4000	Mouse-HRP
α HDAC2	Upstate	05-814	1:1000	Mouse-HRP
α SHP-2	Santa Cruz	Sc280	1:2000	Rabbit-HRP
Streptavidin-HRP	NEL	NEL750	1:10,000	N/A
α STAT3	BD Transduction	610189	1:1000	Rabbit-HRP
α Sall4	Gift from M.Trier	N/A	1:500	Rabbit-HRP
α mouse-HRP	Amersham	NA931	1:2000	N/A
α rabbit-HRP	Amersham	NA934	1:2000	N/A

2.2.8   Immunoprecipitation protocols

2.2.8.1   Standard immunoprecipitation from mammalian cells.

Lysis Buffer (Stored at 4°C)	2x Laemmli Sample Buffer
0.5% NP-40	4% SDS
50mM Tris pH8.0	125mM Tris pH6.8
150mM NaCl	25% Glycerol
(Mini complete protease	0.01% Bromophenol blue
inhibitor tablet; Roche)	5% 2-mercaptoethanol (added fresh)

For transient transfection experiments, 1×10^6 cells were plated 24h post transfection, the media was changed daily and the cells were lysed 72h post transfection. For stable cell lines, 10^7 cells were plated and lysed 24h post plating.

DAY 1

- The cells were grown in 100mm tissue culture plates (Iwaki) and the media changed 2h prior to lysis.

- The media was aspirated and cells were washed twice in ice cold PBS.

- 1.2ml Lysis Buffer was added to plates and rocked for 30min at 4°C.

- The lysates were transferred to non-stick 1.5ml microcentrifuge tube (Alpha Laboratories).

- Lysates were spun at 13,000rpm at 4°C in a microcentrifuge (Micro 24S, Sorvall)

- The supernatant was transferred to a fresh non-stick microcentrifuge tube.

- 25μl lysate was removed and mixed with 25μl 2x Laemmli sample buffer, boiled for 5min and stored at -20°C. This was used as an input sample.

- 5μg antibody added to lysate and incubated overnight at 4°C.

DAY 2

- For each immunoprecipitation experiment 20μl (10μl packed bead volume) protein A or protein G sepharose beads (Amersham) were blocked with 200ng/μl Chicken Egg Albumin (Sigma) in PBS for 1h at room temperature.

- The beads were washed once in lysis buffer before being added to the immunopreciptitates and tumbled at 4°C for 1h to collect the immune complexes.

- The beads were washed 5 times with 500μl lysis buffer.

- After the 5th wash, all traces of lysis buffer were carefully removed before adding 30μl 2x Laemmli Buffer, boiling for 5 min, and either storing at -20°C or subjecting to SDS-PAGE.

2.2.8.2   Dephosphorylation of immunoprecipitated Nanog.

- Nanog was immunoprecipitated from EF4 cell lysates as described (2.2.8.1) using 5μg of α-75x36 (Nanog) antibody.

- Before adding the protein A sepharose beads, the IP mixture was split into 4 equal aliquots of 300μl each.

- 10μl packed bead volume of protein A sepharose beads were added to each tube and rotated at 4°C for 1h.

- The beads were then washed three times with 250μl lysis buffer.

- Three of the tubes were washed with 1x phosphatase buffer (provided by the manufacturer), and 1 tube was washed again with lysis buffer only.

- The beads in the 3 tubes washed in phosphatase buffer were resuspended in 50μl phosphatase buffer and 5μl, 2.5μl, or 0μl of either Shrimp Alkaline

phosphatase (1U/µl; Roche) or Antarctic phosphatase (5U/µl; NEB) were added. The beads in the remaining tube were resuspended in 50µl lysis buffer with no enzyme.

- The beads were incubated at 37°C for 15min. briefly spun down, washed a final time with lysis buffer and boiled in 30µl Laemmli buffer for 5min. before storing at -20°C or subjecting to SDS-PAGE.

2.3   Molecular biology techniques.

2.3.1   Nucleic Acid Isolation.

2.3.1.1   Plasmid preparation from bacterial cells.

Plasmid DNA was prepared from overnight bacterial cultures grown in LB broth containing the appropriate antibiotic. Miniprep and Maxiprep kits (Qiagen) were used to prepare 50µl of ~100ng/µl DNA and 400µl of ~2mg/ml DNA, respectively. DNA concentration was quantified using a ND-1000 spectrophotometer (Nanodrop).

2.3.1.2   RNA extraction from ES cells.

RNA was extracted from ES cells using the RNAEasy kit (Qiagen) according to the manufacturer's instructions.

2.3.1.3   First strand cDNA synthesis.

cDNA was synthesised from RNA using the superscript® II reverse transcriptase (Invitrogen) according to the manufacturer's instructions.

2.3.2   DNA Manipulation.

2.3.2.1   Agarose gel electrophoresis.

- For an analytical 1% TBE gel; 2g agarose (Cambrex) was added to 200ml 0.5X TBE buffer (45mM Tris- Borate, 1mM EDTA). This was heated in a microwave on full power for ~2min until the agarose was completed dissolved.

- The agarose was allowed to cool to below 60°C and 6μl of Ethidium bromide (10mg/ml stock) was added.

- Gels were cast in a gel casting tray, and once set were routinely run in 0.5X TBE buffer at 100V in a gel tank, with the DNA in 1X Ficoll Blue DNA loading buffer (6X stock; 15% w/v Ficoll 400 (Amersham) in dH$_2$O/0.02% bromophenol blue (BDH)).

- Restriction fragments lengths were visualised using a GeneFlash Imager (Syngene).

- Preparative gels were prepared and run in 1X TAE buffer (40mM Tris-acetate, 2mM EDTA).

2.3.2.2   Restriction endonuclease digestion.

DNA digestions were performed using restriction endonucleases from Roche and NEB and are performed in the buffers provided by the manufacturers.

2.3.2.3   Blunt ending of cohesive ends.

- To a 200μl digest volume, 6.6μl 1mM dNTP's (final conc. 33μM) and 4μl (5U/μl) Klenow fragment (NEB) of DNA polymerase were added.

- The reaction was incubated at 25°C for 15min.

- The reaction was inactivated at 75°C for 5 min. and a final concentration of 10mM EDTA was added.

2.3.2.4 Construct Building.

2.3.2.4.1 Purification of restriction DNA fragments

- Routinely, 10µg plasmid DNA was digested with the relevant restriction endonucleases (New England Biolabs and Roche) in the manufacturer's buffer. NB. If the two enzymes used to release a desired fragment were not compatible with a given buffer, an ethanol precipitation step was performed after a single enzyme digestion and the precipitated DNA was resuspended in the second buffer before adding the second restriction enzyme.

- Fragments were separated on TAE agarose gels (1% for fragments >1kb, 2% for fragments < 1kb), visualised via long wavelength UV transillumination (UVP), and the desired fragment was excised with a clean scalpel blade.

- The DNA was extracted from the gel slice using the QIAEX II kit (Qiagen) according to the manufacturer's instructions.

2.3.2.4.2 Ligation.

- Ligation reactions were routinely performed using either T4 ligase or QuickLigase (NEB) in a volume of 20µl, according to the manufacturer's instructions. Ligation reactions were set up with vector: insert in a 1:3 molar ratio using 100ng of the vector (largest) fragment.

- 6μl of the ligation reaction was used to transform 50μl DH5α *E.coli* (Invitrogen).

2.3.2.4.3   Screening for correct ligation products.

- ~ 300ng Miniprep DNA was digested with a panel of (usually 2 or 3) restriction endonucleases which would give an unambiguous restriction fragment length pattern upon agarose gel electrophoresis.

- High purity maxiprep (Qiagen) DNA was produced prior to transfection into mammalian cells.

2.3.3   Transformation of plasmid DNA into *E.coli*.

Firstly, Luria Broth (LB) agar (1.5% w/v agar in LB, 1% w/v tryptone (Difco), 0.5% w/v yeast extract (Difco), 5mM NaCl) was melted, appropriate antibiotics added once agar had cooled to below 60°C, and plates were poured.

- A vial of DH5α *E.coli* (Invitrogen) was thawed on ice.

- ~10ng plasmid (or 6μl of a ligation reaction) DNA was added to 50μl DH5α *E.coli* and flicked gently to mix.

- The mixture was incubated on ice for 30min.

- The bacteria were heat shocked in a 37°C heat block for 30sec., and returned to ice for 2min.

- 950μl LB broth was added and shaken at 200rpm at 37°C for 1h in an orbital shaker.

- For known plasmid DNA; 10μl and 100μl of transformation mix was plated out on 100mm plate containing LB Agar and the appropriate antibiotic (see Table 2.3). For ligation products, 100μl and all of the rest were plated out and incubated at 37°C overnight.

- Transformation efficiency was monitored by transfecting a known amount (5 picograms) pUC19 plasmid.

Table 2.3 Antibiotic concentrations for selection of transformants in *E.coli.*.

ANTIBIOTIC	STOCK CONC.	WORKING CONC.
Ampicillin	100mg/ml in dH$_2$O	50μg/ml
Carbenicillin	100mg/ml in dH$_2$O	50μg/ml
Kanamycin	10mg/ml in dH$_2$O	20μg/ml
Zeocin	100mg/ml in dH$_2$O	25μg/ml

2.3.4   Ethanol precipitation of DNA.

- DNA was precipitated by adding 2.5 volumes 100% Ethanol, 1/10 volume 5M NaCl, and storing at -20°C for at least 30min.

- DNA was spun at full speed in a microcentrifuge for 20min at 4°C.

- The pellet was washed in 70% ice cold Ethanol and allowed to air dry.

- Depending on the experiment, the DNA was either resuspended in a buffer for restriction digestion or in milliQ water if being used for transfection into mammalian cells.

2.3.5   Polymerase Chain Reaction (PCR).

Reactions were performed using Phusion (Finnzymes), a high fidelity *Taq* DNA

polymerase.

For each 50μl reaction the following was used;

DNA template   (200ng cDNA/reaction)
300nM oligo 1
300nM oligo 2
200μM dNTP's
1X PCR Buffer (Supplied by manufacturer)
0.5μl Phusion (5U/μl)
make up to a final volume of 50μl with milliQ dH$_2$O.

The following program was used on a GeneAmp®9700 thermocycler (Applied
Biosystems) to amplify cDNA templates.

98°C 30s
98°C 10s
T_m+3°C lower primer   } 25 cycles
72°C 30s/kb
72°C 10min

See oligonucleotide appendix for details of primers sequences.

3μl of PCR reaction mixture was subjected to TBE agarose gel electrophoresis to

visualise the PCR product. If a single discrete band was seen then TOPO cloning was

performed.

2.3.7   Cloning of blunt end PCR products.

The Zero Blunt® TOPO® cloning kit was used (Invitrogen). TOPO cloning provides

a quick and efficient method to clone blunt ended PCR products such as those

produced by Phusion *Taq* polymerase. The reaction was carried out according to the

manufacturer's protocol. The pCR®-Blunt II- TOPO® vector contains *Eco*RI

flanking the insert site allowing a convenient first screen to analyse correct

transformants via *Eco*RI digestion. Cloned PCR products were then verified by DNA

sequencing at the School of Biological Sciences Sequencing Service- University of

Edinburgh. Sequence data was analysed using Seqman software.

2.3.8   Mutagenesis of plasmid DNA

The template plasmid for the mutagenesis performed in this thesis was IPC 35 (see

plasmid appendix). Single stranded phagemid DNA was produced by M. Robertson

(Sambrook., 2001).

Hybridisation

Annealing Buffer

200mM Tris pH7.4
20mM MgCl$_2$
500mM NaCl

Ratio of template: oligo = 10:1.

- The following were combined on ice in a 1.5ml tube;

  1.3µl   1µM oligo 1
  1.3µl   1µM oligo 2
  1µl.   10X annealing buffer.
  1.3µl   ssDNA IPC35 (200ng)
  5.1µl   milliQ H$_2$O

- The annealing reaction was mixed gently, spun down, and placed in a beaker

  containing 500ml dH$_2$O at 70°C. The beaker was removed from the water

  bath and allowed to cool to room temperature.

Synthesis

- The following were combined on ice, adding the hybridisation mix last.

10µl	Hybridisation mixture
20mM	HEPES pH7.8
2mM	DTT
10mM	$MgCl_2$
500µM	dNTP's
1mM	ATP
2.5 units	T4 DNA ligase (NEB)
2 units	T4 polymerase

  +milliQ dH_2O to a final volume of 100µl.

- The synthesis reaction was incubated for 5 minutes on ice, followed by 5 minutes at room temperature, and finally 2h at 37°C. The reaction was stopped via addition of 3µl 0.5M EDTA.

- DH5α *E.coli* were transformed with 10µl of the synthesis reaction, plated on LB agar plates containing 50µg/ml carbenicillin, and incubated at 37°C overnight.

- The next day, 12 bacterial colonies were picked in to LB broth containing 50µg/ml ampicillin and incubated overnight at 37°C with shaking.

- The next day minipreps were prepared and suitable restriction digests were performed to screen for the putative mutant DNA molecule. These were then verified by DNA sequencing at the School of Biological Sciences Sequencing Service (SBSSS) and sequence data was analysed using Seqman software.

Chapter 3.

Biochemical Characterisation of Nanog.

3.1 Introduction

Nanog has been identified as a key regulator of pluripotency. (Chambers *et al.*, 2003; Mitsui *et al.*, 2003) However, the precise mechanism of Nanog action remains unclear. A complete dissection of the molecular mechanisms by which Nanog acts will involve identification of upstream regulators and downstream target genes. In addition, a thorough biochemical characterisation of Nanog and any associated partner proteins required for Nanog function will also be required.

The ability of a given transcription factor to regulate gene expression depends on a number of biochemical factors including post-translational modification, accessibility to the target DNA sequence and, potentially, inclusion in higher-order multi-protein complex(es). Homeodomain containing proteins bind specifically to DNA sequences with the consensus ATTA and act as sequence specific transcription factors that have myriad roles in development and evolution (reviewed by Gehring, 1987). The homeodomain is a 60 amino acid amino sequence that contains three α helices folded around a hydrophobic core with a flexible arm extending from the N-terminus (reviewed by Wolberger, 1996). Examples are known in which homeodomain proteins bind to DNA as monomers, such Antennapaedia in *Drosophila* (Billeter *et al.*, 1993), hetero-dimers, such as the MATa1 and MATα in

yeast (Li *et al.*, 1995), or homo-dimers, such as the paired class homeodomains (Wilson *et al.*, 1995) .

In simplistic terms Nanog can be thought of as a 305 residue protein containing three domains; a centrally located homeodomain flanked by a serine rich N-terminus and the C-terminus (Figure 3.1). The Nanog homeodomain is most closely related to the NK2 family of homeodomain proteins sharing 50% identity across the homeodomain to mouse members of this family (Chambers *et al.*, 2003). NK2 family members are characterised by the presence of a sequence conforming to the consensus PARRIAVPVLVRDGKPCL located 15 residues C-terminal to the homeodomain as well as a conserved tyrosine residue within DNA binding α-helix of the homeodomain (Lints *et al.*, 1993). As Nanog lacks these signature motifs it cannot be classified as a member of the NK2 family. In fact, outwith the homeodomain there is little homology to other proteins that can be used to infer biological function. The Nanog C-terminal domain can be further subdivided due to the presence of a centrally located pentapeptide repeat in which a tryptophan is present at every 5th residue. The C-terminal domain is likely to be of functional importance as it has been shown to possess two transactivation domains as judged by Gal4-fusion luciferase assays (Pan and Pei, 2003; Pan and Pei, 2005). Furthermore, the tryptophan residues have been shown to be implicit to this transactivation, although these experiments were not performed in ES cells (Pan and Pei, 2005). In addition, Nanog can transactivate a reporter driven by its cognate binding sequence in non-pluripotent cells (Pan and Pei, 2005). Whether Nanog can transactivate endogenous genes in ES cells is not clear. N-terminal to the homeodomain, Nanog contains a sequence which

```
         10            20            30            40
MS VGLPGP HS LPSS EEAS NS GN ASSMP AVFHP ENYS CLQG
                  N-terminal domain
         50            60            70            80
SATEML CTEAAS PRPSS EDLP LQGS PDSS TSP KQKLS SPE

         90           100           110           120
ADKGP EEEENKVLAR KQKMR T V F S Q A Q L C A L K D R F Q K Q K Y
                  Homeodomain
        130           140           150           160
L S L Q Q M Q E L S S  I  L N L S  Y K Q V  K T W F  Q N  Q R M K C  K R W Q KNQWL

        170           180           190           200
KTS NGL I QKGS AP VEYPS I HCS YPQGYLVNAS GS LS MWGS
                      C-N
        210           220           230           240
QTWTNP TWSS QTWTNP TWNNQTWTNP TWSS QAWTAQS WNG
                  Tryptophan repeat
        250           260           270           280
QPWNAAP LHNEGEDFLQPYVQLQQNESASDLEVNLEATRE
                      C-C
        290           300      305
SHAHESTPQALELFLNYSVTPPGEI
```

Figure 3.1- Amino acid sequence of mouse Nanog protein.

The coloured boxes show the boundaries of putative domains within Nanog. The C-terminal domain begins at K156 and is further divided due to the presence of the tryptophan repeat into C-N (K156-M197), tryptophan repeat (W198-W243), and C-C (N244-I305).

has homology to a sequence of unknown function within the linker region of Smad4 (Hart *et al.*, 2004). To begin biochemical characterisation of Nanog, experiments are described below which assess the multimerisation capacity of Nanog, post-translational modification of Nanog protein, as well as a candidate based approach to identify Nanog interacting proteins.

3.2 Nanog multimerises through sequences within the C-terminal domain.

To assess whether Nanog protein could multimerise, transient transfections were performed using two Nanog plasmids with distinct epitope tags fused to the N-terminus (Figure 3.2). This allows subsequent immunoprecipitates collected with an antibody against one of the tags to be examined by immunoblotting for the presence of Nanog tagged by the second epitope. Whole cell lysates were prepared in 0.5% NP-40, 50mM Tris pH8.0, 150mM NaCl and $(Flag)_3$Nanog was immunoprecipitated with α-Flag-M2 antibody. The immunoprecipitates were subjected to SDS-PAGE followed by immunoblotting and detection with α-HA antibody. Experiments in E14/T mouse ES cells (Aubert *et al.*, 2002) show that indeed Nanog is capable of multimerising with itself (Figure 3.3). In addition, as a similar interaction was obtained following transfection of COS-7 cells (Figure 3.3), showing Nanog multimerisation does not depend on additional proteins present in mouse ES cells yet absent from African green monkey kidney fibroblasts. Furthermore, the C-terminus of Nanog is absolutely required for this interaction as $(HA)_3$NanogΔC is not co-immunoprecipitated with $(Flag)_3$Nanog (Figure 3.4). However, truncations lacking the last 20 or 49 residues retain the ability to interact with a second Nanog molecule
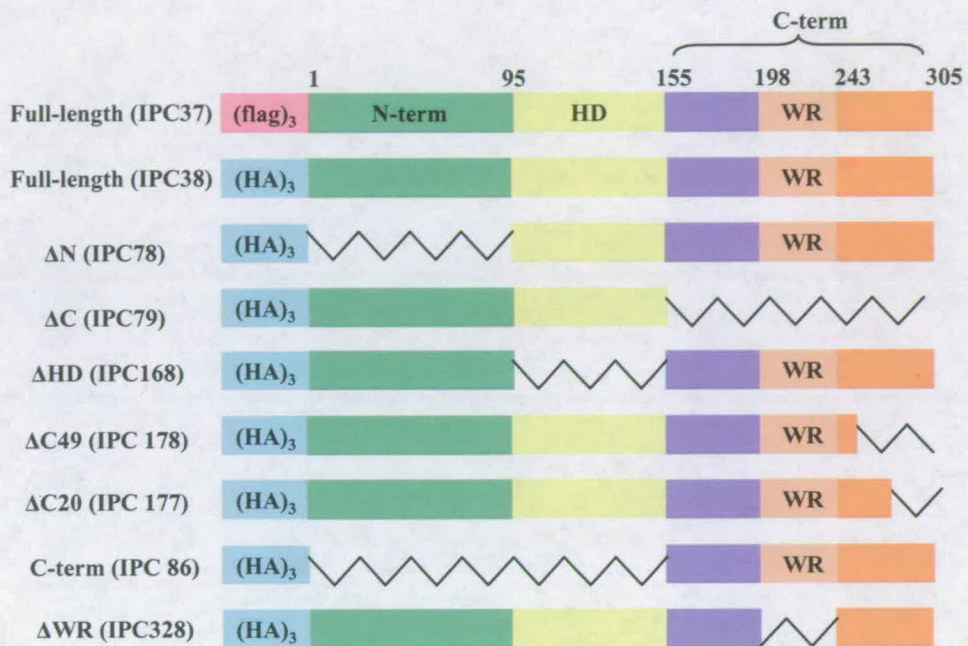
Figure 3.2- Schematic representation of epitope tagged Nanog deletion mutants.

Numbers refer to the amino acid residue. N-term- N-terminal region; HD- homeodomain; WR- Tryptophan repeat; C-term- C-terminal region. IPC numbers are plasmid database reference number.

⋁⋁⋁⋁ Indicates the deleted region.
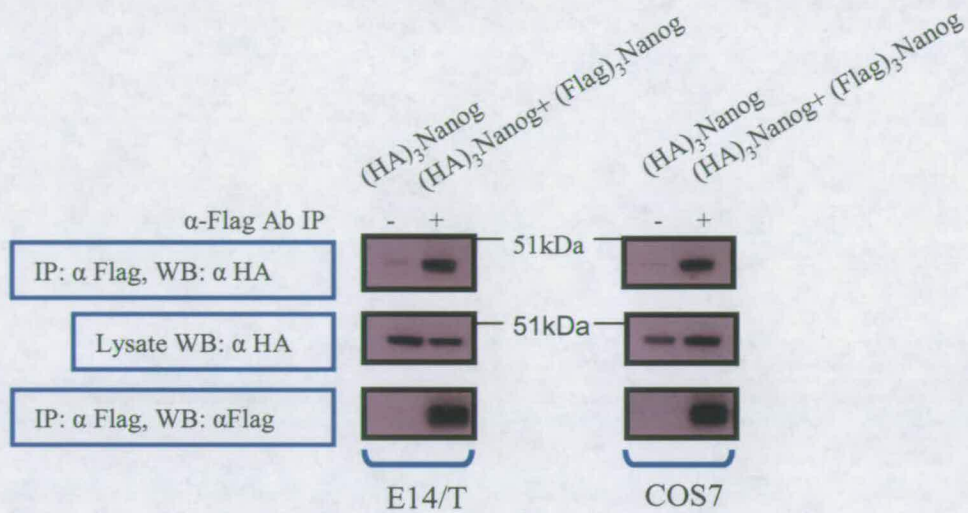
Figure 3.3- Nanog is able to multimerise with itself in ES cells and COS-7 cells

Co-transfections of (Flag)₃Nanog and (HA)₃Nanog into E14/T ES cells and COS-7 cells were performed. α-Flag-M2 immunoprecipitates were prepared and immune complexes separated by SDS-PAGE. The immunoblots were probed with α-HA antibody.. Further details of the protocol are provided in Materials and Methods (section 2.2.8.1).
Constructs are schematically depicted in Figure 3.2
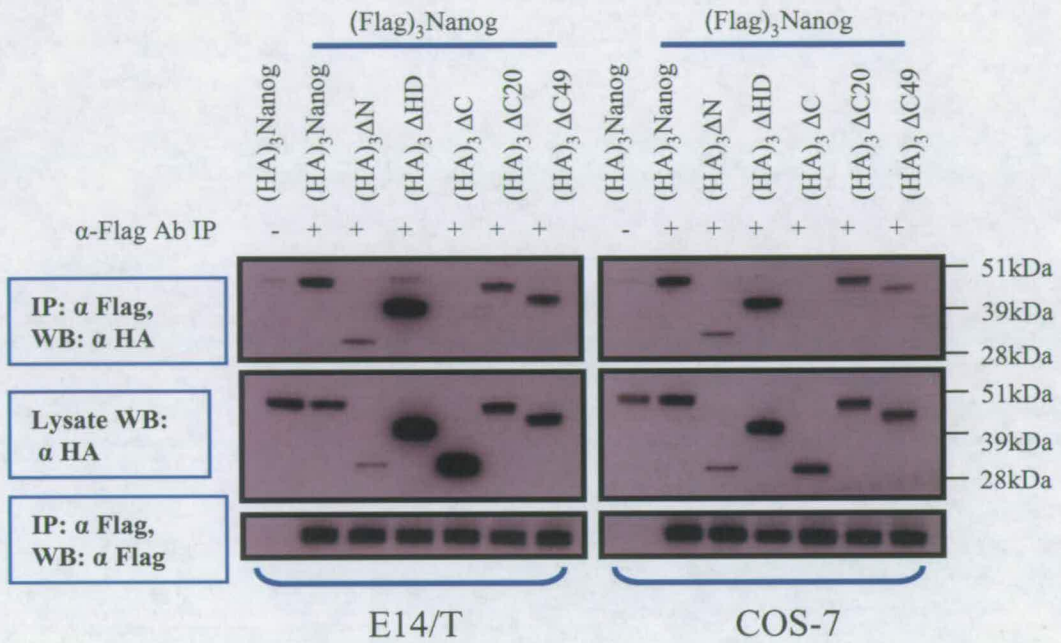
Figure 3.4- The C-terminal domain of Nanog is necessary for multimerisation in ES cells and COS-7 cells in transient co-transfection experiments.

Transfections were performed and processed as described in Figure 3.3 using the deletion mutants indicated and diagrammed in Figure 3.2.

(Figure 3.4), suggesting that the site of interaction lies between the end of the homeodomain at K156 and L256.

To determine whether the multimerisation capacity of Nanog could localise to the C-terminal domain, further transient co-transfections of COS-7 cells were performed. Immunoprecipitation with the α-Flag-M2 antibody followed by SDS-PAGE and immunoblotting with α-HA antibody indicated that under these conditions (0.5%NP-40, 50mM Tris pH8.0, 150mM NaCl), the C-terminal domain can interact with a second Nanog molecule (Figure 3.5). Taken together these data show that the C-terminal domain of Nanog is not only required but also sufficient for Nanog multimerisation.

3.3   The role of the Nanog tryptophan repeat.

3.3.1   The tryptophan repeat is necessary for Nanog multimerisation.

The preceding experiments indicated that the region between K156 and L256 was responsible for Nanog mulitmerisation. As the tryptophan repeat lies in this region, a mutant was generated lacking this repeat sequence; (HA)$_3$NanogΔWr (see plasmid appendix for cloning strategy). Transient co-transfection experiments using E14/T ES and COS-7 cells demonstrated that (HA)$_3$NanogΔWr could not be co-immunoprecipitated with (Flag)$_3$Nanog. This indicates that residues within the tryptophan repeat are required for interaction with a second molecule of Nanog (Figure 3.6).
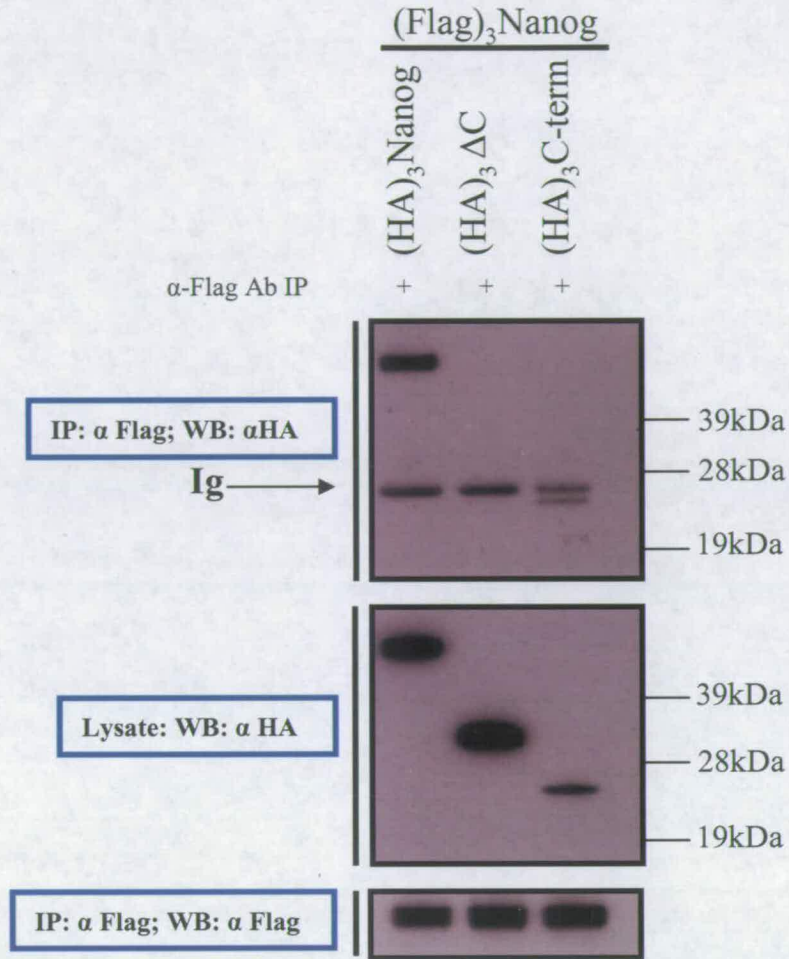
Figure 3.5- The C-terminal domain of Nanog is sufficient to interact with a second molecule of Nanog in transient COS-7 cell transfections.

Co-transfections of (Flag)₃Nanog and (HA)₃Nanog deletion mutants into COS-7 cells were performed as indicated. Cells were lysed (0.5%NP-40, 50mM Tris pH8.0, 150mM NaCl), α-Flag-M2 immunoprecipitates were prepared, immune complexes separated by SDS-PAGE and immunoblotted. The immunoblots were probed with α-HA. Further details of the protocol are found in the materials and methods (section 2.2.8.1).
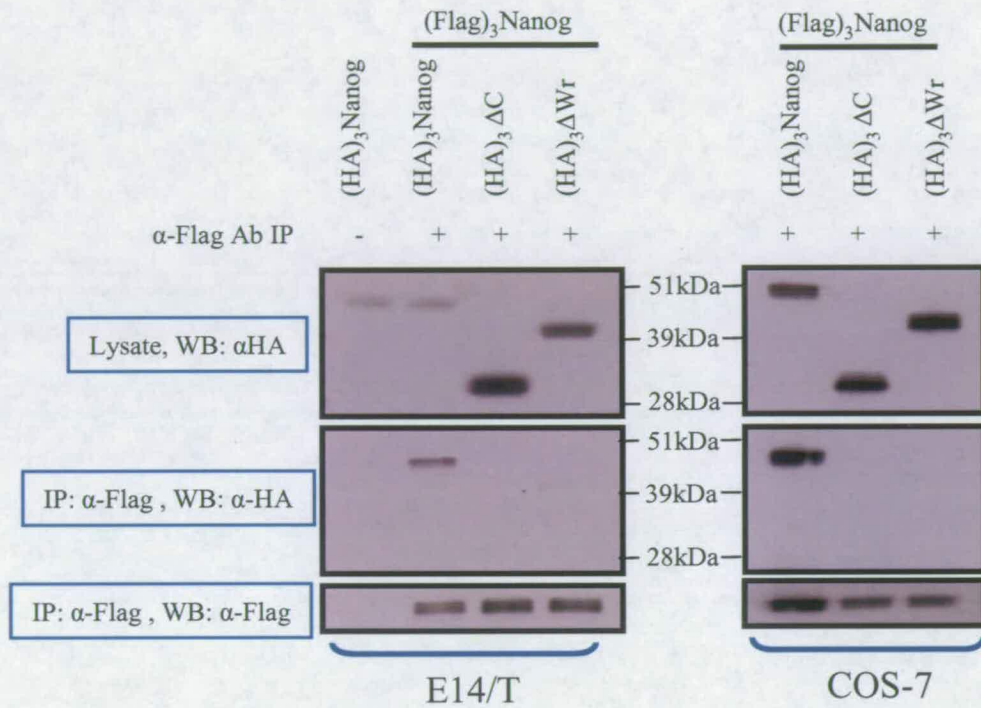
Figure 3.6- The tryptophan repeat of Nanog is necessary for Nanog-Nanog interaction in transient ES and COS-7 cell transfections.

Co-transfections of (Flag)₃Nanog and (HA)₃Nanog deletion mutants into E14/T ES cells and COS-7 cells were performed as indicated. α-Flag-M2 immunoprecipitates were prepared and immune complexes separated by SDS-PAGE and immunoblotted. The immunoblots were probed with α-HA antibody. Details of the immunoprecipitation protocol are found in material and methods 2.2.8.1

3.3.2   The tryptophan repeat is functionally important in mouse ES cells.

In addition to being required for multimerisation, the Nanog tryptophan repeat is required for transactivator function of Nanog (Pan and Pei, 2005). This conclusion rests on experiments conducted in HEK293 cells in which a luciferase reporter under the control of multimerised Gal4 DNA binding sites was co-transfected with fusion proteins formed between the Gal4 DNA binding domain and fragments of Nanog. Mutation of tryptophan residues indicated at least some of the tryptophan residues are required for function. However, the importance of the tryptophan repeat for Nanog function in ES cells is not known. To assess this, E14/T cells were super-transfected (Aubert *et al.*, 2002; Gassmann *et al.*, 1995) with either *(HA)₃Nanog* , *(HA)₃NanogΔC* , *GFP*, or *(HA)₃ NanogΔWr* and transfectants selected in puromycin in the presence or absence of LIF. Colonies were photographed after ten days clonal expansion, and stained for alkaline phosphatase activity the following day. For example undifferentiated, mixed, and differentiated colonies see Figure 4.4a. As expected, full length *Nanog* expression provides the highest proportion (91%) of uniformly undifferentiated colonies in the presence of LIF. A high proportion (>80%) of colonies expressing either *GFP* or *(HA)₃NanogΔWr* are also uniformly undifferentiated in the presence of LIF but these colonies differ qualitatively from those expressing *(HA)₃Nanog*. The latter form raised colonies that are highly refractile under phase contrast microscopy (Figure 3.7a). In contrast, *GFP* expressing colonies appear flatter and *(HA)₃NanogΔWr* expressing colonies are either very small or show some peripheral differentiation when grown in the presence of LIF (Figure 3.7a). Interestingly, in the presence of LIF, *(HA)₃NanogΔC* expressing colonies show a much higher proportion of mixed colonies (57%) compared to the other

constructs suggesting it may have a dominant negative effect on Nanog action. Upon cytokine withdrawal, 57% of colonies expressing *(HA)₃Nanog* remain uniformly undifferentiated whereas the only 7% of *(HA)₃NanogΔWr* colonies are undifferentiated. The inability of the mutant Nanog molecules to efficiently direct LIF independent self-renewal can also be seen by visually inspecting the AP stained plates without microscopy (Figure 3.7b) which shows the gross differences between full-length *(HA)₃Nanog* and the mutant molecules. No uniformly undifferentiated colonies expressing *GFP* and *(HA)₃NanogΔC* were observed in the absence of LIF. Taken together these results indicate that the Nanog C-terminal domain is required for ES cell self-renewal and that the tryptophan repeat constitutes a significant part of this requirement. However, these experiments also suggest that additional motifs within the C-terminal domain but outwith the homeodomain are likely to be important for Nanog function.

(a)

	GFP	Nanog	ΔC	ΔW-repeat
+LIF				
-LIF				

(b)

	ΔC	Nanog	GFP	ΔW-repeat
-LIF				
+LIF				
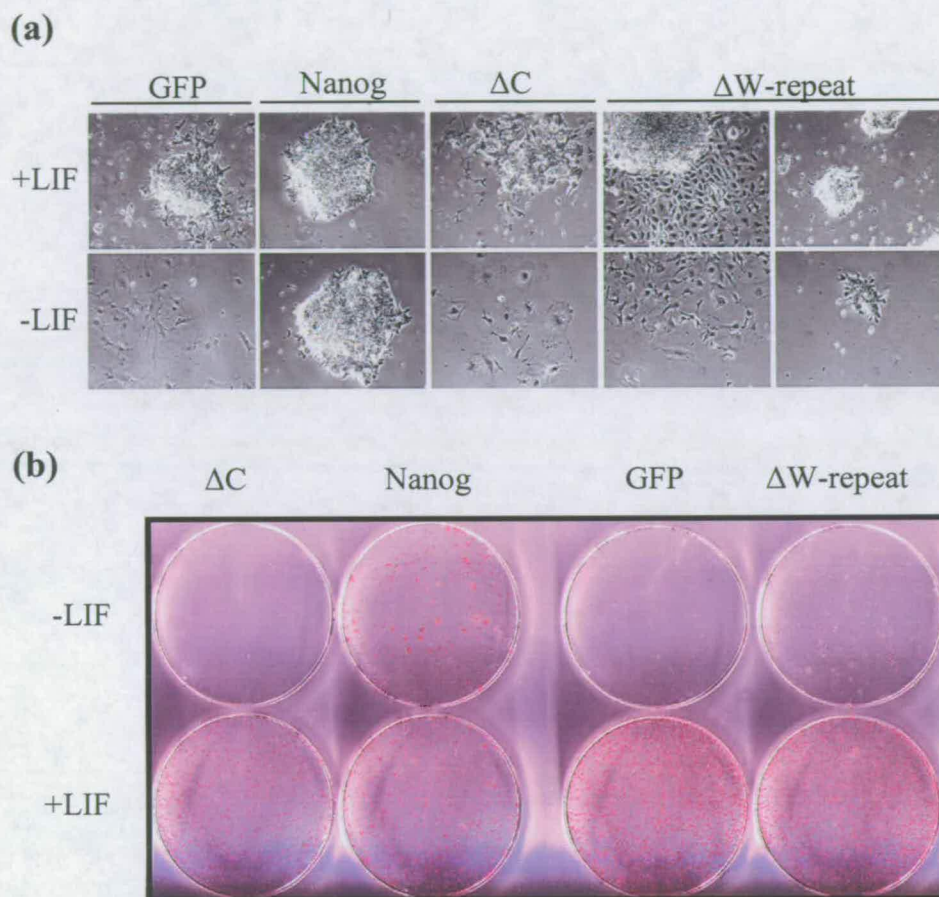
Figure 3.7- Functional assessment of tryptophan repeat in E14/T ES cells.

E14/T ES cells were transiently transfected with the constructs indicated: $(HA)_3\Delta C(IPC\ 79)$, $(HA)_3Nanog$ (IPC 38), GFP (AGS 684), $(HA)_3Nanog\Delta W$-repeat (IPC 328); $5x10^4$ cells plated at a density of 640 cells/cm^2 and cultured in the presence and absence of LIF in 2µg/ml puromycin. After 11 days clonal growth, the cells were photographed under phase microscopy (a), before alkaline phosphatase staining, (b).

3.4   Post-translational modification of Nanog in mouse ES cells

3.4.1   Post-translational modifications

Protein function depends not only on the primary amino acid sequence but also on post-translational modifications and how these affect quaternary structure. Many transcription factors exhibit post-translational modification including ubiquitination, sumoylation, which inhibits transcription factor activity (Gill, 2005), acetylation, methylation, and phosphorylation. Post-translational modification can alter protein function. For example, the phosphorylatable transcription factor Stat3 exists in an inactive state in the cytoplasm when unphosphorylated (Stahl *et al.*, 1995). Following phosphorylation, Stat3 dimerises and translocates to the nucleus, the site of transcriptional activity (Ihle, 1996). In this section an experiment is presented investigating the post-translational modification of Nanog protein.

3.4.2   Nanog is a phosphorylated protein

Nanog has a predicted molecular weight of 35kDa although it migrates much slower than predicted upon SDS-PAGE. An antibody raised against amino acids 2-16 of mouse Nanog (Chambers, 2005) recognises three bands that migrate with an apparent molecular weight (M_r) of ~42kDa on immunoblots of mouse ES cell lysates. It is known that post-translational modifications can affect the migration of proteins during electrophoresis. To investigate whether these bands of retarded migration were due to phosphorylation of Nanog, Nanog immunoprecipitates were prepared from the Nanog over-expressing cell line EF4 (Chambers *et al.*, 2003) using anti-Nanog antibody (Chambers, 2005). These immunopreciptitates were then washed extensively and treated with two independent phosphatases whilst immobilised on
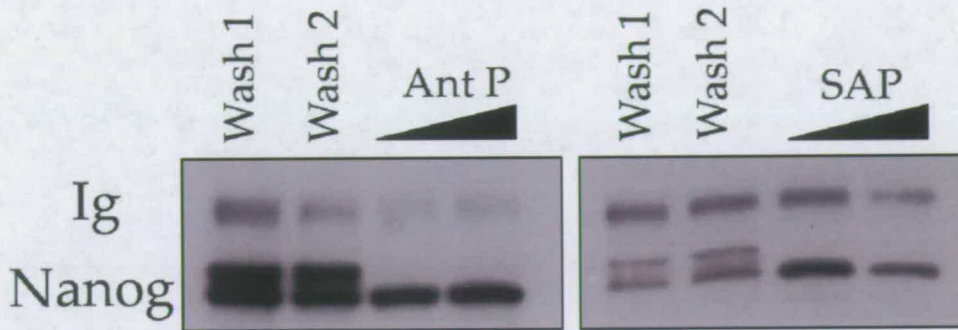
Figure 3.8- Nanog is a phosphoprotein in mouse ES cells.
(Taken from Yates and Chambers, 2005)

EF4 cells (1.2×10^7) cultured in the presence of LIF, were lysed in 1.2ml lysis buffer (0.5%NP-40, 150mM NaCl, 50mM Tris pH8.0), and α-Nanog immunoprecipitates were prepared. The immune complexes were collected on protein A sepharose beads and washed in lysis buffer (wash 1), followed by washing in manufacturer's phosphatase buffer (wash 2). Finally, the immunoprecipitates were subjected to on-bead phosphatase treatment with increasing concentrations of either antarctic phosphatase (Ant P; NEB) or Shrimp alkaline phosphatase (SAP; Roche) as indicated. Note the collapse of the triplet signal to a single band in the phosphatase treated samples; Ig is the heavy chain immunoglobulin. For a detailed protocol see section 2.2.8.2 of this thesis.

the solid phase beads. Subsequent immunoblotting analysis with anti-Nanog antibody revealed that the triplet signal normally observed had collapsed to a single band upon treatment with either antarctic phosphatase or shrimp alkaline phosphatase (Figure 3.8), (Yates and Chambers, 2005).

3.5   Nanog partner proteins: a candidate approach

A major goal towards delineation of the mechanism of Nanog action is to identify Nanog partner proteins in mouse ES cells. At the outset of this project no such partner proteins of Nanog had been identified. Therefore potential partner proteins were examined in a candidate based approach. Interactions were examined between Nanog and two other key regulators of ES self-renewal, Stat3 and Oct4. As Stat3, Oct4, and Nanog are all transcription factors with important roles in ES cell self-renewal, a simple hypothesis is that these molecules might act together in a complex on target genes.

3.5.1   Stat3

The functional expression cloning of Nanog depended upon the ability of Nanog to direct ES cell self-renewal in the absence of LIF. However, the self-renewal efficiency of ES cells over-expressing *Nanog* is enhanced if the cells are treated with LIF. Stat3 is known to be a downstream effector of LIF signalling important for ES cell self-renewal (Matsuda *et al.*, 1999; Niwa *et al.*, 1998). As Stat3 does not appear to be a downstream transcriptional target of Nanog (Chambers *et al.*, 2003) the co-operative effect of Nanog and Stat3 on ES cell self-renewal could result from a direct interaction between Nanog and Stat3. Nanog protein was immunoprecipitated from

EF4 cells using anti-Nanog antibody (Chambers, 2005). The immunoprecipitates were subjected to SDS-PAGE and immunoblots probed with anti-Stat3 antibody. No interaction between Nanog and Stat3 was observed under the conditions tested (Figure 3.9).

3.5.2   Oct4

Over-expression of *Oct4* results in differentiation into a mixture of cell types that express markers of endoderm and mesoderm (Niwa *et al.*, 2000). This cellular differentiation is similar both morphologically and in terms of marker expression to that seen upon LIF withdrawal. This similarity in phenotype caused by these treatments led to the proposition that a binding partner of Oct4 existed that was present in limiting amounts and that was maximally active in cells stimulated by LIF (Niwa, 2001). As Nanog fits both these criteria and as the pro-differentiative effect of *Oct4* over-expression is attenuated in ES cells over-expressing *Nanog* (Chambers unpublished) a direct interaction between Nanog and Oct4 seemed feasible. Nanog protein was immunoprecipitated from EF4 cells using anti-Nanog antibody (Chambers, 2005). The immunoprecipitates were subjected to SDS-PAGE and immunoblots probed with anti-Oct4. No interaction between Nanog and Oct4 was observed under the conditions tested (Figure 3.9).

3.5.3   Generation of *(Flag)₃Oct4: (HA)₃Nanog* expressing cell line.

The failure to detect an interaction between Nanog and either Oct4 or Stat3 described in the previous section could be due to the low affinity between the immunoprecipitating antibody and its antigen. Alternatively, the
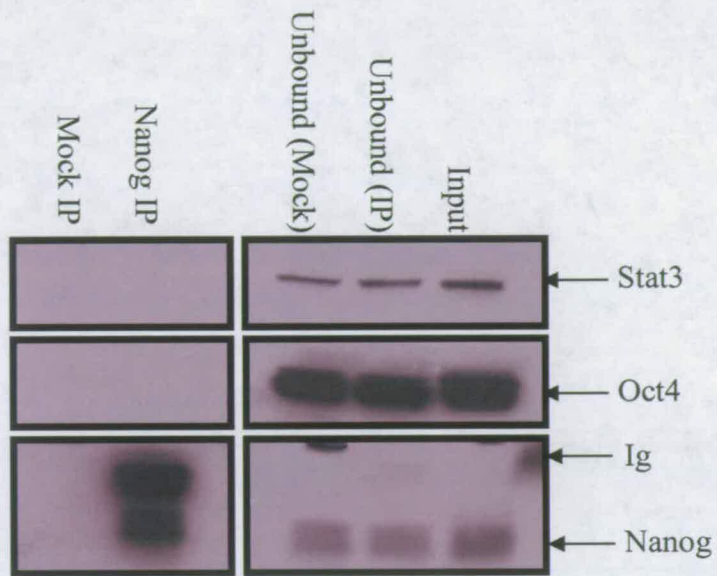
Figure 3.9 Nanog does not interact with Stat3 or Oct4 in mouse ES cells.

Nanog was immumnoprecipitated from from EF4 cells using anti Nanog antibody and subjected to SDS-PAGE. After transfer to nitrocellulose immunoblots were probed with anti -Oct4-C-10, anti-Stat3, and anti-Nanog antibodies. For details of the immunoprecipitation protocol see 2.2.8.1

immunoprecipitating antibody could disrupt the interaction being sought. To circumvent such potential problems, immunoprecipitations were performed using antibodies against haemagglutinin (HA) and the Flag peptide in conjunction with trimerised epitope tags, as these reagents have been optimised for immunoprecipitation and immunoblotting applications. Having the epitope tag in triplicate increases the avidity of the antibody- antigen interaction, a property particularly important in immunoprecipitation procedures. An ES cell line carrying epitope tagged Nanog and Oct-4 transgenes was generated as a reagent to further examine possible Nanog-Oct4 and Nanog-Stat3 interactions. The cell line was based on the ZHBTc4.1 cell line (Niwa *et al.*, 2000) which has two null *Oct-4* alleles but can be sustained in the undifferentiated state by the expression of doxycycline repressible Oct4 transgene. In cell culture, growth without doxycycline activates the transgene and thus the ES cell phenotype is maintained. To generate a derivative cell line expressing (Flag)₃Oct4 protein, ZHBTc4.1 cells were transfected with a *(Flag)₃Oct4* transgene and selected in 1μg/ml puromycin in the presence of doxycycline. Under these conditions cells will differentiate down the trophectodermal lineage unless (Flag)₃Oct4 is expressed at the appropriate level. The fact that undifferentiated clones were obtained shows that the fusion of the (Flag)₃ epitope tag to the N-terminus of Oct4 does not affect Oct4 function. One of the (Flag)₃Oct4 clones obtained, clone c6, was then then transfected with a *(HA)₃Nanog* transgene and selected in G418. A schematic diagram of the *(Flag)₃Oct4:(HA)₃Nanog* cell line is shown (Figure 3.10a). Immunoblotting of *(Flag)₃Oct4* clone c6 whole cell lysates detects (Flag)₃Oct4 protein, and in *(HA)₃Nanog* derivative lines, the size shifted (HA)₃Nanog protein is observed with

(a)
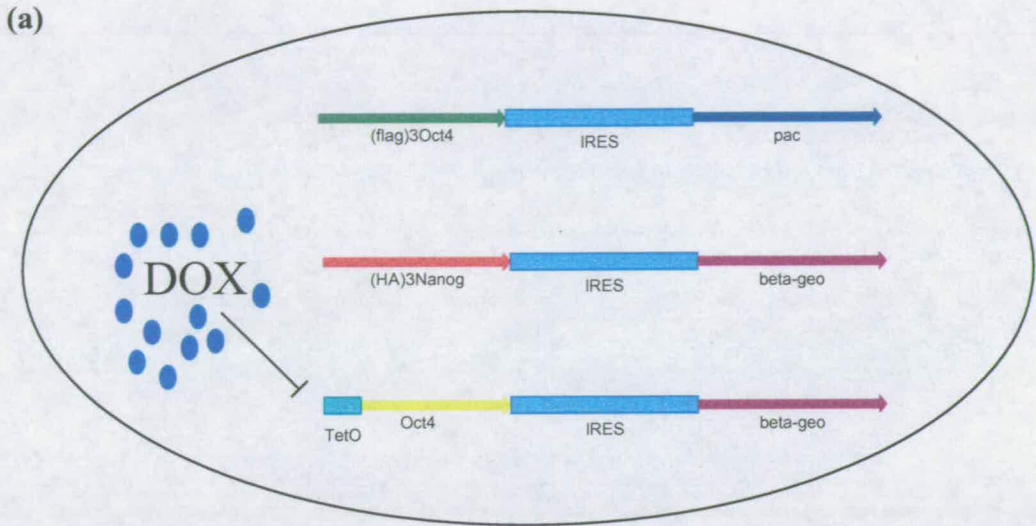
ZHBTc4.1 *(Flag)₃Oct4: (HA)₃Nanog cells*

(b)
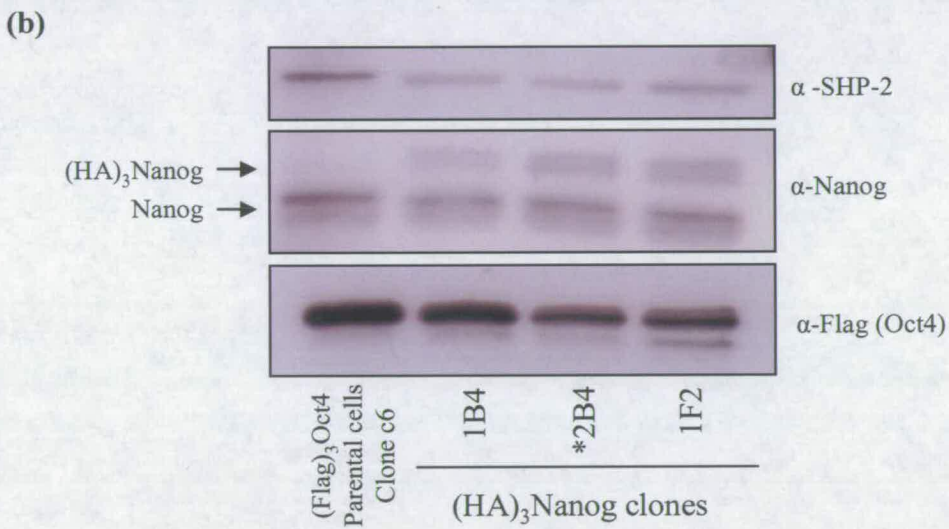
Figure 3.10 *(Flag)₃Oct4: (HA)₃Nanog* **cell line generation.**

(a)   Schematic depiction of *(Flag)3Oct4: (HA)3Nanog* cell line. ZHBTc4.1 cells were stably transfected with *(Flag)₃Oct4IRESpac* and selected with puromycin in the presence of doxycycline to generate (Flag)₃Oct4 parental cells. Clone c6 was then stably transfected with *(HA)₃NanogIRESβ-geo* and selected in G418.

(b)   *(Flag)₃Oct4: (HA)₃Nanog* whole cell lysates were resolved by SDS-PAGE, immunoblotted, and probed with the indicated antibodies. SHP-2 acts as a loading control. Subsequent immunoprecipitation experiments were performed using clone 2B4.

the expected retarded migration relative to the endogenous protein (+ 4kDa) (Figure 3.10b). As the *(Flag)₃Oct4* cells are grown in doxycycline, all the Oct4 protein in the cell is Flag tagged, whereas the HA tagged Nanog represents approximately one third of the total Nanog protein.

3.5.4   Nanog does not interact with Oct4 or Stat3 in *(Flag)₃Oct4: (HA)₃Nanog* ES cells.

(HA)₃Nanog was immunoprecipitated from *(Flag)₃Oct4: (HA)₃Nanog* ES cell lysates using anti-HA antibody. The immunoprecipitates were resolved by SDS-PAGE, immunoblotted, and probed with anti-Stat3 and anti-Flag antibodies. The results show that (HA)₃Nanog is clearly enriched in the anti-HA immunoprecipitates, however (Flag)₃Oct4 and Stat3 do not co-immunoprecipitate with (HA)₃Nanog under these conditions (0.5%NP-40, 50mM Tris pH7.5, 150mM NaCl) (Figure 3.11).

3.6   Sall4 interaction studies in mouse ES cells.

3.6.1   Sall4 physically interacts with Nanog and Oct4

During the course of this study an interaction between Nanog and the Spalt like Zinc finger protein Sall4 was described in ES cells (Wu *et al.*, 2006). This interaction has been confirmed by detecting the presence of Sall4 protein in Nanog immunoprecipitates from EF4 ES cells (Figure 3.12a). The domain within Nanog responsible for the interaction with Sall4 was identified as the homeodomain (Wu *et al.*, 2006). Comparing the Nanog homeodomain sequence with all other known homeodomain sequences identifies a motif (SLQQ), located at the N-terminal end of α-helix 2, that is present in only one other homeodomain, Oct4. This may indicate
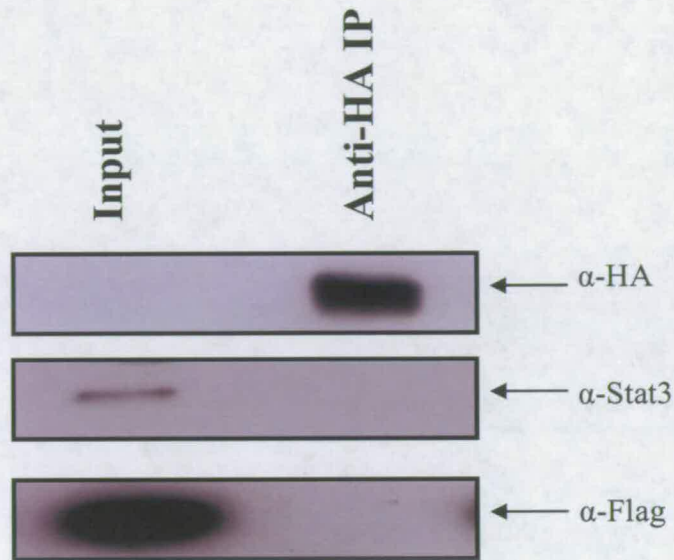
Figure 3.11 Nanog does not interact with Stat3 or Oct4 in
(Flag)3Oct4: (HA)3Nanog cells.

(HA)₃Nanog was immunoprecipitated from *(Flag)₃Oct4: (HA)₃Nanog*
cell lysates using anti-HA antibody and subjected to SDS-PAGE. After transfer
to nitrocellulose, immunoblots were probed with anti -Flag and anti-Stat3. For
details of the immunoprecipitation protocol see 2.2.8.1.
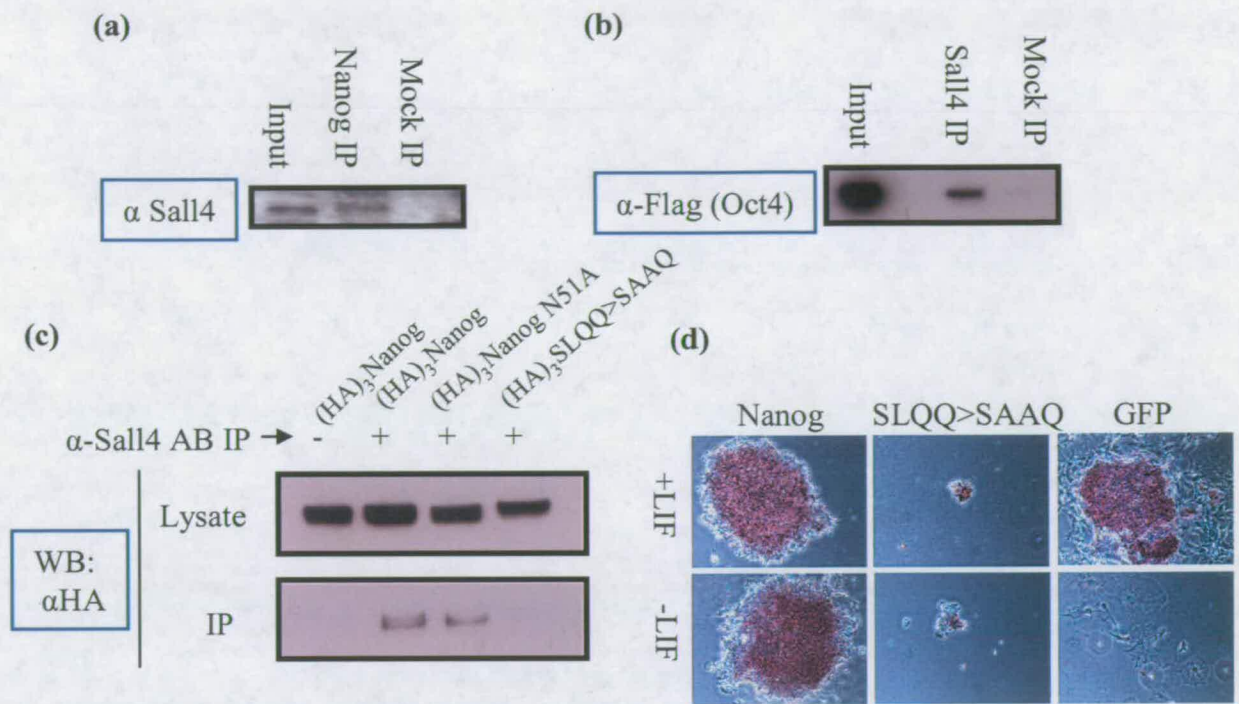
Figure 3.12- Sall4 interacts with Nanog and Oct4

(a) 1×10^7 EF4 ES cells were lysed in 1.2ml lysis buffer (0.5% NP-40, 50mM Tris pH8.0, 150mM NaCl). Nanog was immunoprecipitated using anti-Nanog antibody and immune complexes were collected on protein A sepharose. After extensive washing in lysis buffer, the beads were boiled in Laemmli buffer and bound proteins were resolved by SDS-PAGE prior to immunoblotting with anti- Sall4 antibody.

(b) 1×10^6 Nanog/T (super-transfectable Nanog over-expressing cells) were transiently transfected with pPyCAG(Flag)$_3$Oct4 expression construct. ~72h post-transfection cells were lysed in 1.2ml lysis buffer (0.5% NP-40, 50mM Tris pH8.0, 150mM NaCl). Sall4 was immunoprecipitated using anti-Sall4 antibody and immune complexes were collected on protein A sepharose. After extensive washing in lysis buffer, the beads were boiled in Laemmli buffer and bound proteins were resolved by SDS-PAGE prior to immunoblotting with anti- Flag antibody.

(c) 1×10^6 E14/T ES cells were transiently transfected with (HA)$_3$Nanog or indicated Nanog mutant expression constructs. 72h post-transfection, cells were lysed in 1.2ml lysis buffer (0.5% NP-40, 50mM Tris pH8.0, 150mM NaCl). Sall4 was immuno-precipitated using anti-Sall4 antibody and immune complexes were collected on protein A sepharose. After extensive washing in lysis buffer, the beads were boiled in Laemmli buffer and bound proteins were resolved by SDS-PAGE prior to immuno-blotting with anti- HA antibody. Aliquots of cell lysates were subjected to SDS-PAGE followed by immunoblotting to illustrate (HA)$_3$Nanog expression from all constructs.

(d) 1×10^6 E14/T ES cells were transiently transfected with the constructs indicated-, Nanog (IPC 38), GFP (AGS 684), SLQQ>SAAQ (IPC 120). 24h after transfection the cell were trypsinised and re-plated at 640 cells/cm2 in the presence or absence of LIF with selection in 2µg/ml puromycin. After 11 days clonal growth, the cells were stained for alkaline phosphatase activity.

```
hOct1    RRKKRTSIETNIRVALEKSFLENQKPTSEEITMIADQLNMEKEVIRVWFCNRRQKEKRIN
mNanog   KQKMRTVFSQAQLCALKDRFQKQKYLSLQQMQELSSILNLSYKQVKTWFQNQRMKCKRWQ
mOct4    RRKRKRTSIENRVRWSLETMFLKCPKPSLQQITHIANQLGLEKDVVRVWFCNRRQKGKRSS
hNanog   KQKTRTVFSSTQLCVLNDRFQRQKYLSLQQMQELSNILNLSYKQVKTWFQNQRMKSKRWQ
rNanog   KQKMRTVFSQAQLCALKDRFQRQRYLSLQQMQDLSTILNLSYKQVKTWFQNQRMKCKRWQ
hOct4    RKRKRTSIENRVRGNLENLFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRRQKGKRSS
mNkx2.5  RRKPRVLFSQAQVYELERRFKQQRYLSAPERDQLASVLKLTSTQVKIWFQNRRYKCKRQR
```

(b)

(c)

Serine, (S406)

Glutamate, (E407)

Figure 3.13- Model showing the position of the TSEE motif in human Oct1 homeodomain corresponding to the SLQQ motif in the Oct4 and Nanog.

(a)   Multiple sequence alignment of the human Oct1, mouse Nanog, mouse Oct4, human Nanog, rat nanog, human Oct4, and mouse Nkx2.5 homeodomains. Performed using ClustalW v1.83). SLQQ motif or analogous sequence is Highlighted in the blue box.

(b)   Crystal structure of human Oct1 bound to DNA (adapted from Remenyi et al, 2003). Image produced using PyMol software. The DNA is coloured in purple and protein in green. The TSEE motif can be seen in the second α helix.

(c)   A higher magnification of the Oct1 TSEE motif showing the serine and glutamate side chains projecting into space.

that these sequences are involved in interaction of Nanog and Oct4 with a common partner. The most relevant protein for which a structure has been determined is Oct1 (Remenyi *et al.*, 2003). Alignment of Oct4, Nanog, and Oct1 homeodomain amino acid sequences reveals that the human Oct1 homeodomain contains a TSEE motif at the position of SLQQ in Oct4 and Nanog (Figure 3.13a). Mapping the SLQQ motif onto the three dimensional structure of Oct1 (Remenyi *et al.*, 2003) suggests that these residues protrude from the homeodomain along the helical axis (Figure 3.13b). The position of Oct1 S406 and E407 are labelled on Figure 3.13c and these residues correspond to the central LQ of the SLQQ domain in Oct4 and Nanog. The leucine residue is hydrophobic compared to the hydrophilic serine residue at the analogous position in Oct1, and an uncharged glutamine is found in the SLQQ motif in place of the acidic glutamate residue in TSEE. These considerations predict that if the SLQQ motif is involved in the Sall4 interaction, then Oct4 should also interact with Sall4. To determine whether Sall4 and Oct4 physically interact, a *(Flag)₃Oct4* expression plasmid was transiently transfected into Nanog/T ES cells (supertransfectable EF4 derivatives). Sall4 immunoprecipitates prepared using anti-Sall4 antibody (Elling *et al.*, 2006) were immunoblotted and probed with anti-Flag antibody. A co-immunoprecipitating band corresponding to (Flag)₃Oct4 protein is specifically observed in the Sall4 immunoprecipitate (Figure 3.12b) This suggests that Oct4 can physically interact with Sall4 in mouse ES cells.

3.6.2 The Nanog SLQQ motif mediates Sall4 interaction.

These considerations prompted an analysis of whether the site of Sall4 interaction with Nanog is the SLQQ motif in the homeodomain. To this end E14/T cells were

transiently transfected with *(HA)₃Nanog, (HA)₃Nanog N51A* (a putative DNA binding mutant), or *(HA)₃NanogSLQQ>SAAQ.* Co-immunoprecipitation experiments were carried out (according to protocol 2.2.8.1) with the anti-Sall4 antibody (Elling *et al.*, 2006), immunoprecipitates were subjected to SDS-PAGE, and subsequent immunoblots probed with anti-HA antibody. These experiments show that full length Nanog and the N51A mutant are able to interact with Sall4 with comparable efficiency but the NanogSLQQ>SAAQ mutant is dramatically impaired in this ability (Figure 3.12c).

3.6.3 SLQQ>SAAQ function in mouse ES cells.

To assess whether this perturbation in Nanog-Sall4 interaction has consequences for mouse ES cell self-renewal. E14/T cells were transiently transfected with either *(HA)₃Nanog, (HA)₃ Nanog SLQQ>SAAQ,* or *GFP* and clonally expanded in the presence or absence of LIF for 12 days in puromycin selection. The resultant colonies were then photographed and stained for alkaline phosphatase (Figure 3.12d). As expected the full length Nanog was able to direct clonal ES cell self-renewal in the absence of LIF, whereas the colonies expressing *GFP* terminally differentiated in the absence of LIF. Interestingly, colonies expressing *(HA)₃NanogSLQQ>SAAQ* did not differentiate in the absence of LIF but formed minute rounded colonies similar to those found in the presence of LIF with this mutant. This suggests that the Nanog SLQQ>SAAQ mutation may not affect ES self-renewal *per se* but rather have an effect on ES cell proliferation.

3.7 Discussion

3.7.1 Nanog multimerisation

The Nanog C-terminus is necessary and sufficient for Nanog multimerisation, and the domain mediating interaction has been identified as the tryptophan repeat sequence (198W- 243W) (Figure 3.6). However, it remains to be determined whether Nanog is an obligate multimer. If so, does Nanog solely form homo-multimers or could multimerisation with other proteins have functional significance? Recently a family of proteins has been identified called the pentapeptide repeat proteins which have either a leucine or phenylalanine residue repeated every 5th amino acid (reviewed by Vetting et al., 2006). A crystal structure of one of the family members, MfpA from M. tuberculosum, has revealed that this protein consists of mainly right-handed β-helix that has eight coils of 4 sides each which stack on top of one another (Hegde et al., 2005). MfpA acts to mimic DNA and dock with DNA gyrase, thus preventing the DNA-DNA gyrase interaction and providing resistance to fluroquinilones (Hegde et al., 2005). The tryptophan repeat may be able to adopt this unusual stacked coil structure (2.5 coils) as the tryptophan is a hydrophobic aromatic residue similar to phenylalanine. If this were the case Nanog might act as DNA mimic either to recruit DNA binding proteins to gene targets or to titrate particular transcription factors away from the DNA providing an indirect mechanism for regulating transcription. The latter hypothesis is particularly relevant given the proposal that Nanog may act to limit the action of differentiation inducing transcription factors (Chambers and Smith, 2004). Mutant Nanog proteins lacking only the tryptophan repeat or the C-C domain are not impaired in transactivation of a reporter driven by a multimerised Nanog binding site (Pan and Pei, 2005). However,

deletion of both the tryptophan repeat and the C-C domain abolished transactivation activity in this assay. These data suggests that it is unlikely the tryptophan repeat provides additional DNA binding specificity *per se* at the cognate Nanog binding site. However, it remains unclear whether these conclusions hold true at endogenous genes targets, as these experiments were not performed in ES cells. Moreover, the requirements for transactivation may differ if Nanog binds to sites distal from the promoter of endogenous target genes.

The fact that the Nanog-Nanog interaction can be seen in COS-7 cells (Figure 3.3 and 3.4) suggests that the interaction is independent of any ES cell-specific partner protein or indeed any mouse specific protein. Whether the Nanog-Nanog interaction is direct is however another matter. It is possible that a partner protein is required to bridge two molecules of Nanog, and an orthologous protein provides this link in COS-7 cells. *In vitro* translation experiments using the low complexity of nuclear proteins present in rabbit reticulocyte lysates could be used to further investigate this. It is also possible that the multimerisation of Nanog is dependent on DNA, and future experiments performed in the presence of DNAse or Benzonase to digest the DNA could address this possibility.

In addition to the identification of the Nanog C-terminus as the domain that is both necessary and sufficient for Nanog multimerisation, one can infer from these data that the interaction is homotypic. This is because in each set of co-immunoprecipitates there is one full-length (Flag)$_3$Nanog molecule, therefore if the C-terminus was interacting with the N-terminus or the homeodomain then

(HA)₃NanogΔC would co- immunoprecipitate with full-length Nanog; an interaction that has never been observed.

Considering the functional data (Figure 3.7) in the context of the multimerisation data, one can hypothesize that the decreased self-renewal efficiency of *(HA)₃NanogΔWr* and non-function of *(HA)₃NanogΔC* are due to an inability to multimerise with other Nanog molecules, a process that could be important for Nanog function. This suggests that Nanog may function as a homo-multimer or larger multimer, either alone or as part of a higher order multi-protein complex. Alternatively, if the tryptophan repeat sequence can indeed act as a DNA mimic, the impaired function of the mutant Nanog could be due to inability to sequester differentiation inducing transcription factors away from gene targets. *(HA)₃NanogΔWr* does however direct limited self-renewal in the absence of LIF in contrast to the complete lack of self-renewing colonies obtained upon *GFP* expression. It is possible that (HA)₃NanogΔWr could up-regulate endogenous *Nanog* expression leading to the limited self-renewal observed. Alternatively, it is possible that (HA)₃NanogΔWr acts on target genes in multimers with a small amount of endogenous Nanog protein, not detectable by co-immunoprecipitation studies. By performing the functional assessment of (HA)₃NanogΔWr in *Nanog^{-/-}* ES cells (Chambers unpublished) the importance of the tryptophan repeat can be more robustly monitored, as all the Nanog protein in the cell will be mutant.

3.7.2 Nanog post-translational modification

Nanog is a phosphoprotein but what might the function of this phosphorylation event be? It is possible that phosphorylation may alter the DNA binding properties of Nanog as has been described for Nkx2.5 which exhibits increased DNA binding ability upon phosphorylation of the homeodomain (Kasahara and Izumo, 1999). Alternatively or additionally, phosphorylation may be important for interaction with partner proteins. To address these issues it is first necessary to identify the residue(s) phosphorylated. This could be achieved by immunopurification of Nanog from mouse ES cell lysates followed by mass-spectrometric (MS) analysis of tryptic digests of the purified material. MS can permit identification of which tryptic peptides are phosphorylated and at which residues. This knowledge would inform inhibitor experiments if the phosphorylated residue(s) sit in a well characterised kinase recognition sequences for which an inhibitor is available. More rigorously, site directed mutagenesis of the phosphorylatable residue to alanine and subsequent expression of the mutant Nanog in ES cells would permit investigation of the functional relevance of the particular phosphorylation event.

3.7.3   Nanog partner proteins: a candidate approach.

Oct4 and Stat3 do not interact with Nanog, at least not under the conditions used in the experiments described in this chapter (Figure 3.9). The simplest explanation for this negative data is that Nanog does not interact with either of these molecules. However, co-immunoprecipitation procedures are dependent on the particular lysis and binding conditions used, and it quite possible that by using less stringent lysis procedures, interactions would have been detected, particularly if these were weak or

transient in nature. Recently in contradiction to the data shown here, an MS based screen for Nanog interacting proteins in ES cells identified Oct4 and provided biochemical data to support this interaction (Wang *et al.*, 2006). It is unclear why a Nanog-Oct4 interaction was not also detected in the experiments described here (Figure 3.9). There are a number of technical differences between the experiments described in this chapter and those performed by Wang *et al* (2006) which may explain the difference in the co-immunoprecipitation data. The lysis and purification conditions used vary from 350mM NaCl/0.3% NP-40 (Wang *et al* 2006) to 150mM NaCl/0.5% NP-40 (this chapter). It is possible that the increase in detergent concentration is not compatible with maintenance of an Oct4-Nanog interaction. The experiments to examine a possible Oct4-Nanog interaction performed in this chapter used stable ES cells lines, whereas the presence of Oct4 in Nanog immunoprecipitates detected by Wang *et al* (2006) were achieved via transient transfection of epitope tagged Oct4. Furthermore, the two studies employed different antibodies for both the immunoprecipitation and detection of partner proteins on immunoblots. Also, the experiments in this chapter use cell lines based on E14Tg2a or CGR8 ES cells whereas the experiments of Wang *et al* (2006) use J1 ES cells. Any one, or a combination of these variables, may explain the detection of an Oct4-Nanog interaction by Wang *et al* (2006), yet lack of detection in this thesis.

Oct4 and *Nanog* are both specifically expressed in the pluripotent tissues of the early mouse embryo as well as pluripotent ES cells and are considered markers of pluripotency. However, consequences of their ablation lead to quite different phenotypes both in ES cells and in the mouse embryo. *Oct4* deletion leads to

trophectodermal differentiation of both the ICM and mouse ES cells (Niwa *et al.*, 2000; Nichols *et al.*, 1998). Nanog plays an important role in determination of pluripotent tissues (Mitsui *et al.*, 2003), and the efficiency of self-renewal (Chambers *et al.*, 2003), however Nanog is not absolutely essential for maintenance of pluripotency (Chambers unpublished). These data suggest that Oct4 and Nanog sit at different places in the circuitry responsible for pluripotency. It is therefore possible that Oct4 and Nanog act together on a subset of transcriptional targets involved in initiating a pluripotency program. Indeed recent ChIP based screens show that Oct4 and Nanog co-occupy the promoters of many genes in both mouse and human ES cells (Loh *et al.*, 2006; Boyer *et al.*, 2005). It is possible that Nanog and Oct4 are in the same complex acting at these common target genes, which could explain the interaction seen by Wang *et al* (2006). Alternatively, Oct4 and Nanog could be binding at common targets but in different complexes. However, taking into account the distinct null phenotypes, it is likely that Oct4 and Nanog regulate some genes independently of one another to control distinct cell-fate decisions.

The fact that Stat3 is not found to interact with Nanog in the experiments in this chapter (Figure 3.9) is consistent with the data of Wang *et al* (2006) and the MS based screen described in chapter 5 of this thesis. Considered together, it is likely that Stat3 does not interact with Nanog in mouse ES cells, at least not in a complex that can be purified by standard immunoaffinity purification protocols. It is known that maximal ES cell self-renewal efficiency is achieved upon *Nanog* over-expression and LIF stimulation (Chambers *et al.*, 2003). This is possibly achieved by Nanog and Stat3 acting on either different target genes or different parts of the

promoter of a common target gene(s). Instrumental in Nanog and Stat3 function in ES cells is active Smad1, which has been found to physically interact with Stat3 (Ying *et al.*, 2003) and Nanog (Suzuki *et al.*, 2006b). These two complexes may act in a circuit to maintain maximal self-renewal efficiency with Nanog binding Smad1 to block BMP induced ES cell differentiation (Suzuki *et al.*, 2006b), and active Smad1 interacting with Stat3 to co-regulate LIF and BMP transcriptional targets. This proposed mechanism would explain how elevated Nanog levels coupled with LIF stimulation yields pure ES cell cultures with very few of the normally observed differentiated cells. However, the possibility remains that Nanog, Stat3 and Smad1, are found together in a transient or unstable complex that is not amenable to examination by immunoprecipitation procedures. It may be possible to explore this possibility further using cross-linking agents to stabilise the unstable native complex.

3.7.4 Sall4 interactions in ES cells.

In this chapter the SLQQ motif of Nanog has been mapped as being involved in Sall4 interaction and Oct4 has been preliminarily identified as a Sall4 partner protein (Figure 3.12). This Oct4-Sall4 interaction was also detected in the MS screen for Oct4 partners (Wang *et al.*, 2006). Given that Oct4 is the only homeodomain in the mouse genome other than Nanog to possess such an SLQQ motif in a similar position in the homeodomain, one can hypothesise that the SLQQ also mediates Oct4-Sall4 interactions. Sall4 is known to be a transcriptional regulator of *Oct4* transcription (Zhang *et al.*, 2006). Whether Sall4 acts on target genes as part of an Oct4 or Nanog containing complex has not been demonstrated although data suggesting co-occupancy of many genomic sites by Sall4 and Nanog may make this

a likely possibility (Wu *et al.*, 2006). Likewise, ChIP data suggests Sall4 can be precipitated at the *Nanog* promoter in ES cells and is able to activate the distal *Nanog* enhancer in luciferase assays (Wu *et al.*, 2006). Nanog also acts with Sall4 to activate *Sall4* transcription in ES cells (Wu *et al.*, 2006). Taken together, it appears that Sall4 acts to maintain the expression of *Nanog* and *Oct4* in ES cells. Consistent with this, *Sall4* deletion has been shown to be detrimental to ES cell self-renewal (Elling *et al.*, 2006; Sakaki-Yumoto *et al.*, 2006). In one case ES cells lacking *Sall4* could not be generated by acutely inactivating *Sall4* in ES cells (Elling *et al.*, 2006). A separate study isolated Sall4$^{-/-}$ ES cells at a very low frequency. *Sall4*$^{-/-}$ ES cells proliferated much slower than Sall4 heterozygote ES cells although they appear to retain pluripotentiality as judged by their wide contribution to E7.5 chimaeric embryos. (Sakaki-Yumoto *et al.*, 2006).

The crystal structure of human Oct1 homeodomain in conjunction with the Oct1 DNA binding site was used to further understand the nature of the Nanog-Sall4 interaction mediated by SLQQ motif. The crystal structure of the human Oct1 homeodomain (Remenyi *et al.*, 2003), identifies the position of the TSEE motif (positioned analogously to the SLQQ motif) as being exposed on the protein surface and positioned along the DNA helical axis (Figure 3.13). The evolutionary differences between the TSEE motif of Oct1 and the SLQQ motif of Nanog likely alter both protein interacting capabilities and potential interactions between the SLQQ and the DNA backbone. Notwithstanding the chemical differences between the Oct1 TSEE motif and the Oct4/Nanog SLQQ motif, the spatial information from the Oct1 homeodomain crystal structure predicts that the SLQQ is located in an

exposed position in the nucleoplasm likely to be accessible to interacting partner proteins, such as Sall4 and possibly additional molecules.

At the simplest level it would be predicted that if the (HA)Nanog$_3$SLQQ>SAAQ mutant acted to abolish a Sall4 interaction, ES cell self-renewal would be unaffected by this mutant in the presence of LIF as endogenous Nanog would continue to interact with Sall4. However, the decreased colony size of the SLQQ>SAAQ mutants suggests a different or additional mechanism is at play. It is possible that the (HA)$_3$NanogSLQQ>SAAQ protein is interacting with endogenous Nanog (as this mutant has an intact C-terminal domain) and acting to titrate endogenous Nanog from Sall4 (or other partner protein) containing complexes reducing the functional output of Sall4 (or other partner protein) in the cell. Future experiments, performing the functional assessment of (HA)$_3$NanogSLQQ>SAAQ in *Nanog*$^{-/-}$ ES cells would enable this to be addressed as all the Nanog in the cell would be the SAAQ mutant. In addition, whether the decreased colony size is solely due to abolition of the Sall4 interaction will require identification and mutational analysis of the interacting residues in Sall4. Published data shows that Sall4 does indeed have a role in proliferation of mouse ES cells as *Sall4*$^{-/-}$ ES cells grow slower than wild-type cells (Sakaki-Yumoto *et al.*, 2006). This is intriguing since *Nanog*$^{-/-}$ ES cells also proliferate more slowly than wildtype cells (Chambers unpublished). Therefore the contention that the Nanog-Sall4 complex is a key regulator in mouse ES cells is feasible and that the colonies formed by (HA)$_3$NanogSLQQ>SAAQ over-expression could indeed be due to disruption of this interaction. Again whether the Sall4 interaction with Nanog is mediated directly by the SLQQ motif is unclear. It is

possible that there is a bridging molecule or complex that is SLQQ dependent. *In vitro* experiments using protein synthesised either in *E.coli* or rabbit reticulocyte lysate would allow assessment of whether the interaction is direct.

3.8   Summary.

This chapter has described experiments that begin a biochemical characterisation of Nanog protein in ES cells. Nanog has been shown to be phosphorylated in ES cells and to multimerise through sequences within the tryptophan repeat. Furthermore, the published Nanog-Sall4 interaction has been confirmed and the SLQQ motif within the Nanog homeodomain has been mapped as the key motif in this interaction which has functional significance in ES cells. In addition, preliminary biochemical data shows the identification of an Oct4-Sall4 interaction in ES cells. Negative data from candidate based co-immunoprecipitation experiments have shown that Oct4 and Stat3 do not interact with Nanog under the conditions used, and importantly this instructed an unbiased screening approach to Nanog partner protein identification described in chapter 5 of this thesis.

Chapter 4.

Investigation of the *in vivo* consequences of *Nanog* over-expression during mouse development.

4.1   Introduction

During embryogenesis, Nanog mRNA is expressed from the morula stage and is down-regulated shortly prior to implantation (Chambers *et al.*, 2003). In the early post-implantation embryo (E6.5-E7.5), Nanog mRNA forms a gradient of expression with highest levels at the proximal posterior region of the egg cylinder embryo. Moreover, Nanog mRNA is restricted to the ectoderm and is rapidly eliminated as cells de-laminate and enter the primitive streak (Hart *et al.*, 2004). In this chapter, the consequence of over-expressing *Nanog* on *in vivo* differentiation is addressed. To this end, a mouse ES cell line has been generated in which a *loxP* flanked *Nanog* transgene is expressed from the CAG cassette (Niwa *et al.*, 1991) and in which *Nanog* expression can be monitored through the use of an IRES linked (Mountford *et al.*, 1994) *lacZ* reporter. The generation of chimaeric embryos using this cell line can be visualised by the constitutive expression of a fluorescent protein from a separate additive transgene, thus allowing visualisation of the fate of these *Nanog* over-expressing cells *in vivo*. Exclusion of GFP expressing cells from a particular lineage would suggest that continued *Nanog* expression is incompatible with differentiation into that lineage. This could also be apparent by the lack of β-galactosidase activity; however a β-galactosidase negative phenotype could also result from artifactual silencing of the *Nanog* transgene due to site of integration effects. The presence of a

DsRed2 ORF downstream of the *loxP* flanked *Nanog* transgene allows for control over such artefacts following Cre excision and allows any phenotype to be rigorously assigned to *Nanog* over-expression by *Nanog* removal. If *Nanog* over-expression is directly responsible for a differentiation block, then the restoration of differentiation potential in the Cre-reverted DsRed2$^+$ cells should allow generation of cellular derivatives in the lineage not populated by the *Nanog* over-expressing ES cell line. The information that can be obtained by monitoring β-galactosidase/ DsRed2 expression in the embryo is summarised in Table 4.1. In this way, the effect of elevating *Nanog* expression on epithelialisation of the epiblast and mesoderm formation can be addressed.

Table 4.1- Summary of the information to be gained by tracking *Nanog* expression in ES cell: morula aggregated embryos.

β-GALACTOSIDASE	DSRED2	INFO GAINED
+ve	+ve	No effect of Nanog over-expression
-ve	+ve	Nanog over-expressing cells excluded from lineage 'X'
-ve	-ve	Transgene silenced
+ve	-ve	?

4.2 Generation of *loxP* flanked Nanog expression plasmid.

The construct used in this study consists of the strong CAG cassette (Niwa *et al.*, 1991) driving a *loxP* flanked NanogIRESβ-geo cassette (Figure 4.1). After Cre mediated recombination of the *loxP* flanked cassette, *DsRed2* becomes constitutively expressed. The construct is derived from an analogous plasmid containing a *loxP*
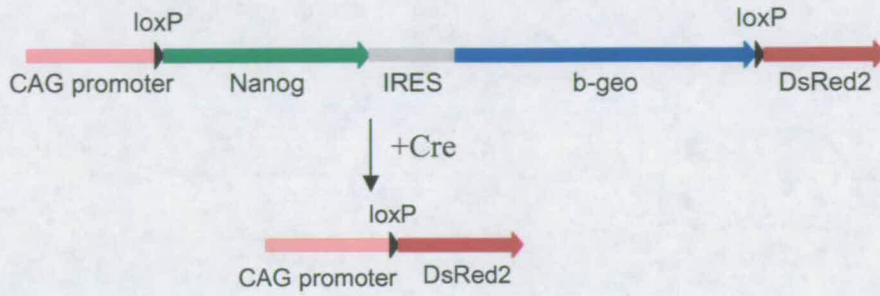
Figure 4.1- Schematic representaion of *loxP* flanked Nanog expression plasmid (IPC 154)

Schematic diagram of IPC 154 and affect of Cre recombinase expression.
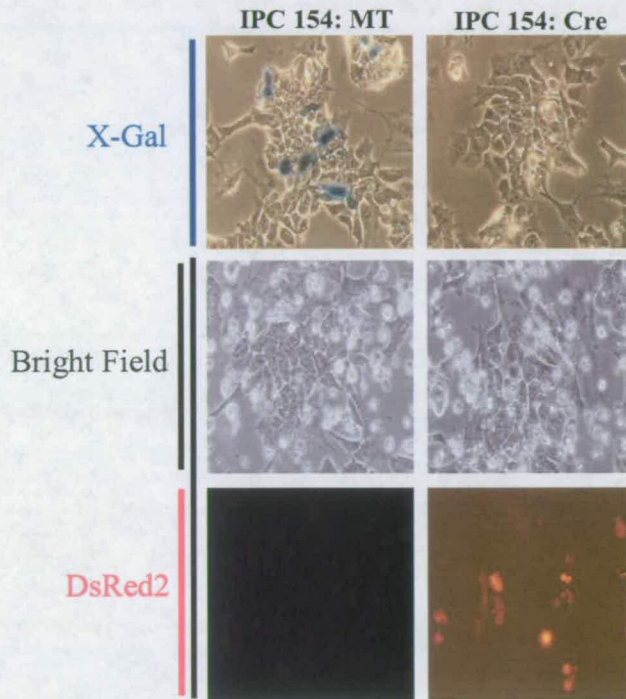
Figure 4.2- Functional test of *loxP* flanked Nanog construct (IPC 154)

Transient co-transfection of IPC 154 with Cre recombinase or empty vector (MT) into LRK-1 ES cells. 24 hours post transfection cells were photographed under fluorescent microscopy prior to staining with X-Gal to visualise the β-galactosidase expression.

flanked NanogIRESpac cassette that expresses GFP upon Cre recombination (Chambers *et al.*, 2003). IPC 154 was constructed in two cloning steps, first to generate IPC 138, and subsequently IPC 154. Details of the cloning strategy and maps are presented in the plasmid appendix.

4.3   Functional assessment of the *loxP* flanked Nanog expression plasmid.

Before generating stable ES cell clones using plasmid IPC 154, a functional test was performed to ensure β-galactosidase expression could be visualised via X-Gal staining and that co-expression with Cre recombinase results in correct excision of the NanogIRESβ-geo cassette to give visible *DsRed2* expression. Plasmid DNA was co-transfected with either a Cre recombinase expression plasmid (AGS 844) or a corresponding empty vector (AGS 564) into super-transfectable ES cells (Chambers *et al.*, 2003). Twenty-four hours after transfection, the cells were examined under fluorescent microscopy, photographed, fixed and then stained with X-Gal (Figure 4.2). IPC154 transfectants expressed *β-galactosidase* in the absence but not the presence of co-transfection with a Cre plasmid. The reciprocal pattern was observed for expression of DsRed2. These data shows that the construct functions as designed and therefore stable transfectants were generated.

4.4   Generation of *taugfp* cells stably expressing IPC 154.

To enable visualisation of cells carrying the *Nanog* transgene in chimaeric mice, a cell line constitutively expressing an easily detectable marker protein is required. One such line is *taugfp* ES cells in which constitutive expression is directed by the CAG cassette (Ying *et al.*, 2003). These cells express a form of GFP that is localised to microtubules

due to fusion of GFP with tau protein (Pratt *et al.*, 2000). *Taugfp* ES cells were stably transfected with IPC 154 plasmid. Selection was then applied to the cells at 600μg/ml G418 in the presence or absence of LIF. To attempt to select for integration sites directing high transgene expression a G418 concentration three times higher than usual was used. The number of undifferentiated colonies obtained was dependent on the addition of exogenous LIF. In the presence of LIF, colonies were obtained at a frequency of $>2x10^{-4}$ whereas in the absence of LIF the frequency was $\sim2x10^{-5}$. Many colonies obtained in the absence of LIF displayed some differentiated cells at the periphery. Fourteen days after transfection, ~70 colonies selected in the absence of LIF were picked and expanded. Clonal expansion of these colonies was performed in absence of LIF in an effort to select only transfectants robustly expressing *Nanog* above the threshold level required for cytokine independent self-renewal.

4.5 LIF independence of IPC 154 stable transfectants.

Several independent clones (TFOG clones; *taugfp*: **f**loxed **O**RF β-**g**eo) were cultured at clonal density (120cells/cm^2) for two passages in the presence of hLIF-05 (Vernallis *et al.*, 1997), a mutant form of human LIF which antagonises LIF function. Three out of the five clones tested maintained an undifferentiated morphology during this rigorous LIF independence assay and were used in further experiments. To assess whether transgenic *Nanog* was expressed uniformly in all cells of a given clone, cultures were stained with X-Gal to visualise β-galactosidase activity (Figure 4.3a). Only clone BB8 shows uniform X-Gal staining and this clone is also the most intensely blue stained. Clones BF3 and X exhibit mosaic transgene expression and

(a)

PREhLIF-05   POSThLIF-05

TFOGBF3

TFOGX

TFOGBB8

taugfp
(parental)

(b)

BB8c2

+LIF   -LIF

DsRed2

Bright Field
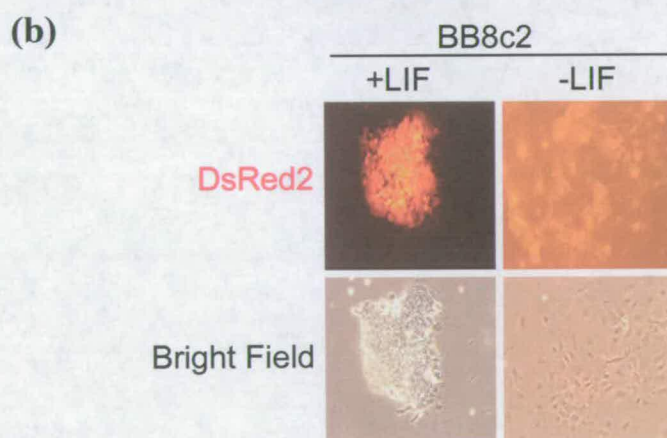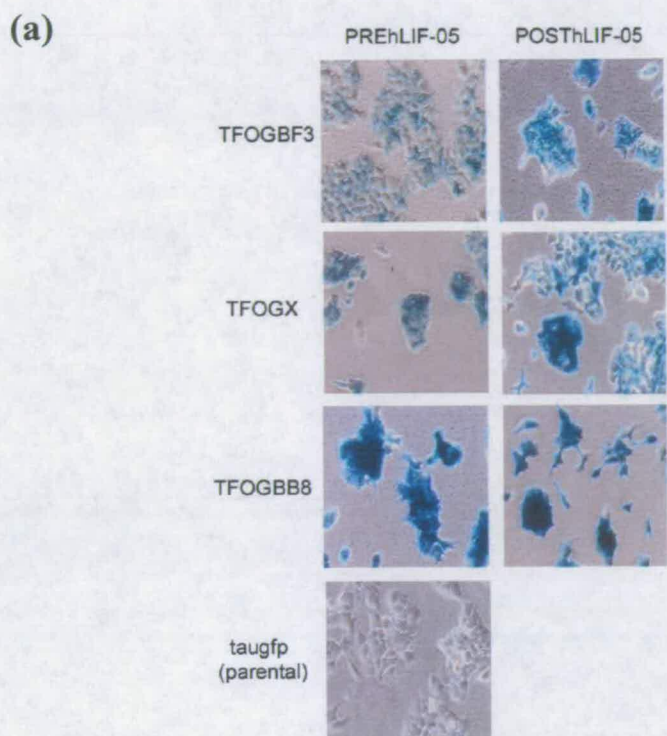
Figure 4.3- X-Gal staining and Cre reversion of TFOG clones

(a) Three independent TFOG clones stained with X-Gal to visualise transgene
expression both before and after culture with hLIF-05.
(b) TFOGBB8c2, a Cre reverted derivative of TFOGBB8 which expresses visible
levels DsRed2, cultured for 6 days at clonal density in the presence and absence
of LIF.

the expression level is lower than clone BB8 (Figure 4.3a). Clone BB8 (pre-hLIF-05) was transiently transfected with Cre recombinase and 24h after transfection, cells were replated at clonal density in the absence of G418. Correctly reverted cell lines were identified by revertion to G418 sensitivity and expression of visible levels of DsRed2 protein (BB8c clones). DsRed2 positive/ G418 sensitive revertant clones were selected and expanded (Figure 4.3b). Further LIF independence assays were performed by plating *taugfp* parental cells, BB8, and Cre reverted BB8 derivatives clones (BB8c) in the presence and absence of LIF at a clonal density for 6 days prior to alkaline phosphatase staining (Figure 4.4). As expected, the *taugfp* cells formed ES cell colonies in a LIF dependent manner, whereas EF4 cells, an independent *Nanog* over-expressing ES cell line (Chambers *et al.*, 2003), could form uniformly undifferentiated alkaline positive colonies in the complete absence of LIF. BB8 cells have a reduced LIF dependence but are not as robustly LIF independent as EF4 cells. When cultured at clonal density in 10U/ml LIF for 6 days, 42% of BB8 (pre hLIF-05) colonies are undifferentiated compared to only 2% of *taugfp* colonies. It has previously been shown that the degree of elevated Nanog protein in EF4 cells is 5-6 times endogenous levels (Yates and Chambers, 2005). Immunoblotting of *taugfp* parental cell and TFOG BB8 cell lysates reveals that Nanog is expressed at 2-3 times parental cell levels (Figure 4.5). Unexpectedly, BB8 cells have differing LIF dependencies, dependent on whether they had been previously passaged with hLIF-05. Importantly however, Cre recombinase treatment of the BB8 cells (clones BB8c2 and BB8c11) reverted them to a LIF dependency comparable to wildtype cells.

(a)

Undifferentiated   Mixed   Differentiated
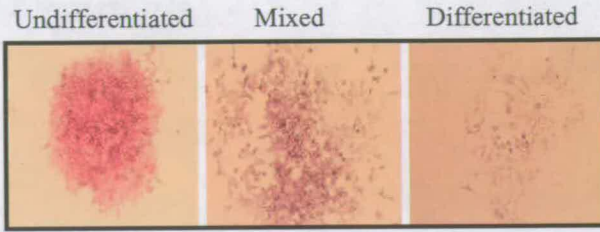
(b)
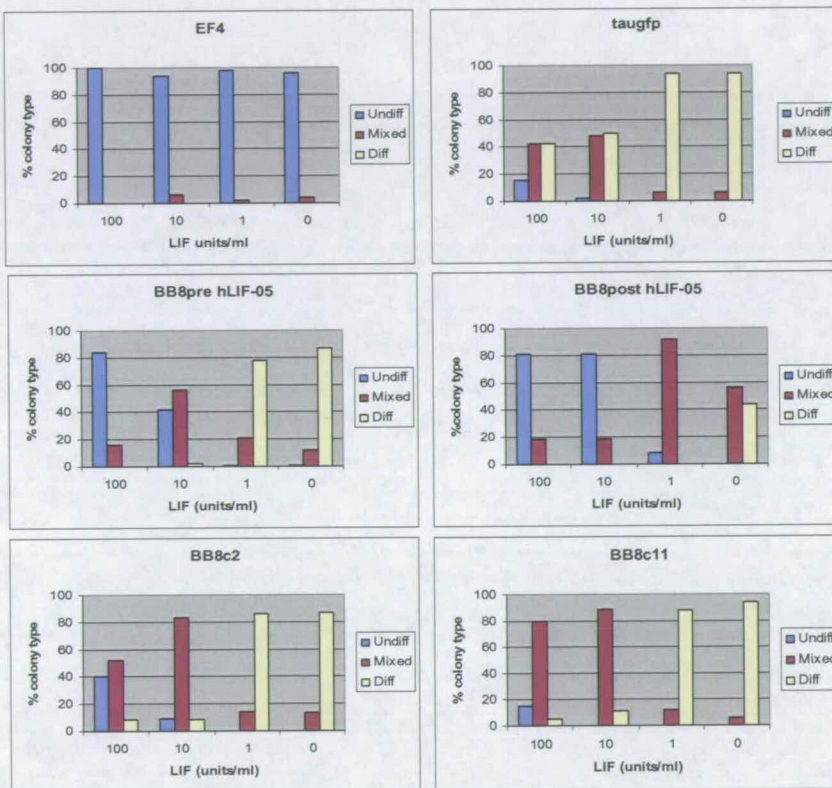
Figure 4.4- LIF dependency of TFOG BB8 clones and Cre derivatives.

(a)   Examples of undifferentiated, mixed, and differentiated colonies stained for alkaline phosphatase activity.

(b)   The indicated cell lines were plated at 64 cells/cm^2 (clonal density) and cultured in varying LIF concentrations for 6 days. Cells were then stained for alkaline phosphatase (AP) activity and the stained colonies scored as AP positive uniformly undifferentiated (Undiff) colonies, mixed colonies containing AP positive and negative cells, or AP negative differentiated (Diff) colonies. This data is from a single experiment.
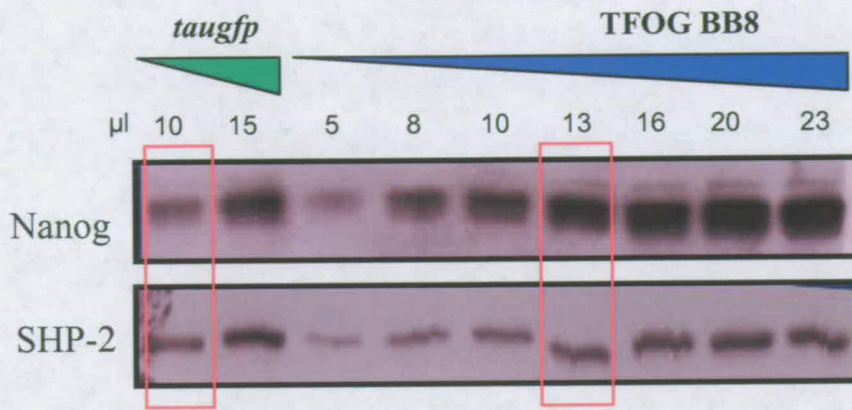
Figure 4.5- Quantification of Nanog protein levels in BB8 cells.

taugfp parental cell lysate and increasing amounts of TFOG BB8 cell lysate (pre hLIF05) were separated by SDS-PAGE and immunoblotted. Blots were probed with anti-Nanog and anti-SHP-2 antibodies. By comparing the levels of SHP-2 and cross-referencing to the Nanog levels, an estimate of Nanog over-expression level can be gained. Red boxes show equivalently loaded lanes. BB8 cells (pre hLIF-05) express 2-3 times the endogenous (*taugfp*) Nanog levels.

TFOG BB8 pre hLIF-05   TFOG BB8 post hLIF-05
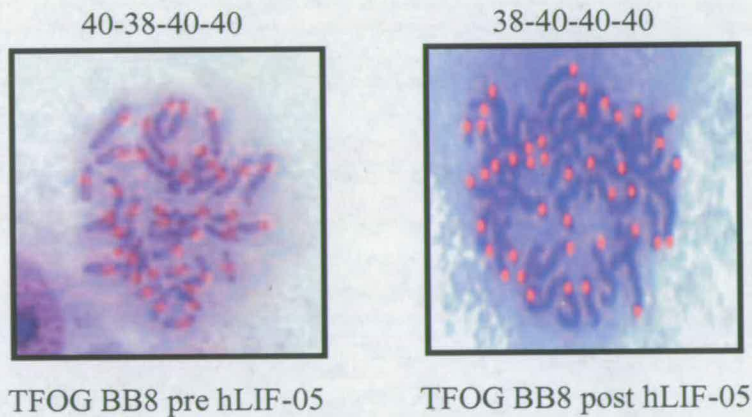
Figure 4.6- Metaphase spreads of BB8 cells
Metaphase spreads of BB8 cells. The chromosomes are stained with Giemsa stain. The red dots mark individual chromosomes.

4.6   BB8 cells are not grossly karyotypically abnormal.

Metaphase spreads of BB8 cells (both pre and post hLIF-05) were performed and the
chromosome number counted (Figure 4.6) These results suggest that the altered LIF
dependence of BB8 cells before and after expansion in hLIF-05 is not due to gross
aneuploidy.

4.7   BB8 cells contribute widely to mouse embryos.

To assess the ability of BB8 cells to form mesoderm in the developing mouse embryo,
BB8 cells were aggregated with E2.5 morulae and aggregated embryos were transferred
to pseudopregnant CBA/vas females as described in Methods (section 2.1.8.1). The
developing embryos were dissected at three time points, E7.5, E8.5, and E9.5, and
chimaerism assessed by visual inspection of GFP expression (Table 4.2). The BB8
chimaeric embryos were stained for β-galactosidase activity to report the location of
Nanog transgene expression and example embryos at E7.5 (Figure 4.7) and E9.5 (Figure
4.8) are shown. No overt difference between the parental cell chimaeras and the BB8
cell chimaeras was observed. One difference that can be seen at E7.5 is that there are no
X-Gal positive cells in the extra-embryonic mesoderm in the BB8 chimaeras whereas
the *taugfp* parental cells readily populate this lineage (Figure 4.7). However, the sample
size for both *taugfp* and BB8 chimaeric embryos is small (n=3) with only one of the
taugfp embryos showing good contribution to the extra-embryonic mesoderm. All the
BB8 chimaeric embryos appear show a lack of GFP$^+$ X-Gal$^+$ cells in the extra-
embryonic mesoderm.
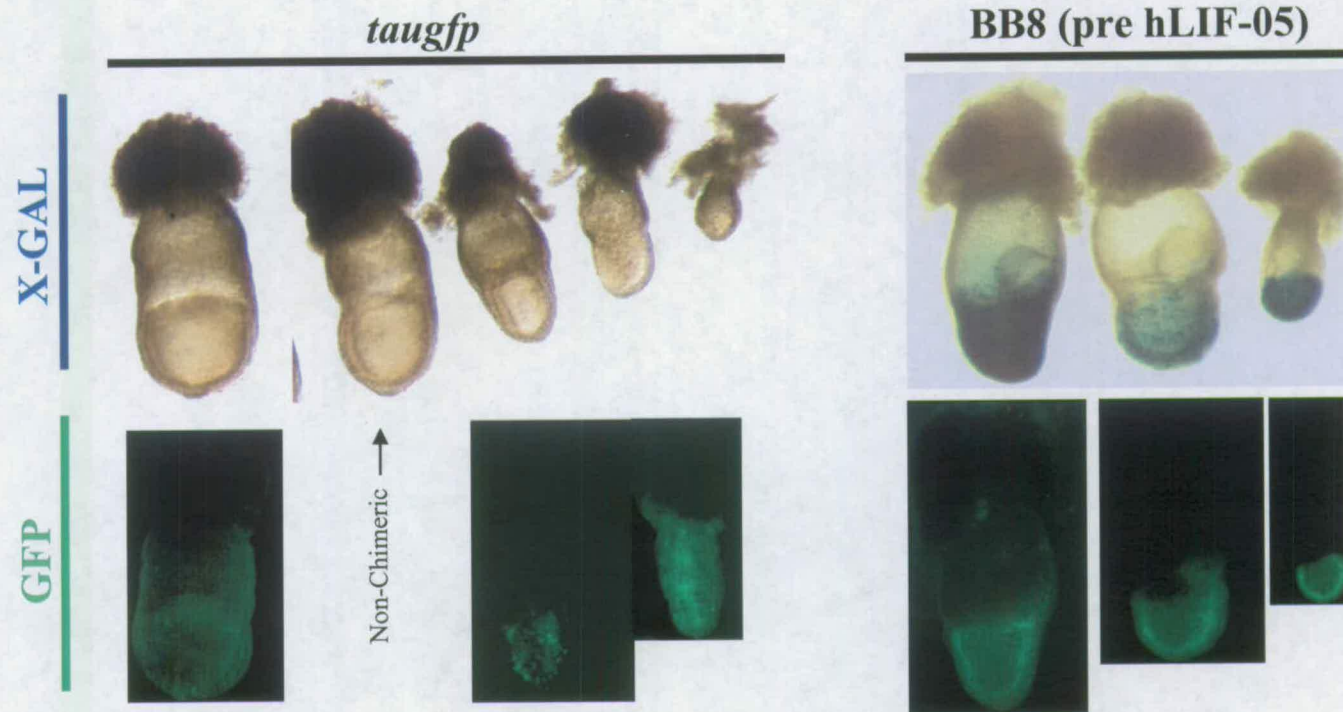
Figure 4.7- Contribution of *taugfp* and BB8 cells to E7.5 mouse embryos.

Taugfp (parental wildtype) ES cells or BB8 (*Nanog* over-expressing) ES cells were aggregated with E2.5 morulae. The next day embryos were transferred to 2.5 days post coitum CBA/ BL/6 (F1 hybrid) pseudopregnant females. At E7.5, embryos were dissected and chimaerism assessed via fluorescence microscopy to visualise GFP expression, prior to staining with X-Gal to report transgenic *Nanog* expression.

| *taugfp* | BB8 (pre hLIF-05) |

X-GAL

Blood Island

GFP
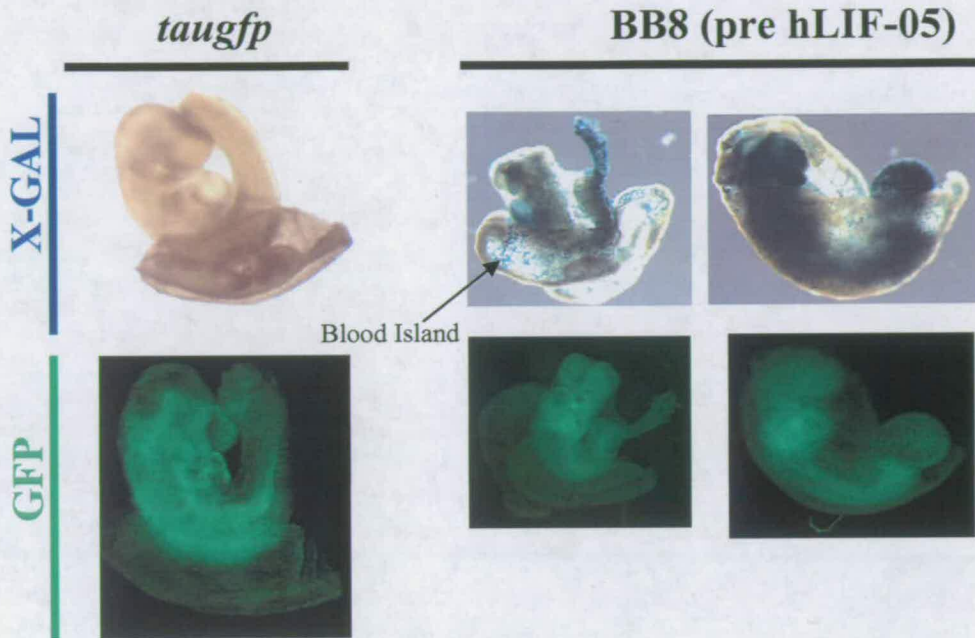
Figure 4.8- Contribution of *taugfp* and BB8 cells to E9.5 mouse embryos.

Taugfp or BB8 ES cells were aggregated with morulae and transferred to pseudopregnant females. At E9.5, embryos were dissected and chimaerism (GFP) visualised by fluorescence microscopy prior to staining with X-Gal to report transgenic *Nanog* expression.

Table 4.2- Table showing number of chimaeric embryos at E7.5, E8.5, and E9.5.

	taugfp	TFOGBB8
	E7.5	
Chimaeric	3	3
Non-chimaeric	2 (1 was "runty")	0
Resorbed	1	4
TOTAL	6	7
	E8.5	
Chimaeric	0	3
Non-chimaeric	0	1
Resorbed	0	1
TOTAL	0	5
	E9.5	
Chimaeric	1	5
Non-chimaeric	3+ 1 empty yolk sac	1
Resorbed	2	10
TOTAL	7	16

To determine whether *Nanog* over-expressing cells were capable of contributing to mesoderm, X-Gal stained embryos were embedded in paraffin wax and 6µm transverse sections were prepared. BB8 *Nanog* over-expressing cells do populate the mesoderm, indicating that forced *Nanog* expression does not prevent movement of cells through the primitive streak (Figure 4.9).
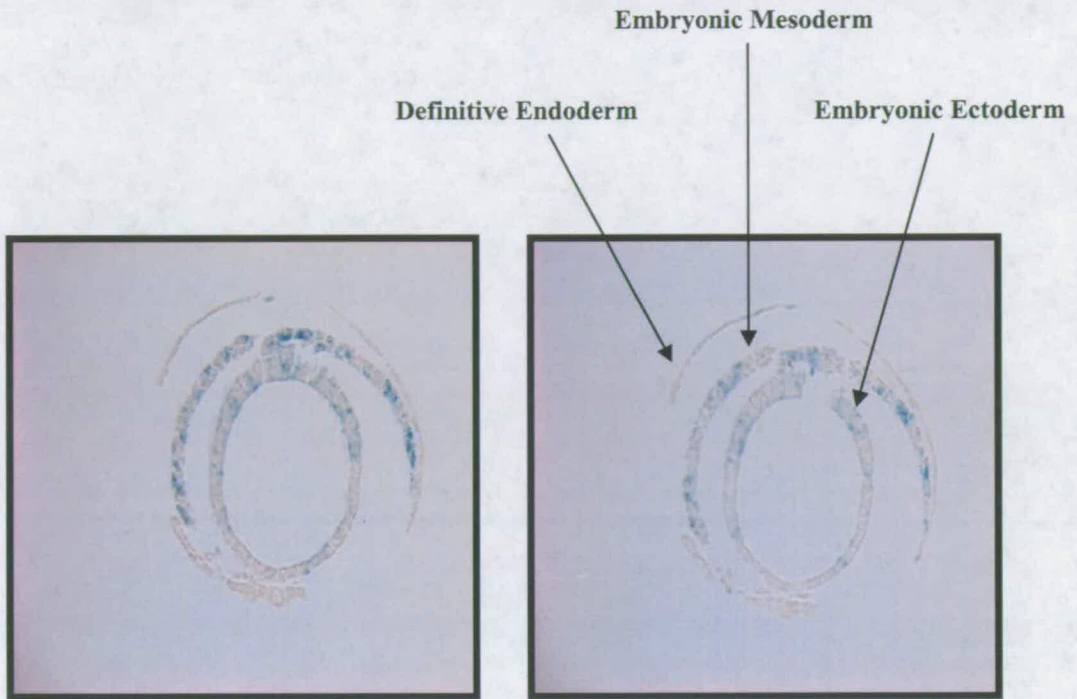
Figure 4.9– Transverse sections of E7.5 BB8 aggregation embryos stained with X-Gal.

The embryos pictured in Figure 4.7 were embedded in paraffin, transversely sectioned (6μm sections), and photographed. Two representative sections clearly showing BB8 cell contribution to the mesoderm are shown.

4.8 Discussion

4.8.1 IPC 154 construct.

The design of the revertible *Nanog* over-expression construct has a number of key attributes. The strong constitutive CAG cassette consists of the cytomegalovirus immediate early enhancer, the chicken β-actin promoter and a chimaeric β-globin intron. This element directs high level transgene expression in mammalian cells (Niwa *et al.*, 1991). Flanking the NanogIRESβ-geo cassette with *loxP* sites allows any observed phenotype to be unambiguously assigned to *Nanog* over-expression if the Cre reverted *DsRed2* expressing cells revert to the wild type phenotype. The stable cell lines generated harbour randomly integrated transgene DNA. Although some loci are constitutively expressed such as *Rosa26* (Zambrowicz *et al.*, 1997), it is known that particular loci can be silenced in a particular somatic lineage or chromosomal location, for example in telomeric regions (Pedram *et al.*, 2006). If *Nanog* over-expressing BB8 cells do not contribute to a particular lineage *in vivo*, the *DsRed2* expressing Cre revertants provide a useful reagent to show that the transgene integration site is indeed transcriptionally active, and therefore any lineage blocking phenotype is due directly to elevated Nanog levels.

4.8.2 LIF dependency of *Nanog* over-expressing BB8 cells.

BB8 cells have a decreased dependency on the normally obligatory LIF signalling pathway, however, they are not robustly LIF independent (Figure 4.4). BB8 colonies cultured at clonal density are not all uniformly alkaline phosphatase positive and morphologically completely undifferentiated. EF4 cells, an independent *Nanog* over-

expressing ES cell line (Chambers *et al.*, 2003) when treated in parallel to the BB8 cells maintain the undifferentiated state in the complete absence of LIF (Figure 4.4). The *Nanog* transgene is expressed uniformly (non-mosaically) in all cells in the cultures (Figure 4.3). Comparison of EF4 and EF4Cre revertant cell lysates reveals Nanog protein is 5-6 times wild-type in the LIF independent EF4 cells (Yates and Chambers, 2005). Nanog protein levels in BB8 cells are ~2-3 the parental *taugfp* levels (Figure 4.5), which, although sufficient for reduced LIF dependency, is not sufficient for complete release from cytokine dependence (Figure 4.4). Crossing a threshold of expression is also important for other homeodomain proteins to effect particular self-renewal or differentiation functions in a stem cell system. For example, in the haematopoetic stem cell (HSC) system, increased *HoxB4* expression has been shown to increase the expansion of HSC's *in vitro* (Antonchuk *et al.*, 2002) and *in vivo* (Sauvageau *et al.*, 1995; Thorsteinsdottir *et al.*, 1999). Furthermore, differing doses of *HoxB4* expression have been shown to effect differentiation into different haematopoetic lineages, with the highest achievable levels of over-expression favouring self-renewal over differentiation (reviewed by Klump *et al.*, 2005). It is not however a general rule that homeodomain proteins important in stem cell regulation increase self-renewal efficiency when over-expressed. This is exemplified by Oct4 which is a master regulator of mouse ES cell self-renewal (Nichols *et al.*, 1998) but when the thrershold of 150% of wildtype level is surpassed, differentiation ensues (Niwa *et al.*, 2000). The immunoblot (Figure 4.5) and clonal expansion assays (Figure 4.4) show the 5-6 fold increase in Nanog protein levels required for true LIF independent self-renewal are not reached in the BB8 cells. This raises the question of how self-renewing colonies were obtained during selection in the absence of LIF. These cultures contain a mixed population of cells which themselves

produce LIF, meaning diffusible LIF is present in the ES cell media. This may reduce the transgenic *Nanog* levels required to generate undifferentiated colonies. Even with regular media changes, (every 2 days during selection) LIF may be present at effective concentrations. This is particularly likely to be the case as the primary colonies will be growing on a matrix associated form of LIF deposited by the initially plated untransfected cells (Rathjen *et al.*, 1990). If this experiment was to be repeated, one possibility may be to include hLIF-05, the LIF antagonist, in the culture media during selection, to increase the probability of only high *Nanog* expressing undifferentiated colonies being obtained in these stringent culture conditions. To date, all published *Nanog* over-expressing, LIF independent cell lines were generated using the strong CAG promoter (Niwa *et al.*, 1991) combined with puromycin drug selection, which selects for high level transgene expression (Chambers *et al.*, 2003; Loh *et al.*, 2006; Mitsui *et al.*, 2003). To illustrate this point, immunoblotting of the *(HA)₃Nanog: (Flag)₃Oct4* cell line (Figure 3.10) reveals that the (HA)₃Nanog, which was selected in G418, constitutes only ~one third of the total Nanog protein in the cell. Furthermore, LIF independency assays reveal that these cell lines are totally LIF dependent (data not shown). It can therefore be appreciated that the choice of drug selection used in this study may not be ideal to generate LIF independent ES cell clones.

One unexpected set of data is that the LIF dependency of the BB8 cells differs dependent on whether they have been cultured in the presence of hLIF-05 (Figure 4.4). BB8 cells post hLIF-05 treatment are less dependent on LIF, with 80% of colonies adopting a uniformly undifferentiated morphology at 10U/ml LIF compared with only

40% of clone BB8 'pre-hLIF-05' colonies. It is possible that a sub-population of higher-expressing cells were selected during culture with hLIF-05, or that a karyotypic change occurred which provided a self-renewal advantage. To assess the possibility of aneuploidy in BB8 cells, metaphase spreads and chromosome counts were performed (Figure 4.6), and it was found that BB8 cells both 'pre' and 'post' exposure to hLIF-05 were karyotypically normal as judged by chromosome counting. It is possible however that a more subtle pro-self-renewal karyotypic alteration is present in the post hLIF-05 BB8 cells that will only be revealed by detailed karyotypic analysis using a high resolution technique such as comparative genomic hybridisation (CGH). In this regard, it is worth noting that some clones derived from selections in the absence of LIF following transfection with *Nanog* transgenes, have apparently accumulated at least three copies of the endogenous *Nanog* gene (Ian Chambers, personal communication).

If these experiments were to be repeated, two alternative approaches could be taken to elevate transgene expression levels, and these are outlined here and in Figure 4.10.

Approach 1: One reason that neo (β-geo) cassette was employed is that it allows expression of the transgene to be visualised, and a second is that the parental *taugfp* cells already have puromycin resistance (Ying *et al.*, 2003). In future experiments, higher level transgene expression may be achieved using a parental GFP[+]: hygromycin[R] ES cell line and use a construct analogous to IPC 154 with a puromycin[R]-fluorescent protein fusion (e.g. CFP- cyan fluorescent protein- puromycin-*N*-acetyl transferase fusion) replacing the β-geo. This would permit high level transgene expression, whilst retaining the key attributes of the construct, namely, a visible report on *Nanog*

expression and Cre-revertabiltiy. However, this may be dependent on the puromycin[R]-fluorescent protein fusion being as active as puromycin[R] (puromycin acetyl transferase).

Approach 2: A second approach, to increase transgenic *Nanog* expression is to include a pre-selection cassette similar to that described in chapter 5 of this thesis for construct IPC 206. An *frt* flanked DsRed2IRESBsdr cassette can be introduced upstream of the 5' *loxP* site in IPC 154. Initial selection of transfectants in blasticidin would be followed by single cell fluorescent activated cell sorting (FACS) of the highest *DsRed2* expressing cells followed by clonal expansion. *Flp* recombinase expression can then be used to excise this 'pre-selection' cassette and induce NanogIRESβ-geo expression. The resultant clones would be analogous to BB8 cells yet may express higher transgene levels. An additional benefit offered by this type of plasmid is that the expression of the fluorescent protein could be monitored in the absence of drug selection to analyse the stability of the transgene integration site. Monitoring fluorescent protein stability during *in vitro* differentiation would help in choosing clones likely to maintain transgene expression in chimaeric embryos. A final advantage of this latter approach is that it avoids direct selection for Nanog activity, which as noted above, can result in accumulation of additional copies of the *Nanog* gene.

Approach 1
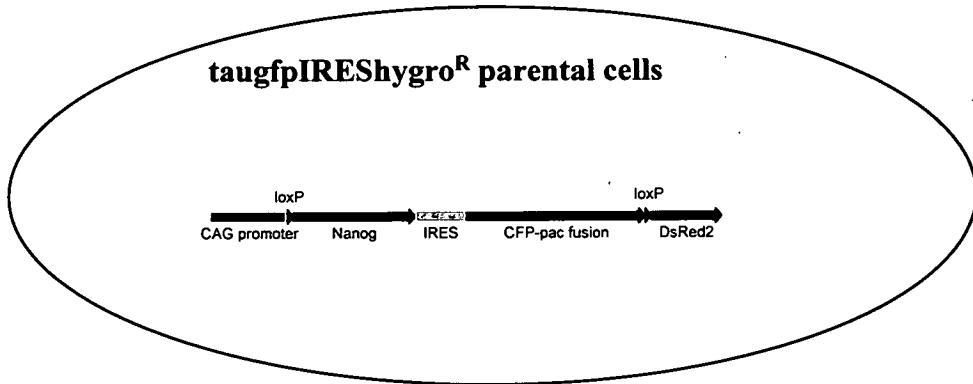
Approach 2
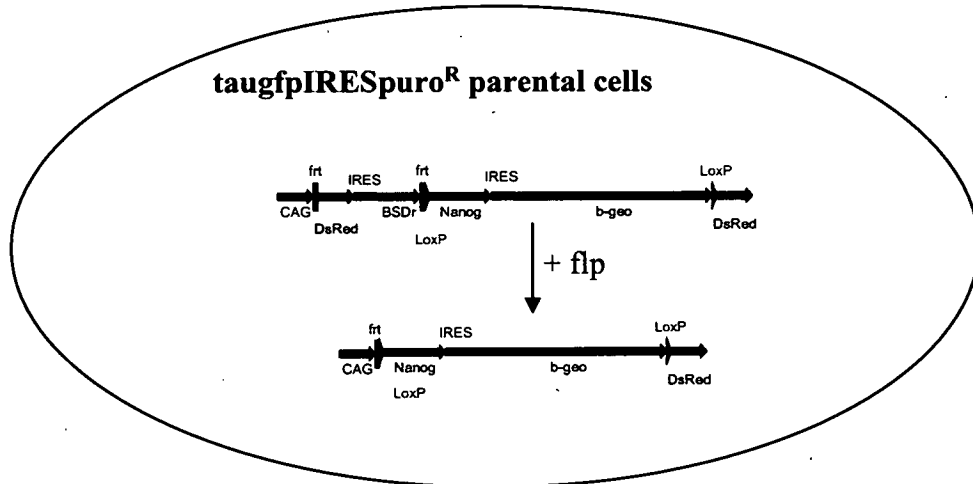
Figure 4.10- Schematic depiction of 2 alternative approaches to generate high level *Nanog* transgene expression.

For details of both approaches see section 4.8.2

128

4.8.3 BB8 cells contribution to mouse embryos.

ES cells are able to contribute widely to all tissues of the developing mouse embryo (Beddington and Robertson, 1989). *In vitro*, elevated *Nanog* expression causes ES cells to be refractory to differentiation even in the presence of potent pro-differentiation stimuli (Chambers *et al.*, 2003). As Nanog mRNA can be detected at E6.5 and E7.5 in the posterior region of the egg cylinder embryo yet not in the primitive streak (Morkel *et al.*, 2003; Hart *et al.*, 2004), it was hypothesised that *Nanog* over-expressing cells in the epiblast may be unable to form mesoderm during gastrulation. A further possibility, based on the decrease in Nanog mRNA expression immediately prior to implantation was that *Nanog* down-regulation might be required for the epithelialisation of the epiblast. The experiments carried out in this chapter aimed to assess whether elevated Nanog levels in the epiblast led to altered developmental potency and, in particular, whether *Nanog* over-expression precluded mesoderm formation. The embryos generated by morula aggregation with BB8 cells (pre-hLIF-05) do not appear overtly different to the *taugfp* parental cell control embryos (Figure 4.7+ 4.8). The visualisation of β-galactosidase activity in the mesoderm (Figure 4.9) shows that 2-3 times endogenous levels of Nanog does not block mesoderm formation. Furthermore, this data shows that *Nanog* down-regulation is not required for formation of the mesoderm. From analysis of *Nanog*$^{-/-}$ chimaeric embryos, it can be seen that *Nanog*$^{-/-}$ cells are able to generate mesoderm (Chambers unpublished) however this ability could be due to wildtype cells rescuing the mesoderm forming ability of *Nanog*$^{-/-}$ cells through release of soluble factors. Little is known about the regulation of *Nanog in vivo*, although both *Wnt3*$^{-/-}$ and *β-catenin*$^{-/-}$ embryos show that Nanog mRNA is abolished in the proximal

posterior region of the embryonic ectoderm at E6.5 (Morkel *et al.*, 2003). Tcf3 is a DNA binding protein effector of Wnt signalling which acts as an activator upon β-catenin stabilisation, but as a repressor in the absence of stabilized β-catenin (Pereira *et al.*, 2006). Recently, Tcf3 has been shown to repress *Nanog* expression in mouse ES cells and this activity is dependent on DNA binding (Pereira *et al.*, 2006). Interestingly, the level of elevation of Nanog protein in Tcf3[-/-] ES cells is around two fold. *Tcf3*[-/-] embryos exhibit defects during gastrulation and anterior-posterior axis formation (Merrill *et al.*, 2004), and it is possible that this is in part due to an inability to repress *Nanog*. One can hypothesise that the abolished Nanog mRNA in *Wnt3*[-/-] and *β-catenin*[-/-] embryos could be due to the repressor activity of Tcf3 in this context. However, given the data in this chapter, additional or alternative mechanism(s) must be functioning to achieve the *Tcf3*[-/-] phenotype, as maintained *Nanog* expression is compatible with mesoderm formation (Figure 4.9). *Brachyury* (*T*), an early marker of mesoderm expressed in the primitive streak, is also a direct transcriptional target of Wnt/ β-catenin pathway (Arnold *et al.*, 2000; Galceran *et al.*, 2001; Yamaguchi *et al.*, 1999). Brachyury has been shown to bind the *Nanog* promoter in ChIP assays in ES cells (Suzuki *et al.*, 2006b), and Nanog has been proposed to protect against precocious mesoderm differentiation *in vitro* (Suzuki *et al.*, 2006a). However, the consensus T-box recognition DNA sequence identified in the *Nanog* promoter may also be bound by other T-box containing proteins such as Eomesodermin (Conlon *et al.*, 2001). It may be that *Brachyury* and *Nanog* are both responsive to Wnt signalling and exist in a feedback loop in the proximal posterior region of the egg cylinder embryo and together control mesodermal differentiation in the primitive streak. The level of Nanog in the BB8 cells however, may not be sufficient to tip the balance in favour retaining epiblast identity. In

this respect, it is possible that there is an upper threshold expression level that needs to be surpassed in order for Nanog to elicit a differentiation blocking phenotype *in vivo*, similar to that observed *in vitro*. To address this question, a further batch of ES cell clones for embryo aggregation could be generated which express *Nanog* at higher levels to give robust LIF independence using either of the approaches described in Figure 4.10. It is also possible that the Nanog protein is non-functional in the BB8 derived cells that ingress through the primitive streak, as the sub-cellular localisation of Nanog may change from nuclear to cytoplasmic. Indeed, there are precedents for homeodomain proteins altering sub-cellular localisation dependent on the environmental signals perceived. For example, HoxA9 is translocated to the nuclear compartment in primitive haematopoetic cells upon thrombopoietin (TPO) stimulation, yet is detected in the cytoplasm upon TPO starvation (Kirito *et al.*, 2004). Whether an analogous mechanism is occurring with Nanog in BB8 derived cells could be addressed by double antibody staining BB8 chimaeric embryo transverse sections at E7.5 with anti-Nanog and anti-GFP antibodies, and examining the localisation of Nanog protein in cells in the embryonic ectoderm compared to the mesoderm.

Although the BB8 cell progeny are clearly able to generate embryonic mesoderm (Figure 4.9), it was noted that at E7.5 no extra-embryonic mesoderm had been formed (Figure 4.7). The *taugfp* parental cells can clearly generate extra-embryonic mesoderm. However, the sample sizes are too small for conclusions to be drawn and due to time constraints the possible extra-embryonic mesoderm phenotype was not further explored. The fact that X-Gal positive blood islands in the yolk sac at E9.5 are visible (Figure 4.8), shows that extra-embryonic mesoderm tissue can ultimately be generated by BB8

cellular progeny, although there may be a delay in formation of this lineage, possibly due to elevated *Nanog* expression. To further address the possibility that increased levels of Nanog cause a delay in extra-embryonic mesoderm formation the DsRed[+] Cre reverted cells (Figure 4.3b) could be employed. Contribution of DsRed2[+] cells to the extraembryonic mesoderm at E7.5 would suggest that elevation of Nanog levels delays this differentiatiation event.

Analysis of maintained expression of another ES cell master regulator of pluripotency, *Oct4,* has been performed (Ramos-Mejia *et al.,* 2005). *Oct4* is expressed in the unfertilised egg, cleavage stage mouse embryos (Palmieri *et al.,* 1994; Scholer *et al.,* 1990a), the epiblast, and is down regulated at E7.5, with the only remaining Oct4[+] cells later in development being primordial germ cells. Transgenic animals engineered to express *Oct4* constitutively throughout embryogenesis do not have an early embryonic phenotype. Rather, mid-hindbrain patterning is altered at E8.0 and forebrain development is compromised at E9.5 (Ramos-Mejia *et al.,* 2005). The lack of early phenotype may be a function of the low expression level of the transgenic *Oct4* (about 50% that of ES cells). Alternatively, it may be that forced *Oct4* expression has no effect on early post-implantation embryos. However this seems unlikely given the differentiation to mesoderm and endoderm that occurs upon modest levels of *Oct4* over-expression in ES cells (Niwa *et al.,* 2000)

4.9   Summary

This chapter has described the generation of a Cre-revertible *Nanog* over-expressing construct (IPC 154), functional testing of the construct, and generation of ES cell lines expressing this construct. The ES cells have been characterised in terms of both LIF dependency *in vitro*, and their developmental potency *in vivo*. No overt differences were observed between *taugfp* parental cell derived embryos and the BB8 cell derived embryos. Alternative strategies have been suggested to further examine this question by increasing *Nanog* expression levels using additive transgenes.

Chapter 5.

A biotin tagging strategy to identify Nanog interacting proteins.

5.1   Introduction

As discussed in section 1.4, a wide range of methodologies are available to examine protein-protein interactions. These technologies generally include purifying the protein of interest using either antibodies to endogenous proteins, or epitope tags with particular binding properties (Terpe, 2003), followed by identification of co-purified partner proteins by mass-spectrometry (Rappsilber and Mann, 2002; Yates, 2000). Alternative methods rely on a genetic system for partner protein identification such as the yeast-2-hybrid (reviewed by Luban and Goff, 1995; Fields and Song, 1989). To identify proteins interacting with Nanog in ES cells, an unbiased proteomic screen for Nanog interacting proteins was performed. The advantages and disadvantages of each potential approach were considered (outlined in section 1.4) before selecting a biotin tagging strategy. This approach was chosen as the interaction between streptavidin and biotin is the strongest non-covalent bond found in nature (K_d~10^{-15}M). Moreover, this approach circumvents the requirement for antibodies which are often insufficient for purification from complex protein mixtures. In addition, the purification can be performed in a simple single step procedure with no requirement for intermediate steps or enzymatic cleavage of the epitope tag. Importantly, the tagged Nanog can be purified under native conditions from ES cells, a context in which Nanog partner proteins are most likely to

be present. Finally, Nanog (and partner proteins) will be appropriately post-translationally modified in ES cells which could be critical for detection of some protein-protein interactions.

A paradigm for the use of this system is provided in a haematopoetic cell culture system in which biotin tagging has been successfully employed to purify a transcription factor, Gata1, and the associated interacting proteins in a single step (de Boer *et al.*, 2003). The identification of partner proteins then instructed further functional experiments to analyse the mechanism of action of Gata1 (Rodriguez *et al.*, 2005). The use of the BIO tag has not been restricted to one study but has been used successfully to purify Ldb1 complexes from MEL cells (Meier *et al.*, 2006) and BIO tagged Oct6 can be purified from mouse ES cells (Driegen *et al.*, 2005).

The experiments described here include the generation of both an inducible *BIO Nanog* expression plasmid and a mouse ES cell line co-expressing *BirA* biotin ligase and the *BIO Nanog* transgene, functional characterisation of the cell line, gel filtration of Nanog containing complexes, pilot experiments to test the streptavidin purification protocol, large scale purification of Nanog complexes for mass spectrometry analysis, and validatory co-immunoprecipitation of three Nanog partner proteins.

5.2   Generation of puromycin sensitive *BirA* ES cells.

A mouse ES cell line expressing the *E.Coli BirA* biotin ligase was kindly provided by D. Meijer (Erasmus MC, Rotterdam; (Driegen *et al.*, 2005)). *BirA* ES cells express *BirA* biotin ligase from the *ROSA26* locus and the targeting construct also contains a *loxP* flanked pgk puro selection cassette (Driegen *et al.*, 2005). Before this cell line could be transfected and clones selected with puromycin, the *loxP* flanked pgk puro cassette was removed by transient Cre recombinase expression, clonal expansion of transfected cells, and duplicate plating in the presence and absence of puromycin to identify puromycin sensitive *BirA* ES cell clones. The puromycin sensitive clones were used for subsequent transfections with the BIO Nanog construct (IPC206).

5.3   Generation and functional validation of a BIO Nanog expression plasmid.

The BIO Nanog expression plasmid (Figure 5.1) was constructed as described in the plasmid appendix. To verify that plasmid IPC206 had the desired properties, E14/T ES cells were transiently transfected either with IPC206 alone or in combination with with a Cre recombinase expression plasmid. The next day, the single and double transfected cells were replated and selected with 2µg/ml puromycin or 10µg/ml blasticidin, respectively. After 10 days culture in either the presence or absence of LIF, colonies were photographed (Figure 5.2). IPC206:Cre co-transfections gave rise to blasticidin resistant, DsRed2⁻ LIF independent ES cell colonies, whereas IPC206 single transfectants were DsRed2⁺, puromycin resistant, and LIF dependent (Figure 5.2). These data indicate that IPC206 functions as designed, with the *loxP* flanked DsRed2IRESpac cassette being efficiently excised upon Cre recombination, inducing
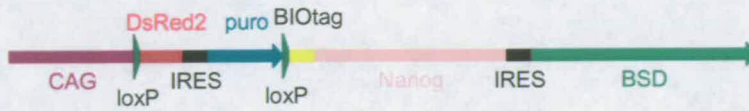
Figure 5.1 A schematic representation of IPC206.
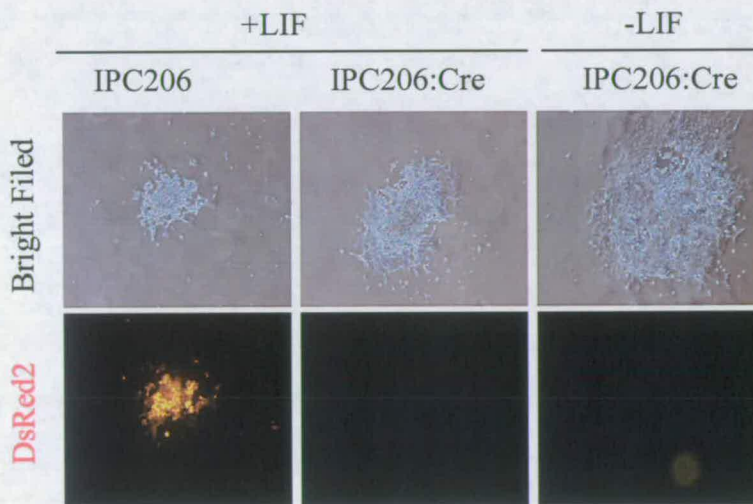
Figure 5.2 Functional test of IPC206 in E14/T ES cell transient transfections.

IPC206 was transfected either alone or in combination with a Cre recombinase plasmid into E14/T ES cells and selected in puromycin (single transfectants) or blasticidin (co-transfectants). Photographs were taken 10 days post transfection.

BIO Nanog expression above the threshold (~5 times endogenous) necessary to direct self-renewal independently of LIF stimulation.

5.4   Generation of *BirA: BIO Nanog* ES cells.

Puromycin sensitive *BirA* ES cells were stably transfected with plasmid IPC206 and selected in 1.0μg/ml puromycin. Fourteen days after transfection, *DsRed2* expression was visualised by fluorescent microscopy. Colonies with strong and uniform expression of DsRed2 protein in all cells were picked and expanded. FACS analysis revealed that of three clones analysed (BirA:206 clones B3, B4, and B5), all expressed *DsRed2* and clone B5 had the highest and most uniform expression level (Figure 5.3). To induce *BIO Nanog* expression, clone B5 was transiently transfected with a Cre recombinase expression plasmid. Twenty-four hours after transfection cells were trypsinised, replated at very low (~6 cells/cm^2), and selected with 5μg/ml blasticidin in the absence of LIF. After 10 days clonal expansion, individual self-renewing DsRed2⁻ clones were picked (B5Cre clones), expanded and stock vials frozen.

5.5   Functional assessment of BIO Nanog.

The selection of B5Cre (*BirA: BIO Nanog*) clones in the absence of LIF shows that the level of *BIO Nanog* expression is sufficient to permit LIF independent self-renewal. To confirm this LIF independent phenotype, cells were plated at clonal density and selected in the presence or absence of LIF for 6 days followed by alkaline phosphatase staining (Figure 5.4b). Qualitatively, the colonies formed by *BIO Nanog* expressing cells are intensely alkaline phosphatase positive (a marker of
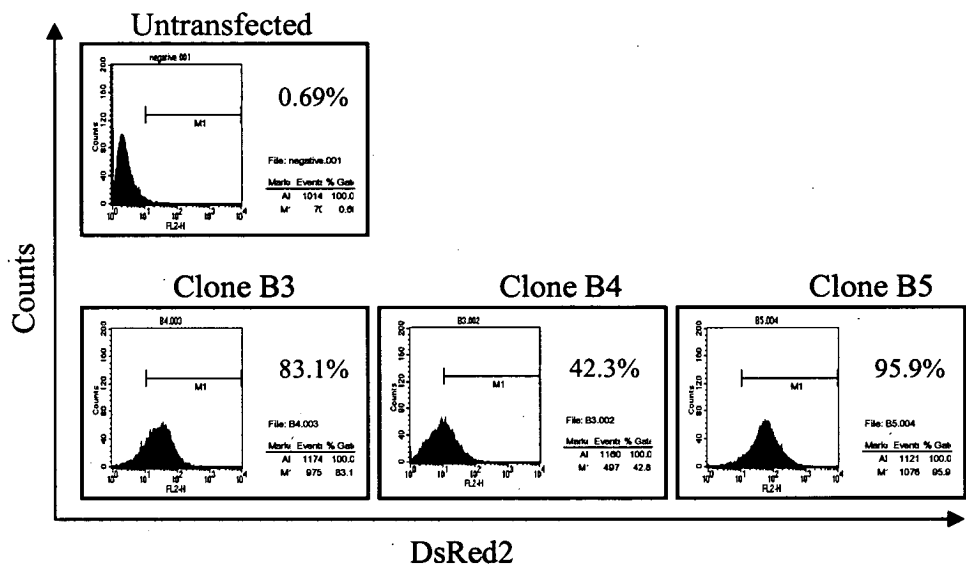
Figure 5.3- FACS analysis of stable IPC206 clones.

FACS analysis shows that BirA:IPC206 stably transfected clones B3, B4, and and B5 are DsRed2 positive. Percentages show the proportion of cells falling within the indicated gate.

undifferentiated cells) and exhibit no differentiation at the periphery of colonies in the presence of LIF, and very little differentiation upon LIF withdrawal. In contrast, *BirA* parental cells growing in LIF supplemented media form colonies containing some differentiated cells, and in the absence of LIF the colonies were flatter in morphology, differentiated, and do not express alkaline phosphatase (Figure 5.4b). Colony counting show that close to 100% of the colonies of all the *BIO Nanog* cell lines generated remain uniformly undifferentiated after 6 days clonal expansion in the absence of LIF (Figure 5.4c). This suggests that *BIO Nanog* would be expressed at ~5-6 times endogenous levels as this level of over-expression is reported to provide robust cytokine independence (Yates and Chambers, 2005). Indeed, immunoblotting of *BirA* and *BirA:BIO Nanog* nuclear extracts shows the presence of an intense band of retarded migration specifically in the BIO Nanog lane corresponding to the expected size of BIO Nanog protein (endogenous Nanog + ~4kDa; Figure 5.4a). This functional data further indicates that the fusion of the BIO tag to the N-terminus of Nanog does not affect its function, as judged by LIF independence assays. For the rest of this chapter experiments involve the use of clone BirA:IPC206B5Cre2, hereafter referred to as *BirA:BIO Nanog* cells.

5.6   Nanog is present in complexes of a broad molecular weight range.

At the outset of this project, it was not known if Nanog had any partner proteins, whether it could multimerise, or whether it was present in higher order multi-protein complexes. To ascertain whether Nanog is present in high molecular weight complexes in mouse ES cells, gel filtration experiments were performed. Nuclear extracts from both the *BirA* parental cells and *BirA:BIO Nanog* cells were subjected
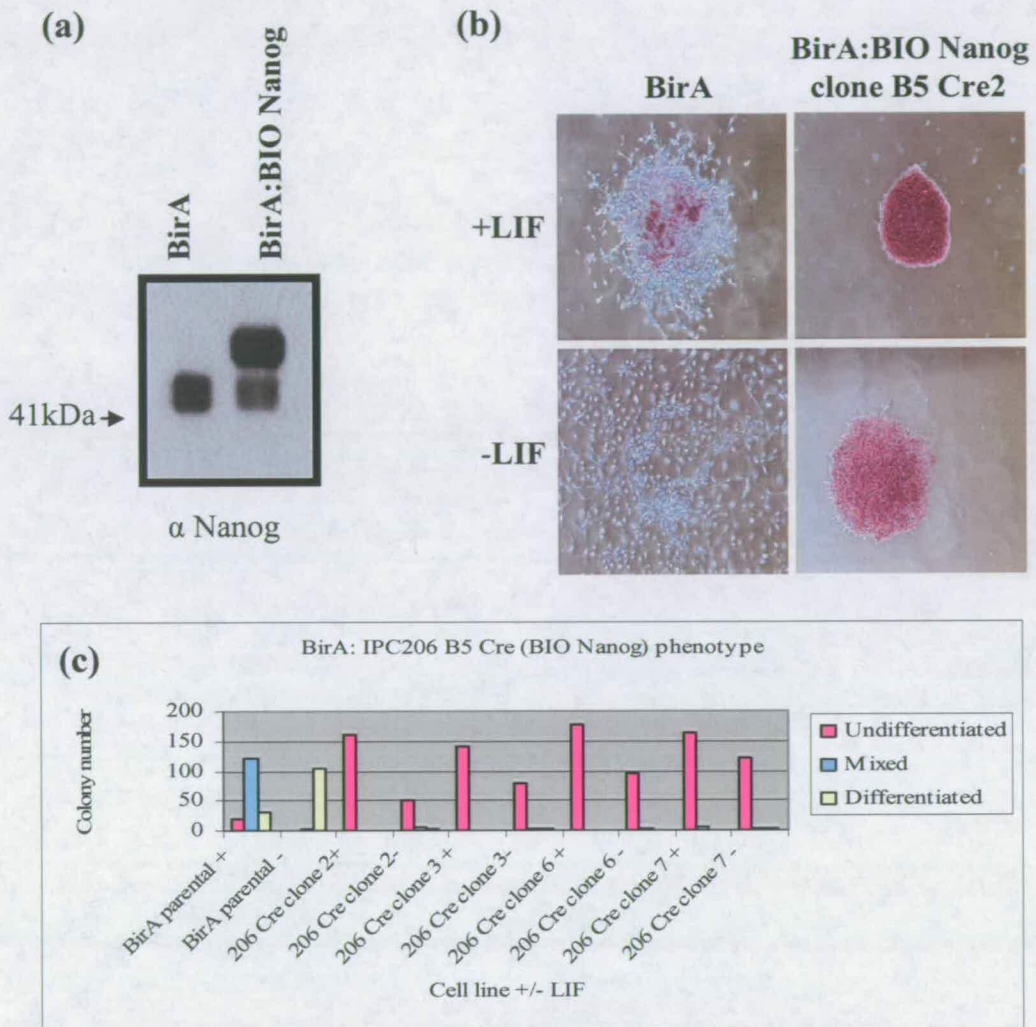
Figure 5.4 Functional assessment of BirA:BIO Nanog cells.

(a) Immunoblot analysis of *BirA* and *BirA:BIO Nanog* nuclear extracts with anti-Nanog antibody reveals a size shifted band specifically in the *BirA:BIO Nanog* lane.

(b) *BirA* and *BirA:BIO* Nanog clones were plated at clonal density, cultured in the presence or absence of LIF, and stained for alkaline phosphatase activity after 6 days. Representative colonies are shown.

(c) The number of uniformly undifferentiated, mixed, and completely differentiated colonies were counted after 6 days clonal expansion.
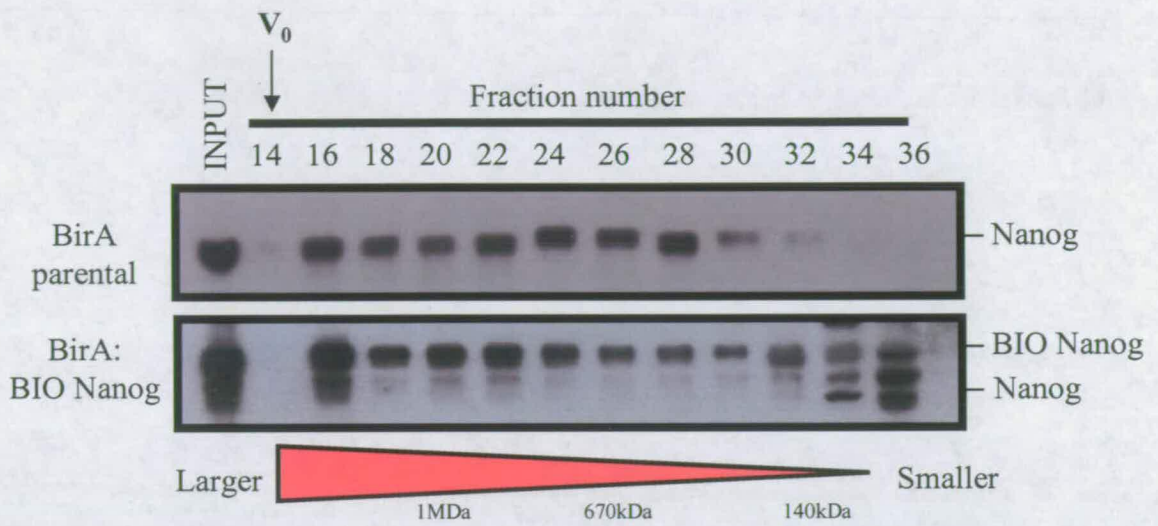
Figure 5.5 Size exclusion chromatography of Nanog in mouse ES cell nuclear extracts.

Nuclear extracts from *BirA* and *BirA:BIO Nanog* cells were fractionated using a Superose 6 column (Amersham). 500µl fractions were collected, precipitated using TCA and resuspended in Laemmli sample buffer. The first 14 fractions correspond to the void volume, V_0. Even numbered fractions from 14-36 were subjected to SDS-PAGE followed by immunoblotting and probing with anti-Nanog antibody.

to size exclusion chromatography using a Superose 6 column. Fractions were collected, proteins precipitated with TCA, and resolved by SDS-PAGE. Immunoblotting with an anti-Nanog antibody reveals that Nanog is present in complexes of a broad range of molecular weights from approximately 140kDa to upwards of 1MDa in weight (Figure 5.5). Crucially, the addition of the BIO tag to Nanog does not disrupt its ability to integrate into these large complexes. However, the presence of BIO Nanog in fractions close to the size of monomeric Nanog suggests that available interaction sites are titrated out by the level of Nanog over-expressed in *BirA:BIO Nanog* cells.

5.7   BIO Nanog is efficiently biotinylated and can be captured on streptavidin coated beads.

To assess whether BIO Nanog could be purified using the high affinity of streptavidin for biotin, *BirA* and *BirA:BIO Nanog* nuclear extracts were incubated with streptavidin coated paramagnetic beads as described in Methods. Immunoblotting with both an anti-Nanog antibody (Chambers, 2005) and streptavidin-HRP was used to analyse the input sample, bound material and the unbound fraction. As these samples were equivalently loaded, an estimate of binding efficiency could be made. Routinely, approximately 50% of the biotinylated BIO Nanog can be captured (Figure 5.6b), although the captured fraction can reach 100% (Figure 5.6a). As the BIO Nanog purification was rarely 100% efficient, gel filtration analysis on the unbound fraction was performed. This permits visualisation of whether either a proportion of all the different Nanog containing complexes are being captured, or whether some complexes are fully captured whilst others are

(a)

BirA | BirA:BIO Nanog

I B U | I B U

α Nanog

Streptavidin-HRP

(b)

BirA:BIO Nanog

Input | Bound | 3x Bound | Unbound

Streptavidin- HRP

(c)

INPUT | V_0 | Fraction number

14 16 18 20 22 24 26 28 30 32 34 36

BIO Nanog
Nanog

Larger — Smaller

1MDa   670kDa   140kDa
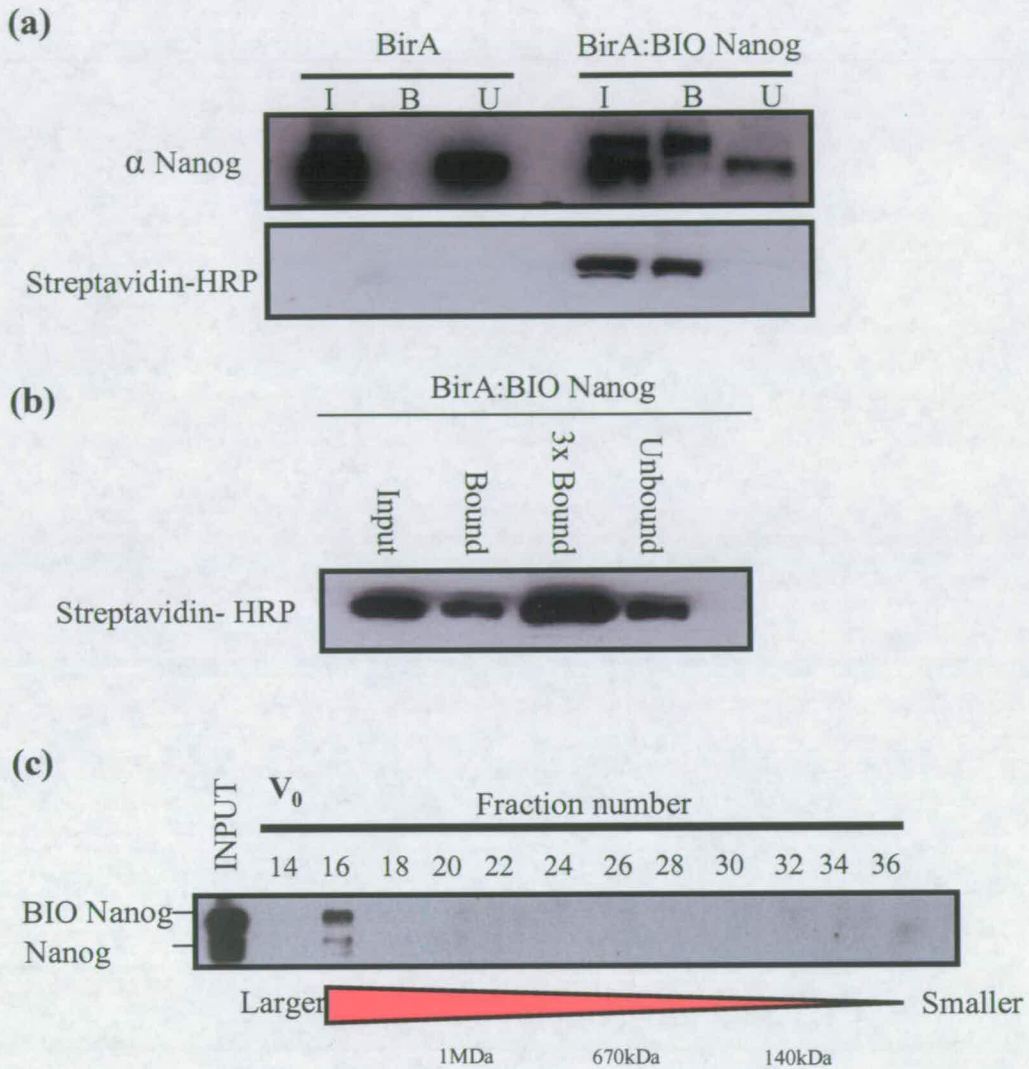
Figure 5.6 BIO Nanog is efficiently biotinylated *in vivo* and can be captured on streptavidin coated beads.

(a,b) Nuclear extracts were prepared from *BirA* (wildtype parental) and *BiAr: BIO Nanog* (Nanog over-expressing) ES cells and incubated with streptavidin coated beads as described in Materials and Methods (Section 2.2). Bound proteins were resolved via SDS-PAGE prior to immunoblotting. Equivalent starting amounts of material were losaded into adjacent lanes labelled input (I), bound (B) and unbound (U). It is possible to capture 100% of BIO Nanog (a). Routinely however, around a 50% capture efficiency is obtained (b). Immunoblots were probed with the indicated antibodies.

(c) Size exclusion chromatography (Superose 6) of the unbound fraction following a streptavidin capture of *BirA: BIO Nanog* nuclear extracts was performed. Fractions were precipitated with TCA, boiled in Laemmli buffer, and subjected to SDS-PAGE. Subsequent immunoblotting with α Nanog antibody reveals that the majority of unbound protein is in fraction 16 (>1MDa). V_0 represents the void volume.

144

refractory to capture (Figure 5.6c). This data indicates that it is predominantly the very high molecular weight complexes in fraction 16 (>1MDa) that have not been captured. These data show that the BIO Nanog capture should allow identification of Nanog partner proteins in a wide range of complex molecular weights.

5.8 Large Scale BIO Nanog purification for mass spectrometry analysis .

The streptavidin capture protocol was scaled up to purify enough material for MS analysis. Routinely the purifications are performed on nuclear extract containing 15mg protein with a final concentration of 100mM KCl and 0.3% NP-40 (see Methods). After the purification procedure, ~95% of the bound material is subjected to either SDS-PAGE and stained with colloidal blue protein stain (Figure 5.7) or to on-bead trypsinisation. Prior to submitting the purified material for MS, a small amount of bound material is equivalently loaded with input and unbound material for immunoblot analysis to retrospectively monitor the purification efficiency (example is shown in Figure 5.7a). If $\geq 50\%$ of the BIO Nanog has been captured this is deemed sufficient to warrant MS analysis. In addition, the efficiency of nuclease treatment (RNAseA/ Benzonase) treatment is assessed by analysing nuclear extracts before and after nuclease treatment via agarose gel electrophoresis and ethidium bromide staining. A typical analysis shows that the nucleic acid has been digested (Figure 5.7c) meaning that any interactions detected by MS are likely to be protein- protein interactions that are stable in the absence of DNA/ RNA.
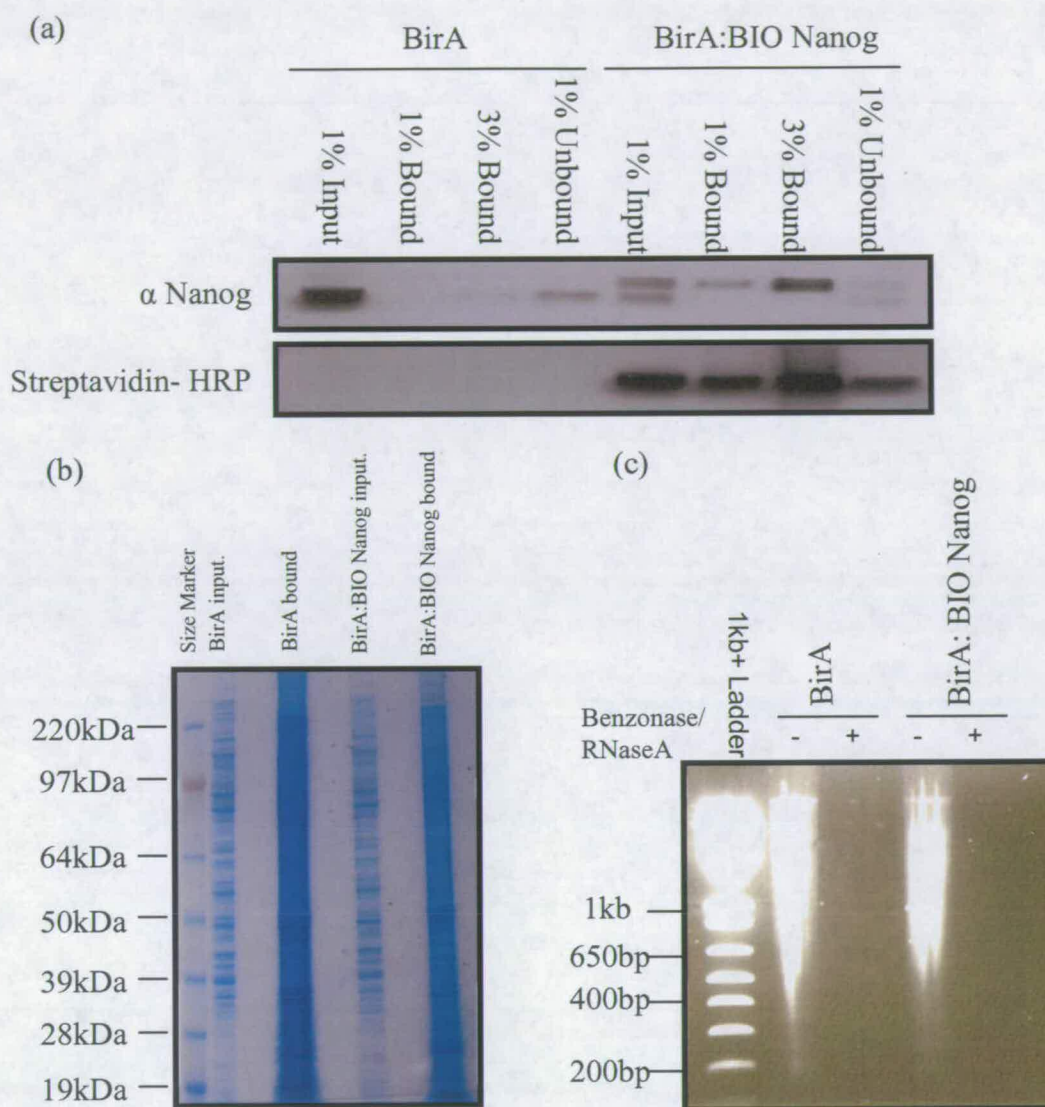
Figure 5.7 Example of large scale purification of BIO Nanog containing complexes.

(a)   Large Scale capture (15mg starting extract) of *BirA and BirA: BIO Nanog* nuclear
        extracts was performed according to the Materials and Methods (Section 2.2). Bound
        proteins were equivalently loaded, input (1%), bound (1% and 3%), and unbound (1%)
        fractions and resolved via SDS-PAGE. After immunoblotting, identical membranes were
        probed with anti- Nanog and streptavidin-HRP.

(b)   Large Scale capture (15mg starting extract) of *BirA and BirA: BIO Nanog* nuclear
        extracts was performed according to the Materials and Methods (Section 2.2). ~50µg
        nuclear extract (input lane) and bound material (from ~15mg starting nuclear extract)
        was resolved by SDS-PAGE prior to staining with colloidal blue.

(c)   Nuclear extracts were treated with Benzonase and RNAseA during streptavidin capture.
        0.33% of binding reaction was removed both before and after addition of nucleases.
        Phenol:chloroform extraction of nucleic acid was performed and samples were
        resuspended in DNA loading dye. Nucleic acids were visualized via agarose gel
        electrophoresis/ Ethidium Bromide staining.

146

5.9 Mass spectrometry analysis of BIO Nanog purifications.

The large scale purification of BIO Nanog was performed four times (details in Materials and Methods 2.2.2), and the conditions for each of the purifications are summarised in Table 5.1. Nano LC-MS/MS was performed on the purified material as described in Materials and Method 2.2.3. The primary criteria for considering a mass spectrometry identified protein as a putative Nanog partner is that it is present in the list of proteins purified from the *BirA:BIO Nanog* nuclear extracts yet absent from the *BirA* parental cell control purifications. Encouragingly, it was found that in all four purification experiments (MS1-4), Nanog peptides are only detected from the *BirA:BIO Nanog* purifications and are never detected in the *BirA* parental cell controls (Table 5.1). This gives confidence that the subtraction of *BirA* parental cell background from the *BirA:BIO Nanog* protein lists should not remove potential partner proteins from consideration. Ideally these peptides should be identified in more than one independent purification

Table 5.1- A summary of conditions used for each large scale BIO Nanog purification.

EXPERIMENT	ON BEAD/ IN GEL TRYPSINISATION	NUCLEASE TREATMENT	WASH BUFFER	NANOG PEPTIDE NO.
MS1	IN GEL	RNAseA	250mM HENG/ 0.3% NP-40	12
MS2	IN GEL	RNAseA	250mM HENG/ 0.3% NP-40	9
MS3	ON BEAD	RNAseA	500mM HENG/ 0.3% NP-40	2
MS4	IN GEL	RNAseA and Benzonase	250mM HENG/ 0.3% NP-40	9

5.10 Preliminary co-immunoprecipitation experiments to confirm interactions of MS identified proteins.

Details of the peptides identified in the mass-spectrometry analysis of *BirA:BIO Nanog* cells, yet absent in the *BirA* control purifications are provided in Table 5.2. This table contains details of proteins identified in all four, three out of four, or two out of four of the BIO Nanog purifications. Due to the fact that Nanog itself was the only protein specifically identified in all four purifications, further investigation of potential partner proteins could not be limited to these proteins co-purifying in all experiments. Therefore, proteins that were identified in more than one of the BIO Nanog purifications (either absolutely or enriched compared to background), that were identified by multiple peptide hits, or had biological rationale for being a Nanog partner were further investigated. Nanog has been shown to possess transactivation potential (Pan and Pei, 2003; Pan and Pei, 2005) and as such may associate with other proteins associated with transcriptional activation. To ascertain whether the proteins identified are genuine Nanog partners, co-immunoprecipitation experiments were performed as this remains the gold-standard for biochemical validation of protein-protein interactions. These experiments are however often limited by the availability of good antibodies

Firstly, a putative interaction with Esrrb was investigated as it acts as a transcription factor and appears to be transcriptionally regulated by Nanog. Furthermore, *Esrrb* knock-down via RNAi leads to differentiation of ES cells (Loh *et al.*, 2006) suggesting a possible role in maintaining pluripotency. Secondly, HDAC2 (histone deacetylase 2) was followed up due to the fact that HDAC2 is able to modify chromatin and is involved in gene repression. Finally, a possible interaction with the WD40 repeat

Table 5.2- List of proteins identified in BIO Nanog purifications.

The peptides listed in these tables are identified in the BIO Nanog purifications indicated and absent from the BirA background controls.

Peptides identified in 4 purifications

Gene Identifier	Name	Present on purifications
gi|31338864	homeobox transcription factor Nanog [Mus musculus]	MS1, MS2, MS3, MS4

Peptides identified in 3 of the 4 purifications

Gene identifier	Name	Present on purification
gi|45272186	Nanog variant protein 1b [Mus musculus]	MS1,MS2,MS3
gi|56201283	NANOG/STM1 [Mus musculus molossinus]	MS1,MS2,MS3
gi|20987847	Hnrpl protein [Mus musculus]	MS1,MS2,MS3
gi|26378644	unnamed protein product [Mus musculus]	MS1,MS2,MS4
gi|2760632	protein L [Mus musculus]	MS1,MS2,MS3
gi|71681342	Helicase, lymphoid specific [Mus musculus]	MS1,MS2,MS4
gi|31560812	breast cancer 1 [Mus musculus]	MS1,MS2,MS4
gi|4097808	Brca1 [Mus musculus]	MS1,MS2,MS4
gi|46329772	Breast cancer 1 [Mus musculus]	MS1,MS2,MS4
gi|969172	breast/ovarian cancer susceptibility protein homolog	MS1,MS2,MS4

Peptides identified in 2 of the 4 purifications

Gene identifier	Name	Present on purification
gi|32966256	nanog [Mus musculus]	MS1,MS2
gi|12859782	unnamed protein product [Mus musculus]	MS1,MS2
gi|26354819	unnamed protein product [Mus musculus]	MS1,MS2
gi|12847921	unnamed protein product [Mus musculus]	MS1,MS3
gi|29126987	Hypothetical protein LOC66184 [Mus musculus]	MS1,MS2

gi|54912	unnamed protein product [Mus musculus]	MS1,MS4
gi|200246	pyruvate carboxylase	MS1,MS4
gi|26347787	unnamed protein product [Mus musculus]	MS1,MS3
gi|63547292	PREDICTED: similar to 40S ribosomal protein S2 [Mus musculus]	MS1,MS2
gi|63594466	PREDICTED: similar to 40S ribosomal protein S2 isoform 1 [Mus musculus]	MS1,MS4
gi|15919908	unnamed protein product [Mus musculus]	MS1,MS4
gi|28302223	Hypothetical protein LOC432987 [Mus musculus]	MS1,MS3
gi|495128	mCBP [Mus musculus]	MS1,MS4
gi|1568657	testicular antigen [Mus musculus]	MS1,MS2
gi|16554627	WD repeat domain 5 [Homo sapiens]	MS1,MS4
gi|18204699	Nucleolar protein 5A [Mus musculus]	MS1,MS2
gi|26348689	unnamed protein product [Mus musculus]	MS1,MS2
gi|26353478	unnamed protein product [Mus musculus]	MS1,MS2
gi|26353514	unnamed protein product [Mus musculus]	MS1,MS2
gi|38614350	Cp protein [Mus musculus]	MS1,MS2
gi|12232371	helicase, lymphoid specific [Mus musculus]	MS1,MS4
gi|15489014	Nsun2 protein [Mus musculus]	MS1,MS4
gi|1903238	type II intermediate filament of hair keratin [Mus musculus]	MS1,MS2
gi|19353961	Wdr74 protein [Mus musculus]	MS1,MS4
gi|26335199	unnamed protein product [Mus musculus]	MS1,MS2
gi|26338343	unnamed protein product [Mus musculus]	MS1,MS2
gi|26341620	unnamed protein product [Mus musculus]	MS1,MS2
gi|26344922	unnamed protein product [Mus musculus]	MS1,MS4
gi|27883856	low density lipoprotein receptor-related protein 4 [Mus musculus]	MS1,MS4
gi|30047836	MutS homolog 2 [Mus musculus]	MS1,MS4
gi|31542461	NOL1/NOP2/Sun domain family 2 [Mus musculus]	MS1,MS4
gi|37994713	Krt1-13 protein [Mus musculus]	MS1,MS2
gi|47117264	Low-density lipoprotein receptor-related protein 4 precursor (LDLR dan)	MS1,MS2
gi|55153885	Similar to glyceraldehyde-3-phosphate dehydrogenase [Mus musculus]	MS1,MS4
gi|726086	MutS homolog 2	MS1,MS4
gi|12847667	unnamed protein product [Mus musculus]	MS1,MS2
gi|12848426	unnamed protein product [Mus musculus]	MS1,MS4
gi|12862002	unnamed protein product [Mus musculus]	MS1,MS2
gi|1304155	pokeweed agglutinin-binding protein [Mus musculus]	MS1,MS4
gi|13272554	cytokeratin KRT2-6HF [Mus musculus]	MS1,MS3
gi|13386238	keratin complex 1, acidic, gene 4 [Mus musculus]	MS1,MS2
gi|13529464	Nucleolin [Mus musculus]	MS1,MS2
gi|13542826	Cdc2a protein [Mus musculus]	MS1,MS4
gi|15214716	Rps9 protein [Mus musculus]	MS1,MS3
gi|21759414	Exosome complex exonuclease RRP43 (Ribosomal RNA-processing protein 43) (Exosome component 8)	MS1,MS4
gi|2228746	Dlgh1 homolog [Mus musculus]	MS1,MS4
gi|2407195	putative RNA helicase and RNA dependent ATPase [Mus musculus]	MS1,MS3
gi|26325114	unnamed protein product [Mus musculus]	MS1,MS2
gi|26327461	unnamed protein product [Mus musculus]	MS1,MS2
gi|26327475	unnamed protein product [Mus musculus]	MS1,MS2
gi|26341628	unnamed protein product [Mus musculus]	MS1,MS2
gi|26986202	SMC5 protein [Mus musculus]	MS1,MS4
gi|27777677	deoxyribose-phosphate aldolase-like [Mus musculus]	MS1,MS4

150

gi|2809537	beta prime coatomer protein [Mus musculus]	MS1,MS4
gi|28548933	PREDICTED: exosome component 7 isoform 1 [Mus musculus]	MS1,MS4
gi|29611663	exosome complex exonuclease RRP41 [Mus musculus]	MS1,MS4
gi|30962875	Exosome component 7 [Mus musculus]	MS1,MS4
gi|31542366	cell division cycle 2 homolog A [Mus musculus]	MS1,MS4
gi|32822838	1700021I09Rik protein [Mus musculus]	MS1,MS2
gi|37359764	mKIAA0078 protein [Mus musculus]	MS1,MS4
gi|37359778	mKIAA0116 protein [Mus musculus]	MS1,MS4
gi|39104489	mKIAA0594 protein [Mus musculus]	MS1,MS4
gi|4097873	eIF3-p44 [Mus musculus]	MS1,MS4
gi|45595709	Exosc8 protein [Mus musculus]	MS1,MS4
gi|50360	unnamed protein product [Mus musculus]	MS1,MS4
gi|50510377	mKIAA0136 protein [Mus musculus]	MS1,MS4
gi|51704904	PREDICTED: similar to Elongation factor 1-gamma (EF-1-gamma) (eEF-1B gamma) isoform 1 [Mus musculus	MS1,MS4
gi|53454	nucleolin [Mus musculus]	MS1,MS2
gi|58477264	Unknown (protein for MGC:102174) [Mus musculus]	MS1,MS2
gi|60359866	mKIAA4013 protein [Mus musculus]	MS1,MS4
gi|6754488	keratin complex 2, basic, gene 6b [Mus musculus]	MS1,MS3
gi|67972435	DEAD (Asp-Glu-Ala-Asp) box polypeptide 49 [Mus musculus]	MS1,MS4
gi|72679679	Krt1-24 protein [Mus musculus]	MS1,MS2
gi|8393684	keratin, hair, acidic, 5 [Mus musculus]	MS1,MS2
gi|988214	breast/ovarian cancer susceptibility homolog	MS1,MS2
gi|28972888	mKIAA3005 protein [Mus musculus]	MS2,MS3
gi|53733821	Tubulin, alpha 1 [Mus musculus]	MS2,MS3
gi|17512384	BC031593 protein [Mus musculus]	MS2,MS3
gi|22096212	Transcription factor 20 (Stromelysin 1 PDGF-responsive element-binding protein) (SPRE-binding prote	MS2,MS3
gi|387399	epidermal keratin type I	MS2,MS4
gi|12849997	unnamed protein product [Mus musculus]	MS2,MS4
gi|26330478	unnamed protein product [Mus musculus]	MS2,MS4
gi|26353010	unnamed protein product [Mus musculus]	MS2,MS4
gi|30931339	Transcriptional repressor NAC1 [Mus musculus]	MS2,MS4
gi|37675525	AHNAK [Mus musculus]	MS2,MS4
gi|3851579	ribosomal protein L8 [Mus musculus]	MS2,MS4
gi|50510329	mKIAA0020 protein [Mus musculus]	MS2,MS4
gi|50510873	mKIAA1298 protein [Mus musculus]	MS2,MS4
gi|62185779	Nucleoporin 107 [Mus musculus]	MS2,MS4
gi|62201695	Rpl14 protein [Mus musculus]	MS2,MS4
gi|12853471	unnamed protein product [Mus musculus]	MS2,MS3
gi|1401337	Mre11b	MS2,MS4
gi|23274262	Orc1l protein [Mus musculus]	MS2,MS4
gi|26340418	unnamed protein product [Mus musculus]	MS2,MS3
gi|27693623	Unknown (protein for IMAGE:4481097) [Mus musculus]	MS2,MS4
gi|27808676	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1 (SWI/	MS2,MS4
gi|46399229	Ewing sarcoma breakpoint region 1 [Mus musculus]	MS2,MS4
gi|50510319	mKIAA0002 protein [Mus musculus]	MS2,MS4
gi|51261317	Smarcd1 protein [Mus musculus]	MS2,MS4
gi|5295992	chaperonin containing TCP-1 theta subunit [Mus musculus]	MS2,MS4
gi|56206934	Ewing sarcoma homolog [Mus musculus]	MS2,MS4

151

gi|56206935	Ewing sarcoma homolog [Mus musculus]	MS2,MS4
gi|37748194	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21 [Mus musculus]	MS3,MS4
gi|6677757	ribonucleic acid binding protein S1 [Mus musculus]	MS3,MS4
gi|110625918	homeobox transcription factor Nanog [Mus musculus]	MS3,MS4
gi|12852157	unnamed protein product [Mus musculus]	MS3,MS4
gi|26349459	unnamed protein product [Mus musculus]	MS3,MS4
gi|26334773	unnamed protein product [Mus musculus]	MS3,MS4
gi|62471421	Orfx [Mus musculus]	MS3,MS4

protein Wdr5 was investigated. Wdr5 is a part of the MLL/ Trithorax complex capable of modifying histone H3 at residue K4, a modification associated with transcriptional activation (Wysocka et al., 2005).

5.10.1 Nanog-Esrrb interaction

E14/T ES cells were transiently transfected with (Flag)$_3$Nanog expression vector (IPC 37) and after 3 days in culture whole cell lysates prepared. A rabbit polyclonal antibody to Esrrb (obtained from Abcam- ab19331) was used to immunoprecipitate (IP) Esrrb protein. The (Flag)$_3$Nanog expression vector allowed the use of mouse anti-Flag antibody for Nanog detection. This circumvents cross reactivity between the rabbit anti-Esrrb and rabbit anti-Nanog antibody routinely used for Nanog detection obscuring potentially interesting data. Esrrb immunoprecipitates were subjected to SDS-PAGE electrophoresis, immunoblotted and probed with mouse anti-Flag antibody. The immunoblot shows there is a discrete band corresponding to (Flag)$_3$Nanog specifically in the Esrrb immunoprecipitate that is absent from the mock immunoprecipitate (Figure 5.8). This promising preliminary data suggests that Esrrb is a *bona fide* Nanog partner protein.

5.10.2 Nanog- HDAC2 interaction.

Nanog immunoprecipitates were prepared from EF4 (Nanog over-expressing) cells (Chambers et al., 2003) using anti-Nanog antibody (Chambers, 2005). The immunoprecipitate was subjected to SDS-PAGE, immunoblotted, and probed with an anti-HDAC2 antibody. Again a specific signal was observed only in the Nanog IP lane which suggests that Nanog and HDAC2 can be found in complex together (Figure 5.9).

5.10.3 Nanog-Wdr5 interaction

To investigate the potential Wdr5-Nanog interaction, two approaches were undertaken. To assess whether Wdr5 and Nanog interact in ES cells, (Flag)$_3$Wdr5 (IPC 328) was constructed (see plasmid appendix for cloning strategy) and transiently transfected into E14/T cells. After 72h in culture, lysates were made, Nanog was immunoprecipitated, and immunoprecipitates were resolved by SDS-PAGE. Immunoblotting revealed that (Flag)$_3$Wdr5 is specifically present in Nanog immunoprecipitates suggesting confirmation of this protein-protein interaction (Figure 5.10b). A histone tail peptide binding experiment was performed as Wdr5 has been shown to bind preferentially to histone H3 dimethylated at lysine 4 (H3K4(me)$_2$) (Wysocka *et al.*, 2005). A similar binding specificity for Nanog would support a model in which Nanog is bridged by Wdr5 to H3K4(me)$_2$. Streptavidin immobilised biotinylated histone H3 tail peptides either unmodified or dimethylated at lysine 4 were incubated with ES cell nuclear extracts and assayed for their ability to interact with both Nanog and Wdr5. The immunoblot data (Figure 5.10a) shows that Nanog can bind the histone tails, and is indiscriminate in binding preference for unmodified or dimethylated H3K4. Wdr5 however shows preference for dimethylated H3K4 in accordance with published data in other cell types (Wysocka *et al.*, 2005).
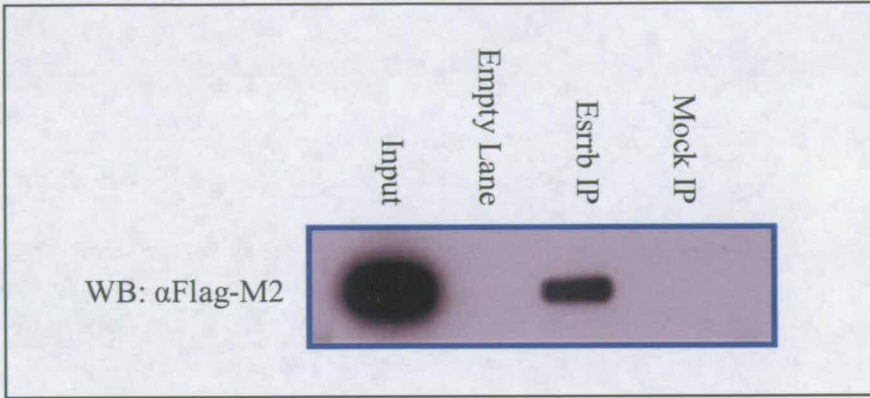
Figure 5.8- Co- immunoprecipitation of Nanog with Esrrb.

Whole ES cell lysate was prepared from E14/T ES cells expressing
(Flag)₃Nanog. Esrrb immunoprecipitates were prepared using anti-Esrrb
antibody, subjected to SDS-PAGE and immunoblotted with anti-Flag-M2
antibody.
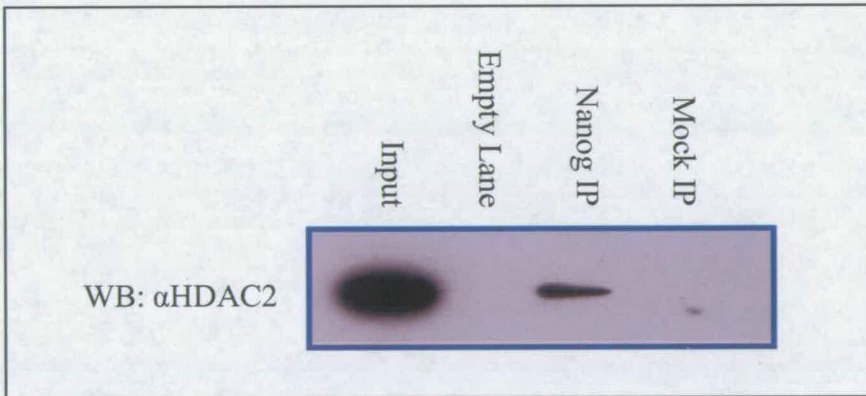
Figure 5.9- Co- immunoprecipitation of Nanog with HDAC2.

Whole cell lysate was prepared from EF4 ES cells. Nanog
immunoprecipitates were prepared using anti-Nanog antibody,
subjected to SDS-PAGE, immunoblotted, and probed with anti-
HDAC2 antibody.
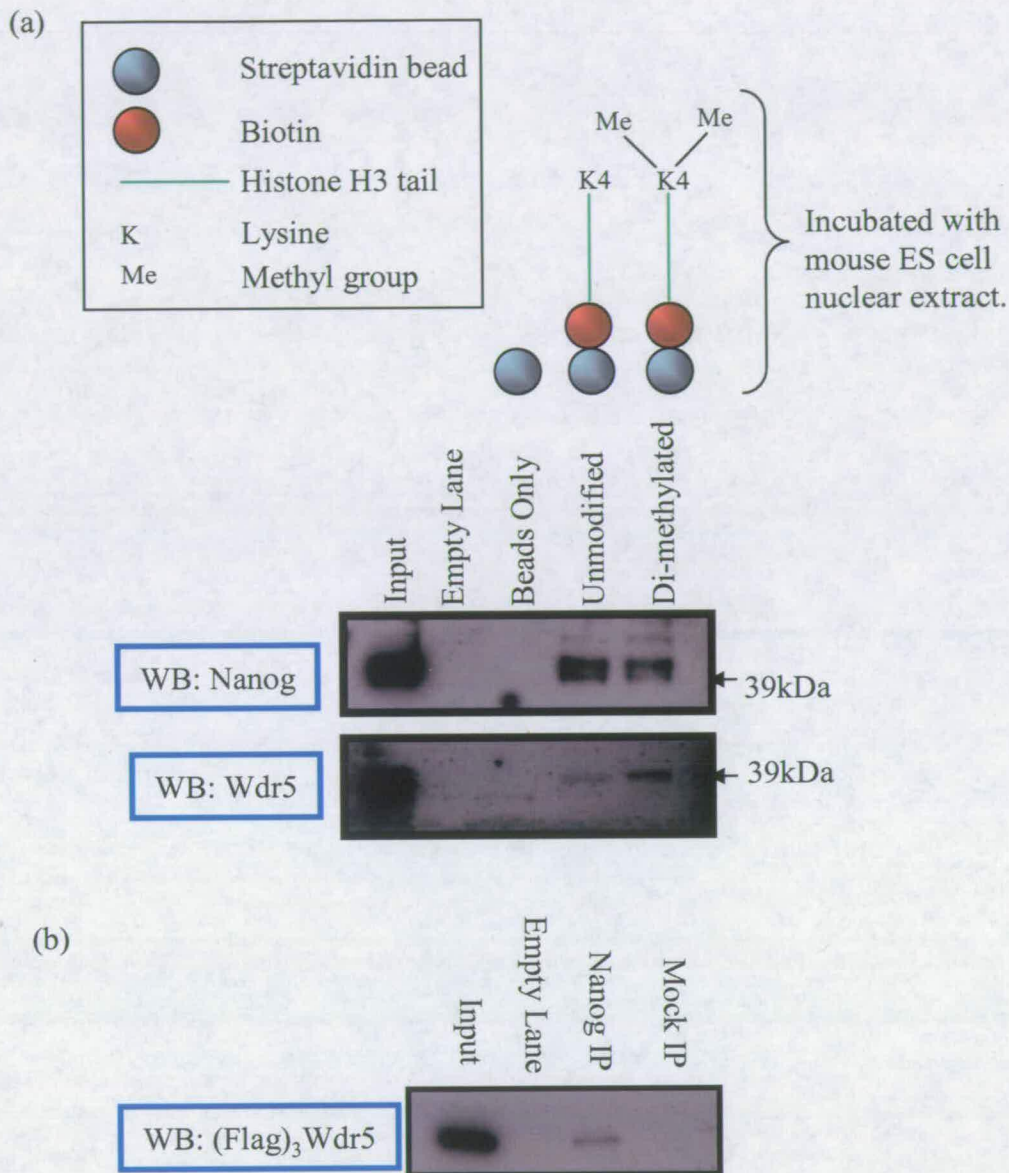
Figure 5.10- Nanog interacts with histone H3 tails and Wdr5.

(a) Incubation of streptavidin immobilised biotinylated histone H3 tails with ES cell nuclear extract reveals Nanog can bind H3 tails regardless of the K4 modification, whereas dimethylated H3K4 is the preferred binding substrate for Wdr5 as judged by immunoblotting with anti-Wdr5 and anti-Nanog antibody.

(b) Nanog immunoprecipitates were prepared from ES cells transiently expressing (Flag)₃Wdr5. The immunoblot was probed with α-Flag-M2 antibody.

5.11   Discussion

5.11.1   Background

At the outset of the experiments described in this chapter there were no known Nanog interacting proteins. Given the failure of the candidate approach to partner identification described in Chapter 3, a biotinylation strategy was employed as a single step procedure to identify Nanog partner proteins in mouse ES cells in an unbiased manner. This chapter has described the design, generation and use of a biotinylation tagging strategy to identify Nanog partner proteins in mouse ES cells. Here the technical aspects of the experiments will be discussed prior to discussion of the results in the context of relevant published data.

5.11.2   Technical aspects.

5.11.2.1   Plasmid design.

The design of plasmid IPC 206 (Figure 5.1) has 2 major advantages over a more simple expression plasmid such as pPyCAG BIONanogIRESpac (IPC203). Firstly, the level of transgene expression can be quantified by FACS analysis to identify the highest expressing clones, which should then correlate to *Nanog* over-expression and cytokine independence upon transient Cre recombinase co-expression. Having a cell line expressing the maximal amount of *BIO Nanog* possible should yield the highest possible number of tryptic Nanog peptides and partner protein peptides by MS analysis. The *BirA:BIO Nanog* ES cell line therefore provides a useful reagent for the proposed proteomic screen. This type of construct with a "pre-selection cassette" would also be useful in other studies in which maximal expression of a cDNA is required. Secondly,

based on the work in chapter 4 of this thesis, it can be seen that transgenic *Nanog* expression can be mosaic in nature. Using IPC206, fluorescent microscopy can be employed to identify colonies uniformly expressing DsRed2 protein in all cells.

5.11.2.2   Nanog is present in complexes of a broad molecular weight range.

Many eukaryotic proteins do not exert their action alone but rather act as part of complexes containing many different proteins. Some proteins, such as components of the spliceosome machinery are found in multi-protein complexes of discrete molecular weights (Reed *et al.*, 1988) whereas others such as the haematopoetic transcription factor Gata1 are found in a number of distinct complexes (Rodriguez *et al.*, 2005). Size exclusion chromatography reveals Nanog is found in a range of different molecular weight complexes, a profile also seen by others (Wang *et al.*, 2006). The *BirA:BIO Nanog* ES cell line generated in this study expresses *BIO Nanog* at elevated levels in the population compared to genetically unmanipulated ES cells, thus permitting cytokine independent self-renewal. Importantly, Nanog is present in the same profile of molecular weight complexes in both the parental *BirA* cell line and the *BIO Nanog* cell line from ~140KDa to complexes upwards of 1MDa in size (Figure 5.5). This shows the BIO Nanog protein can form the same complexes as endogenous Nanog. Moreover, these data give confidence that there are a number of Nanog interacting molecules awaiting identification. Fraction 14 (V_0, void volume) contains no Nanog protein suggesting that Nanog is not forming large non-specific aggregates in solution and, one can infer therefore, that the Nanog protein present in fractions 16-36 is in *bona fide* functionally relevant multi-protein complexes.

5.11.2.3   BIO Nanog is efficiently biotinylated and captured.

Pilot experiments show that BIO Nanog protein can be efficiently purified by streptavidin coated magnetic beads. Routinely, the purification efficiency was found to be 50-60% of total and that this represents a wide range of Nanog containing complexes, but with the highest molecular weight complexes been inefficiently captured (Figure 5.5c). The sub-optimal purification efficiency is not due to low BirA activity (biotinylation efficiency), as biotinylated Nanog is observed in the unbound fraction. It is possible that the high molecular weight complexes are refractory to capture as the BIO epitope could be buried within the complex rendering it inaccessible to the streptavidin coated beads. In future experiments, purification efficiency could be increased by (a) increasing detergent concentration although this may lead to the loss of some associated partner proteins, or (b), producing a C-terminal BIO tag fusion which may be more amenable to capture by streptavidin coated beads. The BIO Nanog purifications were performed in the presence of nucleases for a number of reasons. Previous streptavidin purifications of BIO tagged transcription factors identify RNA binding proteins as abundant background proteins (de Boer *et al.*, 2003; Meier *et al.*, 2006). Furthermore, false positive identification of interacting proteins may occur if the two proteins are simply bound to same fragment of DNA but at a distance of several kilobases apart. Of course, DNA may be a crucial factor in the formation and stability of the ternary multi-protein complex, therefore ideally, purifications should be performed in both the presence and absence and absence of nucleases.

5.11.2.4 Large scale BIO Nanog purification and MS analysis.

One would expect the banding pattern of bound material from *BirA* and *BirA: BIO Nanog* cell nuclear extracts to vary dramatically upon colloidal blue protein staining of SDS-PAGE gel as specific Nanog partners should be enriched from only the *BIO Nanog* nuclear extracts. This however is not the case with a huge background visible in the *BirA* nuclear extracts (Figure 5.7b). Comparing the background bound material (*BirA* parental cells) from ES cells (this study), to the background bound material from MEL cells (de Boer *et al.*, 2003), one can see background binding by ES cell nuclear extracts is more significant. Notwithstanding this high background, samples were processed by mass-spectrometry as immunoblotting revealed that Nanog was specifically purified in the *BirA: BIO Nanog* pull-downs (Figure 5.7a). Indeed, each of the four sets of MS data (MS1-4) also confirms that Nanog is only purified from *BirA: BIO Nanog* nuclear extracts (see Table 5.3). Methods to reduce the background include increasing the salt concentration or detergent concentration during the purification procedure, or introducing a second step in the purification procedure. For this reason, the construct used in this thesis has a TEV protease cleavage site cloned at the 3' end of the BIO tag. In the future, a second purification step could be introduced in which TEV protease can be used to cleave BIO Nanog and associated proteins from the beads, leaving behind the non-specifically bound background material. Four independent BIO Nanog purifications were performed on four independently prepared nuclear extracts. The conditions used for each experiment are overviewed in Table 5.1. If the procedure had worked efficiently, and the background was clear, a simple subtraction of the proteins identified in the *BirA* parental cell purification from the proteins identified in the *BirA:BIO Nanog* purification would yield a list of potential Nanog interacting proteins. Performing the

experiment several times and analysing the overlapping proteins should focus the investigator on which proteins are likely to be specific Nanog partners. The mass-spectrometry appendix shows the proteins which are identified either in all four purifications, three of the four purifications, and two of the four purifications but not in the *BirA* parental cell background. These tables are produced using the Genbank identifier (gi) numbers which are not unique, meaning a potential partner should be also checked manually by name in *BirA* background list. The tables contain only proteins that are either present or absent in a given purification. Firstly it can be seen that the only protein identified in all 4 purifications is Nanog itself, therefore the stringent criteria described above are not sufficient to analyse this MS data. The criteria for further investigation of identified proteins had to be modified; peptides identified in more than one purification (either absolutely or enriched compared to the background) were considered, particularly if their biological function was related to transcriptional regulation. There are two published positive controls for Nanog partner proteins i.e. active Smad1 (Suzuki *et al.*, 2006b) and Sall4 (Wu *et al.*, 2006). Active Smad1 was never identified as a Nanog partner in this study or that of Wang *et al* (2006) and Sall4 was only identified in the 4th purification and was enriched in the BIO Nanog pull-down but not absolutely absent from the *BirA* background (7 peptides and 3 peptides respectively), suggesting these complexes are either transient or unstable under the purification conditions used.

5.11.3   Data discussion.

5.11.3.1   Nanog-Wdr5 interaction.

Gene regulation is dependent on accessibility of the transcriptional machinery to the target DNA. In eukaryotic systems, this is governed not only by sequence specific transcription factors, but also on modifications made to the histone tails with the writing of the so called "histone code" which adds a further level of complexity to transcriptional regulation (Strahl and Allis, 2000). One example of a chromatin modification effecting gene expression is provided by the Set1 family of methyltransferases, which are involved in catalysing di- to tri- methylation of lysine 4 (K4) of histone H3 tails, a modification which is found at actively transcribed genes (Santos-Rosa *et al.*, 2002). Set1/MLL/trithorax complexes are capable of catalyzing di- to tri- methylation reaction of H3K4 in a diverse range of organisms ranging from yeast to human (Sims *et al.*, 2003). The protein in the MLL complex responsible for reading the histone code and providing specificity for dimethylated H3K4 substrate was identified as the WD40 domain containing protein Wdr5 (Dou *et al.*, 2005; Wysocka *et al.*, 2005). Structural studies show that the interaction of Wdr5 with dimethylated H3K4 is stabilised by a pair of hydrogen bonds that are not formed with the unmodified H3 tail due to distance constraints (Han *et al.*, 2006). Furthermore, when interacting with Wdr5, the dimethylated H3K4 is readily accessible to histone methyltransferases that catalyse the di- to tri-methylation reaction (Han *et al.*, 2006). Wdr5 was identified specifically in two independent BIO Nanog purifications (3 peptides in MS1 and 2 peptides in MS4) and is unrepresented in the background binding proteins. If Wdr5 is a *bona fide* Nanog partner one might expect to find further MLL/trithorax complex

components such as Ash2 and MLL itself co-purifying with Nanog and Wdr5. That this was not the case, could reflect complex instability during purification. However, characterised MLL complexes are reportedly stable under similar extraction conditions to those used here (Wysocka *et al.*, 2005; Dou *et al.*, 2005). A preliminary co-immunoprecipitation experiment shows that Wdr5 is found in Nanog immunoprecipitates prepared from mouse ES cells (Figure 5.10b) suggesting Nanog and Wdr5 physically interact in ES cells. Histone tail peptide pull-down experiments using mouse ES cell nuclear extracts revealed that Nanog can bind to H3 tails and exhibits no preference for di-methylated K4 as opposed to unmodified H3 tails. MLL also shows no binding preference for dimethylated H3 tails although MLL does bind histone H3 (Wysocka *et al.*, 2005). As expected, Wdr5 binds H3 tails and exhibits a substrate preference for dimethylated K4 (Figure 5.10a). Wysocka *et al* (2005) suggest that the majority of Wdr5 is in free-form (non-complexed), thus allowing visualisation of the binding specificity difference on immunoblots. These data are encouraging as it shows Nanog and Wdr5 can bind the same histone tails *in vitro*. Core histones are detected as apparent background binding proteins in the mass-spectrometry data but as the chromatin is precipitated during the nuclear extraction protocol it is unlikely these histones represent intact nucleosomes. It is unlikely therefore that the Nanog-Wdr5 interaction is mediated non-specifically via nucleosomes. It will however be important to analyse whether Nanog can bind additional core histone tails, and to repeat the co-immunoprecipitation on lysates containing disrupted nucleosomes. In addition, further confirmatory co-immunoprecipitation experiments will be required to more fully explore the interaction between Wdr5 and Nanog. It may also be interesting to perform *in vitro* methyltransferase assays on Nanog immunoprecipitates to identify whether

Nanog containing complexes possess methyltransferase activity. Wdr5 was not reported in the screen for Nanog interactors performed by Wang et al (2006). However, analysis of the supplementary MS data reveals Wdr5 was indeed specifically identified in one of the single step BIO Nanog purifications (Wang et al., 2006). A model of MLL function proposed by Wysocka et al (2005) requires a sequence specific transcription factor to recruit the MLL complex to chromatin; it is possible that Nanog acts as a recruiter of the MLL complex to activate target genes or mark them as poised for activation upon differentiation. Indeed, in mouse ES cells, tracts of chromatin have been identified which are dually marked with the opposing trimethylated H3K27 and trimethylated H3K4 modifications which are suggested to poise genes for activation upon differentiation (Bernstein et al., 2006; Azuara et al., 2006). These so called "bivalent domains" have been identified at ~50% of the genomic regions identified as Nanog bound regions by Loh et al (2005). This suggests Nanog could be key in establishing a repressive "bivalent domain" in ES cells which is released upon differentiation concomitant with spreading of the trimethylated H3K4 modification (Bernstein et al., 2006). Of further interest, the H3K4 tri-methylation modification is increased sharply as PGC's enter the genital ridge (E11) and persists throughout the time window (E8.5-E12.5) during which EG cells can be derived (Seki et al., 2005; reviewed by Surani et al., 2007). If this modification is a critical requirement of the re-programming events occurring during pre-implantation development as well as during PGC maturation, and Nanog recruits the Wdr5-MLL complex to chromatin, then this could explain why mature PGC's cannot be generated by Nanog[-/-] ES cells (Chambers unpublished).

5.11.3.2   Nanog-Esrrb interaction.

Esrrb (estrogen related receptor beta, also known as Err2, Errβ) is an orphan nuclear receptor, that is to say, it is a ligand dependent transcription factor with the identity of the ligand remaining unknown (Robinson-Rechavi *et al.*, 2003). *Esrrb* is highly expressed in ES and EC cells, and *in vivo*, post-implantation expression is restricted to the extra-embryonic ectoderm that forms the chorion (Pettersson *et al.*, 1996). *Esrrb*[-/-] embryos die at E10.5 as a result of early placentation defects (Luo *et al.*, 1997). Examination of *Esrrb* expression later in development reveals Esrrb mRNA and protein is seen specifically in the primordial germ cells (PGCs) but not the surrounding mesonephros at the time PGCs arrive in the genital ridge (~E11), with expression persisting until ~E16.0 (Mitsunaga *et al.*, 2004). Functional analysis of Esrrb in the PGCs was carried out by aggregating diploid *Esrrb*[-/-] embryos with tetraploid wild-type embryos to rescue the placental defect. This revealed that the PGC number was significantly reduced in rescued *Esrrb*[-/-] embryos compared to the wild-type controls (Mitsunaga *et al.*, 2004). This suggests that Esrrb has a proliferative effect on PGCs. Nanog is also specifically expressed in PGCs (Chambers *et al.*, 2003; Yamaguchi *et al.*, 2005). Moreover, *Nanog*[-/-] cells exhibit a more severe phenotype during PGC development. *Nanog*[-/-] cells can enter the germ-cell programme and migrate to the genital ridges. However, the expression of *mvh*, a germ cell marker activated as cells enter the genital ridge at E11, is reduced at E11.5, and by E12.5 Mvh[+] *Nanog*[-/-] cells are no longer detectable in chimaeric genital ridges (Chambers unpublished). *Esrrb* expression is increased in *Nanog* over-expressing cells (Loh *et al.*, 2006), and RNAi knock-down of *Esrrb* in ES cells leads to morphological differentiated colonies suggesting it may be important for the maintenance of pluripotency (Loh *et al.*, 2006;

Ivanova *et al.*, 2006). Esrrb was identified in two of the four BIO Nanog purifications performed in this thesis with four peptides versus zero background peptides in MS2, and seven peptides versus two background peptides in MS4. This interaction was also reported by Wang *et al* (2006) by mass-spectrometry but no confirmatory biochemistry was performed. The transcriptional relationship between Nanog and Esrrb, taken together with differentiation of *Esrrb* knock-down ES cells and the PGC phenotypes led to Esrrb being investigated in preliminary experiments to validate a Nanog interaction. A specific band corresponding to Nanog can be seen in Esrrb immunoprecipitates from mouse ES cell lysates suggesting Nanog and Esrrb interact in ES cells (Figure 5.8). A future experiment to epitope tag Esrrb and perform the reciprocal co-immunoprecipitation will be important to confirm the Nanog-Esrrb interaction. As well as further biochemical characterisation of the interaction, it will be interesting to examine the expression of *Esrrb* in pre-implantation embryos as previous studies focussed upon early post-implanatation expression (Luo *et al.*, 1997). The Esrrb antagonist (DES) has been shown to cause self-renewing trophoblast stem (TS) cells to differentiate into giant cells, and furthermore the same differentiation event occurs *in vivo* with DES injection antagonising placental development (Tremblay *et al.*, 2001). To assess whether Esrrb may be a key partner of Nanog it will be interesting to investigate whether ES cells can be derived from *Esrrb*[-/-] blastocysts, and whether the Esrrb antagonist diethylstilbesterol (DES) can antagonise ES cell self-renewal. It has been demonstrated that Nanog over-expression can over-ride the differentiation of *Esrrb* knock-down ES cells although the colonies obtained are only weakly alkaline phosphatase positive (Ivanova *et al.*, 2006). To further examine whether Esrrb and Nanog are functionally linked, one could examine the effect of the Esrrb antagonist

DES on LIF independent self-renewal of *Nanog* over-expressing cells. Additional experiments could determine if *Esrrb*$^{-/-}$ ES cells can be derived from blastocysts and whether *Nanog* over-expression can still guide LIF independent self-renewal of *Esrrb*$^{-/-}$ ES cells. As Nanog and Oct4 are suggested to share a common subset of partner proteins, it may also be interesting to examine a possible Oct4-Esrrb interaction using the (Flag)$_3$Oct4 expressing cells.

5.11.3.3   Nanog-HDAC2 interaction

Histone deacetylase 2 (HDAC2) catalyses removal of acetyl group modifications from histone tails (Taunton *et al.*, 1996). Acetylation of histone tails is generally associated with gene activation, therefore removing the acetyl group acts to repress genes. HDAC2 was identified as a Nanog co-purifying protein in two of the BIO purifications (MS2+MS4) in this thesis and also in the Wang *et al* (2006) study although they did not confirm this by co-immunoprecipitation. Later BIO Nanog purification (MS3+MS4) revealed HDAC2 was also found in the background fraction. As the function of transcription factors are often intimately linked to chromatin modifications, the putative Nanog-HDAC2 complex was probed further. A preliminary co-immunoprecipitation experiment revealed the presence of HDAC2 protein in Nanog containing complexes that had been immunoprecipitated from mouse ES cell lysates (Figure 5.9). It will be important in the future to further confirm this interaction and assay deacetylase activity of Nanog immunoprecipitates. Wang *et al* (2006) suggest Nanog interacts with the NuRD complex via Nac1 and Sall4, and although this may be the case, one should not rule out that it may act in other HDAC2 containing complexes such as the Sin3a/Mad-Max complex which also act to repress transcription (Laherty *et al.*, 1997). Indeed,

affinity purification of two NuRD complex components followed by mass-spectrometry does not identify Nanog as a NuRD associated protein suggesting it may be present in an alternative HDAC2 containing complex(es) (I.Costello, personal communication). To dissect which HDAC2 complex Nanog is present in, Nanog could be immunoprecipitated from ES cell lysates and probed with antibodies against proteins present in a distinct repressor complex e.g. Sin3a, or Mi2 for the NurD complex. A possible mechanism of Nanog-HDAC2 is that Nanog could recruit HDAC2 to generate repressive chromatin at lineage specific genes during the establishment of the pluripotent state. It may seem an apparent paradox that Nanog should interact with Wdr5 (associated with gene activation) and HDAC2 (associated with gene repression) but there are examples of transcription factors functioning in distinct complexes to effect both activation and repression, for example the haematopoetic transcription factor Gata1 (Rodriguez *et al.*, 2005).

5.11.3.4   Nac1 and Zfp281

Nac1 is a BTB/POZ transcriptional repressor previously found to be important in preventing neural gene expression (Mackler *et al.*, 2000). Nac1 was identified in MS2 and MS4 in this thesis, with two and three Nac1 co-purifying tryptic peptides being identified, respectively, with zero peptides detected in *BirA* control purifications. However due to time constraints, the Nac1 interaction was not followed up by co-immunoprecipitation experiments. Nac1 was also identified in the published single step purifications of Nanog partner proteins and validated by co-immunoprecipitation experiments (Wang *et al.*, 2006). Furthermore, Wang *et al* (2006) performed ChIP experiments showing Nanog and Nac1 together with a third protein Zfp281, bound the

Gata6 promoter, and that ES cells either heterozygous or knocked down for Zfp281 and Nac1 have increased *Gata6* expression, although the extra-embryonic differentiation that occurs upon forced *Gata6* expression (Fujikura *et al.*, 2002) does not occur. It has been hypothesised that Nanog functions to repress *Gata6* (Mitsui *et al.*, 2003; Ralston and Rossant, 2005; Chambers, 2004). In addition, *Nanog* over-expressing cells cannot differentiate into *Gata6* expressing primitive endoderm cells in embryoid bodies (Hamazaki *et al.*, 2004). Wang *et al* (2006) provide evidence that Nanog binds the *Gata6* promoter and it may be that Nac1 and Zfp281 are critical proteins in this repression. However, there still remains no evidence that Nanog is directly required for *Gata6* repression. Indeed, in contradiction to this hypothesis, recent data shows that continued *Nanog* is not required for maintenance of pluripotency; *Gata6* derepression and endodermal differentiation do not ensue upon acute *Nanog* deletion from ES cells (Chambers unpublished). It may be that a simultaneous reduction in the expression levels of Nanog plus one of either Nac1 or Zfp281 (or both) is required for *Gata6* derepression and endodermal differentiation.

5.11.3.5   Biotinylation tagging approach to Nanog partner identification.

The BIO Nanog purifications performed in this thesis have revealed a number of candidate partner proteins including Wdr5, HDAC2, Esrrb, and Nac1. Preliminary co-immunoprecipitations in this thesis have shown that Wdr5, HDAC2 and Esrrb may be Nanog interactors, although further more rigorous confirmation is required. Furthermore, Sall4 was identified in MS4 as highly enriched compared to the background (14 peptides versus 3 peptides), and this published interaction (Wu *et al.*, 2006) is

confirmed in chapter 3 of this thesis. During the course of this study, Wang *et al* (2006) performed a similar biotin tagging strategy in which they identified the molecules above along with an additional cohort of putative Nanog partner proteins, Oct4, Zfp281, Rif1 and Dax1 by mass-spectrometry followed by preliminary confirmation by co-immunoprecipitation (Wang *et al.*, 2006). Wang *et al* (2006) performed a tandem affinity BIO:Flag Nanog purification in addition to single step BIO Nanog purifications. It is interesting to note that seven out of seventeen putative partner proteins including Oct4 and Zfp281, were not detected by mass-spectrometry analysis of tandemly purified Nanog complexes. This may suggest that these interactions are weak or transient in nature. A further group of putative partner proteins were included in the Nanog "mini-interactome" without any validation of the mass-spectrometry data. Given the abundance of background binding proteins, and the fact that many of the putative specific partners found by Wang *et al* (2006) were found in the background fraction of experiments performed in this thesis (see Table 5.3), caution should be taken when drawing the conclusion that these are genuine physiologically meaningful interactions. To further increase the likelihood of identifying Nanog partner proteins using such a proteomic screen, future experiments could have a second purification step introduced to effectively use a tandem affinity tagging approach, or to cleave the epitope tag from the solid phase support to release only Nanog itself, and Nanog containing complexes. Differences between the screen in this thesis and that of the Wang *et al* (2006) study could also be due in part to different salt concentrations (350mM versus 100mM in this thesis) used for extract binding to streptavidin. It may also be useful to perform a parallel and complementary screen using a yeast-2-hybrid approach. Although this approach may suffer problems of high false positive rate and possible mis-modification

Table 5.3- Comparison of the proteins categorised as Nanog partner proteins by Wang *et al* (2006) to those identified by the BIO Nanog purifications in this thesis. Numbers indicate the total number of peptides identified.

	Wang *et al* (2006) MS1+MS2+MS3		This thesis MS1+MS2+MS3+MS4	
	BIO Nanog	BirA parental	BIO Nanog	BirA parental
Nanog	**4+7+8**	**0+0+0**	**12+9+2+9**	**0+0+0+0**
Sall1	3+2+1	0+0+0	0+0+0+0	0+0+0+4
Sall4	4+2+14	0+0+0	0+0+0+7	0+0+0+3
Rif1	12+5+5	0+0+0	0+8+0+6	0+1+0+5
Tif1β	11+16+13	0+0+0	23+7+1+48	22+7+1+48
Mybbp	8+14+15	0+0+0	11+28+5+18	4+33+4+13
Dax1	0+1+2	0+0+0	0+0+0+0	0+0+0+0
Nac1	2+0+2	0+0+0	0+2+0+2	0+0+0+0
Zfp281	1+2+4	0+0+0	0+0+0+4	0+0+1+11
Esrrb	7+2+1	0+0+0	0+4+0+7	0+0+0+2
Elys	1+2+1	0+0+0	5+7+0+16	6+0+0+22
Oct-4	2+0+1	0+0+0	0+0+0+4	0+0+0+1
Zfp198	0+2+5	0+0+0	0+0+0+4	3+0+0+14
NF45	4+0+1	0+0+0	0+0+0+0	0+0+0+0
HDAC2	2+0+1	0+0+0	0+1+0+5	0+0+1+7
REST	0+0+1	0+0+0	0+0+0+0	1+0+0+0
Sp1	0+1+0	0+0+0	0+0+0+0	0+0+0+0
Wapl	0+0+1	0+0+0	0+0+0+0	0+0+0+0
Wdr5	1+0+0	0+0+0	3+0+0+2	0+0+0+0

of mouse proteins, it may identify direct protein-protein interactions, which are likely to be the most important and functionally relevant. In addition to the *BirA:BIO Nanog* ES cell line described here, a *BirA: BIO Nanog* co-expressing neural stem (NS) cell line has been derived (data not shown) (Conti *et al.*, 2005). As NS cells do not express *Nanog*, this cell line will be a useful tool to identify which Nanog partners only interact with Nanog in an ES cell context and not in NS cells, as these are likely to be functionally significant.

5.12   Summary

This chapter has described experiments to design and generate a reagent that facilitates the purification of Nanog containing complexes from mouse ES cells. An inducible BIO Nanog construct was made, *BirA:BIO Nanog* expressing ES cell lines generated, and pilot experiments to characterise the purification efficiency were performed. Four large scale BIO Nanog purifications were performed and analysed via mass-spectrometry to identify associated proteins. The data revealed that the system could not reproducibly identify specific co-purifying proteins, however three proteins were further investigated based on their being present (either specifically or enriched) in BIO Nanog purifications, and there being biological rationale for these interactions. These proteins, Wdr5, Esrrb, and HDAC2 were preliminarily confirmed as Nanog partners by co-immunoprecipitation experiments. In the future it will be important to perform further confirmation of these interactions and begin functional analysis of these proteins in ES cells.

Chapter 6

Concluding Remarks

Nanog is a divergent homeodomain protein that is required to establish pluripotent cell types and is able to direct ES cell self-renewal in the absence of cytokine signalling. However, the mechanism by which Nanog acts in ES cells is incompletely understood. This thesis has presented experiments that begin biochemical characterisation of Nanog protein, in addition to experiments addressing the consequence of *Nanog* over-expression on mesoderm formation in the mouse embryo.

It has been demonstrated that Nanog can form multimers in ES cells and that the tryptophan repeat in the C-terminal domain of Nanog is required to mediate this interaction. That such multimerisation is functionally significant has been demonstrated by the inability of a mutant Nanog molecule lacking the tryptophan repeat to efficiently direct ES cell self-renewal.

During normal mouse embryogenesis, *Nanog* expression in the post-implantation embryo is highest in the proximal posterior region of the epiblast and is rapidly down-regulated as cells delaminate and ingress through the primitive streak. Cellular reagents were designed and generated to address the consequence of *Nanog* over-expression in the gastrulating mouse embryo. It was found that an elevation of Nanog protein to 2-3 times the endogenous level does not cause an overt phenotype at this stage of mouse

development, and *Nanog* over-expressing cells can be clearly visualised in the mesoderm. Therefore *Nanog* down-regulation is not required for mesoderm differentiation.

Experiments designed to investigate potential interactions between Nanog and the candidate molecules Oct4 and Stat3 did not establish any such physical link. Therefore an unbiased proteomics approach was employed to identify Nanog partner proteins in ES cells. A biotinylation tagging system was established, allowing biotinylated Nanog and associated proteins to be purified from ES cell nuclear extracts. Analysis of the purified material by mass-spectrometry identified putative Nanog partner proteins. Preliminary biochemical confirmation of three of these identified proteins, Esrrb, HDAC2, and Wdr5 was performed. An interaction between Nanog and the spalt family member Sall4 was confirmed and found to be mediated by the SLQQ motif within the Nanog homeodomain. Furthermore, mutation of the SLQQ motif to SAAQ was shown to compromise ES cell growth. The only other protein containing an SLQQ at the same position within the homeodomain is Oct4. Consistent with a fundamental role of this motif in directing efficient ES cell self-renewal, a physical interaction between Oct4 and Sall4 was demonstrated.

References

Ambrosetti, D.-C., Basilico, C., and Dailey, L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by sox2 and oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *MCB* **17**, 6321-6329.

Andrews, P. W. (2002). From teratocarcinomas to embryonic stem cells. *Philos Trans R Soc Lond B Biol Sci* **357**, 405-417.

Antonchuk, J., Sauvageau, G., and Humphries, R. K. (2002). HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. *Cell* **109**, 39-45.

Arnold, S. J., Stappert, J., Bauer, A., Kispert, A., Herrmann, B. G., and Kemler, R. (2000). Brachyury is a target gene of the Wnt/beta-catenin signaling pathway. *Mech Dev* **91**, 249-258.

Aubert, J., Dunstan, H., Chambers, I., and Smith, A. (2002). Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nat Biotechnol* **20**, 1240-1245.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126-140.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., and Fisher, A. G. (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**, 532-538.

Beckett, D., Kovaleva, E., and Schatz, P. J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* **8**, 921-929.

Beddington, R. S., and Robertson, E. J. (1999). Axis development and early asymmetry in mammals. *Cell* **96**, 195-209.

Beddington, R. S. P., and Robertson, E. J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733-737.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326.

Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W., and Wuthrich, K. (1993). Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. *J Mol Biol* **234**, 1084-1093.

Booth, H. A., and Holland, P. W. (2004). Eleven daughters of NANOG. *Genomics* **84**, 229-238.

Boulton, T. G., Stahl, N., and Yancopoulos, G. D. (1994). Ciliary neurotrophic factor/leukemia inhibitory factor/interleukin-6/oncostatin M family of cytokines induces tyrosine phosphorylation of a common set of proteins overlapping those induced by other cytokines and growth factors. *JBC* **269**, 11648-11655.

Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P. O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., *et al.* (2004). A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat Cell Biol* **6**, 97-105.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-956.

Bradley, A., Evans, M. J., Kaufman, M. H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255-256.

Brinster, R. L. (1974). The effect of cells transferred into the mouse blastocyst on subsequent development. *J Exp Med* **140**, 1049-1056.

Brook, F. A., and Gardner, R. L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. *PNAS* **94**, 5709-5712.

Buehr, M., Nichols, J., Stenhouse, F., Mountford, P., Greenhalgh, C. J., Kantachuvesiri, S., Brooker, G., Mullins, J., and Smith, A. G. (2003). Rapid loss of Oct-4 and pluripotency in cultured rodent blastocysts and derivative cell lines. *Biol Reprod* **68**, 222-229.

Buehr, M., and Smith, A. (2003). Genesis of embryonic stem cells. *Philos Trans R Soc Lond B Biol Sci* **358**, 1397-1402; discussion 1402.

Burdon, T., Smith, A., and Savatier, P. (2002). Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol* **12**, 432.

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.

Canon, S., Herranz, C., and Manzanares, M. (2006). Germ cell restricted expression of chick Nanog. *Dev Dyn*.

Carlin, R., Davis, D., Weiss, M., Schultz, B., and Troyer, D. (2006). Expression of early transcription factors Oct4, Sox2 and Nanog by porcine umbilical cord (PUC) matrix cells. *Reprod Biol Endocrinol* **4**, 8.

Chambers, I. (2004). The molecular basis of pluripotency in mouse embryonic stem cells. *Cloning and Stem Cells* **6**, 386-391.

Chambers, I. (2005). Mechanisms and factors in embryonic stem cell self-renewal. *Rend Fis Acc Lincei* **9**, 83-87.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-655.

Chambers, I., and Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**, 7150-7160.

Chapman-Smith, A., and Cronan, J. E., Jr. (1999). The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends Biochem Sci* **24**, 359-363.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell* **10**, 615-624.

Chew, J. L., Loh, Y. H., Zhang, W., Chen, X., Tam, W. L., Yeap, L. S., Li, P., Ang, Y. S., Lim, B., Robson, P., and Ng, H. H. (2005). Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Mol Cell Biol* **25**, 6031-6046.

Conlon, F. L., Fairclough, L., Price, B. M., Casey, E. S., and Smith, J. C. (2001). Determinants of T box protein specificity. *Development* **128**, 3749-3758.

Conti, L., Pollard, S. M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q. L., Cattaneo, E., and Smith, A. (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol* **3**, e283.

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.

Daheron, L., Opitz, S. L., Zaehres, H., Lensch, W. M., Andrews, P. W., Itskovitz-Eldor, J., and Daley, G. Q. (2004). LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* **22**, 770-778.

Dani, C., Chambers, I., Johnstone, S., Robertson, M., Ebrahimi, B., Saito, M., Taga, T., Li, M., Burdon, T., Nichols, J., and Smith, A. (1998). Paracrine induction of stem cell renewal by LIF-deficient cells: a new ES cell regulatory pathway. *Dev Biol* **203**, 149-162.

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.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A* **100**, 7480-7485.

Diwan, S. B., and Stevens, L. C. (1976). Development of teratomas from ectoderm of mouse egg cylinders. *JNatlCancer Inst* **57**, 937-942.

Do, H. J., Lim, H. Y., Kim, J. H., Song, H., and Chung, H. M. (2006). An intact homeobox domain is required for complete nuclear localization of human Nanog. *Biochem Biophys Res Commun.*

Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W., and Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *JEmbryolExpMorphol* **87**, 27-45.

Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121**, 873-885.

Drakas, R., Prisco, M., and Baserga, R. (2005). A modified tandem affinity purification tag technique for the purification of protein complexes in mammalian cells. *Proteomics* **5**, 132-137.

Driegen, S., Ferreira, R., van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., and Meijer, D. (2005). A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Res* **14**, 477-482.

Elling, U., Klasen, C., Eisenberger, T., Anlag, K., and Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. *Proc Natl Acad Sci U S A* **103**, 16319-16324.

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

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246.

Finch, B. W., and Ephrussi, B. (1967). Retention of multiple developmental potentialities by cells of a mouse testicular teratocarcinoma during prolonged culture in vitro and their extinction upon hybridisation with cells of permanent lines. *Proc NatlAcad Sci USA* **57**, 615-621.

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**, 784-789.

Galceran, J., Hsu, S. C., and Grosschedl, R. (2001). Rescue of a Wnt mutation by an activated form of LEF-1: regulation of maintenance but not initiation of Brachyury expression. *Proc Natl Acad Sci U S A* **98**, 8668-8673.

Gardner, R. L. (1983). Origin and differentiation of extra-embryonic tissues in the mouse. *IntRevExpPath* **24**, 63-133.

Gassmann, M., Donoho, G., and Berg, P. (1995). Maintenance of an extrachromosomal plasmid vector in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **92**, 1292-1296.

Gehring, W. J. (1987). Homeo boxes in the study of development. *Science* **236**, 1245-1252.

Gill, G. (2005). Something about SUMO inhibits transcription. *Curr Opin Genet Dev* **15**, 536-541.

Gu, P., Lemenuet, D., Chung, A. C., Mancini, M., Wheeler, D. A., and Cooney, A. J. (2005). Orphan Nuclear Receptor GCNF Is Required for the Repression of Pluripotency Genes during Retinoic Acid-Induced Embryonic Stem Cell Differentiation. *Mol Cell Biol* **25**, 8507-8519.

Guo, Y., Costa, R., Ramsey, H., Starnes, T., Vance, G., Robertson, K., Kelley, M., Reinbold, R., Scholer, H., and Hromas, R. (2002). The embryonic stem cell transcription factors Oct-4 and FoxD3 interact to regulate endodermal-specific promoter expression. *Proc Natl Acad Sci U S A* **99**, 3663-3667.

Hamazaki, T., Kehoe, S. M., Nakano, T., and Terada, N. (2006). The Grb2/Mek pathway represses Nanog in murine embryonic stem cells. *Mol Cell Biol*.

Hamazaki, T., Oka, M., Yamanaka, S., and Terada, N. (2004). Aggregation of embryonic stem cells induces Nanog repression and primitive endoderm differentiation. *J Cell Sci* **117**, 5681-5686.

Han, Z., Guo, L., Wang, H., Shen, Y., Deng, X. W., and Chai, J. (2006). Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. *Mol Cell* **22**, 137-144.

Hanna, L. A., Foreman, R. K., Tarasenko, I. A., Kessler, D. S., and Labosky, P. A. (2002). Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev* **16**, 2650-2661.

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.

Hegde, S. S., Vetting, M. W., Roderick, S. L., Mitchenall, L. A., Maxwell, A., Takiff, H. E., and Blanchard, J. S. (2005). A fluoroquinolone resistance protein from Mycobacterium tuberculosis that mimics DNA. *Science* **308**, 1480-1483.

Humphrey, R. K., Beattie, G. M., Lopez, A. D., Bucay, N., King, C. C., Firpo, M. T., Rose-John, S., and Hayek, A. (2004). Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. *Stem Cells* **22**, 522-530.

Ihle, J. (1996). STATs: signal transducers and activators of transcription. *Cell* **84**, 331-334.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I. R. (2006). Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-538.

Johnson, M. H., and Ziomek, C. A. (1981). The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71-80.

Kanatsu-Shinohara, M., Inoue, K., Lee, J., Yoshimoto, M., Ogonuki, N., Miki, H., Baba, S., Kato, T., Kazuki, Y., Toyokuni, S., *et al.* (2004). Generation of pluripotent stem cells from neonatal mouse testis. *Cell* **119**, 1001-1012.

Kasahara, H., and Izumo, S. (1999). Identification of the in vivo casein kinase II phosphorylation site within the homeodomain of the cardiac tisue-specifying homeobox gene product Csx/Nkx2.5. *Mol Cell Biol* **19**, 526-536.

Keegan, L., Gill, G., and Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231**, 699-704.

Kirito, K., Fox, N., and Kaushansky, K. (2004). Thrombopoietin induces HOXA9 nuclear transport in immature hematopoietic cells: potential mechanism by which the hormone favorably affects hematopoietic stem cells. *Mol Cell Biol* **24**, 6751-6762.

Kleinsmith, L. J., and Pierce, G. B. (1964). Multipotentiality of single embryonal carcinoma cells. *Cancer Res* **24**, 1544-1552.

Klump, H., Schiedlmeier, B., and Baum, C. (2005). Control of self-renewal and differentiation of hematopoietic stem cells: HOXB4 on the threshold. *Ann N Y Acad Sci* **1044**, 6-15.

Knuesel, M., Wan, Y., Xiao, Z., Holinger, E., Lowe, N., Wang, W., and Liu, X. (2003). Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. *Mol Cell Proteomics* **2**, 1225-1233.

Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R., and Grosveld, F. (1999). The transcription factor GATA6 is essential for early extraembryonic development. *Development* **126**, 723-732.

Kuhnlein, R. P., Frommer, G., Friedrich, M., Gonzalez-Gaitan, M., Weber, A., Wagner-Bernholz, J. F., Gehring, W. J., Jackle, H., and Schuh, R. (1994). spalt encodes an evolutionarily conserved zinc finger protein of novel structure which provides homeotic gene function in the head and tail region of the Drosophila embryo. *Embo J* **13**, 168-179.

Kunath, T., Arnaud, D., Uy, G. D., Okamoto, I., Chureau, C., Yamanaka, Y., Heard, E., Gardner, R. L., Avner, P., and Rossant, J. (2005). Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. *Development* **132**, 1649-1661.

Kunath, T., Gish, G., Lickert, H., Jones, N., Pawson, T., and Rossant, J. (2003). Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nat Biotechnol* **21**, 559-561.

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

Labosky, P. A., Barlow, D. P., and Hogan, B. L. (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* **120**, 3197-3204.

Laherty, C. D., Yang, W. M., Sun, J. M., Davie, J. R., Seto, E., and Eisenman, R. N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell* **89**, 349-356.

Lane, D. P., and Crawford, L. V. (1979). T antigen is bound to a host protein in SV40-transformed cells. *Nature* **278**, 261-263.

Lee, J. W., and Lee, S. K. (2004). Mammalian two-hybrid assay for detecting protein-protein interactions in vivo. *Methods Mol Biol* **261**, 327-336.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science* **270**, 262-269.

Lin, T., Chao, C., Saito, S., Mazur, S. J., Murphy, M. E., Appella, E., and Xu, Y. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nat Cell Biol* **7**, 165-171.

Lints, T. J., Parsons, L. M., Hartley, L., Lyons, I., and Harvey, R. P. (1993). Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development* **119**, 419-431.

Linzer, D. I., and Levine, A. J. (1979). Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* **17**, 43-52.

Liu, S., Qu, Y., Stewart, T. J., Howard, M. J., Chakrabortty, S., Holekamp, T. F., and McDonald, J. W. (2000). Embryonic stem cells differentiate into oligodendrocytes and myelinate in culture and after spinal cord transplantation. *Proc Natl Acad Sci U S A* **97**, 6126-6131.

Liu, X., Sun, Y., Constantinescu, S. N., Karam, E., Weinberg, R. A., and Lodish, H. F. (1997). Transforming growth factor beta-induced phosphorylation of Smad3 is required for growth inhibition and transcriptional induction in epithelial cells. *Proc Natl Acad Sci U S A* **94**, 10669-10674.

Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet.*

Luban, J., and Goff, S. P. (1995). The yeast two-hybrid system for studying protein-protein interactions. *Curr Opin Biotechnol* **6**, 59-64.

Luo, J., Sladek, R., Bader, J. A., Matthyssen, A., Rossant, J., and Giguere, V. (1997). Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. *Nature* **388**, 778-782.

Lutticken, C., Wegenka, U. M., Yuan, J., Buschmann, J., Schindler, C., Ziemiecki, A., Harpur, A. G., Wilks, A. F., Yasukawa, K., Taga, T., *et al.* (1994). Association of transcription factor APRF and protein kinase Jak1 with the interleukin-6 signal transducer gp130. *Science* **263**, 89-92.

Mackler, S. A., Korutla, L., Cha, X. Y., Koebbe, M. J., Fournier, K. M., Bowers, M. S., and Kalivas, P. W. (2000). NAC-1 is a brain POZ/BTB protein that can prevent cocaine-induced sensitization in the rat. *J Neurosci* **20**, 6210-6217.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA* **78**, 7634-7638.

Massague, J., and Wotton, D. (2000). Transcriptional control by the TGF-beta/Smad signaling system. *Embo J* **19**, 1745-1754.

Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M. G., Sharov, A. A., and Ko, M. S. (2006). Dissecting oct3/4-regulated gene networks in embryonic stem cells by expression profiling. *PLoS ONE* **1**, e26.

Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., and Yokota, T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *Embo J* **18**, 4261-4269.

Matsui, Y., Zsebo, K., and Hogan, B. L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841-847.

Meier, N., Krpic, S., Rodriguez, P., Strouboulis, J., Monti, M., Krijgsveld, J., Gering, M., Patient, R., Hostert, A., and Grosveld, F. (2006). Novel binding partners of Ldb1 are required for haematopoietic development. *Development* **133**, 4913-4923.

Merrill, B. J., Pasolli, H. A., Polak, L., Rendl, M., Garcia-Garcia, M. J., Anderson, K. V., and Fuchs, E. (2004). Tcf3: a transcriptional regulator of axis induction in the early embryo. *Development* **131**, 263-274.

Miller, R. A., and Ruddle, F. H. (1977). Teratocarcinoma X friend erythroleukemia cell hybrids resemble their pluripotent embryonal carcinoma parent. *Dev Biol* **56**, 157-173.

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

Mitsunaga, K., Araki, K., Mizusaki, H., Morohashi, K., Haruna, K., Nakagata, N., Giguere, V., Yamamura, K., and Abe, K. (2004). Loss of PGC-specific expression of the orphan nuclear receptor ERR-beta results in reduction of germ cell number in mouse embryos. *Mech Dev* **121**, 237-246.

Monti, M., Orru, S., Pagnozzi, D., and Pucci, P. (2005a). Functional proteomics. *Clin Chim Acta* **357**, 140-150.

Monti, M., Orru, S., Pagnozzi, D., and Pucci, P. (2005b). Interaction proteomics. *Biosci Rep* **25**, 45-56.

Morkel, M., Huelsken, J., Wakamiya, M., Ding, J., van de Wetering, M., Clevers, H., Taketo, M. M., Behringer, R. R., Shen, M. M., and Birchmeier, W. (2003). Beta-catenin

regulates Cripto- and Wnt3-dependent gene expression programs in mouse axis and mesoderm formation. *Development* **130**, 6283-6294.

Morrisey, E. E., Ip, H. S., Lu, M. M., and Parmacek, M. S. (1996). GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev Biol* **177**, 309-322.

Mountford, P., Zevnik, B., Duwel, A., Nichols, J., Li, M., Dani, C., Robertson, M., Chambers, I., and Smith, A. (1994). Dicistronic targeting constructs: reporters and modifiers of mammalian gene expression. *Proc Natl Acad Sci USA* **91**, 4303-4307.

Nagy, A. G., M.; Vintersten, K. (2002). Manipulating the mouse embryo: A laboratory guide, 3 edn).

Nagy, A. R., J. (2001). Gene targetting: a practical approach 2nd Edition).

Nakashima, K., Yanagisawa, M., Arakawa, H., Kimura, N., Hisatsune, T., Kawabata, M., Miyazono, K., and Taga, T. (1999). Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. *Science* **284**, 479-482.

Narazaki, M., Witthuhn, B. A., Yoshida, K., Silvennoinen, O., Yasukawa, K., Ihle, J. N., Kishimoto, T., and Taga, T. (1994). Activation of JAK2 kinase mediated by the interleukin 6 signal transducer gp130. *PNAS* **91**, 2285-2289.

Nichols, J., Chambers, I., and Smith, A. (1994). Derivation of germline competent embryonic stem cells with a combination of interleukin-6 and soluble interleukin-6 receptor. *Exp Cell Res* **215**, 237-239.

Nichols, J., Chambers, I., Taga, T., and Smith, A. (2001). Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Development* **128**, 2333-2339.

Nichols, J., Davidson, D., Taga, T., Yoshida, K., Chambers, I., and Smith, A. (1996). Complementary tissue-specific expression of LIF and LIF-receptor mRNAs in early mouse embryogenesis. *Mech Dev* **57**, 123-131.

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

Nishimoto, M., Fukushima, A., Okuda, A., and Muramatsu, M. (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol* **19**, 5453-5465.

Nishimoto, M., Miyagi, S., Yamagishi, T., Sakaguchi, T., Niwa, H., Muramatsu, M., and Okuda, A. (2005). Oct-3/4 maintains the proliferative embryonic stem cell state via

specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. *Mol Cell Biol* **25**, 5084-5094.

Niwa, H. (2001). Molecular mechanism to maintain stem cell renewal of ES cells. *Cell Struct Funct* **26**, 137-148.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12**, 2048-2060.

Niwa, H., Masui, S., Chambers, I., Smith, A. G., and Miyazaki, J. (2002). Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. *Mol Cell Biol* **22**, 1526-1536.

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**, 917-929.

Niwa, H., Yamamura, K.-I., and Miyazaki, J.-I. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193-200.

Oh, J. H., Do, H. J., Yang, H. M., Moon, S. Y., Cha, K. Y., Chung, H. M., and Kim, J. H. (2005). Identification of a putative transactivation domain in human Nanog. *Exp Mol Med* **37**, 250-254.

Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005). Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *JBC* **280**, 5307-5317.

Paling, N. R., Wheadon, H., Bone, H. K., and Welham, M. J. (2004). Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. *JBC* **279**, 48063-48070.

Palmieri, S. L., Peter, W., Hess, H., and Scholer, 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**, 259-267.

Pan, G., Li, J., Zhou, Y., Zheng, H., and Pei, D. (2006). A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *Faseb J* **20**, 1730-1732.

Pan, G., and Pei, D. (2005). The stem cell pluripotency factor NANOG activates transcription with two unusually potent subdomains at its C terminus. *JBC* **280**, 1401-1407.

Pan, G. J., and Pei, D. Q. (2003). Identification of two distinct transactivation domains in the pluripotency sustaining factor nanog. *Cell Res* **13**, 499-502.

Parrott, M. B., and Barry, M. A. (2000). Metabolic biotinylation of recombinant proteins in mammalian cells and in mice. *Mol Ther* **1**, 96-104.

Parrott, M. B., and Barry, M. A. (2001). Metabolic biotinylation of secreted and cell surface proteins from mammalian cells. *Biochem Biophys Res Commun* **281**, 993-1000.

Pedram, M., Sprung, C. N., Gao, Q., Lo, A. W., Reynolds, G. E., and Murnane, J. P. (2006). Telomere position effect and silencing of transgenes near telomeres in the mouse. *Mol Cell Biol* **26**, 1865-1878.

Peng, J., and Gygi, S. P. (2001). Proteomics: the move to mixtures. *J Mass Spectrom* **36**, 1083-1091.

Pereira, L., Yi, F., and Merrill, B. J. (2006). Repression of Nanog Gene Transcription by Tcf3 Limits Embryonic Stem Cell Self-Renewal. *Mol Cell Biol*.

Pesce, M., Wang, X., Wolgemuth, D. J., and Scholer, H. (1998). Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. *Mech Dev* **71**, 89-98.

Pettersson, K., Svensson, K., Mattsson, R., Carlsson, B., Ohlsson, R., and Berkenstam, A. (1996). Expression of a novel member of estrogen response element-binding nuclear receptors is restricted to the early stages of chorion formation during mouse embryogenesis. *Mech Dev* **54**, 211-223.

Pratt, T., Sharp, L., Nichols, J., Price, D. J., and Mason, J. O. (2000). Embryonic stem cells and transgenic mice ubiquitously expressing a tau tagged green fluorescent protein. *Dev Biol* **228**, 19-28.

Prelle, K., Zink, N., and Wolf, E. (2002). Pluripotent stem cells--model of embryonic development, tool for gene targeting, and basis of cell therapy. *Anat Histol Embryol* **31**, 169-186.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218-229.

Qi, X., Li, T. G., Hao, J., Hu, J., Wang, J., Simmons, H., Miura, S., Mishina, Y., and Zhao, G. Q. (2004). BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proc Natl Acad Sci U S A* **101**, 6027-6032.

Ralston, A., and Rossant, J. (2005). Genetic regulation of stem cell origins in the mouse embryo. *Clin Genet* **68**, 106-112.

Ramos-Mejia, V., Escalante-Alcalde, D., Kunath, T., Ramirez, L., Gertsenstein, M., Nagy, A., and Lomeli, H. (2005). Phenotypic analyses of mouse embryos with ubiquitous expression of Oct4: effects on mid-hindbrain patterning and gene expression. *Dev Dyn* **232**, 180-190.

Rappsilber, J., and Mann, M. (2002). What does it mean to identify a protein in proteomics? *Trends Biochem Sci* **27**, 74-78.

Rathjen, P. D., Toth, S., Willis, A., Heath, J. K., and Smith, A. G. (1990). Differentiation inhibiting activity is produced in matrix-associated and diffusible forms that are generated by alternate promoter usage. *Cell* **62**, 1105-1114.

Reed, R., Griffith, J., and Maniatis, T. (1988). Purification and visualization of native spliceosomes. *Cell* **53**, 949-961.

Remenyi, A., Lins, K., Nissen, L. J., Reinbold, R., Scholer, H. R., and Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17**, 2048-2059.

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**, 550-551.

Reubinoff, B. E., Pera, M. F., Fong, C. Y., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotechnol* **18**, 399-404.

Reya, T., Duncan, A. W., Ailles, L., Domen, J., Scherer, D. C., Willert, K., Hintz, L., Nusse, R., and Weissman, I. L. (2003). A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409-414.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17**, 1030-1032.

Robertson, M., Stenhouse, F., Colby, D., Marland, J. R., Nichols, J., Tweedie, S., and Chambers, I. (2006). Nanog retrotransposed genes with functionally conserved open reading frames. *Mamm Genome* **17**, 732-743.

Robinson-Rechavi, M., Escriva Garcia, H., and Laudet, V. (2003). The nuclear receptor superfamily. *J Cell Sci* **116**, 585-586.

Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B., Ng, H. H., and Robson, P. (2005). Transcriptional Regulation of Nanog by OCT4 and SOX2. *JBC* **280**, 24731-24737.

Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K. E., Guyot, B., Heck, A. J., Vyas, P., de Boer, E., Grosveld, F., and Strouboulis, J. (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. *Embo J* **24**, 2354-2366.

Rodriguez, P., Braun, H., Kolodziej, K. E., de Boer, E., Campbell, J., Bonte, E., Grosveld, F., Philipsen, S., and Strouboulis, J. (2006). Isolation of transcription factor complexes by in vivo biotinylation tagging and direct binding to streptavidin beads. *Methods Mol Biol* **338**, 305-323.

Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M., Poirier, F., Rigby, P., and Staudt, L. M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**, 686-692.

Rybak, J. N., Ettorre, A., Kaissling, B., Giavazzi, R., Neri, D., and Elia, G. (2005). In vivo protein biotinylation for identification of organ-specific antigens accessible from the vasculature. *Nat Methods* **2**, 291-298.

Saijoh, Y., Fukii, H., Meno, C., Sato, M., Hirota, Y., Nagamatsu, S., Ikeda, M., and Hamada, H. (1996). Identification of putative downstream genes of Oct-3, a pluripotent cell-specific transcription factor. *Genes to Cells* **1**, 239-252.

Sakaki-Yumoto, M., Kobayashi, C., Sato, A., Fujimura, S., Matsumoto, Y., Takasato, M., Kodama, T., Aburatani, H., Asashima, M., Yoshida, N., and Nishinakamura, R. (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* **133**, 3005-3013.

Sambrook. (2001). Molecular Cloning: A Laboratory Manual, CSHL Press).

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-411.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A. H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* **10**, 55-63.

Sauvageau, G., Thorsteinsdottir, U., Eaves, C. J., Lawrence, H. J., Largman, C., Lansdorp, P. M., and Humphries, R. K. (1995). Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo. *Genes Dev* **9**, 1753-1765.

Scholer, H. R., Dressler, G. R., Balling, R., Rohdewohld, H., and Gruss, P. (1990a). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J* **9**, 2185-2195.

Scholer, H. R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990b). New type of POU domain in germ line-specific protein Oct-4. *Nature* **344**, 435-439.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278**, 440-458.

Sheng, Z., Knowlton, K., Chen, J., Hoshijima, M., Brown, J. H., and Chien, K. R. (1997). Cardiotrophin 1 (CT-1) inhibition of cardiac myocyte apoptosis via a mitogen activated protein kinase-dependent pathway. *JBC* **272**, 5783-5791.

Shi, W., Wang, H., Pan, G., Geng, Y., Guo, Y., and Pei, D. (2006). Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. *JBC* **281**, 23319-23325.

Silva, J., Chambers, I., Pollard, S., and Smith, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441**, 997-1001.

Sims, R. J., 3rd, Nishioka, K., and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. *Trends Genet* **19**, 629-639.

Smith, A. (2005). The battlefield of pluripotency. *Cell* **123**, 757-760.

Smith, A. G. (1991). Culture and differentiation of embryonic stem cells. *J Tiss Cult Meth* **13**, 89-94.

Smith, A. G. (2001). Embryo-derived stem cells: of mice and men. *Ann Rev Cell Dev Biol* **17**, 435-462.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-690.

Smith, A. G., and Hooper, M. L. (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev Biol* **121**, 1-9.

Smith, T. A., and Hooper, M. L. (1983). Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. *ExpCell Res* **145**, 458-462.

Solter, D., Skreb, N., and Damjanov, I. (1970). Extrauterine growth of mouse egg cylinders results in malignant teratoma. *Nature* **227**, 503-504.

Stahl, N., Boulton, T. G., Farruggella, T., Ip, N. Y., Davis, S., Witthuhn, B. A., Quelle, F. W., Silvennoinen, O., Barbieri, G., Pellegrini, S., *et al.* (1994). Association and activation of Jak-Tyk kinases by CNTF-LIF-OSM-IL-6 beta receptor components. *Science* **263**, 92-95.

Stahl, N., Farrugella, T. J., Boulton, T. G., Zhong, Z., Darnell, J. E., and Yancopoulos, G. D. (1995). Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* **267**, 1349-1353.

Stevens, L. C. (1964). Experimental Production of Testicular Teratomas in Mice. *Proc Natl Acad Sci U S A* **52**, 654-661.

Stevens, L. C. (1968). The development of teratomas from intratesticular grafts of tubal mouse eggs. *J Embryol Exp Morphol* **20**, 329-341.

Stevens, L. C., and Little, C. C. (1954). Sponataneous testicular teratomas in an inbred strain of mice. *PNAS* **40**, 1080-1087.

Stewart, C. L., Gadi, I., and Bhatt, H. (1994). Stem cells from primordial germ cells can reenter the germ line. *Dev Biol* **161**, 626-628.

Stewart, C. L., Kaspar, P., Brunet, L. J., Bhatt, H., Gadi, I., Kontgen, F., and Abbondanzo, S. J. (1992). Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature* **359**, 76-79.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* **403**, 41-45.

Strumpf, D., Mao, C. A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093-2102.

Surani, M. A., Hayashi, K., and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747-762.

Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F. H., Rodriguez-Esteban, C., and Belmonte, J. C. (2006a). Maintenance of embryonic stem cell pluripotency by Nanog-mediated reversal of mesoderm specification. *Nat Clin Pract Cardiovasc Med* **3 Suppl 1**, S114-122.

Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F. H., Rodriguez-Esteban, C., and Izpisua Belmonte, J. C. (2006b). Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. *Proc Natl Acad Sci U S A* **103**, 10294-10299.

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.

Tada, T., Tada, M., Hilton, K., Barton, S. C., Sado, T., Takagi, N., and Surani, M. A. (1998). Epigenotype switching of imprintable loci in embryonic germ cells. *Dev Genes Evol* **207**, 551-561.

Takahashi, K., Mitsui, K., and Yamanaka, S. (2003). Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature* **423**, 541-545.

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.

Takao, Y., Yokota, T., and Koide, H. (2006). beta-Catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. *Biochem Biophys Res Commun.*

Tam, P. P., and Behringer, R. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. *Mech Dev* **68**, 3-25.

Tanaka, S., Kunath, T., Hadjantonakis, A.-K., Nagy, A., and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF-4. *Science* **282**, 2072-2075.

Taunton, J., Hassig, C. A., and Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408-411.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **60**, 523-533.

Thomas, K. R., and Capecchi, M. R. (1987). Site directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-512.

Thompson, S., Clarke, A. R., Pow, A. M., Hooper, M. L., and Melton, D. W. (1989). Germ line transmission and expression of a corrected gene produced by gene targetting in embryonic stem cells. *Cell* **56**, 313-321.

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147.

Thorsteinsdottir, U., Sauvageau, G., and Humphries, R. K. (1999). Enhanced in vivo regenerative potential of HOXB4-transduced hematopoietic stem cells with regulation of their pool size. *Blood* **94**, 2605-2612.

Tokuzawa, Y., Kaiho, E., Maruyama, M., Takahashi, K., Mitsui, K., Maeda, M., Niwa, H., and Yamanaka, S. (2003). Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. *Mol Cell Biol* **23**, 2699-2708.

Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M., and Okuda, A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res* **30**, 3202-3213.

Tremblay, G. B., Kunath, T., Bergeron, D., Lapointe, L., Champigny, C., Bader, J. A., Rossant, J., and Giguere, V. (2001). Diethylstilbestrol regulates trophoblast stem cell differentiation as a ligand of orphan nuclear receptor ERR beta. *Genes Dev* **15**, 833-838.

Vernallis, A. B., Hudson, K. R., and Heath, J. K. (1997). An antagonist for the leukemia inhibitory factor receptor inhibits leukemia inhibitory factor, cardiotrophin-1, ciliary neurotrophic factor, and oncostatin M. *JBC* **272**, 26947-26952.

Vetting, M. W., Hegde, S. S., Fajardo, J. E., Fiser, A., Roderick, S. L., Takiff, H. E., and Blanchard, J. S. (2006). Pentapeptide repeat proteins. *Biochemistry* **45**, 1-10.

Vidal, M., and Legrain, P. (1999). Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res* **27**, 919-929.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W., and Orkin, S. H. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364-368.

Ware, C. B., Horowitz, M. C., Renshaw, B. R., Hunt, J. S., Liggit, D., Koblar, S. A., Gliniak, B. C., McKenna, H. J., Papayannopoulou, T., Thoma, B., *et al.* (1995). Targeted disruption of the low-affinity leukemia inhibitory factor receptor gene causes placental, skeletal, neural and metabolic defects and results in perinatal death. *Development* **121**, 1283-1299.

Williams, D. C., Jr., Cai, M., and Clore, G. M. (2004). Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex. *JBC* **279**, 1449-1457.

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**, 684-687.

Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell* **82**, 709-719.

Wolberger, C. (1996). Homeodomain interactions. *Curr Opin Struct Biol* **6**, 62-68.

Wu da, Y., and Yao, Z. (2005). Isolation and characterization of the murine Nanog gene promoter. *Cell Res* **15**, 317-324.

Wu, Q., Chen, X., Zhang, J., Loh, Y. H., Low, T. Y., Zhang, W., Sze, S. K., Lim, B., and Ng, H. H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *JBC* **281**, 24090-24094.

Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., Roeder, R. G., Brivanlou, A. H., and Allis, C. D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859-872.

Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N., and Tada, T. (2005). Nanog expression in mouse germ cell development. *Gene Expr Patterns* **5**, 639-646.

Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N., and McMahon, A. P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev* **13**, 3185-3190.

Yan, Q. J., Chen, X. M., Zhang, Y. M., Xie, Y., Shi, S. Z., Fu, B., Hong, Q., Xu, G. S., Zhang, X. G., Zhu, H. Y., *et al.* (2005). NANOG changes in mouse kidneys with age. *Rejuvenation Res* **8**, 248-253.

Yang, Y., Liu, W., Liu, Z. L., Hu, X. X., and Li, N. (2004). Mapping NANOG to chromosome 5 in swine. *Anim Genet* **35**, 411.

Yates, A., and Chambers, I. (2005). The homeodomain protein Nanog and pluripotency in mouse embryonic stem cells. *Biochem Soc Trans* **33**, 1518-1521.

Yates, J. R., 3rd (2000). Mass spectrometry. From genomics to proteomics. *Trends Genet* **16**, 5-8.

Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H. R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-894.

Yin, T., and Yang, Y.-C. (1994). Mitogen-activated protein kinases and ribosomal S6 protein kinases are involved in signaling pathways shared by interleukin-11, interleukin-6, leukemia inhibitory factor, and oncostatin M in mouse 3T3-L1 cells. *JBC* **269**, 3731-3738.

Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-292.

Yoshida, K., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T., and Kishimoto, T. (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. *Mech Dev* **45**, 163-171.

Yoshida, K., Taga, T., Saito, M., Suematsu, S., Kumanogoh, A., Tanaka, T., Fujiwara, H., Hirata, M., Yamagami, T., Nakahata, T., *et al.* (1996). Targeted disruption of gp130, a common signal transducer for the interleukin 6 family of cytokines, leads to myocardial and hematological disorders. *PNAS* **93**, 407-411.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1996). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *GD* **9**, 2635-2645.

Zambrowicz, B. P., Imamoto, A., Fiering, S., Herzenberg, L. A., Kerr, W. G., and Soriano, P. (1997). Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci U S A* **94**, 3789-3794.

Zhang, J., Tam, W. L., Tong, G. Q., Wu, Q., Chan, H. Y., Soh, B. S., Lou, Y., Yang, J., Ma, Y., Chai, L., *et al.* (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol* **8**, 1114-1123.

Zhang, J. G., Owczarek, C. M., Ward, L. D., Howlett, G. J., Fabri, L. J., Roberts, B. A., and Nicola, N. A. (1997). Evidence for the formation of a heterotrimeric complex of leukaemia inhibitory factor with its receptor subunits in solution. *Biochem J* **325 (Pt 3)**, 693-700.

Zhao, S., Nichols, J., Smith, A. G., and Li, M. (2004). SoxB transcription factors specify neuroectodermal lineage choice in ES cells. *Mol Cell Neurosci* **27**, 332-342.

Oligonucleotide Appendix

OLIGO NAME	PURPOSE	SEQUENCE 5'-3' (P= 5' PHOSPHATE GROUP MODIFICATION)
NheI	Tryptophan repeat mutagenesis	PGATGCGTTCACCAGGCTAGCATAGCCCTGGGGAT
Nhe2	Tryptophan repeat mutagenesis	PTCCCCGAAGTTATGGCTAGCGAGCGGAGCAGCAT
Wdr5amp f	Amplification of Wdr5 cDNA	TTGGATCCGGCGGCGGCGGCGGCGCCATGGCCACAGAGGAGAAGAA (annealing temp- 56°C)
Wdr5amp r	Amplification of Wdr5 cDNA	TTGCGGCCGCTTAGCAGTCACTCTTCCACA (annealing temp- 56°C)
Wdr5seq f	Sequencing of Wdr5 cDNA	GATATGGGACGTGAAGACAG
Wdr5seq r	Sequencing of Wdr5 cDNA	AGAGGCGGTGTCCCAGATTC
IC3	Nanog sequencing	GTACCTCAGCCTCCAGCAGAT
IC4	Nanog sequencing	AGGCTTCCAGATGCGTTCAC
M13f	Sequencing in TOPO vector	GTAAAACGACGGCCAG
M13r	Sequencing in TOPO vector	CAGGAAACAGCTATGAC

Plasmid Appendix

IPC9

Lab Stock Plasmid

IPC25
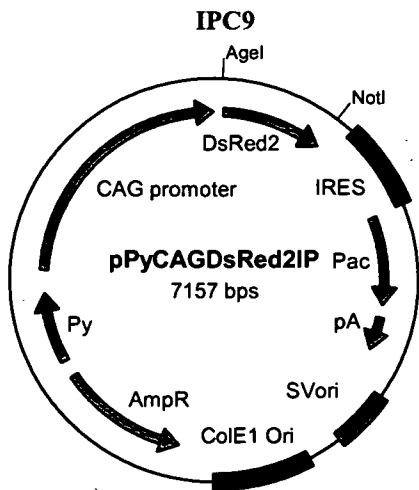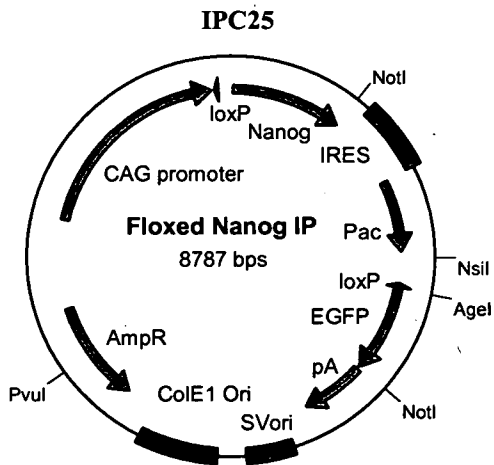
Lab Stock Plasmid
(Chambers et al, 2003)

IPC33

Lab Stock Plasmid

IPC35

pCRII(HA)3Nanog
5001 bps

Xhol
SP6 promoter
HA-Nanog
NotI
T7 promoter
f1 ori
Kan R
Amp R

Lab Stock plasmid

IPC37

pPyCAG(Flag)3NanogIP
7436 bps

Py
CAG promoter
AmpR
(flag)3Nanog
ColE1 Ori
SVori
pA Pac IRES

Lab Stock plasmid

IPC38

pPyCAG(HA)3NanogIP
7460 bps

Py
CAG promoter
AmpR
(HA)3Nanog
ColE1 Ori
SVori
pA Pac IRES
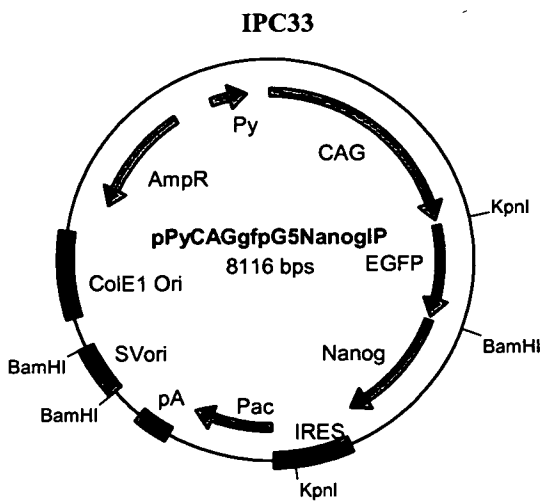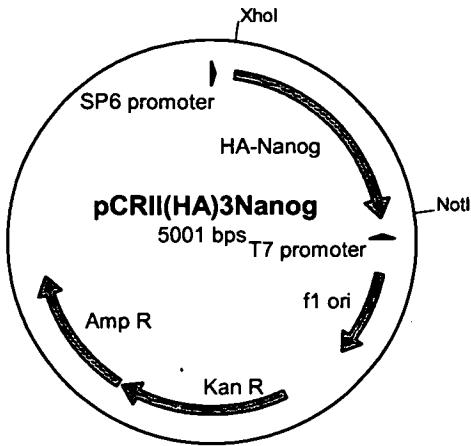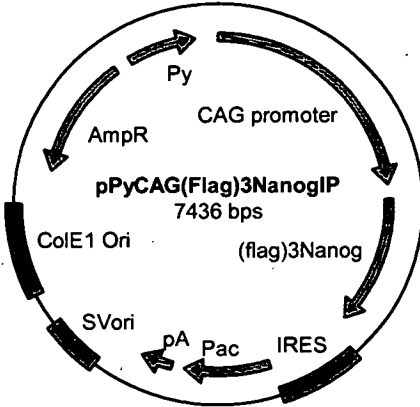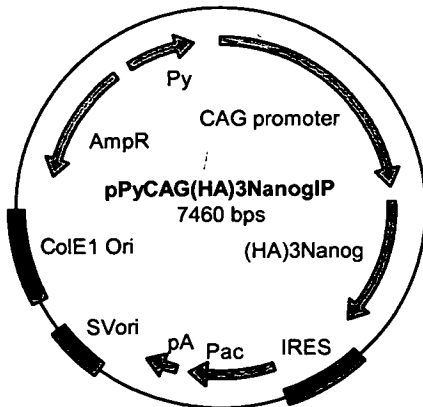
Lab Stock plasmid. Deletion mutant constructs described in chapter 2 of this thesis are based on this plasmid

198

IPC138
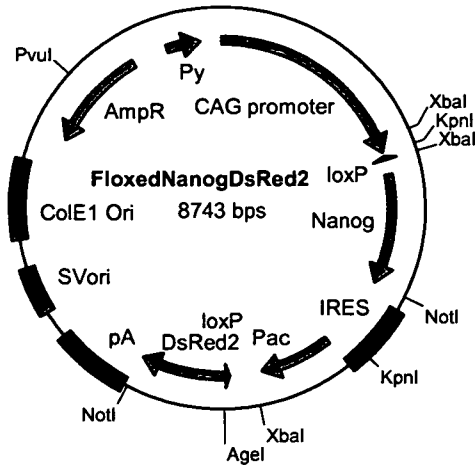
Made via 3 fragment ligation of;

AgeI/ NotI of IPC9- use 700bp fragment.

+

AgeI/ PvuI of IPC25- use 5.5kb fragment.

+

NotI/ PvuI of IPC25- use 2.5kb fragment.

Diagnose via PstI digestion; correct ligation products have the following restriction pattern; 0.7kb, 2kb, 0.9kb, and 5.1kb.

IPC154
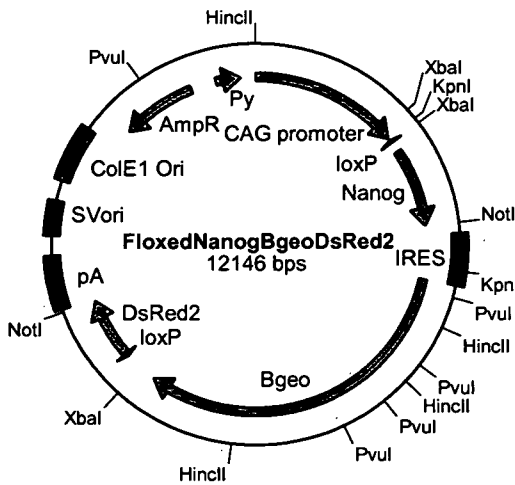
Made by 4 fragment ligation of;

PvuI/ XbaI of IPC138- use 3.5kb fragment.
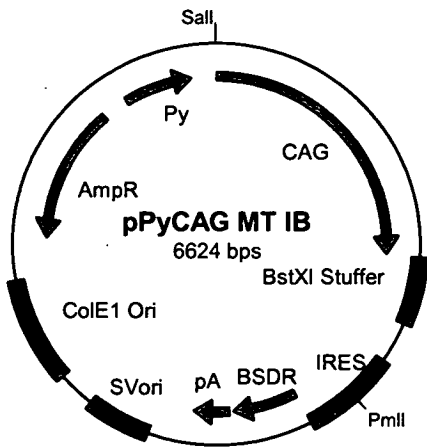
+

KpnI/ NotI of IPC138- use 0.5kb fragment.

+

NotI/ PvuI of IPC138- use 4kb fragment.

+

KpnI/ XbaI of AGS335- use 4.2kb fragment.

Diagnose via HincII digestion; correct ligation products have the following restriction pattern is 0.6kb, 1.8kb, 5.8kb, 3.9kb.
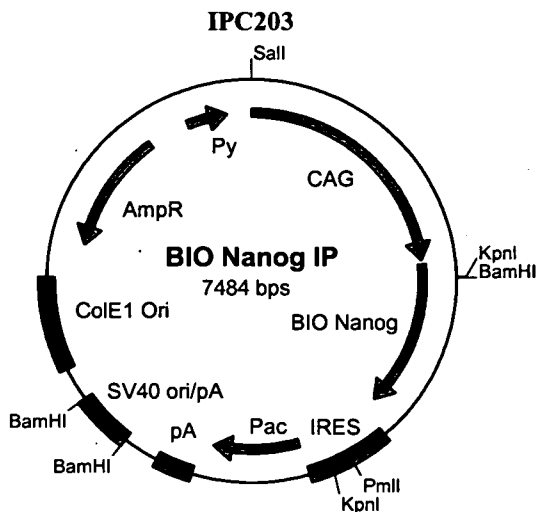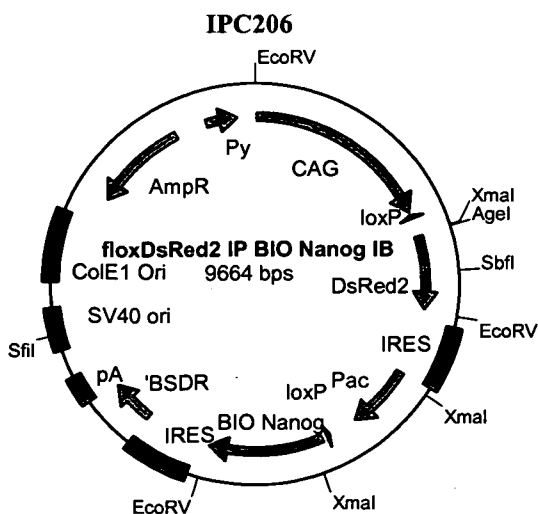
IPC183

pPyCAG MT IB
6624 bps

Sall, Py, CAG, AmpR, BstXl Stuffer, ColE1 Ori, IRES, SVori, pA, BSDR, Pmll

Lab stock plasmid

IPC194

floxDsRedIP Nanog IB
9538 bps

Py, CAG, AmpR, loxP, Agel, ColE1 Ori, DsRed2, Sbfl, SVori, IRES, Sfil, pA, 'BSDR, Pac, Nanog, loxP, IRES, Agel

Lab stock plasmid

IPC203

BIO Nanog IP
7484 bps

Py
CAG
AmpR
ColE1 Ori
BIO Nanog
SV40 ori/pA
BamHI
BamHI
pA
Pac
IRES
Pmll
Kpnl
Sall
Kpnl
BamHI
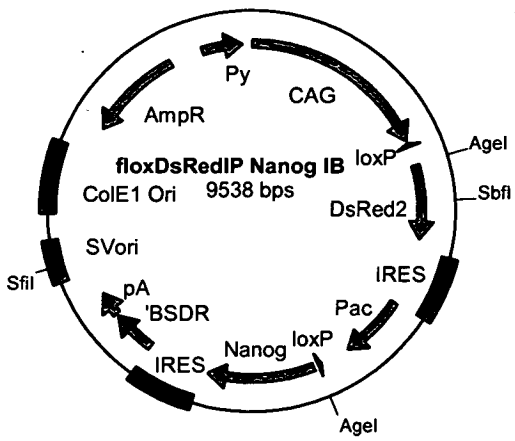
Generate by 3 fragment ligation of;

110bp NotI to BamHI fragment from pTRE BIO TEV provided by J.Strouboulis (Erasmus MC)
+

BamHI/ KpnI of IPC 33 use 1.4kb fragment
+

NotI/ KpnI of AGS 576 use 5.9kb fragment.

Diagnose via NcoI digestion; correct ligation products have the following restriction pattern- 3.2kb, 2.8kb, 1.4kb.

IPC204

BIO Nanog IB
7216 bps

Sall
EcoRV
Py
CAG
AmpR
ColE1 Ori
BIO Nanog
SVori
Sfil
pA
'BSDR
IRES
Pmll
EcoRV
EcoRV
Xmal
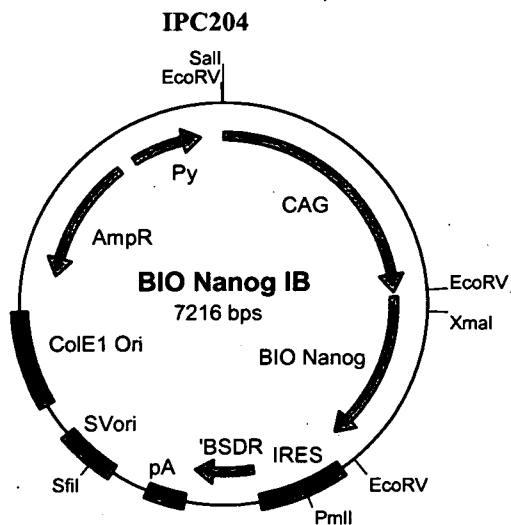
Make by 2 way ligation;

SalI/ PmlI of IPC203 use 3.1kb fragment
+

SalI/ PmlI of IPC183 use 4kb fragment.

Diagnose via EcoRV digestion; correct ligation products have the following restriction pattern- 1.1kb, 4.4kb, 1.7kb.

IPC206

floxDsRed2 IP BIO Nanog IB
ColE1 Ori   9664 bps

EcoRV
Py
CAG
AmpR
loxP
DsRed2
SV40 ori
Sfil
pA
'BSDR
IRES
loxP
Pac
BIO Nanog
IRES
EcoRV
Xmal
Xmal
Agel
Sbfl
EcoRV
Xmal

Make by 4 fragment ligation of;
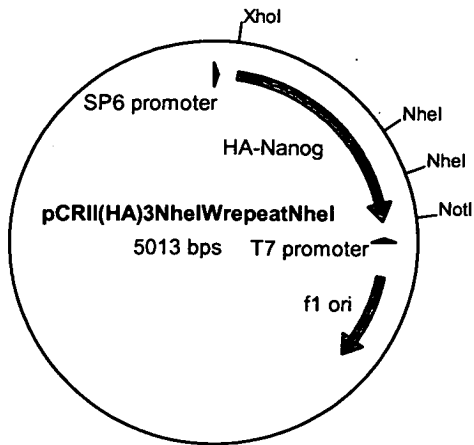
EcoRV/ XmaI of IPC204 use 121bp fragment.
+

SfiI/ XmaI of IPC 204 use 2.5kb fragment.
+

Digest IPC194 with AgeI and Klenow fill in the products. Recut with SbfI and use 1.9kb fragment.
+

SbfI/ SfiI of IPC194 use 5.1kb fragment.

Diagnose via NcoI digestion; correct ligation products have the following restriction pattern- 1.6kb, 1.9kb, 2.5kb, 3.2kb, 0.5kb.
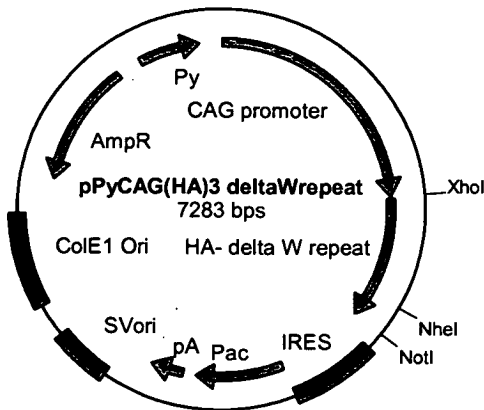
IPC332

Oligonucleotide primers (NheI and Nhe2) were annealed to single stranded IPC35 phagemid DNA, and following polymerisation and ligation, double stranded DNA was produced in DH5α *E.coli* as described in Methods 2.3.8.

The oligonucleotides introduced NheI restriction sites flanking the W repeat. NheI digestion was performed to screen for mutated DNAs; these were verified by sequencing with oligos M13F/R, IC 3 and IC4.

DNA was then NheI digested, diluted to 6ng/ml and religated. Molecules with the W repeat deleted were screened for via EcoRI digest. Mutated DNA releases a 704bp fragment whereas non-mutated DNA releases a 893bp fragment.
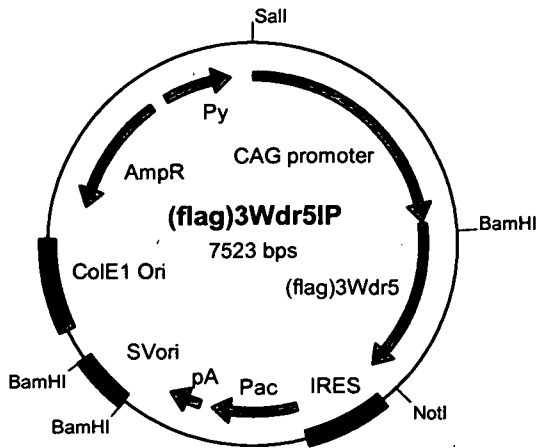
IPC328

Plasmid made by 2 way ligation of;

XhoI/ NotI of IPC332 use 900bp fragment
+

XhoI/ NotI of AGS564 use 6.4kb fragment.

202

IPC335

(flag)3Wdr5IP
7523 bps

Make via 3 fragment ligation of;
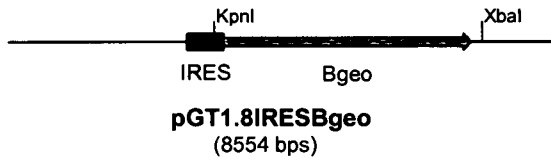
BamHI/ SalI of IPC37- use 1.8kb fragment
+

SalI/ NotI of IPC37- use 4.7kb fragment
+

PCR amplify Wdr5 from ES cell cDNA using Wdr5amp f and r primers. TOPO clone the PCR product and sequence verify with Wdr5seq f and r primers along with M13 F/R primers. Then digest TOPOWdr5 with BamHI and NotI and use 1kb fragment.
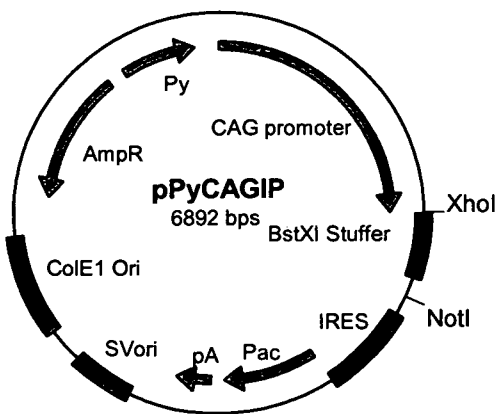
Correctly ligated plasmids can be analysed with an NcoI diagnostic digest yielding 1.3kb, 0.09kb, 2.8kb, and 3.2kb fragments.
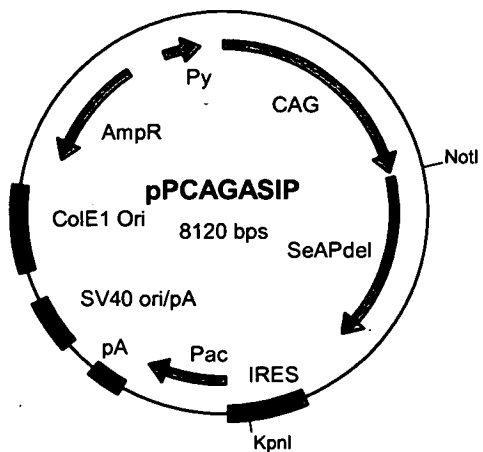
AGS335

pGT1.8IRESBgeo
(8554 bps)
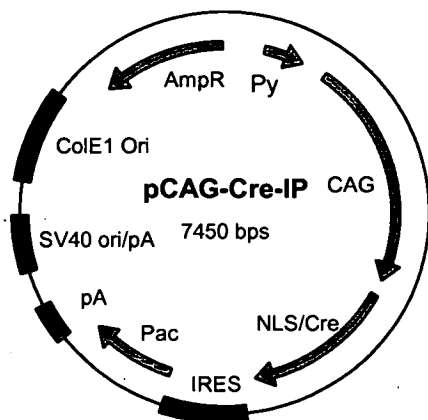
Lab stock plasmid

AGS564

pPyCAGIP
6892 bps

Lab stock plasmid

IPC576

pPCAGASIP
8120 bps

Py
CAG
AmpR
NotI
ColE1 Ori
SeAPdel
SV40 ori/pA
pA
Pac
IRES
KpnI

Lab Stock plasmid

AGS844

pCAG-Cre-IP
7450 bps

AmpR
Py
CAG
ColE1 Ori
SV40 ori/pA
NLS/Cre
pA
Pac
IRES

Lab Stock plasmid