Expression of the Homeobox Transcription Factor Hex in Embryonic Stem Cells

Maurice A. Canham

PhD

The University of Edinburgh

2008

Declaration

I declare that:

1. This thesis was composed by me.
2. The work presented is of my own unless otherwise stated.

Maurice Canham

Acknowledgments:

Firstly, I thank my supervisor, Dr. Josh Brickman, for giving me the opportunity to undertake this work in his lab and for all his help, support and reassurance during the past 4 years.

For Andrea and Gillian, I thank you both dearly for your kindness, understanding and encouragement, particularly during the early years when things were difficult.

To the other members of our lab, past present, I thank you all for your help and friendship; you have all made my time here a happy one.

For the many people who have giving me technical help, I'm indebted to you for helping to make things possible, particularly, Andrea for the Xenopus injections and Jan Ure, Rosa and Jeanette for mouse dissections. Also I thank all the ancillary staff for keeping the wheels of the department turning, particularly Jan Vrana for his assistance with FACs and Ron for the cryostat sectioning.

Thanks to everyone else at the ISCR, past and present, from whom I've been giving numerous constructs, antibodies and primers as well as the general guidance that I have needed to think more critically and undertake experiments with more confidence.

Thanks to our collaborators, Minoru Ko and Alexi Sharov at the NIH for undertaking the microarray experiment and the analysis.

To my family, thank you for your support, encouragement and belief in my abilities.

Finalmente, para mi cariño Leo. Durante todos estos años tu amor me ha dado el poder para hacer eso, gracias. Gracias tambien para tu paciencia durante los momentos en que me he vuelto un poco loco. Te quiero para siempre.

This work was supported by the MRC and Pfizer.

Abstract

Hex is one of the earliest markers of anterior-posterior asymmetry in vertebrates. It is a homeodomain transcription factor expressed in different endoderm tissues during embryogenesis that regulates early lineage specification and may have important additional roles in progenitor populations in the blood, skin and liver. The aim of my PhD is to understand the significance of Hex expression, sub-cellular localization and molecular function during development.

To this end I have generated an extremely sensitive reporter cell line that unveils a new domain of Hex expression in embryonic stem (ES) cells. The introduction of a cDNA encoding a tagged version of Hex upstream of an internal ribosomal entry site and a variant of the Yellow Fluorescent Protein, Venus, into the first exon of *Hex*, has revealed a heterogeneous expression pattern among ES cell cultures. Manipulation of Fgf signalling alters the percentage of venus positive cells and suggests that this subpopulation maybe fated to become primitive endoderm, the earliest domain of *Hex* expression in the mouse embryo. Although there is an equivalence of *Oct3/4* in both venus positive and negative subpopulations, *Nanog* and venus appear mutually exclusive. Indeed, forced expression of *Nanog*, which is a negative regulator of primitive endoderm, reduces venus positivity. Microarray and quantitative PCR analyses show an enrichment of primitive endoderm specific genes in venus positive ES cells while markers of pluripotency are comparatively reduced. While clonal density plating of these subpopulations demonstrate interconvertability, venus positive ES cells have a reduced ability for clonal growth and contribution to the embryo in chimera analyses.

Based on these experiments in ES cells, I also attempt to overexpress *Hex* in ES cell cultures. Establishment of stable clones overexpressing *Hex* in ES and other cell types is difficult, suggesting intolerance. Construction of an inducible system to characterize the phenotype of ectopic *Hex* expression in ES cells reveals the onset of apoptosis upon induction by a mechanism that depends on its ability to bind DNA. These observations reflect previous studies which suggest that *Hex* is a key regulator in maintaining a balance between immediate early cell lineage decisions and proliferation. These studies suggest that *Hex* maybe an important marker of early cell fate decisions, but is probably not the primary mediator of early blastocyst/ES cell asymmetry.

Abbreviations

AVE	Anterior Visceral Endoderm
ADE	Anterior Definitive Endoderm
bp	base pair
BSA	Bovine Serum Albumin
°C	Degrees centigrade
CO_2	Carbon dioxide
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
DVE	Distal Visceral Endoderm
EDTA	Ethylenediamine tetra-acetate
EGTA	Ethylene Glycol Tetra acetic acid
EtOH	Ethanol
FCS	Fetus Calf Serum
g	Grams
GFP	Green fluorescent protein
GMEM	Glasgow's modified Eagle's medium
H_2O	Water
hr	Hour
kb	Kilobase
kD	KiloDaltons
l	Litre
LB	Luria-Broth
m	Milli
μ	Micro
M	Molar
MgCl2	Magnesium chloride
min	Minutes
n	Nano
NaCl	Sodium choride
NaOAc	Sodium acetate

NaOH	Sodium hydroxide
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline-Tween20 soltion
PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
rpm	revolutions per minute
RT	room temperature
TAE	Tris-acetate-EDTA
TB	Terrific Broth
UV	Ultraviolet
VE	Visceral Endoderm
w/v	Weight/Volume

List of Figures

List of Tables

Table of Contents

Chapter 1

General Introduction

The goal of developmental biology is to gain an understanding of how a group of indifferent cells, arising from the first few divisions after fertilisation, can expand and organise themselves into the diverse array of tissues and organs that are seen in the adult. The dramatic cell movements following these initial cell divisions are called gastrulation, giving rise to the three fundamental germ layers, ectoderm, mesoderm and endoderm. The inner most layer, the endoderm, gives rise to all gut tissue and associated organs. The ectoderm generates the nervous system and the epidermis. Between these two, the mesoderm develops into connective tissue, the cardiovascular system, the muscles, bones and kidneys.

Recent advances in embryological and molecular techniques have shed light on how cells communicate and influence each other in order to adopt certain fates in a spatial-temporal context. Pioneering work by Spemann and Mangold (1924) demonstrated how a particular group of cells in the gastrulating amphibian embryo can serve as an organising tissue, generating a second axis when grafted into ectopic regions of a second embryo. Similar organising tissue has been found in zebrafish, chick, rabbit and mouse indicating a basic conserved developmental mechanism between vertebrates.

The mouse is the organism of choice when coming to study development in mammals given the wealth of genetic knowledge available and the relatively short turnover rate of breeding.

1.1 Early mouse development

After entry of the sperm into the egg two polar bodies arising from the first and second meiotic divisions are ejected from the cell on the opposite side and remain attached following subsequent cell divisions. These divisions, known as cleavage, continue to a 16 cell stage where the cells compact together to produce a morula. At the 32 cell stage (blastulation) two mutually exclusive populations arise, those around the outside comprise the trophoectoderm which give rise to the trophoblast

and extraembryonic ectoderm (ExE) and those on the inside, the inner cell mass (ICM), form the embryo and extraembryonic endoderm and mesoderm. At day 4.5 the implanting blastocyst forms a third layer on the underside of the ICM. This tissue is the primitive endoderm (or hypoblast) and gives rise to parietal and visceral endoderm (VE) which envelops the remainder of the ICM is known as the epiblast. The epiblast forms the embryo proper as well as extraembryonic mesoderm. Following continued growth of the polar trophoectoderm the epiblast and surrounding VE are pushed to the distal end of the conceptus (distal VE (DVE)) and at the same time a cavity within the epiblast is formed giving rise to a cup shape. To this point the conceptus is bilaterally symmetrical but by day 6.5 the first morphological sign of asymmetry becomes apparent on the future posterior side with the emergence of the primitive streak. This marks the beginning of gastrulation. The streak extends towards the distal end of the embryo culminating in an organising structure called the node. The node is considered to be the mammalian equivalent to the amphibian organiser. Epiblast cells involuting into the streak and emerging through the node become the axial mesendoderm comprising the future midline mesoderm (prechordal and notochord plates) and the anterior endoderm (foregut endoderm). The axial mesendoderm extends anteriorly replacing the VE, on the opposite side to the streak, to underlie the anterior ectoderm. Posterior to the node, paraxial and lateral plate mesoderm emerge from different levels of the streak (Beddington and Robertson, 1999). Figure 1.1 depicts the stages of mouse development from conception to the beginning of gastrulation.

1.2 Organising centres of the mouse embryo

As in the frog, the mouse node has been shown to induce a second axis following heterotopic transplantation experiments. However, unlike grafts with the amphibian organiser, the node does not induce as full axis as it lacks forebrain (Beddington, 1994). Interestingly, as a follow-up to the original amphibian organiser experiments,

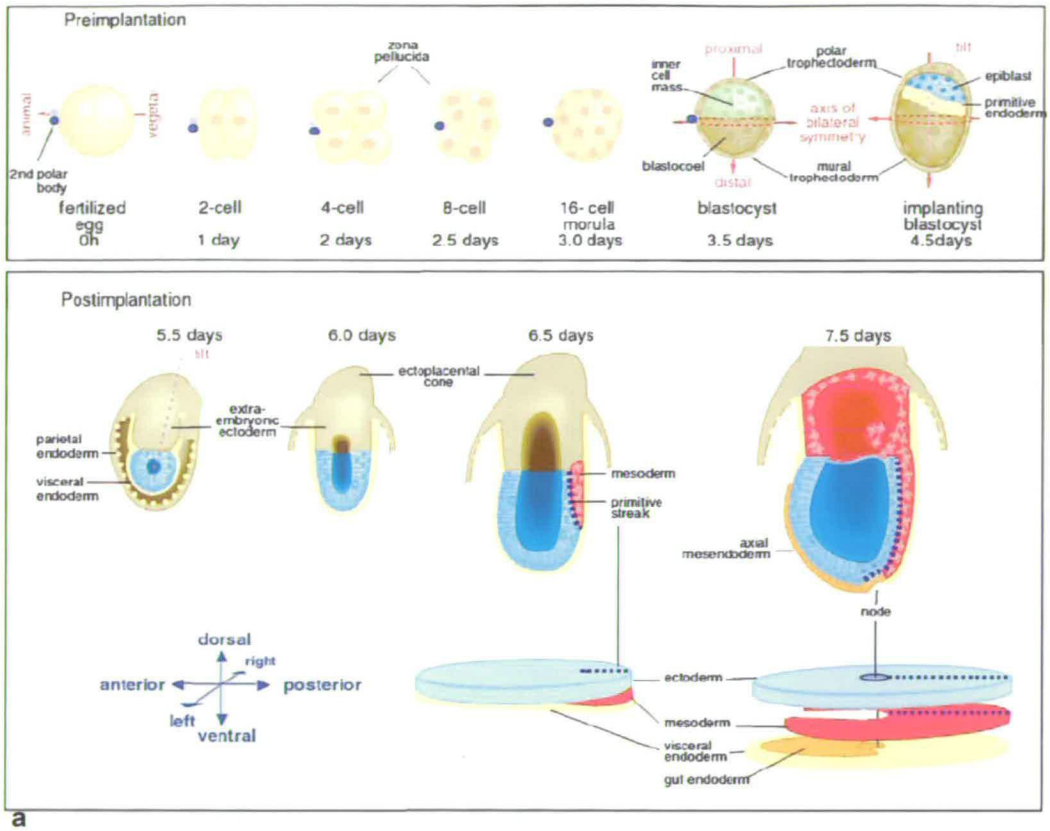


Fig. 1.1 Early development of mouse.

a, Schematic diagram depicting the first 7.5 days of murine development. The upper box shows the developmental stages prior to uterine implantation. The lower box represents those stage after implantation. Taken from Beddington and Robertson, 1999. **b,c & d**, Expression domain of *Hex* in the primitive endoderm at E4.5 (**b**), the distall tip visceral endoderm at E5.5 (**c**), and the AVE (arrows) and EGO (arrowhead) at E7.0 (**d**). Taken from Thomas et al. 1998.

it was shown that only transplantation of the early and not the late organiser could induce the full axis in *Xenopus* (Spemann 1931). Thus, the organiser appeared to contain stage dependent separable head and trunk inducing activities. However, in mouse, even the precursor tissue of the node, known as the anterior primitive streak or Early Gastrula Organiser (EGO), cannot induce a full axis upon transplantation (Tam and Steiner, 1999). As with the original node grafts, the forebrain is missing indicating that in mouse, neither the node nor its precursors are sufficient to be head organisers (Beddington, 1994).

The asymmetric expression of a number of mouse homologues of "head organizer," genes from other species in the anterior domain of the extra-embryonic visceral endoderm point to a requirement of the VE surrounding the epiblast in contributing to forebrain development. At day 5.5, a full day before gastrulation begins, gene expression analysis has shown the presence of VE-1 antigen (Rosenquist and Martin, 1995), *Cerberus related 1*(Belo et al., 1997), *Hex* (Thomas et al., 1998), and *Otx2* (Ang et al., 1996) on the anterior side in the VE. This asymmetric expression marks the anterior VE (AVE) and analysis of *Hex* expression has shown that this population is derived from the distal tip of the embryo (Thomas et al., 1998). The anterior movement of distal tip cells to become the AVE precedes the formation of the primitive streak and therefore defines the first signs of anterior-posterior asymmetry in mouse (Kimura et al., 2000; Srinivas et al., 2004). Moreover, both embryological and genetic ablation experiments suggest that the AVE and DVE, functions as an anterior organizing centre. Thus, removal of the tissue results in the loss of forebrain tissue as judged by loss of the anterior neural marker, *Hesx1* (Thomas and Beddington, 1996). Genetic ablation of key determinants in chimeric mice has confirmed the necessity of the AVE in forebrain development. This technique involves production of embryos where the extraembryonic tissues (including the VE) are of a particular genotype, different from that of the injected epiblast cells which give rise to the definitive germ layers. The transcription factors *Otx2* (Rhinn et al., 1998), *FoxA2* (Dufort et al., 1998) and *Lim1* (Shawlot et al., 1999) have been shown to be required in the VE for correct forebrain specification. These results indicate that this extraembryonic tissue is required to establish anterior identity in the neural

plate. However, despite this requirement for the AVE, its inductive activity in heterotrophic grafting experiments is not like the original Spemman experiments. Thus the AVE has some neural inducing activity but only in combination with the EGO and anterior epiblast can it give rise to the expression of anterior neural genes in grafting experiments (Tam and Steiner, 1999). Taken together, these results suggest that, in mouse, a single separate head organiser does not exist. Rather, synergistic activities between the AVE and EGO are necessary for initiation of anterior neural induction. These results are also consistent with data from further chimera studies that suggest that the activity of key early transcription factors such as *Otx2*, are required for the initiation of forebrain identity in the AVE, but are also required for the maintenance and further development of this region, in a second anterior signalling centre derived from the node, the anterior definitive endoderm (ADE). Interestingly, during gastrulation the ADE migrates anteriorly pushing the VE away from the embryo, but maintaining a similar signalling centre for the establishment of the anterior neural plate (Martinez-Barbera and Beddington, 2001).

Genetic evidence supporting the dual role for two anterior signalling centres in the production of the anterior neural plate has been demonstrated by the analysis of mutant mice. In *β-catenin* and *Wnt3* deficient embryos the EGO is not formed as seen by the lack of mesoderm and the posterior markers *Goosecoid* and *Brachyury*. Despite correct patterning of the AVE, which remains distal in the case of the *β-catenin* mutant, the anterior side of the epiblast fails to form head structures (Huelsken et al., 2000; Liu et al., 1999). Embryos lacking *Fgf8* or *cripto* have a properly formed EGO and AVE. However, the AVE in both cases also fails to rotate and this results in the expansion of forebrain markers at the distal tip, indicating that the AVE has an essential role in initial anterior specification (Ding et al., 1998; Sun et al., 1999). Taken together, these results show that in mammals head organizing activity has multiple components separated in space and time, the AVE induced prior to gastrulation and never contributing to the embryo proper, the EGO, situated on the posterior side of the embryo, then gives rise to a second signalling centre, the ADE, that will migrate anteriorly and replace the AVE.

The role of the AVE in establishing the anterior character was demonstrated by chimeric and tissue explant analysis of *Otx2* which indicated that this gene is required in the VE for AVE movement and for the suppression of posterior markers (Kimura et al., 2000; Perea-Gomez et al., 2001a). Likewise, *Lim1:FoxA2* double mutant mice show ectopic expression of posterior epiblast markers (Perea-Gomez et al., 1999). These genes are also required in the VE and it is proposed that they control the expression of inhibitors that restrict activity of posterior signals in order to impart anterior identity onto the overlying anterior ectoderm (Perea-Gomez et al., 2001b).

Many of the same genes that are expressed in the AVE are also expressed later in the ADE including, *Hex*, *Lim1*, *FoxA2*, *Otx2*, *Cer1* and *Lefty1* (Perea-Gomez et al., 2002; Rodriguez et al., 2001).

1.3 Anterior-posterior axis formation in mouse

Anterior-posterior axis formation occurs before and during gastrulation. It involves the interaction of various signalling pathways and the establishment of different localised gene expression areas in the mouse embryo. Generally, anterior cues are established at one end of the embryo which suppresses the extension of posterior signals from the opposite end. These effects occur simultaneously, where each end of the axis reinforces its own status while limiting that of the other. As discussed in the previous section, the patterning of the neural axis in mouse requires multiple signalling centres including the AVE, posterior epiblast, and ADE. Here we discuss the molecular basis for axis induction.

1.4 TGF-β signalling

1.4.1 Role in Mesoderm Induction

The induction of the early axes' is difficult to separate from embryonic induction, the process by which key signalling pathways induce the formation of mesoderm and definitive endoderm. One of the key signalling molecules governing the establishment and action of the AVE and ADE in anterior-posterior patterning is *Nodal,* which as the name implies, was identified based on an insertional mutation that exhibited a failure in node formation and problems in mesoderm induction (Conlon et al., 1994). This member of the transforming growth factor-beta (TGF-β) family of ligands has a highly dynamic pattern of expression in the developing embryo. Expression is found prior to gastrulation in the proximal epiblast that then becomes progressively restricted towards the posterior side of the embryo. It is also expressed at low levels through out the visceral endoderm. By late gastrulation expression is confined to the node (Conlon et al., 1994). *Nodal* related homologues have been found in Xenopus, zebrafish and chick (Feldman et al., 1998; Levin et al., 1995; Smith et al., 1995). Mice homozygous for a mutation in mouse *Nodal* gene fail to form a primitive streak and lack evidence of mesoderm (Conlon et al., 1994). Likewise, combined mutation of two *Nodal* related genes in zebrafish *Cyclops* and *Squint*, fail to form mesoderm (Feldman et al., 1998). A similar phenotype is seen in *Xenopus* when approaches are employed to specifically inhibit Nodal signalling (Agius et al., 2000). Additionally ectopic expression studies in zebrafish and Xenopus have shown that Nodal related factors can induce the formation of mesoderm (Jones et al., 1995; Toyama et al., 1995). Together these results show that among vertebrates Nodal signalling is required for mesoderm induction at the start of gastrulation.

Interestingly, while *Nodal* was initially identified by genetics in mouse, functional studies examining the role of TGF-β signalling in embryonic induction had focused on activin, the first molecule with mesoderm inducing activities to be identified (Smith et al., 1990). Activin signalling occurs by the binding to type II serine/threonine kinase receptors ActRIIA and ActRIIB which phosphorylate the type I receptor, ActRIB. This in turn phosphorylates the receptor mediated Smad molecules, *Smad2* and *Smad3*. Upon phosphorylation, these R-Smad molecules combine with the co-Smad, Smad4 and the FAST family of forkhead transcription

factors to mediate expression of downstream genes (Schier and Shen, 2000). Genetic evidence has shown that *Nodal* is thought to act through the same signalling pathway as Activin. *Smad2* and ActRIB deficient mice show similar gastrulation defects to those which occur in *Nodal* mutants (Gu et al., 1998; Heyer et al., 1999; Waldrip et al., 1998). In *Xenopus* and zebrafish, overexpression of *Nodal* and Activin have similar effects in the production of mesoderm (Jones et al., 1995; Rebagliati et al., 1998). Furthermore, more recent biochemical analysis has shown that Nodal signalling is mediated by both *Smad2* and *Smad3* (Kumar et al., 2001). Unlike other TGF-β signalling, Nodal signalling has been shown to require the presence of additional cofactors (Kumar et al., 2001). *Cripto* deficient mice phenocopy aspects of the *Nodal* mutant in the failure to produce a primitive streak and embryonic mesoderm (Ding et al., 1998). Figure 1.2 summarizes the main components of the TGF-β/Nodal pathway as well as those for the Wnt and FGF pathways discussed later.

1.4.2 Role in Anterior Induction

Further mutational analysis in mouse of *Nodal* and downstream signalling factors has also begun to tease out the means by which *Nodal* specifically regulates axis formation independently from its role in mesoderm induction. The anterior movement of distal tip cells to become the AVE immediately precedes a shift in expression of *Brachyury* from the proximal epiblast to the posterior side. Therefore a pre-existing proximal-distal (P-D) axis is converted into an anterior-posterior one (Beddington and Robertson, 1999). Interestingly, embryos mutant for components in the Nodal pathway display P-D defects including the absence proximal marker expression in the extra-embryonic ectoderm (ExE) and proximal VE such as *Bmp4* and *Eomes* (Arnold et al., 2008; Ben-Haim et al., 2006; Brennan et al., 2001). Moreover, embryos lacking *Smad2* phenocopy the VE defects in *Nodal* mutants such that the AVE is not specified and there is an expansion of proximal-posterior genes (Brennan et al., 2001; Waldrip et al., 1998). Therefore *Nodal* and its downstream

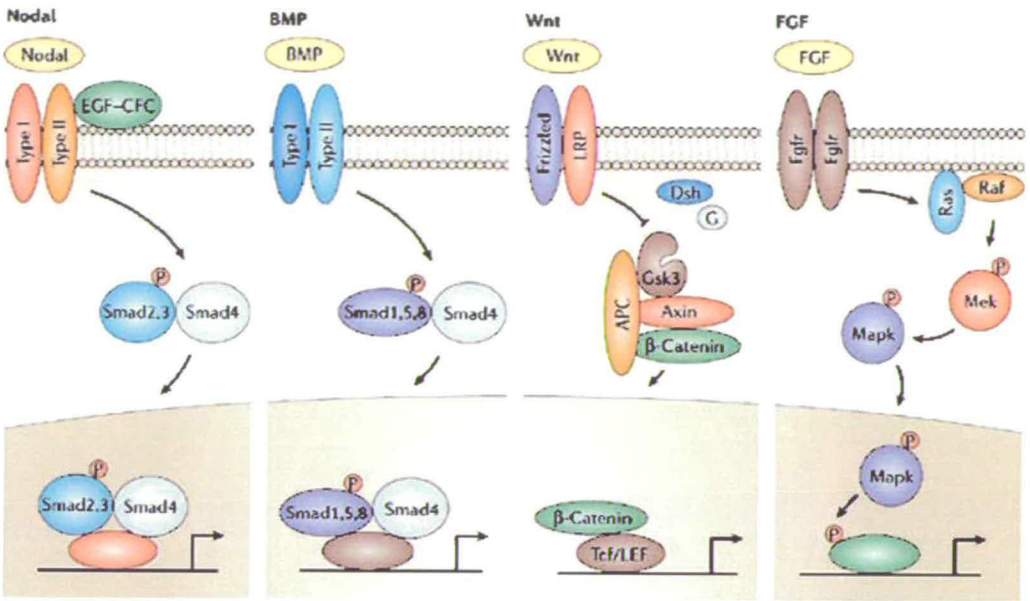


Fig. 1.2 Major Signalling Pathways during Embryonic Development.

Schematic diagram summarizing TGF β (Nodal/Activin and BMP), canonical Wnt and Fgf pathways. Binding of Nodal or Activin to type I and type II receptors leads to the phosphorylation of the R-Smads, Smad2 and Smad3 which can then associate with the co-Smad, Smad4. This complex can then translocate to the nucleus to affect transcription of downstream targets. BMP signalling occurs similarly but uses the R-Smads, Smad1, Smad5 and Smad8. Binding of Wnts to the Frizzled/LRP receptors activated the Dishevelled protein which leads to the ubiquitination and degredation of the β-Catenin disruption complex. Genetic ablation of any members of this disruption complex leads to free β-Catenin which can translocate to the nucleus where it can bind to members of the Tcf/LEF family of transcription factors and lead to activation of downstream targets. One of the effects of Fgf binding to its receptors is the activation of the Raf/Ras/Mek/Erk pathway. Upon ligand binding, receptors are dimerized, leading to a series of sequential phosphorylation events by a cascade of kinases. Phosphorylation of Raf at the cytolplasmic side of the membrane is mediated by the Grb2 protein (not shown). The result is the translocation of phospho-Erk (shown as Mapk-P) to the nucleus which can then affect transcription of downstream targets. Taken from the review by David Kimelman. Nature Reviews Genetics 7, 360-372 (May 2006)

effecter, *Smad2*, are responsible for maintaining a P-D polarity that is essential to the establishment of the subsequent A-P polarity.

Interestingly, not only is *Nodal* required for the induction of these anterior signalling centres, it also behaves in a dose dependent morphogen, with the highest level of signalling inducing anterior identity (Lewis and Tam, 2006). The study of mechanisms in which these levels can be achieved has shown that *Nodal* has a positive autoregulatory loop. Analysis of the ASE enhancer element in the *Nodal* gene has shown that the downstream mediator *Smad2*, in conjunction with the forkhead protein FoxH1 (FAST1), bind to this site and promote production of *Nodal* in the AVE (Norris et al., 2002). Additionally, various studies into the reduction of *Nodal* and its downstream effectors *Smad2* and *Smad3* through this autoregulatory loop have demonstrated a second requirement for high activity of this signalling pathway in the specification of the ADE during gastrulation (Dunn et al., 2005; Dunn et al., 2004; Lowe et al., 2001; Lu and Robertson, 2004; Norris et al., 2002; Tremblay et al., 2000; Vincent et al., 2003). Reduced signalling in these mutants lead to incorrect specification of the ADE and resulted in forebrain truncations despite normal AVE activity. Thus, the highest levels of Nodal signalling are required for the induction of definitive endoderm and, in particular, the most anterior definitive endoderm, the ADE.

While Nodal signalling is important for the establishment of AVE patterning and the highest levels of Nodal signalling for the establishment of ADE, these centres also express *Nodal* antagonists *Cerberus related 1* (*Cer1*) and *Lefty1* (Yamamoto et al., 2004). Indeed compound mutants of these two antagonists lead to the development of an ectopic primitive streak (Perea-Gomez et al., 2002). Thus these signalling centres may promote anterior identity by antagonizing the signals that continue to mediate poster mesoderm induction throughout gastrulation. Evidence *Nodal* antagonists promote anterior identity is demonstrated by ectopic head induction following injection of *Cerberus* into *Xenopus* (Bouwmeester et al., 1996). Given that both *Cer1* and *Lefty1* expression are lost in either *Nodal* or *Smad2* mutant mice, they appear to be downstream targets of this pathway (Brennan et al., 2001). So why does

this highly Nodal responsive tissue express these secreted antagonists? Perhaps these secreted antagonists are required to insulate the anterior epiblast from the mesoderm inducing properties of Nodal produced in the primitive streak region, thus allowing to undergo the early stages of neural differentiation.

Thus, evidence from this important signalling pathway demonstrates that the establishment of anterior-posterior axis in mouse embryos is a complicated process in which different levels of *Nodal* and its downstream effectors perform different roles which culminate in donating both posterior identity (formation of mesoderm and primitive streak at the posterior end) and anterior identity (formation of the AVE to initiate anterior character by suppressing mesoderm and subsequent specification of the ADE to reinforce these initial cues). Interestingly, the *Hex* null mouse shares many similar defects to those seen with the reduced Nodal signalling mutants and both genes thought to regulate each others expression (see later).

1.5 Wnt Signalling

A second important pathway involved in the establishment of anterior-posterior identity in different vertebrates during gastrulation is the canonical Wnt pathway. There are around fifteen different Wnt genes expressed in various embryonic and adult tissues (Cadigan and Nusse, 1997). The first of which to be characterised, Wnt1, was identified as a proto-oncogene activated by integration of the mouse mammary tumour virus and was identified as the orthologue to the drosophila gene *wingless* (Nusse and Varmus, 1992). The are a number of pathways downstream of Wnt ligands, however, the predominant pathway regulating axis formation is know as canonical.

In the canonical Wnt pathway, the absence of a Wnt signal leads to the formation of a complex between the downstream mediator, β-catenin together with Axin and the tumour suppressor gene product adenomatous polyposis coli (APC). These proteins act as scaffolds for β-catenin which allows its phosphorylation by Gsk3β resulting in

ubiquitination and degradation. Upon binding of a Wnt ligand to Frizzled receptors and the low density lipoprotein receptor-related protein coreceptors LRP5 and LRP6, Dishevelled becomes activated which leads to the destruction of the APC/axin/GskB complex resulting in free hypophosphorylated β-catenin which can freely translocate into the nucleus. β-catenin in the nucleus can then act with members of the Lef/Tcf transcription factors to activate gene transcription (Huelsken and Behrens, 2002). In the absence of β-catenin, of the Lef/Tcf proteins are able to bind to Groucho-related co-repressors and therefore exhibit an inhibitory affect on target genes (Brantjes et al., 2001). Thus, with the complexity of this pathway, mutations in any of the components can lead to gain or loss of Wnt activity. Loss of function mutants in any Wnt ligand, frizzled/LRP receptors, *β-catenin* or *Lef/Tcf* genes would lead to a reduction in signal, whereas mutations in the *APC*, *Axin* or *Gsk3β* would lead to constitutive activation.

As with Nodal signalling, the Wnt pathway is crucial for the induction of mesoderm. Mutation of the mouse *Wnt3* gene results in embryos completely lacking the primitive streak (Liu et al., 1999). Conversely, overexpression of chick *Wnt8C* in mouse leads to an ectopic primitive streak but fails to induce a complete a full second axis containing anterior structures (Popperl et al., 1997). While genes of the AVE are normally expressed in *Wnt3*-/- mice, mutation of the downstream effecter β-catenin results in the inability of the AVE to rotate correctly concomitant with the absence of some markers normally expressed there such as *Hex* (Huelsken et al., 2000). These data suggest that while *Wnt3* affects axis formation by influencing the formation of the primitive streak, the downstream effecter β-catenin may have an additional earlier effect on induction and migration of the early AVE (Kimura-Yoshida et al., 2005).

Interestingly, when β-catenin is constitutively activated throughout the embryo (as opposed to only in the visceral endoderm) in the APCmin mice, ectopic primitive streak is formed and AVE lost. Thus, APC-/- mutant embryos ectopically express markers of the primitive streak and epithelial-mesenchyme transition such as *T*, *Eomes* and *S nail* while expression of the distal VE such as *Hex* and *Cerl* are absent

(Chazaud and Rossant, 2006). Together these reports show that Wnt signalling is important in axis specification and can act in a context dependent fashion to specify both anterior and posterior fates.

Interestingly, as with Nodal signalling, inhibitors of Wnt signalling such as *Dickkopf* (Dkk1) and the secreted frizzled related proteins 5 and 8 (Sfrp5 and Sfrp8) are found to be expressed in both the AVE and ADE tissues (Finley et al., 2003; Glinka et al., 1998; Lu et al., 2004). Indeed, mutation of *Dkk1* shows severe anterior truncations in mutant mice suggesting that inhibition of Wnt signalling plays a role in anterior specification during development (Mukhopadhyay et al., 2001). This supports the general notion that the function of the AVE/ADE is to produce secreted antagonists that can insulate the anterior neural plate from the mesoderm inducing signals found at the posterior side of the embryo.

Together these data show that TGF-β and Wnt signalling are important for establishing anterior-posterior asymmetry and that both the AVE and ADE are important tissues for ensuring anterior character. The homeobox transcription factor *Hex* is also expressed in both of these organising centres and has recently been shown to be both induced by and modulate the TGF-β and Wnt pathways (Zamparini et al., 2006).

1.6 The Homeobox gene *Hex*

Hex (haematopoietically expressed homeobox), also known as *Prh* (proline rich homeodomain), was first identified in screens to isolate novel homeobox genes in haematopoietic and primary transformed cell lines in chicken, mouse and human (Bedford et al., 1993; Crompton et al., 1992; Hromas et al., 1993). Homologues have also been found in rat, zebrafish and *Xenopus* (Newman et al., 1997; Tanaka et al., 1999; Yatskievych et al., 1999). The mouse and human genes have been mapped to chromosomes 10 and 19, respectively (Ghosh et al., 1999; Hromas et al., 1993). Its location in the genomes of both mouse and human lies outside regions occupied by clustered homeobox (HOX) genes, therefore making it a divergent or orphan

homeobox gene. The mouse *Hex* gene spans 5.75 kb and consisting of four exons, the central two of which encode the homeodomain (Ghosh et al., 1999). The protein contains an N-terminal proline rich region and a C-terminal acidic region which flank the homeodomain region (Crompton et al., 1992). The homeodomain shares 46% identity with that of the *Drosophila Antennapedia* gene (Bedford et al., 1993).

1.6.1 Expression of *Hex* during Early Development

During development, *Hex* is initially expressed in the primitive endoderm underlying the ICM at day 4.5dpc. By day 5.5 it becomes restricted to a small population of cells at the very distal tip of the VE and these cells begin to move towards the future anterior side to become the AVE a full day prior to emergence of the primitive streak. A second site of *Hex* expression is detected in the anterior most part of the axial mesendoderm (ADE) emerging from the distal end of the primitive streak. This site of expression moves in an anterior direction as the axial mesendoderm extends to merge with and replace the AVE (Thomas et al., 1998). In other vertebrates, *Xenopus*, chick and zebrafish, Hex is also expressed in analogous anterior endodermal tissues in (Ho et al., 1999; Jones et al., 1999; Yatskievych et al., 1999). By headfold and early somite stages, Hex expression has expanded within the endoderm to encompass the descendents of the ADE, the ventral foregut endoderm (Thomas et al., 1998). By day 10, *Hex* is found in the 3rd pharyngeal pouch from which the thymus originates, the endodermal cells of the liver that are invading the septum transversum, the pancreatic bud, the thyroid and gall-bladder primordium (Bogue et al., 2000). In addition to Hex's expression in the endoderm, it is also expressed in the nascent blood islands and transiently in endothelial precursors. This endothelial precursor expression is also conserved in other vertebrate models (Ho et al., 1999; Thomas et al., 1998; Yatskievych et al., 1999).

In the adult mouse, *Hex* is expressed in lung, liver, thyroid and the myelomonocytic and B cell lineages of the haematopoietic system (Bogue et al., 2000; Crompton et al., 1992; Keng et al., 1998; Keng et al., 2000; Manfioletti et al., 1995). It is not

expressed in T-lymphocytes and downregulation of *Hex* is essential for normal development of this lineage (Mack et al., 2002).

Given the expression of Hex in several progenitor populations in blood, a number of studies have looked at the role of *Hex* in proliferation and/or differentiation. Overexpression of *Hex* in *Xenopus* and zebrafish leads to an increase in vascular endothelial cells (Liao et al., 2000; Newman et al., 1997) and misexpression in mouse haematopoietic precursors lead to neoplastic transformation (Mack et al., 2002). In mouse *Hex* is downregulated in terminally differentiated B cells suggesting a role in their development (Manfioletti et al., 1995). Other overexpression studies in various endothelial and haematopoeitic cell lines were indicative of a negative effect of Hex on the proliferation or maturation of these lineages (George et al., 2003; Jayaraman et al., 2000; Mack et al., 2002; Nakagawa et al., 2003; Ying et al., 2002). This inhibitory effect on proliferation was also observed in cell types representative of vascular-endothelium and early haematopoeitic progenitors in embryoid body differentiation from murine ES cells, while neuronal fate was unaffected (Kubo et al., 2005).

1.6.2 Loss of function Hex phenotypes in mouse

Models for the means by which Hex acts on lineage specification in embryonic progenitor populations need to be assessed alongside the *Hex* mutant phenotypes. The *Hex* knockout mouse is devoid of liver, thyroid and the majority of mutants also display varying degrees of forebrain truncations (Martinez Barbera et al., 2000). With result, no embryos survive to birth. In $Hex^{-/-}$ mice the AVE is patterned normally as seen by *Cer1*, *Lim1* and *FoxA2* expression. Examination of these same genes at late streak stages shows a loss of *Cer1* in the ADE and prechordal plate while *Lim1* and *FoxA2* are increased in the node. These results indicate that the AVE functions normally in these mutants but there is a defect in the anterior axial mesendoderm tissues (the ADE and prechordal plate but not notochord). Forebrain defects arise after day 7.5 and range from none to severe truncations rostral to the

zona limitans intrathalamica (ZLI). The restriction of these truncations are proposed to be reflective of the restriction of the phenotype observed in the axial mesendoderm (Martinez-Barbera and Beddington, 2001). Chimeric experiments using *Hex*$^{-/-}$ ES cells with a wild type blastocysts showed similar forebrain defects indicating a requirement for *Hex* in the definitive endoderm (Martinez Barbera et al., 2000) and supporting the notion that *Hex* is not required in the AVE. This was the first concrete evidence that a signal from the ADE is required for correct forebrain patterning. The mouse *Hex* gene has been knocked out by another group (Keng et al., 2000). Their mice also display liver defects but interestingly none show any forebrain defects.

Analysis of the liver defect shows initial hepatic gene expression but a failure in localised endoderm proliferation leads to poor liver bud formation (Bort et al., 2004). In the same study Hex is shown to control the specification of ventral pancreas formation by an indirect mechanism; by allowing a subset of endoderm cells to grow past the inhibitory effects of the cardiac mesoderm. The role of *Hex* during anterior endoderm development may be at the morphogenetic level rather than directly specifying tissue-specific genes.

1.6.3 Control of Hex Expression

While Hex may not be required during the earliest events in lineage specification, it is clearly expressed there. Different enhancer elements that govern *Hex* expression have been identified upstream of the first exon and within the first and third introns. That present in the third intron controls the expression of *Hex* in the AVE and ADE (Rodriguez et al., 2001). In *Xenopus* and zebrafish, Nodal related signals (Osada and Wright, 1999; Zorn et al., 1999) and Wnt/β-catenin (Ho et al., 1999; Zorn et al., 1999) signals have been implicated in *Hex* regulation. *β-catenin* mutants in mouse result in a loss of *Hex* expression (Huelsken et al., 2000). Similarly in *Smad2* mutant mouse embryos, *Hex* is lost in both the AVE and ADE (Heyer et al., 1999; Vincent et al., 2003). Sequence analysis of the AVE/ADE enhancer reveals the presence of a LEF/Tcf binding site. This is the DNA binding element of the Wnt/β-catenin

pathway. Also present in this enhancer are 6 SIP1 binding sites. SIP1 is shown to be a partner for *Smad2* (Rodriguez et al., 2001; Verschueren et al., 1999). Taken together, these results would imply that Hex is downstream of the Wnt and Nodal signalling pathways. Indeed, the forebrain defects as a result of unspecified ADE seen in Hex mutants are reminiscent of those seen with reduced Nodal levels in the epiblast (as discussed above). During late gastrulation BMP signalling in chick and mouse overlap with *Hex* expression in the ventral foregut and it has been shown in chick that BMP signalling regulates *Hex* (Zhang et al., 2002).

Studies *in vitro* also indicate the influence of endoderm specific transcription factors in the regulation of *Hex*. Both *Gata4* and *FoxA2* have been shown to upregulate Hex expression (Denson et al., 2000b). It is thought that these factors act to permit chromatin accessibility to other transcription factors such as Bmp2 that is necessary for Hex expression during later development in chick (Zhang et al., 2004).

Such an observation of control by the primitive endoderm marker, *Gata4*, may also have implications for the expression of Hex in this tissue (see later).

1.6.4 Targets of Hex

To understand the means by which Hex regulates early cell fate decisions it is essential to understand the network downstream of *Hex*. Generally, homeodomain proteins recognise DNA sequences containing the core tetranucleotide ATTA (TAAT on the complementary strand) (Gehring, 1996). A Hex target sequence was defined by binding site selection as 5'-ATTAA-3' and this interaction are dependent on the homeodomain (Brickman et al., 2000; Crompton et al., 1992; Guiral et al., 2001; Pellizzari et al., 2000; Tanaka et al., 1999)

Several studies have examined modulation of expression by *Hex*. *Hex* represses the activity of the mouse thyroglobulin promoter in transient transfection assays (Pellizzari et al., 2000). Using similar methods, *Hex* represses luciferase production

from the Xenopus *Goosecoid* promoter, one of the first genes to be expressed in the organiser. An Antimorphic Hex protein (Hex-λVP2), consisting of Hex tethered to two domains of the potent minimum VP16 activation domain module interspaced with the flexible hinge sequence from bacteriaphage lambda repressor (Brickman et al., 2000). This and work by others has shown that the proline rich domain plays a role in the repressive activities of Hex (Brickman et al., 2000; Guiral et al., 2001; Tanaka et al., 1999). Indeed, this domain has been shown to bind to a co-repressive protein, Tle1, hence providing a mechanism by which Hex represses transcription (Swingler et al., 2004). In mouse and rat, it has been argued that Hex can activate the *sodium-dependent bile acid cotransporter* gene promoter in hepatocytes and that this activity is conferred by the C-terminal acidic region (Denson et al., 2000a; Kasamatsu et al., 2004). However, these data is solely based on low level luciferase assays and to date there is no clear evidence that *Hex* can activate transcription.

Xenopus Hex has been shown to be expressed in two centres equivalent to the AVE and ADE in mouse, and knock down of Hex in these centres points to a conserved role for these centres in anterior development (Smithers and Jones, 2002). The use of the Hex-λVP2 molecule has shown that *Cerberus* is indirectly upregulated by Hex while markers of the organiser, *Goosecoid* and *Chordin* are downregulated (Brickman et al., 2000). These results show that Hex promotes anterior identity by suppressing the posteriorising effects of the organiser. This is consistent with the findings in the *Hex* mutants that *Cer1* is lost, while axial mesendoderm (pan organizer markers, *FoxA2* and *Shh* are expanded).

Recently in our lab it has been shown in *Xenopus* and ES cells that Hex promotes the early Wnt signals that are required for axis formation while at a later stage it inhibits the posteriorising effects of continued Nodal signalling (Zamparini et al., 2006). Given the coexpression of *Hex*, *Cerl* and *Lefty1* in the AVE and ADE in mouse, and the role of *Hex* in Xenopus, it is thought that *Hex* might also play a role in repressing the effects of Nodal signalling. Despite being expressed in the ADE where high levels of *Nodal* are required, *Hex* might be acting on the autoregulatory loop of *Nodal* in order to define a boundary of anterior most endoderm.

Taken together, these data have shown that *Hex* appears to have many roles during development, including: haematopoeitic and endothelial proliferation; anterior-posterior axis formation by influencing Nodal and Wnt signalling; and development of the liver and thyroid in later development. While these studies have suggested a role for *Hex* in influencing anterior character through lineage specification of endoderm, it remains unclear whether *Hex* plays an important role in the earliest cell fate decisions involved in establishing the anterior axis. Chapter 4 of this thesis addresses the significance of early *Hex* expression in the primitive endoderm, while chapter 5 attempts to examine the consequences of ectopic Hex expression for early cell fate decision in embryonic stem (ES) cells.

1.7 Embryonic Stem Cells

1.7.1 Molecular basis for pluripotency and self-renewal

ES cells are an in vitro cell line derived from the inner cell mass (ICM) of the early embryo (Evans and Kaufman, 1981; Martin, 1981). In mouse they are defined functionally as a karyotypically normal immortal cell line that can give rise to all the can give rise to all the future lineages of the conceptus (Smith, 2001). They have the ability to self-renew under defined culture conditions indefinitely while retaining the capacity to differentiate *in vitro* or contribute to embryonic lineages upon injection into blastocysts. Importantly, they can be genetically manipulated by ablation of specific genes or introduction of a gene of interest into specific loci. Thus, such changes to the genome of ES cells can be propagated through the germ line of mice and therefore allow analysis of the effects of these genetic alterations *in vivo*.

The molecular basis for ES cell self-renewal relies on both extrinsic and intrinsic factors to suppress differentiation and maintain pluripotency. Conventional culturing conditions rely on the presence of serum and a feeder layer of fibroblasts. The key components supplied by these complex and undefined culture supports have been

identified. Fractionation of media conditioned by fibroblasts revealed that the active component produced by feeders and promoting self-renewal as leukaemia inhibitory factor (LIF) (Smith et al., 1988). Upon binding of LIF to its receptor gp130, dimerization occurs resulting in STAT3 dependent transcription. Indeed, constitutive activation of STAT3 permits ES cell self-renewal in the absence of LIF (Matsuda et al., 1999). Attempts to grow cells under serum free conditions revealed that members of the bone morphogenic protein family (BMP4 and BMP2) could support ES cell growth and renewal in the absence of serum (Ying et al., 2003). BMP signalling acts by stimulating members of the inhibitors of differentiation (Id) family resulting in a block to neural commitment (Ruzinova and Benezra, 2003).

In addition to the extrinsic requirements, expression of the three transcription factors *Oct3/4*, *Sox2* and *Nanog* is crucial for the pluripotent state (Chambers and Smith, 2004). *Oct3/4* is a POU domain homeobox transcription factor that is specifically expressed in all pluripotent cells during mouse embryogenesis. Heterodimerization with the HMG-box factor *Sox2* leads to control of downstream targets containing an octamer binding motif such as *Fgf4* (ATGCAAAT) (Ambrosetti et al., 1997; Avilion et al., 2003). *Oct3/4* deficient mouse embryos fail to develop beyond the blastocyst stage due to the lack of pluripotent ICM, showing an absolute critical role *in vivo* (Nichols et al., 1998). Reduction of *Oct3/4* transcription *in vitro* leads to the differentiation of ES cells to a trophectodermal (Niwa et al., 2000) or primitive endoderm (Hay et al., 2004; Morrison and Brickman, 2006) fate, while its overexpression to levels greater than 50% causes cells to progress into mesoderm and endoderm precursor state reminiscent of epilast stage embryos (Morrison and Brickman, 2006) . These reports demonstrate the nature of *Oct3/4* as a master regulator of the pluripotent state. The third of this 'trinity' is *Nanog*, a homeodomain transcription factor that confers LIF independent self-renewal upon its overexpression in ES cells (Chambers et al., 2003; Mitsui et al., 2003). As with *Oct3/4* and *Sox2*, *Nanog* is also expressed in the early blastocyst, however it is downregulated at an earlier stage of development prior to implantation (Hart et al., 2004). Attempts to derive *Nanog* null mice fail at an early stage whereby the epiblast is not formed correctly and extraembryonic tissue predominates (Mitsui et al., 2003).

This and other *in vivo* and *in vitro* studies have implied a role for *Nanog* in repressing the emergence of a primitive endoderm fate by inhibiting *Gata6* transcription in ICM and ES cells (Chazaud et al., 2006; Hamazaki et al., 2006).

Recently, numerous attempts to unveil the complexity of transcriptional networks involved in pluripotency of both mouse and human ES cells have demonstrated combinatorial occupancy of target promoters by *Oct3/4*, *Sox2* and *Nanog*. Indeed the binding of these transcription factors to their own and each others promoters reveals both feed-forward and autoregulatory circuitry and suggests a role for other genes in contributing the pluripotent state (Boyer et al., 2005; Chickarmane et al., 2006; Kim et al., 2008; Loh et al., 2006). Indeed, in both mouse and human contexts, forced expression of *Oct3/4* and *Sox2* together with *Klf4* and *c-Myc* can reprogram fibroblasts to become induced pluripotent cells (iPS cells), the quality of which is enhanced by selection of cells that express high levels of *Nanog* (Maherali et al., 2007; Okita et al., 2007; Park et al., 2008; Takahashi et al., 2007; Yu et al., 2007).

While mouse and human ES cells share some common features with regard to the expression of *Oct3/4*, *Sox2* and *Nanog*, they require very different culturing conditions in order to permit self-renewal. LIF is not sufficient maintain human ES cells and BMPs cause rapid differentiation. Instead, a balance between TGF-β and BMP signalling in combination with the presence of basic Fgf is necessary for their propagation in serum free conditions (James et al., 2005; Vallier et al., 2005; Xu et al., 2005).

Recently, a new class of ES cell-like cells have been isolated from the epiblast of postimplantation mouse embryos (E6.5). These Epiblast Stem Cells (EpiSCs) share culturing requirements and gene expression profiles more similar to human ES cells than their murine counterpart (Brons et al., 2007; Tesar et al., 2007). These differences between mouse ES cells and EpiSCs are reflective of their developmental origin. While EpiSCs inefficiently colonize the developing embryo following blastocyst injection, this restricted character may prove beneficial for directing

unidirectional commitment during *in vitro* differentiation protocols, an accomplishment difficult to achieve with ES cells.

Current work in Austin Smith's lab has demonstrated a redundancy for traditional mouse ES cell culturing conditions, whereby LIF and BMP signalling are dispensable (Ying et al., 2008). FGF signalling normally provokes the tendency of ES cells to differentiate and this appears dependent on the Erk1 and Erk2 branch of the FGF pathway (Kunath et al., 2007). Suppression of this pathway with various chemical inhibitors permits ES cell growth, albeit inefficiently. There is also accumulating evidence implying a role for Wnt signalling in maintaining the pluripotent state. Thus, inhibitors of Gsk3β that lead to activation of the canonical Wnt pathway have been shown to support both human and mouse ES cell self renewal (Sato et al., 2004). The stimulation of the canonical Wnt signalling alongside antagonists of the FGF pathway overcomes the efficiency problems of maintaining pluripotency by FGF suppression alone when culturing ES cells without LIF and serum (Ying et al., 2008). As these studies were done in a completely defined system, it can be concluded that the inhibition of FGF signalling appears to be a dominant mechanisms for regulating self renewal. In conjunction with the induction of β-catenin, this pathway is able to completely support ES cells and that BMP and Lif signalling maybe acting on a cell state downstream or epistatic to these principle cues.

1.7.2 Heterogeneity of ES cells

While ES cells can be described based on morphology, the presence of cell surface markers such as SSEA1 and PECAM, as well as expression of the key transcription factors such as *Oct3/4*, *Sox2*, *Nanog* and all the ECATs, they are defined based on the retrospective function (Chambers and Smith, 2004; Cui et al., 2004; Mitsui et al., 2003). Thus, a culture of cells from the early embryo can be referred to as ES cells if some component of that culture is capable of continually producing cells that are able to differentiate into all the lineages of the conceptus. Upon injection into the

blastocyst only a small number ever contribute to the embryo raising the question of functional equivalence among ES cells. Quantification of the number of founder ES cells, suggest only around 20% of the injected cells actually contribute at all and yet the majority of ES cell cultures are *Oct3/4* and SSEA1 positive (Wang and Jaenisch, 2004). Thus as with other adult stem cell populations, ES cells are not clearly defined based on marker expression, but rather a fraction of cells that contain functional cell types.

So what determines the functionalality of an ES cell? While well maintained undifferentiated ES cell colonies generally express *Oct3/4*, *Sox2* and SSEA1 relatively homogenously, certain pluripotency factors appear to be expressed in a sub-population of the culture. Thus, ES cells are known to express *Nanog* and *Rex1* heterogeneously (Chambers et al., 2007; Singh et al., 2007; Toyooka et al., 2008). This heterogeneity does not appear to reflect spontaneous differentiation, but rather it reflects a range of gene expression levels in morphologically homogenous colonies or cultures. Moreover, this heterogeneity is also reversible. As a result, one might ask whether this heterogeneity of gene expression has functional significance. Chapter four of my thesis in part addresses this issue.

ES cells are derived from ICM cells. In embryonic development, ICM cells of the expanded blastocyst stage can contribute to the primitive endoderm as well as the fetus (Chazaud et al., 2006; Gardner, 1985a), however, a day later, the pluripotent cells of the 5th day blastocyst (now referred to as primitive ectoderm) are then unable to form primitive endoderm (Gardner, 1982; Gardner, 1985a; Gardner and Rossant, 1979). Primitive endoderm cells of the peri-implantation blastocyst (E4.5-E5.0) are restricted to their own lineage by colonizing the visceral and mostly parietal endoderm in chimera experiments (Gardner, 1982; Gardner, 1985b; Hogan and Tilly, 1981). While derived from the early ICM, ES cells are thought to behave like primitive ectoderm when injected into blastocyst, i.e. they predominantly contribute to the definitive germs layers of the embryo but can not contribute to the primitive endoderm (Rossant, 2008). However, careful inspection of the literature suggests they can also colonize the descendants of the primitive endoderm lineage following

blastocyst injection and might therefore be considered the equivalent to early ICM cells (Beddington and Robertson, 1989). *In vitro*, ES cells can generate primitive endoderm like cells by the removal of LIF or by the overexpression of *Oct3/4*, which leads to a moderate upregulation of *Gata4* (Niwa et al., 2000), a transcription factor restricted to primitive endoderm derived lineages of the early embryo (Arceci et al., 1993). Indeed, forced expression of *Gata4* or the related upstream factor, *Gata6*, results in primitive endoderm differentiation leading to terminal differentiation of this lineage to parietal endoderm (Fujikura et al., 2002). Indeed, this phenotypic change results in a morphological similarity to extraembryonic endoderm (Xen) cells that are derived from the primitive endoderm and display many characteristics of parietal endoderm . Interestingly, ES cell cultures appear to contain a basal level of *Gata4* and *Gata6* expression, suggesting background levels of primitive endoderm gene expression (Fujikura et al., 2002; Koutsourakis et al., 1999). Thus, expression of lacZ from the Gata6 locus is reported to have a heterogeneous expression pattern in ES cell clones that maybe suggestive of a subpopulation with tendencies towards the primitive endoderm fate. Another marker of the primitive endoderm as seen by *in situ* hybridisation in E4.5 embryos is the homeobox transcription factor *Hex* (Chazaud et al., 2006; Thomas et al., 1998) and like the GATA factors, low level *Hex* transcripts are also detectable in ES cell cultures (Kubo et al., 2004).

1.7.3 Expression of Hex in ES cells

In this thesis, I look at the consequences of *Hex* expression in ES cell culture. In chapter 3 I describe the construction of a unique set of reagents that allows me to both follow and induce Hex expression in ES cell culture. In chapter 4 I describe some exciting findings I have made using one of these reagents and focusing on the functional significance of ES cell heterogeneity. Here I used a translational amplifier coupled to a bright (Venus) fluorescent protein to give a read out of extremely low levels of Hex expression in ES cells. This fluorescent read out of transcriptional noise in ES cells suggests the morphologically normal undifferentiated ES cells may fall into at least three classes. The Venus positive sub-fraction of these colonies

expresses relatively higher levels of many specific primitive endoderm markers such as *Gata6* and reduced levels of a wide array of pluripotency markers. However, both Venus positive and negative cells express equivalent levels of the ES cell marker *Oct3/4*. FGF signalling can be used to manipulate the level of the Venus positive population and overexpression of *Nanog*, greatly reduces it. Clonal analysis shows that these populations interconvert efficiently *in vitro*, while when placed back into the blastocyst, only the Venus negative population is able to efficiently contribute to the somatic lineages in chimeras. Taken together, these data suggest that low level transcriptional noise defines several interconvertable states in ES cell cultures and that these states have functional significance for the mechanism by which these cell maintain pluripotency. Additionally, genes downstream of TGFβ/Nodal signalling as well as components of the Wnt pathway appear to be associated with the Venus positive population.

Given the association of primitive endoderm marker expression with Venus (Hex) positive cells, there may be a 'pre-primitive endoderm' subpopulation existing in ES cell cultures. In chapter 5 of this thesis I attempt to force the expression of *Hex* in ES cells in order to look for a phenotypic change similar to that observed with *Gata4* and *Gata6* overexpression (Fujikura et al., 2002). While low levels of Hex do not induce any morphological change, high level expression produced from an inducible system lead to apoptosis.

Chapter 2

Materials and Methods

2.1 Cloning

2.1.1 Restriction Digests

1μg of DNA was incubated in a 20μl reaction with 2μl of appropriate digest buffer and 1μl (usually 10 units) for 3 hours at 37°C. All restriction enzymes were from New England Biolabs or Roche Diagnostics.

2.1.2 Dephosphorylation of DNA

At the end of restriction digestion of cloning vectors, 1μl (10 units) of Calf Intestinal Phosphatase (CIP) (New England Biolabs) was added to the reaction and incubation at 37°C was continued for 30 minutes. This ensured removal of 5' phosphates from DNA ends preventing recircularization.

2.1.3 Ligation of DNA fragments

Generally, ligation of purified insert and cloning vector fragments took place in a 3:1 ratio. Ligation reaction were performed in a 20μl volume at r.t. for 10 minutes using the Quick Ligation Kit (New England Biolabs) by adding 10μl of 2x Quick Ligation buffer and 1μl (2,000 units) of Quick ligase enzyme.

2.1.4 Plasmid transformation

For subcloning, around 10 ng of DNA was transformed into DH5α subcloning efficiency chemically competent *Escherichia coli* (*E.coli*) (Invitrogen). For cloning, 5μl of ligation reaction were transformed into One Shot TOP10 chemically competent *E.coli* (Invitrogen). For generation of contructs greater than 10kb those containing loxP sites, 5μl of ligation reaction were transformed into MAX Effiency

Stbl2 competent cells (Invitrogen) to prevent undesired recombination events. In each case, the transformation procedure was carried out according to the manufacturers instructions. Transformants were plated out on LB agar plates containing eiither 50μg/ml kanamycin or 100μg/ml ampicillin (according to resistant gene present on each vector) and incubated overnight at 37 °C or (30°C for Stbl2 cells) .

2.1.5 Preparation of plasmid DNA

For small scale preparation, a single bacterial colony was inoculated and grown overnight at 37 °C in 3 ml of LB, or at 30 °C in TB (for Stbl2 cells), containing 100μg/ml of ampicillin or 50 μg/ml kanamycin. 1.5 ml of was collected by centrigugation and plasmid DNA was extracted using the Qiagen Mini-prep Kit (Qiagen) to extract DNA according to the manufacturer's instructions.

For large scale preparation, a single bacterial colony was inoculated into 100 ml of LB or TB containing 100μg/ml of ampicillin or 50μg/ml kanamycin and shaken overnight at 37 °C or at 30 °C. the Qiagen Maxi-prep kit (Qiagen) was used to extract DNA according to the maufacturer's instructions.

2.2 General Molecular Biology

2.2.1 Preparation of Genomic DNA

Genomic DNA was extracted from approximately $1x10^6$ ES cells using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions.

2.2.2 Preparation of RNA

RNA was extracted from usually $1x10^6$ ES cells (or differentiated derivatives) grown in culture or following separation by Flow Cytometry. Extraction was performed using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. During this protocol a step of DNase1 treatment for 15 minutes at r.t. was carried out to prevent contamination of downstream rtPCR reactions by residual genomic DNA.

2.2.3 Purification of DNA

For small scale nucleic acid clean-up of enzymatic reactions, samples were purified through QIAquick Gel Extraction Kit columns, or through QIAquick PCR Purification Kit columns (following PCR reactions) according to manufacturer's instructions (Qiagen).

For large scale clean-up, phenol/chloroform/isoamyl-alcohol (Sigma) was added in a 1:1 volume ratio to enzymatic reaction and mixed thoroughly. After 5 minutes of centrifugation in a microfuge at 13,000 r.p.m., the upper aqueous phase was transferred to a fresh eppendorf tube. DNA was precipitated from the aqueous phase by adding 1:10 volume of 3M NaOAc and 2 volumes of 100% EtOH and incubation at -20°C for 30 minutes. Following centrifugation for 30 minutes at 13,000 r.p.m. the DNA pellet was washed in 70% EtOH, dried and resuspended in nuclease free water (Ambion).

2.2.4 Agarose gel electrophoresis of DNA

1% agarose gels were prepared in 1 X TAE solution containing ethdium bromide (0.5μg/ml)(Sigma). DNA samples containing 1x loading buffer solution were loaded onto gels and electrophoresis in an electrolyte of 1 X TAE. Voltage chosen depended on distance between electrodes: thus, 5-8Volts/cm. DNA was visualised using UV light and the size was estimated using a 1kb ladder (Invitrogen).

2.2.5 Purification of DNA fragments from agarose gels

DNA bands of interest were excised from gels using a scalpel under UV light. Purification DNA was carried out using the QIAquick Gel Extraction Kit (Qiagen) following the manufacture's instructions.

2.2.6 Quantification of Nucleic acids

1.5µl samples od DNA or RNA were directly pipeted onto a Nanodrop spectrophotemeter (Thermo Scientific) to measure concentration (A260 reading of 1 = 40ng/µl for RNA and 50ng/µl for DNA) and to judge purity (expected A260/A280 ratio is 2.0 if nucleic acid is clean).

2.2.7 Reverse transcription of mRNA

Reverse transcription of mRNA was carried out using the Superscript II RNaseH-Reverse Transcriptase system (Invitrogen). Briefly, 1µg of total RNA was incubated for 5 minutes at 65°C with 1µl of 1µM random primers and 1µl of 10mM dNTPs. 4 µl of first strand synthesis buffer together with 1µl 0.1M Dithiothreitol and 10 units of RNaseOut solution, were added together with nuclease free water to a total volume of 19µl. 1µl of Superscript II enzyme or 1µl of water was then added and the mix was incubated at 50°C for 50 minutes. Resultant cDNAs were then treated with 10 units .RNaseH for 20 minutes at 37°C. Following heat inactivation at 70°C for 10 minutes cDNA samples were diluted 1:5 and stored for PCR analysis.

2.2.8 Polymerase Chain Reaction

Diagnostic polymerase chain reaction (PCR) was carried out using Taq DNA Polymerase (Qiagen) together with 200μM (each) of dNTPS, 0.5μM each primer (forward and reverse), in 1 X PCR buffer with 0.5units of Taq together with 100ng (for genomic PCR) or 1ng (amplification from plasmid DNA) or 5μl (of 1:5 dilution of cDNA reaction for rtPCR) of template DNA in a total volume of 50μl. Cycling conditions were carried out according to manufacturer's instructions with annealing temperature varying depending on Tm of primers.

For preparative PCR, Phusion High Fidelity DNA Polymerase (Finnzymes) was used to amplify sequences in order to minimise potential mutational events. Concentrations of components in a reaction were similar to those for diagnostic PCR.

2.2.9 TOPO cloning of PCR products

PCR products amplified were cloned using the TOPO TA or ZeroBlunt TOPO (for Phusion generated amplicons) cloning Kits according to the manufacturer's instructions. Products cloned were those with restriction sites incorporated onto the ends of primers. Additionally, as template controls for quantitative PCR, ORFs of various genes were amplfied by rtPCR were cloned. Primers used to amplify each gene ORF were those also used for quantitative rtPCR measurement of those genes (see below).

2.2.10 Quantitative rtPCR

Primers for quantitative rtPCR (qPCR) were designed using the Universal ProbeLibrary (UPL) Assay Design Centre (Roche, see: https://www.roche-applied-science.com/sis/rtpcr/upl/acenter.jsp?id=030000) . Primers were designed to span an intron to ensure reduced likelyhood of amplification from residual genomic DNA. All primers are chosen to have an anneling temperature of 60°C For each set of primers designed, a specific probe from a mouse UPL set (Roche) was suggested to

be used. This set consists of around 165 8mer or 9mer probes which together offer coverage of the entire mouse transcriptome. In this system measurement of increase of fluorescence release from UPL probes reflects production of de novo amplicons. In combination with gene specific primers, amplicon specific UPL probes ensure a further level of specificity. This reduces the chance of artifact measurement to an extremely low probability. 3µl of diluted cDNA, together with 5µl LightCycler480 Probes Master mix (Roche) and 1µl of each primer (10pmol each) were used for each reaction. Additionally, 0.1µl of specific probes from a Mouse UPL set were used for each amplicon. Cycling using the standard UPL default conditions took place in a LightCycler480 (Roche). All reactions were done in triplicate to ensure against pipetting errors. Estimation of copy number readout during qPCR was made possible by inclusion of a dilution series of known amounts control plasmid DNA containing an ORF relating to the gene of interest.

Some qPCR was performed using the SyberGreen System (Roche). This system does not use the UPL probes but measures the incorporation of SyberGreen into *de novo* DNA. Similar quantities of cDNA and primers were used for this method A dilution series of standards was also included. Below are the list of primers (shown as 5' to 3')used for qPCR:

Gene	Forward Primer	Reverse Primer	UPL probe number
Tbp	ggggagctgtgatgtgaagt	ccaggaaataattctggctca	11
Hex	ctacacgcacgccctactc	cagaggtcgctggaggaa	50
Cerberus	gactgtgcccttcaaccag	agcagtgggagcagaagc	105
Gata6	ggtctctacagcaagatgaatgg	tggcacaggacagtccaag	40
Oct3/4	gttggagaaggtggaaccaa	ctccttctgcagggctttc	95
Nanog	cctccagcagatgcaagaa	gcttgcacttcatcctttgg	25
Rex1	gatgcacaacgatccagattt	tggaaattagaacgtacatcctccaa	18
Klf4	tgcagtcacaagtcccctct	gaccttcttcccctctttgg	82
Actin	gtgggccgctctaggcaccaa	ctctttgatgtcacgcacgatttc	n/a
Bf,1r	ggtggcctgaatgacatctt	gcatacagcgggactcccacg	n/a

Quantitative analysis was performed with the assoiciated LightCycler software such that cycle crossing point values were translated into absolute copy numbers. Copy number values obtained for the housekeeping genes Tbp or Actin were used to normalise the values obtained for the gene of interest.

2.2.11 Sequencing of DNA

Sequencing of DNA was performed using the Big-Dye Terminator Cycle Sequencing kit. Purified plasmid DNA (400ng) or purified amplicon DNA following genomic PCR (100ng) was sent together with 1µl sequencing primers (concentration of 1.6pmol/µl) to the sequencing service at Edinburgh University (Ashworth buikding). Sequence analysis was performed using Lasergene v7.0 software (DNASTAR, Madison, USA).

2.2.12 Southern Analysis

Digested genomic DNA was subjected to electrophoresis o/n in 0.6% agarose gels. Gels were first treated in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 30 minutes followed by treatment in neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl, pH5.5) for a further 30 minutes. Blotting onto HyBond membrane (Amersham) was achieved by solution by capillary transfer in 10X SSC (20X stock, 3 M NaCl, 0.3 M sodium citrate, 1 mM EDTA) o/n.

Membranes were baked at 120°C for 1 hour to ensure crosslinking, followed by prehybridisation for 2 hours at 65°C in Hybridisation solution (5X SSC, 5X Denhardt's solution (100X solution, 2% (w/v) BSA, 2% (w/v) Ficoll, 2% (w/v) polyvinylpyrrolidone), 0.5% SDS).

Probes used for southern analysis were generated by PCR amplification of 5' and 3' *Hex* genomic genomic regions and Venus sequence (as an internal probe). Primers (shown as 5' to 3') used were as follows:

5' external probe; F ccccttatctttcccttgg, R cagttgaagggaatgcacct

3' external probe; F cctttcatccagtgggagag, R ctgccccttcaacatgaaat

Internal probe; F cacatgaagcagcacgactt, R gaactccagcaggaccatgt

Products from the above amplifications were cleaned by column purification (as described above). 25ng of purified DNA were labelled using the MegaPrime labelling kit (Amersham) together with 50μCi of ^{32}Pα-dCTP (Amersham) according to the manufacturer's instructions. Labelled probes were cleaned through G-50 sepharose size exlcusion columns (Amersham), denatured at 100°C for 5 minutes before adding to the hybridisation solution. Following hybridisation o/n at 65°C, membranes were washed in 2% SSC, 0.1% SDS solution until counts were reduce to around 20 c.p.m. Membranes were then exposed to autoradiography film (Kodak).

2.2.13 Extraction of Protein and Western Analysis

Protein was prepared from approximately 1x10^6 ES cells by boiling in 30 μl Laemlli buffer (BioRad) for 20 minutes. Following centrifugation to remove cell debris, 30μl samples were loaded onto precast NuPAGE Novex 10% Bis-Tris Gels (Invitrogen). Following electrophoresis at 200 volts for 45 minutes, protein was tranferred electrophoretically onto ECL membranes at 360mA constant current for 1 hour.

Membranes were blocked for 1 hour at 4°C in 1 X TBS solution (25mM Tris, pH 7.4, 3.0mM KCl, 140mM NaCl and 0.05% Tween 20) containing 3% BSA. Antibodies were then added at a dilution according to manufacuter's instructions for o/n incubation at 4°C. Antibodies used were as follows: Rabbit anti caspase 3 (NEB, 9662), Rabbit anti GFP (Invitrogen, A6455), Mouse anti α Tubulin (Santa Cruz, sc-5286), Strepavidin-HRP (Cambridge BioSciences). Membranes were washed in TBST solution and incubated with HRP-conjugated secondary antibodies (anti

mouse and rabbit, Cambridge Biosciences) in blocking solution for 1 hour at r.t. Detection was performed with ECL kit (Pierce) according to manufacturer's instructions.

2.2.14 Precipitation of Bio-tagged proteins

$1x10^6$ ES cells were collected and lysed with lysis buffer (Pierce) containing protease inhibitors (Roche). Lysates were bound to M-280 Dynabeads (Invitrogen) in solution according to reported method (de Boer et al., 2003). Following several washing steps, bound proteins were released by boiling in Laemmli buffer and subjected to western analysis with anti-GFP antibody or Strepstavidin-HRP.

2.2.15 Site Directed Mutagenesis

Site directed mutagenesis on vector pTLC BioHex (see Chapter 5.9) was carried out using the QuickChange XL Site Directed Mutagenesis Kit (Stratagene) according to the manufacturer's instructions. The following pairs of mutagenic oligonucleotides were used for the generation of two different mutants:

pTLC BioHex Ebm,

5' tgcgcccacgccggcggcgcagcccgctca 3'

5' gtgagcgggctgcgccgccggcgtgggcgc 3'

pTLC BioHexDbm,

5' tcagtgagagacaggtcaaaacctggtttgcggctcgccgagctaaatg 3'

5' catttagctcggcgagccgcaaaccaggttttgacctgtctctcactg 3'

2.2.16 Other Oligos used for cloning

Biotin tag sequence oligos (see Chapter 3.1)

5' -TCGAGGCCACCATGGCTGGTGGCCTGAATGACATCTTTGAGGCCC
AGAAGATCGAGTGGCATGAGAACCTGTACTTCCAGGGAGCCGG 3'

5' AATTCCGGCTCCCTGGAAGTCCAGGTTCTCATGCCACTCGATCTT
CTGGGCCTCAAAGATGTCATTCAGGCCACCAGCCATGGTGGCC- 3'

Cloning of Hex ORF (see Chapter 3.1)

5'-GCGTCGACTATGCAGTTCCCGCACCCGGGG-3'

5'-TACCGGTCCTCCTCCTCCTCCAGCATTAAAGTAGCCTTT-3'

Cloning of Inducible pTLC vectors (see Chapter 3.6)

5'- GTTAATTAAGCCACCATGGCTGGTGGC-3'

5'-GGCTAGCTTAGGCTCCCTGGAAGTACAG-3'

5'-GGCTAGCTTATCCAGCATTAAAGTAGCC-3'

Modification of Bio tag sequence for targeting vector (see Chapter 3.7)

5' –ATTATTAAGATGGCTGGTGGCCTGAAT-3'

2.2.17 Constructs

pEGFPN1 (Clontech); pNEB193 (New England Biolabs); pGTIV2 (by William Stanford); pCAGSIH and pCAGSIP (by Hitoshi Niwa); pCAG-Hex (ME204) (by Osmany Larralde/Josh Brickman); pβ-catPCAGSIP (a gift from Austin Smith's lab); pCAGSNan-IP (a gift from Ian Chamber's lab); Hex targeting vector, pT1pA (a gift from S. Srinivas and T. Rodriguez); TOPFlash and FOPFlash reporters (Upstate Biotechnology); pSP64T-XB (J. Brickman lab); pRL-SV40 Renilla (Promega). pBirA (a gift from Ita Costello).

2.3 Xenopus Experiments

2.3.1 RNA for *Xenopus* injections

Transcription of RNA was performed in: 5µl 10X transcription buffer (Ambion); 6µl 100nM DTT (Promega); 5µl each of rATP (10mM), rCTP (10mM) , rUTP (10mM) and rGTP (1mM), 2.5µl RNasin (Promega), and 6 µl nuclease free water. Incubation occurred at 37°C for 10 minutes. Following centrigugation to collect reaction

atbottom of the tube, the following components were added: 3µl of 1µg/µl linearised DNA plasmid template, 5µl of 5mM RNA Cap structure analog (New England Biolabs) and 2.5µl of appropriate RNA polymerase (Ambion). Reaction was incubated for 30 minutes at 37°C upon which 2.5 10mM rGTP was added and the reaction continued for 1 further hour. 50µl of nuclease free water and 5µl RQ1 RNase free DNase 1were added and incubated for 30 minutes at r.t. RNA was purified using Chroma Spin 100 columns (Clontech).

2.3.2 Injection of RNA

RNA was kindly injected into single dorso-vegetal cells of 4 cell stage embryos by Andrea Zamparini in our lab (Thank you Andrea).

2.3.3 X-gal staining and analysis of embryos

Following injections, embryo were washed in 1 X PBS (80mM sodium phosphate, 15mM potassium phosphate, 27mM KCl and 1.37M NaCl) and then fixed in fixing solution (1 X PBS, 2mM $MgCl_2$, 5mM EGTA, 15% paraformaldehyde, 0.2% Glutaradehyde, 0.02% NP-40) and observed for phenotypic change. To confirm the site of injection of βGal RNA, embryos were then stained in X-Gal staining solution (5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 2 mM MgCI, 0.01% sodium deoxycholate, 0.02% Nonidet P-40 (NP-40) in PBS) o/n at r.t. Embryos were washed twice in PBS before fixing in MEMFA solution (0.1M MOPS, pH 7.4, 2mM EGTA, 1mM $MgSO_4$, 4% paraformaldehyde) for 1 hour at r.t.

2.4 Cell culture

2.4.1 Cell Lines

The following Mouse ES cells lines were used in this study: CGR8 and R26CreErt2 E14Tg2a (Austin Smith's lab), R26BirA E14Tg2a (Driegen et al., 2005), Hex-Redstar (HRS) (Morisson et al., in press). Additionally, Human Embryonic Kidney (HEK293T) and p19 Embryonal Carcinoma (EC) cells were kindly supplied by Austin Smith's lab. IM8A1 Extraembryonic endoderm (Xen) cells were kindly supplied by Tilo Kunath.

2.4.2 Cell culture conditions and maintenance

ES Cells were cultured on 0.1% gelatin flasks or plates (IWAKI) in Glasgow modified Eagle's medium (Gibco) containing, non-essential amino-acids, glutamine and sodium pyruvate, 01.mM mercaptoethanol and 10% Fetal Calf Serum (FCS) together with LIF. For long term storage, cells were stored in the above media supplemented with 10% DMSO in liquid nitrogen. Culturing occurred at 37°C and 7% CO_2. Xen cells were cultured similalrly as were HEK293 and P19 EC cells with the exception that no LIF was used.

Cells were passaged by removal in 0.1% trypsin solution or Cell Dissociation buffer (Gibco). Trypsonised cells were pelleted by centrifugation at 1200 rpm for 3 minutes. Following a single wash in PBS, cells were pelleted again and resuspended in fresh media. Around one fifth of the cells were plated into pre-gelatinised flasks. Generally, all cell lines were successively passaged every 3-4 days, with media changes every 48 hours.

2.4.3 Transient transfection of DNA into HEK293 cells

To quickly monitor fluorescence expression of various vectors following their construction, vector DNA was transfected into HEK293 cells using Lipofectamine2000 (Invitrogen) according to the manufacturer's instructions. Following 48 hours of culture, cells were visualised by fluorescence microscopy.

For TOPFlash/FOPFlash assay (Upstate Biotechnology), HEK293 cells were plated into 24 well plates at a density of $2x10^5$ cells/ml. Following growth for 24 hours, cells were transfected using Lipofectamine 2000 with the following vectors: 10ng each of TOPFlash or FOPFlash reporter, 10 ng of internal control plasmid pRL-SV40-Renilla to gether with a total of 100ng test vectors. Total DNA was adjusted to 900ng with pBSK (Bluescript) for each transfction. 48 hours after transfection, cells were collected and monitored for luciferase activity (from both reporter and control vectors) using the Dual Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions. Exceptions were 10µl of lysate (instead of 20µl) and 50µl each of Luciferase Assay Reagent II (LAR II) and STOP and Glo Reagent (instead of 100µl).

2.4.4 Transfection of Cells for gene targetting or stable integration

Prior to electroporation, ES cells were grown to approach confluency ($2.0\text{-}10 \times 10^7$). Cells were collected using trypsin, washed twice in PBS and resuspended in 800µl of PBS. 100µg of linearized DNA was transferred to an electroporation cuvette (BioRad) together with the cells. Cells/DNA mixture was left at RT for 3 minutes. Electroporation waas carried out with a BioRad GenePulser using the following settings: Capacitance, 3 µF; Voltage, 800V; and time constant, 0.1 seconds. Cells were transferred into prewarmed media before plating at a density of 5x105 cells per pre-gelatinised 10cm dish. Following o/n incubation, appropriate antibiotics were added to select for transfectants. After 2 weeks growth, single surviving colonies were physically removed and transferred into individual wells of a pre-gelatinised 96 well plates. Expansion of clones was achieved by successful passaging into wells of greaater size.

2.4.5 Differentiation of ES cells to ADE

Aggregate suspension cultures were achieved by plating cells at a density of 5000 cells per ml in 10cm bacterial petri dishes (Sterllin). Cells were plated in 10ml standard ES cell media with serum (as above) but without LIF. Following 2 days growth, embryoid bodies were collected and resuspended in 8ml N2B27 Culticell media (Stem Cell Sciences) with or without 20ng/µl Activin and cultured for a further 2 days upon which cells were changed to the same media conditions. Embryoid bodies were usually collected at day 7 of the diferentiation protocol.

2.4.6 Immunocytochemistry

Cells grown in 12 well plates were washed 2 X in PBS before fixation in 4% paraformaldehyde. Cells were then permeabilised in PBST (1 X PBS, 0.1% Triton X (Sigma)). Blocking was performed by adding 1% Bovine serum albumin (Sigma) in PBST solution to the fixed cells for 30 minutes at r.t. Primary antibodies were added at a dilution of 1:1000, and incubation continued o/n at 4°C. Following 3 X 10 minute washes in PBST, Alexa568 conjugated secondary antibodies diluted (1:1000) in block solution were added to the cells and incubation took place at r.t. for 1 hour. Also included at this step was DAPI solution (1:1000). Finally, cells were washed 3 times, then stored in PBS. Primary antibodies used were: mouse anti-Oct3/4 (Santa Cruz) and rabbit anti-Nanog peptide specific antibodies (α75+35, a kind gift from Ian Chamber's lab). Secondary conjugated antibodies (Alexa568) against mouse and rabbit were obtained from Invitrogen.

2.4.7 Fluorescence Activated Cell Sorting (FACs)

ES cells or embryoid bodies were collected into Cell Dissociation Buffer (Gibco) and incubated at 37°C for 10 minutes. Single cells suspension was achieved by gentle repeated pipetting. Following washes in PBS, cells were resuspended in 500µl FACs buffer (1 X PBS, 10% FCS) and 7AAD solution (BD Pharmingen, 5µl/1x10^6 cells) which marks dead cells. Analysis of fluorescence took place in a FACSCalibur flow

cytometer (BD Biosciences). Dotplots were generated using CellQuest software (BD Biosciences).

In the case of additional labelling of specific cell surface proteins, primary antibodies were added at a dilution of 1:1000 to cells resuspended in FACs buffer. Incubation took place for 10 minutes on ice. Following 3 washes in FACs buffer, cells were resuspended in frsh FACs buffer containing appropriate conjugated antibody at a dilution of 1:1000 and incubated as before. After 3 washes in FACs buffer, cells were finally resuspended in 500µl FACs buffer and analysed as above.

For collection of populations, cells were prepared similarly and subjected to flow cytometry using the MoFlo MLS high speed sorting aparatus (DakoCytomation). Cells were collected in FACs buffer and stored on ice for further analysis.

2.5 Mouse Experiments

2.5.1 Mouse Embryo Analysis

Embryos were dissected in Dulbecco's modified Eagle's medium (DMEM) containing 10mM sodium Hepes buffer (pH 7.4) and 5% fetal bovine serum using a dissection microscope. Visualisation was performed using a Zeiss Axiovert fluorescence microscope for bright and YFP fields.

X-gal staining of embryos occurred as follows. Embryos were washed in PBS solution (80mM sodium phosphate, 15mM potassium phosphate, 27mM KCl and 1.37M NaCl) then fixed with X-gal fix solution (1 X PBS, 2mM MgCl2, 5mM EGTA, 15 paraformaldehyde, 0.2% Glutaradehyde, 0.02% NP-40) at 4°C for 20 minutes. Following 3 X 20 minute washes in PBS they were then stained with X-gal staining solution (5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 2 mM MgCl, 0.01% sodium deoxycholate, 0.02% Nonidet P-40 (NP-40) in PBS) o/n

in the dark at r.t. Following 3 X 5 minute washes in PBS, stained embryos were then fixed in 4% paraformaldehyde.

2.5.2 Generation of chimeras

Chimera mouse generation was performed by morula aggregation with or injection of ES cells into host blastocysts. Injected or aggregated blastocysts were then tranferred into psuedopregnant recipient mothers. All methods and husbandry of animals were performed by Jan Ure (Transgenic mouse service, ISCR, University of Edinburgh) and Carolyn Manson (Animal House facilty, ISCR, University of Edinburgh).

2.5.3 Histology

Embryos were fixed in 4% paraformaldehyde for 30 minutes to 2hr at 4°C before embedding. Tissue was then sunk in 30%sucrose/PBS, frozen in Tissue Teck, and sections were cut in a cryostat. Sections were collected on Polysine microscope slides (VWR International), air-dried for 30 minutes to 1hr, and stored at -20°C until used.

2.6 Microarray analysis

RNA prepared from FACS sorted populations was sent to collaborators, Minoru Ko and Alexei Sharov at the NIH in Baltimore. Biological and technical replicates for each population were hybridised to NIA Mouse 44K Microarray v2.1 (Carter et al., 2005). Pairwise comparisons were performed by A. Sharov using standard statistical conditions (FDR<0.05, > 1.5-fold expression levels) to unveil genes upregulated or downregulated between the populations. Log intensity plots for each gene were created to find pattern matches between those of similar tissue origin.

Chapter 3

Generation and Analyses of DNA Constructs

Introduction

The aim of my PhD was to generate materials to follow *Hex* expression at both the transcript level and at the protein level and if possible to characterize Hex function in specific cell types and sub-cellular domains. In the long run this might extend to the identification of cell type specific transcription factor targets and Hex associated complexes.

The strategy I chose to follow Hex protein and as a potential platform for further biochemistry employed in vivo biotinylation as a means to tag the Hex protein. To accomplish this, I used BirA system developed by deBoer and colleagues. This system involves the expression of bacterial BirA ligase in mammalian cells and leads to the in *vivo* biotinylation of a protein of interest that is fused to a small (23aa) artificial peptide tag. The resulting biotinylated protein has an affinity for streptavidin several orders of magnitude stronger than that existing in any epitope-antibody interaction and can therefore be isolated with a great degree of specificity (de Boer et al., 2003). This technique has been successfully employed in chromatin immunoprecipitation (ChIP) procedures to identify the DNA target sequences bound by several biotinylated transcription factors that are important in ES cell self-renewal (Kim et al., 2008). Additionally, this system has been used to examine *in vivo* protein-protein interactions for the transcription factors *GATA1* and *Nanog* (Rodriguez et al., 2005; Wang et al., 2006). Recently, an ES cell line (R26BirA) has been produced with the bacterial BirA ligase gene inserted into the ROSA26 allowing its ubiquitous expression in mice derived from this line (Driegen et al., 2005; Zambrowicz et al., 1997). Expression of a bio-tagged Hex (BioHex) protein in these cells would ensure its biotinylation whether they are maintained under self-renewing conditions, allowed to differentiate *in vitro* or used to generate mice. Thus these cell lines can be used to follow Hex protein localization during ES cell differentiation or in embryonic development and for biochemical techniques such as ChIP and protein-protein interaction studies in specific cell types, both *in vivo* and *in vitro*.

While the BirA tag is useful of for biochemical techniques, it does not give a real time, live cell read out of Hex expression. The ability to use flow cytometry to purify populations of cells expressing BioHex would allow downstream functional analysis more straight foreword. To this end, cloning strategies were designed to allow co-expression of a fluorescent reporter with BioHex. These included the construction of BioHexGFP fusion proteins as well as the incorporation of internal ribosomal entry site (IRES) (Chappell et al., 2000; Mountford et al., 1994) to generate bicistronic messages in which the fluorescent protein would be expressed at the same time as BioHex. Fluorescent reporters included dsred2, EGFP or the YFP variant Venus (Nagai et al., 2002; Shaner et al., 2005). This chapter will describe the generation and analysis of a set of BioHex constructs with the aim of either expressing it at physiological levels from the *Hex* locus following a knock-in strategy (chapter 4) or overexpressing the protein in ES cells (chapter 5).

3.1 Generation of BioHexGFP cDNA.

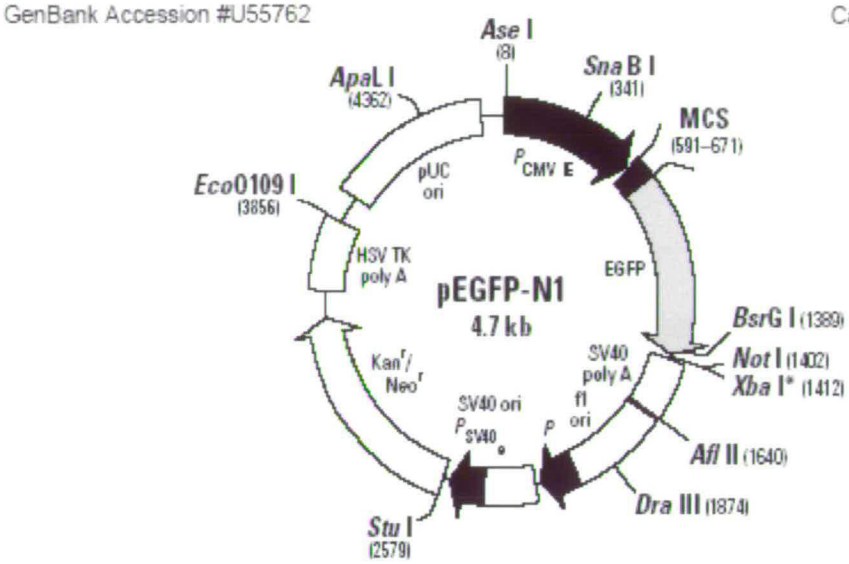
An oligonucleotide linker with *Xho*I and *EcoR*I overhangs containing a Kozak sequence and coding for the biotinylation tag was introduced into the multiple cloning site of pEGFP (Fig. 3.1a,b). This produced pBioGFP in which the biotinylation tag was N-terminal to and in frame with the EGFP sequence. Mouse Hex cDNA was amplified with *Sal*I and *Age*I ends and cloned into the respective sites of pBioGFP to give pBioHexGFP such that all three elements were in frame. In order to reduce the chance of steric hindrance of Hex by GFP in the context of a fusion protein five additional glycine residues were included at the end of the Hex cDNA sequence (Fig. 3.1b). Following the sequencing of pBioHexGFP to check the reading frame had been maintained during the PCR and cloning steps (data not shown) both plasmids were transiently transfected into HEK293 cells to check for GFP expression (Fig. 3.1c). Both pBioGFP and pBioHexGFP showed green fluorescence suggesting intact ORFs in each case.

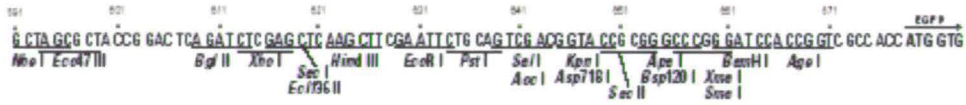
Fig. 3.1 Cloning of a Biotinylation Tag with the mouse Hex cDNA sequence.

a, Map of pEGFPN1. **b**, An oligonucleotide linker coding for the 23 aa biotinylation tag was inserted into the *Xho*I and *Eco*RI sites of pEGFP-N1, upstream of and in frame with the EGFP sequence to give pBioGFP. The asterisk denotes the lysine residue that becomes biotinylated in the presence of *E.* coli BirA ligase. Mouse Hex cDNA was amplified and inserted between the tag and EGFP sequences such that all three were in frame to give pBioHexGFP. **c**, Following transient transfection into HEK293 cells, GFP expression from pBioGFP, pBioHexGFP and the parental pEGFPN1 plasmid was monitored after 48 hours by fluorescence microscopy. *E* = *Eco*RI, *S* = *Sal*I. Scale bars represent 50μm.

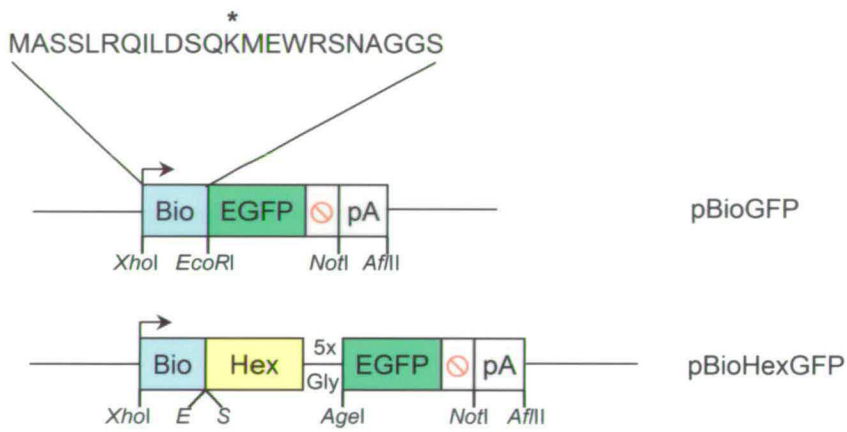
pEGFP-N1 Vector Information PT3027-5

GenBank Accession #U55762 Catalog #6085-1

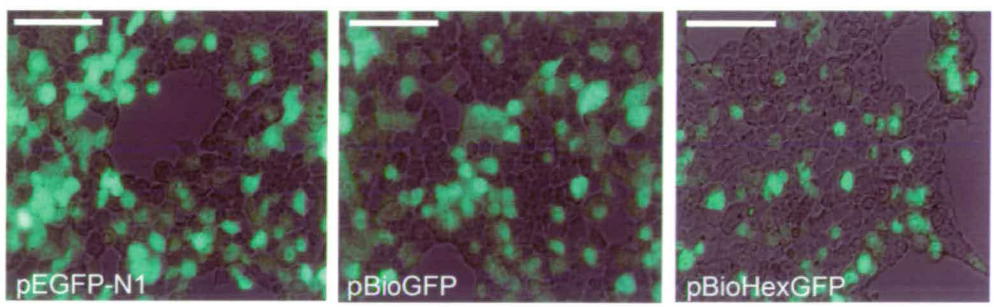




a



b



c

3.2 BioHexGFP Displays Wild-Type Activity *In Vivo*.

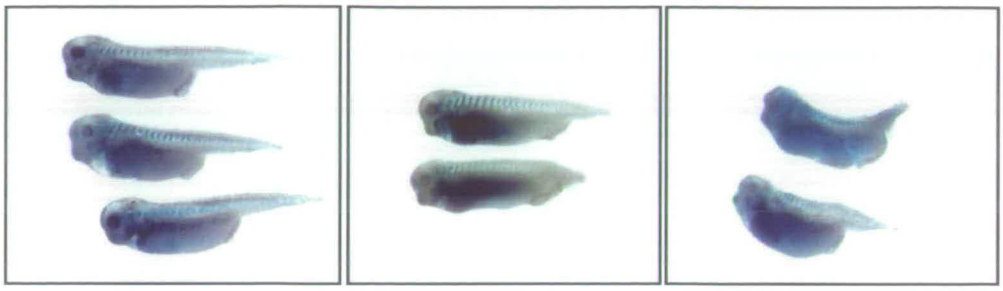
It has been shown previously that overexpression of murine Hex cDNA (mHex) on the dorsal side of *Xenopus* embryos leads either to mild axial distortions or a more extreme phenotype in which both anterior and posterior structures are affected (Brickman et al., 2000). As a means to check that the BioHexGFP protein could display wild-type activity, its RNA was injected into *Xenopus* embryos to see if the mutant phenotypes could be recapitulated.

In order to generate BioHexGFP RNA it was first necessary to subclone the *Xho/Not* fragment of pBioHexGFP into a variant of the pSp64T vector to give pBHGSp6 (Fig. 3.2a,b). Following in vitro transcription, 250pg of BioHexGFP RNA was injected into single dorso-vegetal blastomeres of 4-cell stage embryos. β-Gal RNA (160pg) was co-injected as a lineage tracer. Control sibling embryos were injected with wild-type mHex and β-Gal RNA for comparison or β-Gal RNA alone. Following conditions that allow gastrulation, embryos were left to grow to stage 37 and then stained for β-Gal activity. Figure 3.2c shows that the numbers of mild and severe embryos generated from the overexpression of wild-type mHex or BioHexGFP RNA were comparable. BioHexGFP RNA did not produce any other phenotype. This result would suggest that the Hex protein in the context of a fusion with a biotinylation tag and GFP displays wild type activity.

3.3 Generation and analysis of a BioHexGFP overexpression construct with a hygromycin selectable marker.

(i) Subcloning of BioHex GFP into the expression vector pCAGSIH.

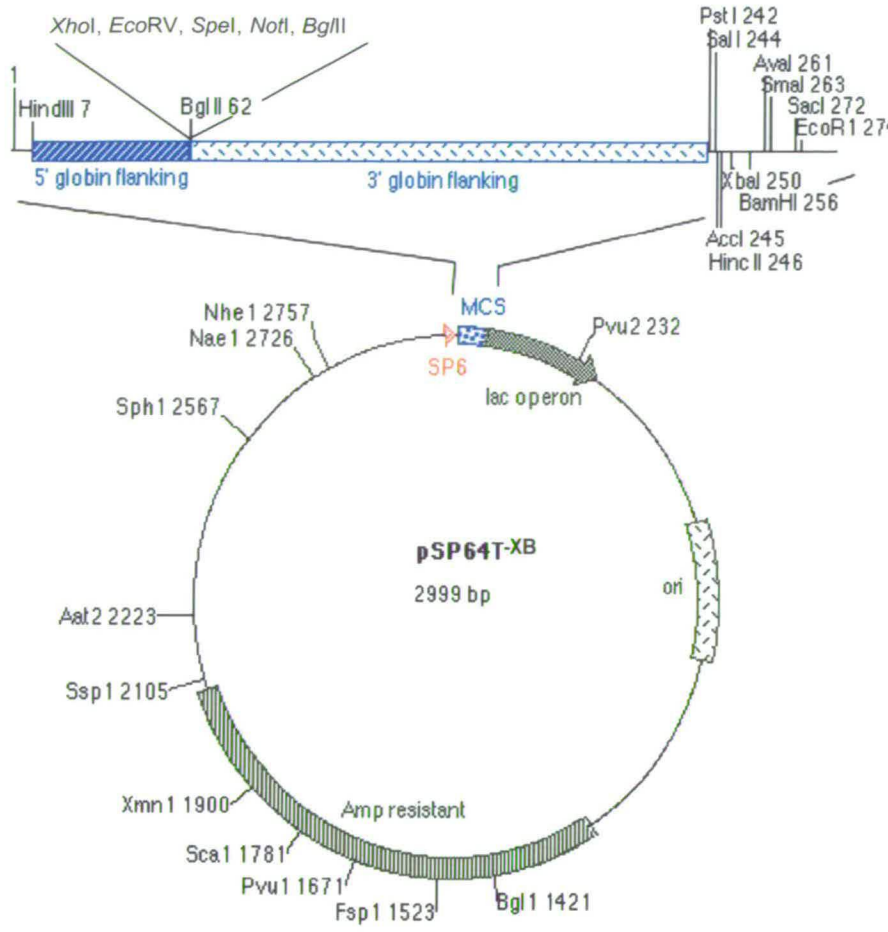
To achieve stable expression of BioHexGFP, the *Xho/Not* fragment of pBioHexGFP was subcloned into pCAGSIH vector. This vector contains a strong CAG promoter which contains the chicken β-*actin* promoter and cytomegalovirus enhancer, β-*actin*



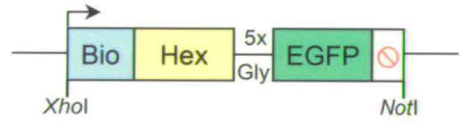
d

Fig. 3.2 Overexpression of BioHexGFP or Hex produces similar phenotypes in *Xenopus*.

a, Map of the expression vector pSp64T-XB (previously modified from pSp64T). **b**, The *Xho*I/*Not*I fragment of pBioHexGFP was cloned into the corresponding sites of pSp64T-XB in to generate pBHGsp6. **c**, Single dorso-vegetal injections of four cell stage blastulae were carried out with either mHex (250pg) or BioHexGFP (250pg) RNA. 200pg of β-Gal RNA was coinjected as a lineage tracer. Injection with β-Gal alone produced normal embryos only whereas Hex or BioHexGFP generated similar numbers of mild or severe phenotypes. Embryos were fixed and stained for β-Gal activity at stage 37. **d**, Examples of normal, mild and severe phenotypes.



a



pBHGsp6

b

	Normal	Mild	Severe
β-Gal	18 (100%)	0 (0%)	0 (0%)
β-Gal + Hex	10 (29%)	21 (60%)	4 (11%)
β-Gal + pBHGsp6	4 (13%)	23 (74%)	4 (13%)

c

intron and bovine β-*globin* poly-adenylation signal (Niwa et al., 1991). This cloning produced pBHGiH in which BioHexGFP was expressed as a fusion transcript with an encephalomyocarditis virus (ECMV) IRES sequence and a hygromycin resistance gene (Fig. 3.3a,b). To serve as a control for downstream experiments, a variant of pBHGiH was produced, pBHiGiH, which contained an added ECMV internal ribosome entry site (IRES) sequence between BioHex and GFP, ensuring their separation during translation (Fig. 3.3b). As a Hex negative control the *Xho*/*Not* fragment of pBioGFP was also subcloned into the pCAGSIH vector to give pBGiH (Fig.3.3b).

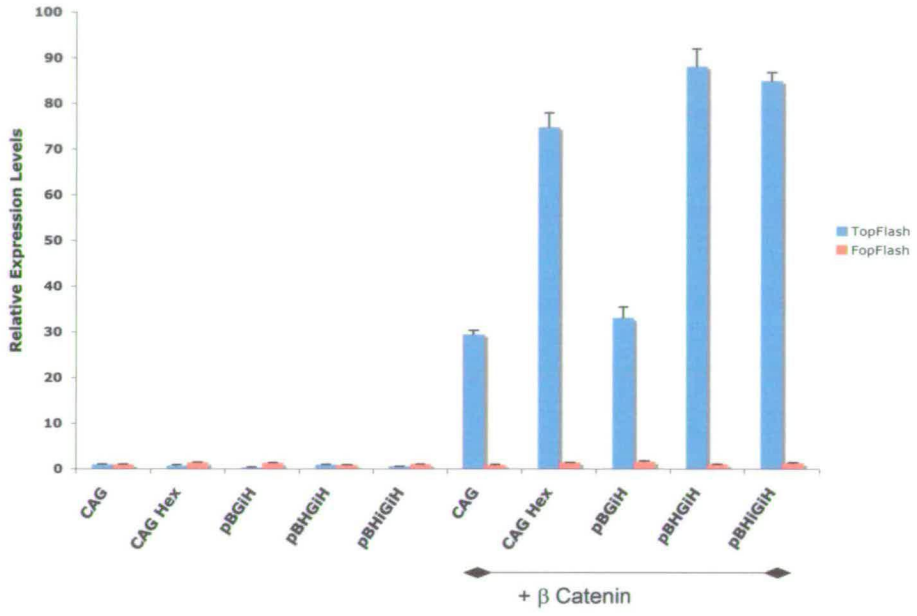
(ii) Expression of hygromycin constructs in HEK293 cells.

All three hygromycin constructs were transiently transfected into HEK293 cells to check for GFP expression by fluorescence microscopy. Both pBGiH and pBHGiH showed strong fluorescence whereas pBHiGiH was weaker, probably due to a translation gradient effect of the IRES element (Fig 3.3c) (Hobbs et al., 1998).

(iii) pBHGiH and pBHiGiH enhance β-catenin transcriptional activity *in vitro*.

Previous work in our lab has shown that Hex can enhance β-catenin induced activation of a reporter construct containing reiterated Tcf binding sites (the DNA binding components of the canonical Wnt pathway) (Zamparini et al., 2006). Therefore, to check that pBHGiH and pBHiGiH could display wild-type Hex activity *in vitro*, reporter constructs containing either Tcf binding sites (TopFlash) or mutated versions of these sites (FopFlash) were cotransfected into HEK293 cells with the indicated constructs (Fig. 3.3d,e). β-catenin increased the expression level of the TopFlash reporter almost 30 fold in the presence of an empty vector containing the CAG promoter. This induction was amplified a further 3-fold (to a total of 80 to 90 fold stimulation compared to the absence of b-cat) when the plasmid expressing b-cat was co-transfected with plasmids expressing the different Hex derivatives; CAGHex

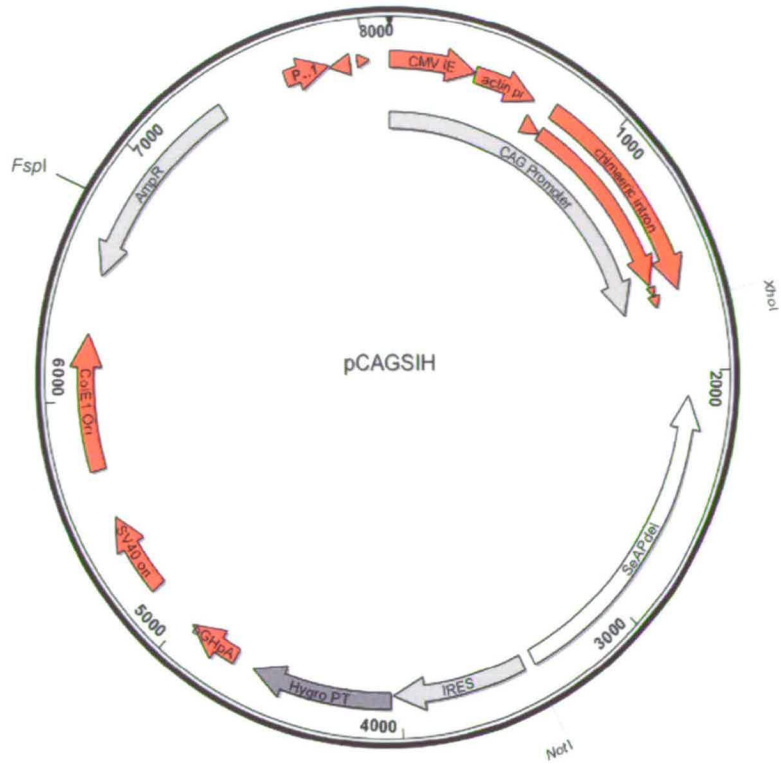
d



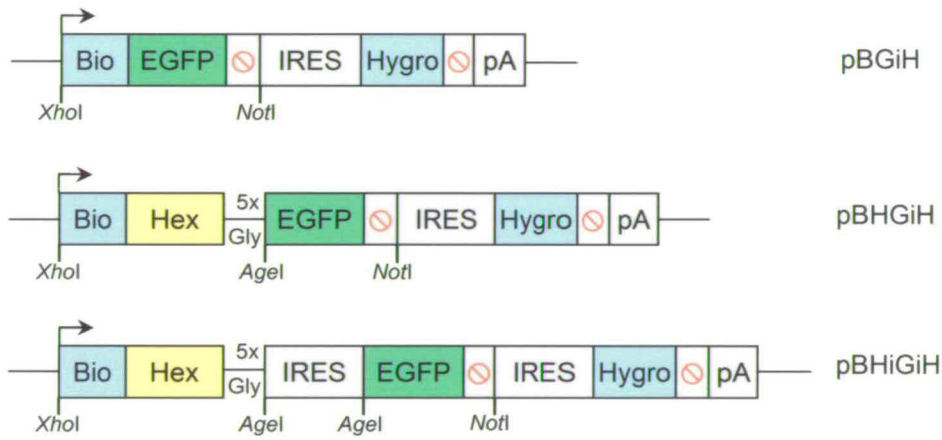
e

Fig. 3.3 Cloning, expression and functional analysis of pBHGIH.

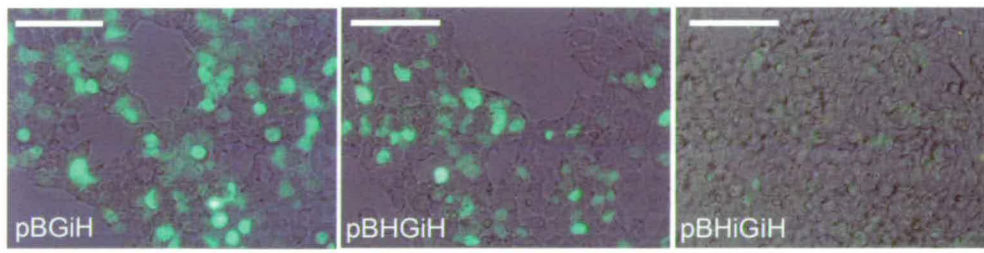
a, Map of pCAGSIH. **b**, The *Xho*I/*Not*I fragments of pBioGFP and pBioHexGFP were cloned into the pCAGSIH expression vector to give pBGiH and pBHGiH, respectively. A third CAG IRES Hygro construct, pBHiGiH, was generated by cloning an amplified IRES sequence into the *Age*I site of pBHGiH. **c**, Following transient transfection into HEK293 cells, GFP expression from pBGiH, pBHGiH and pBHiGiH was monitored after 48 hours by fluorescence microscopy. Scale bars represent 50μm. **d**, Diagram depicting the reporter constructs used in a luciferase assay to determine the functionality of pBHGIH. **e**, A reporter construct containing reiterated Tcf binding sites (TOPflash) or mutated versions of these sites (FOPflash) was transfected into HEK293 cells with the indicated constructs (100ng of each). In the presence of a mutated version of β-catenin (stabilised β-catenin), the two bio-tagged Hex constructs (pBHGiH and pBHiGiH) and wild-type Hex (CAG Hex) comparably increase the activity of TOPflash beyond that obtained from an empty vector with a CAG promoter (CAG). The control vector, pBGiP does not amplify this signal.


8000
1000
2000
3000
4000
5000
6000
7000
CMV IE
actin pr
chimeric intron
CAG Promoter
pCAGSIH
SeAP/neI
IRES
Hygro PT
rGHpA
SV40 ori
ColE1 Ori
AmpR
FspI
XhoI
NotI

a


Bio
EGFP
IRES
Hygro
pA
XhoI
NotI
pBGiH
Bio
Hex
5x
Gly
EGFP
IRES
Hygro
pA
XhoI
AgeI
NotI
pBHGiH
Bio
Hex
5x
Gly
IRES
EGFP
IRES
Hygro
pA
XhoI
AgeI
AgeI
NotI
pBHiGiH

b


pBGiH
pBHGiH
pBHiGiH

c

(wild-type), pBHGiH or pBHiGiH. This result was specific for the TopFlash reporter as no activity was seen with FopFlash. Amplification of β-catenin induced activation was not seen with the control vector, pBGiH. These results indicated that the modified Hex proteins expressed from pBHGiH and pBHiGiH displayed wild-type activity *in vitro* and supports the result of the observed wild-type effect of BioHexGFP in *Xenopus*.

3.4 Generation and analysis of a BioHexGFP overexpression construct with a puromycin selectable marker.

(i) Subcloning of BioHexGFP into the expression vector pCAGSIP

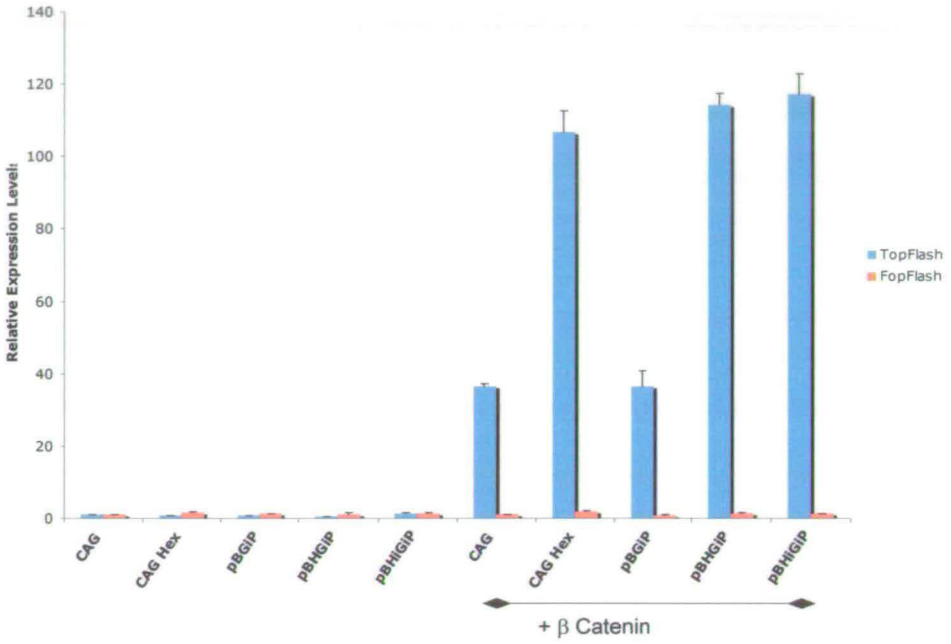
The expression vector pCAGSIP is similar to pCAGSIH aside from the presence of a gene conferring resistance to puromycin rather than hygromycin. The *Xho*/*Not* fragments of the three hygromycin constructs were subcloned into pCAGSIP to give, pBGiP, pBHGiP and pBHiGiP (Fig. 3.4a,b).

(ii) Expression of puromycin constructs in HEK293 cells.

All three puromycin constructs were transiently transfected into HEK293 cells to check for GFP expression. Both pBGiP and pBHGiP showed strong fluorescence whereas pBHiGiP was weaker, again probably due to a translation gradient effect of the IRES element (Fig. 3.4c).

(iii) pBHGiP and pBHiGiP enhance β-catenin transcriptional activity *in vitro*.

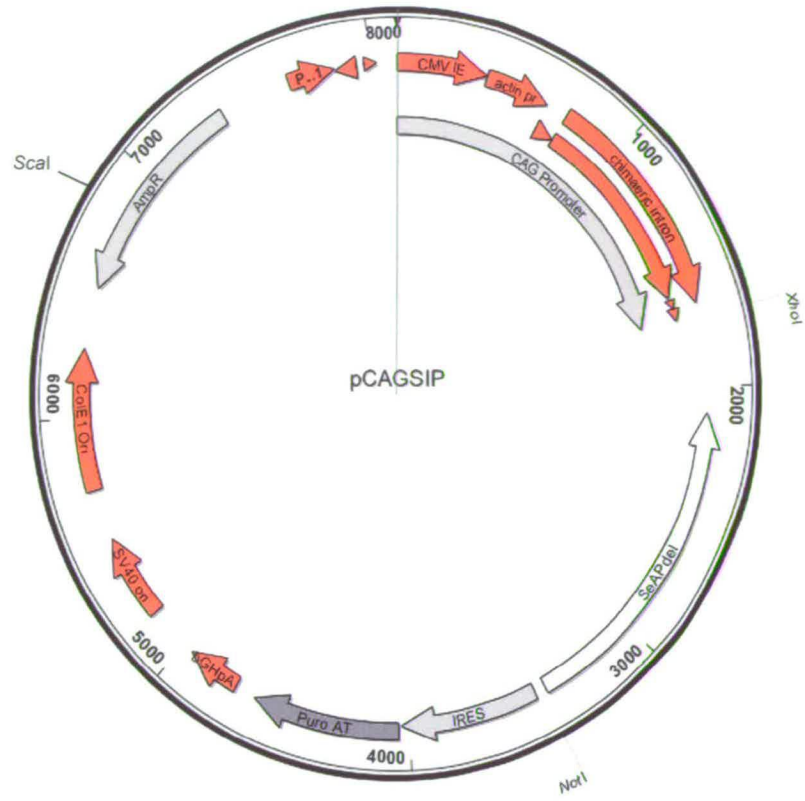
As with the hygromycin constructs, the three puromycin constructs were tested for Hex function using the TopFlash assay (Fig. 3.4d). Again, the two Hex containing vectors, pBHGiP and pBHiGiP, and wild type Hex expressed proteins that comparably enhanced β-catenin induced activation of the TopFlash reporter. This effect was specific to the TopFlash reporter as no effect was seen with FopFlash. The



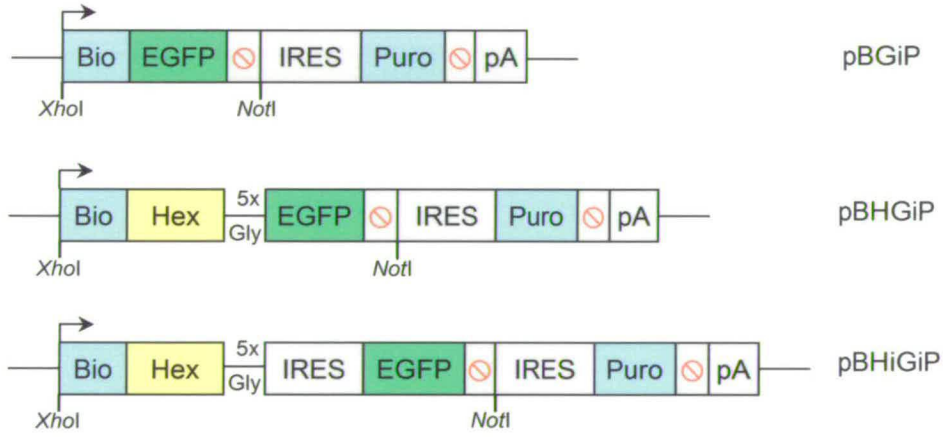
d

Fig. 3.4 Cloning, expression and functional analysis of pBHGIP.

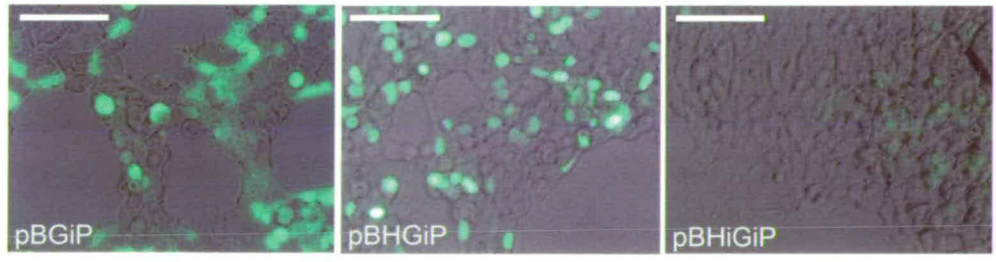
a, Map of pCAGSIP. **b**, The XhoI/*Not*I fragments of pBGiH, pBHGiH and pBHiGiH were cloned into the pCAGSIP expression vector to give pBGiP, pBHGiP and pBHiGiP, respectively. **c**, Following transient transfection into HEK293 cells, GFP expression from pBGiP, pBHGiP and pBHiGiP was monitored after 48 hours by fluorescence microscopy. Scale bars represent 50μm **d**, A reporter construct containing reiterated Tcf binding sites (TOPflash) or mutated versions of these sites (FOPflash) was transfected into HEK293T cells with the indicated constructs (100ng of each). In the presence of a mutated version of β-catenin (stabilised β-catenin), the two BioHex constructs (pBHGiP and pBHiGiP) and wild-type Hex (CAG-Hex) comparably increase the activity of TOPflash beyond that obtained from an empty vector with a CAG promoter (CAG). The control vector, pBGiP does not amplify this signal.


8000
CMV IE
actin IX
1000
chimaeric intron
CAG Promoter
pA
Xho I
2000
SeAP del
3000
IRES
Puro AT
4000
Not I
rGHpA
5000
SV40 on
6000
ColE1 On
7000
AmpR
ScaI
pCAGSIP

a


Bio
EGFP
IRES
Puro
pA
XhoI
NotI
pBGiP
Bio
Hex
5x
Gly
EGFP
IRES
Puro
pA
XhoI
NotI
pBHGiP
Bio
Hex
5x
Gly
IRES
EGFP
IRES
Puro
pA
XhoI
NotI
pBHiGiP

b


pBGiP
pBHGiP
pBHiGiP

c

Hex negative vector, pBGiP did not enhance β-catenin induced activation of TopFlash. These results indicated that pBHGiP and pBHiGiP displayed wild-type activity *in vitro*.

3.5 Western analysis of pBHGiH and pBHGiP.

To confirm that BioHexGFP is indeed an intact fusion protein when expressed from either CAG driven vector, western analysis with an anti-GFP antibody was carried out on lysates of HEK293 cells that had been transiently transfected with the hygromycin and puromycin constructs (Fig. 3.5). Both Hex negative control constructs, pBGiH and pBGiP, produced a band corresponding to a predicted size of 31kda for a BioGFP fusion protein. Both BioHexGFP constructs, pBHGiH and pBHGiP, produced a band corresponding to the predicted size of 61kda for a fusion protein. pBHiGiH and pBHiGiP which contained an IRES element between Hex and GFP should have produced a band of the same size as that obtained with the positive control pEGFPN1. Only a very faint band can be seen for pBHiGiH and a slightly stronger one for pBHiGiP, reflecting the faint GFP fluorescence seen in figures 3.3c and 3.4c.

3.6 Generation and analysis of a BioHex inducible overexpression vector.

A third set of CAG driven overexpression vectors bearing a bio-tagged Hex cDNA was generated in order to permit its inducible expression. The parental vector used (pTLC, a kind gift from John Hall) consisted of a floxed DsRed IRES Puromycin stop cassette upstream of IRES-GFP. Expression from this vector would generate both a red fluorescent marker of expression as well as conferring resistance to puromycin. In this way, puro resistant colonies can be screened for high levels of homogenous expression prior to using lines for studies of inducible Hex over expression (chapter 5). Only upon the removal of the floxed DsRed IRES puromycin

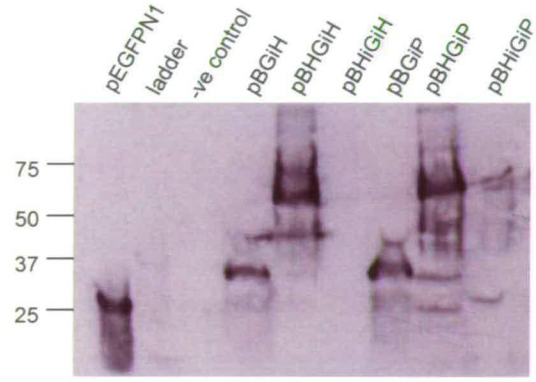


Fig. 3.5 Western Analysis of pBHGiH and pBHGiP.

Equal quantities (4µg) of pBHGiH, pBHGiP or the control constructs (pBGiH, pBHiGiH, pBGiP, pBHiGiP) were transiently transfected into HEK293 cells. 48 hours after transfection, lysates were analysed by western blotting with an anti-GFP antibody. As a positive control, cells were transfected with the vector pEGFP-N1. Non-transfected cells were included as a negative control. The predicted sizes for GFP, Bio-tagged GFP and BioHex-GFP are 27kda, 31kda and 61kda, respectively.

stop cassette in the presence of Cre recombinase (Cre), expression of the downstream components would be permitted. The presence of *Pac*I and *Nhe*I restriction sites between the removable cassette and IRES-GFP allowed the insertion of a cDNA of interest.

(i) Cloning of BioHex into pTLC.

Primers with *Pac*I and *Nhe*I ends were used to amplify either the Bio-tagged sequence or the Bio-tagged Hex cDNA from pBioHexGFP. Each amplicon was cloned into the respective sites of pTLC generating pTLC Bio and pTLC BioHex, respectively (Fig. 3.6a,b).

(ii) Expression of pTLC BioHex in HEK293 cells.

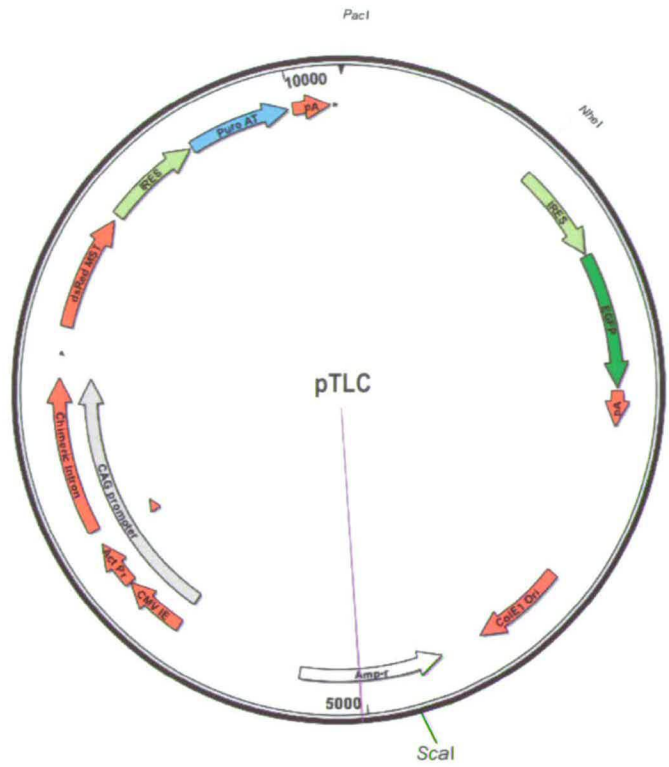
Both pTLC Bio and pTLC BioHex were transiently transfected into HEK293 cells in the presence or absence of a vector expressing Cre from a CAG promoter. Both inducible vectors when transfected alone produced strong red fluorescence, while GFP could not be seen. However, when co-transfected with Cre each produced a marked reduction of red fluorescence while strong GFP fluorescence became apparent (Fig. 3.6c,d)

(iii) pTLC BioHex in the presence of Cre enhances β-catenin transcriptional activity *in vitro*.

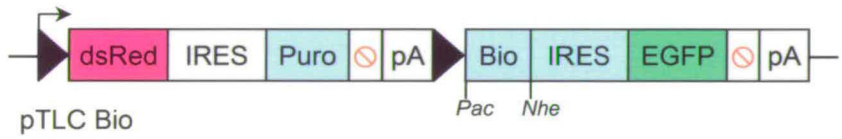
To test the function of Hex in the context of the TLC inducible overexpression vector, pTLC BioHex was co-transfected with a CAG driven Cre vector into HEK293 cells alongside b-catenin and the TopFlash reporter. While co-transfection of pTLC Bio (with out the Hex insert) had no effect on b-catenin activity, co-transfection of Cre, alongside pTLC BioHex has a similar affect on b-catenin activity as did co-transfection of the parent CAG-hex expression plasmid (Fig. 3.6e). This

Fig. 3.6 Cloning, expression and functional analysis of pTLC BioHex.

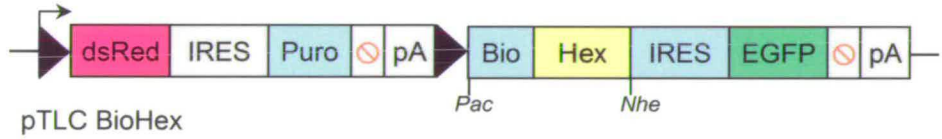
a, Map of the inducible overexpression vector pTLC. **b,** The bio-tagged Hex cDNA sequence or the bio-tagged sequence alone was amplified from pBioHexGFP with primers containing *Pac*I and *Nhe*I restriction sites. Resultant digested products were cloned into the corresponding sites of pTLC to give pTLC Bio and pTLC BioHex. Following transient transfection into HEK293 cells expression was monitored after 48 hours by fluorescence microscopy. **c**, pTLC Bio was transfected alone or with a Cre expressing vector. **d**, pTLC BioHex was transfected alone or with a Cre expressing vector. Bright field, red filter and green filter photographs are shown for each transfection. Scale bars represent 50μm. **e**, A reporter construct containing reiterated Tcf binding sites (TOPflash) or mutated versions of these sites (FOPflash) was transfected into HEK293T cells with the indicated constructs (100ng of each). In the presence of a mutated version of β-catenin (stabilised β-catenin), wild-type Hex (CAG Hex) and pTLC BioHex + Cre comparably increase the activity of TOPflash beyond that obtained from an empty vector with a CAG promoter (CAG). The control vector, pBGiP + Cre does not amplify this signal.



a

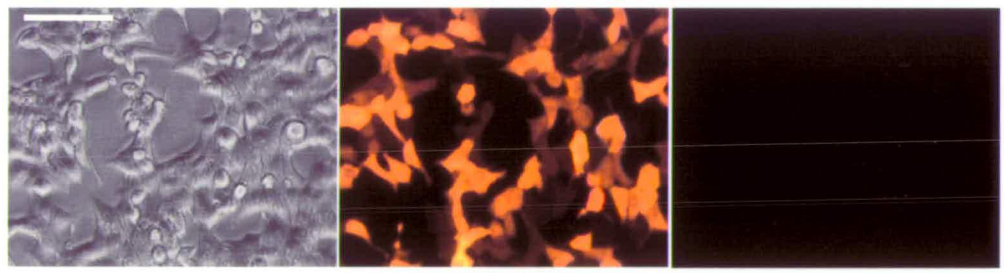


pTLC Bio

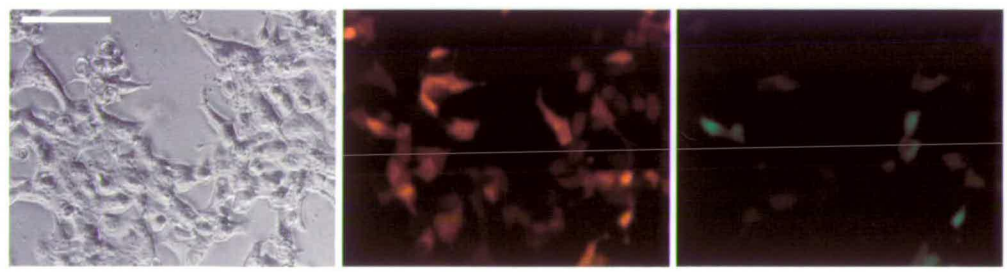


pTLC BioHex

b

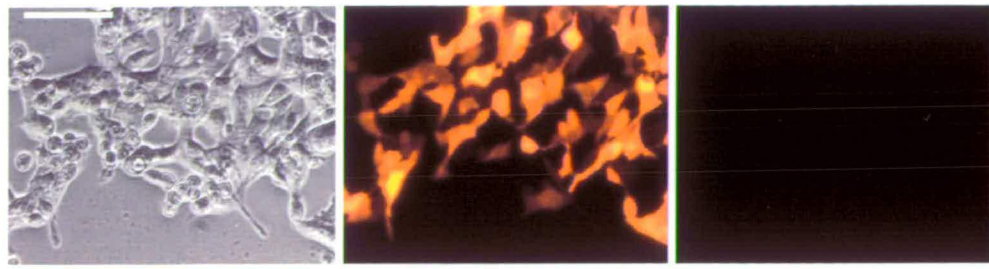


pTLC Bio



pTLC Bio +Cre

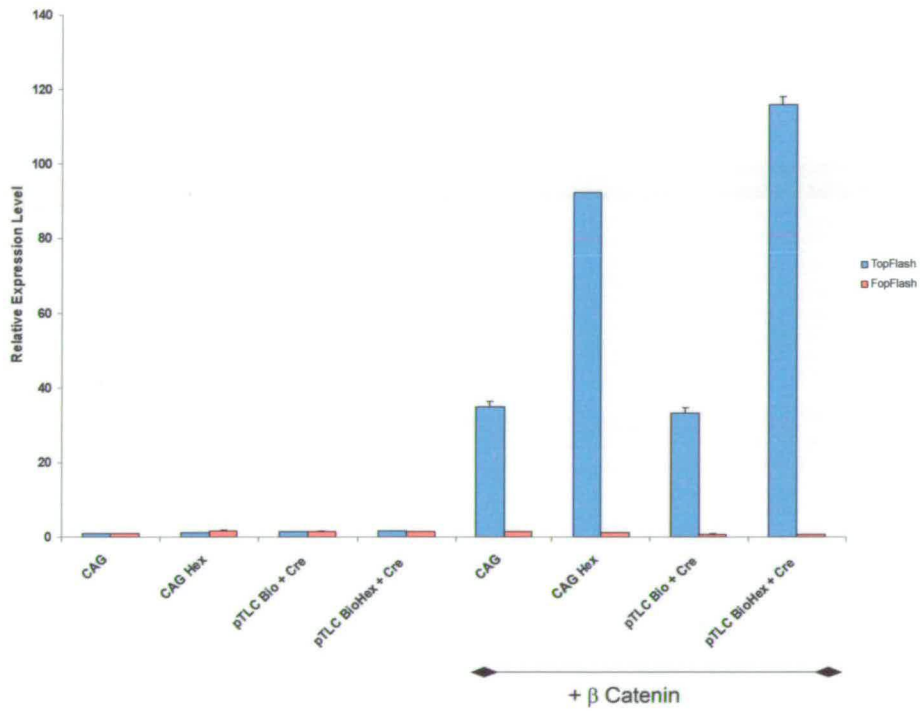
c



pTLC BioHex

pTLC BioHex +Cre

d



e

result indicated that pTLC BioHex in the presence of Cre displays wild-type Hex activity.

3.7 Cloning strategies to generate knock-in targeting vectors to introduce BioHex cDNA into the *Hex* Locus.

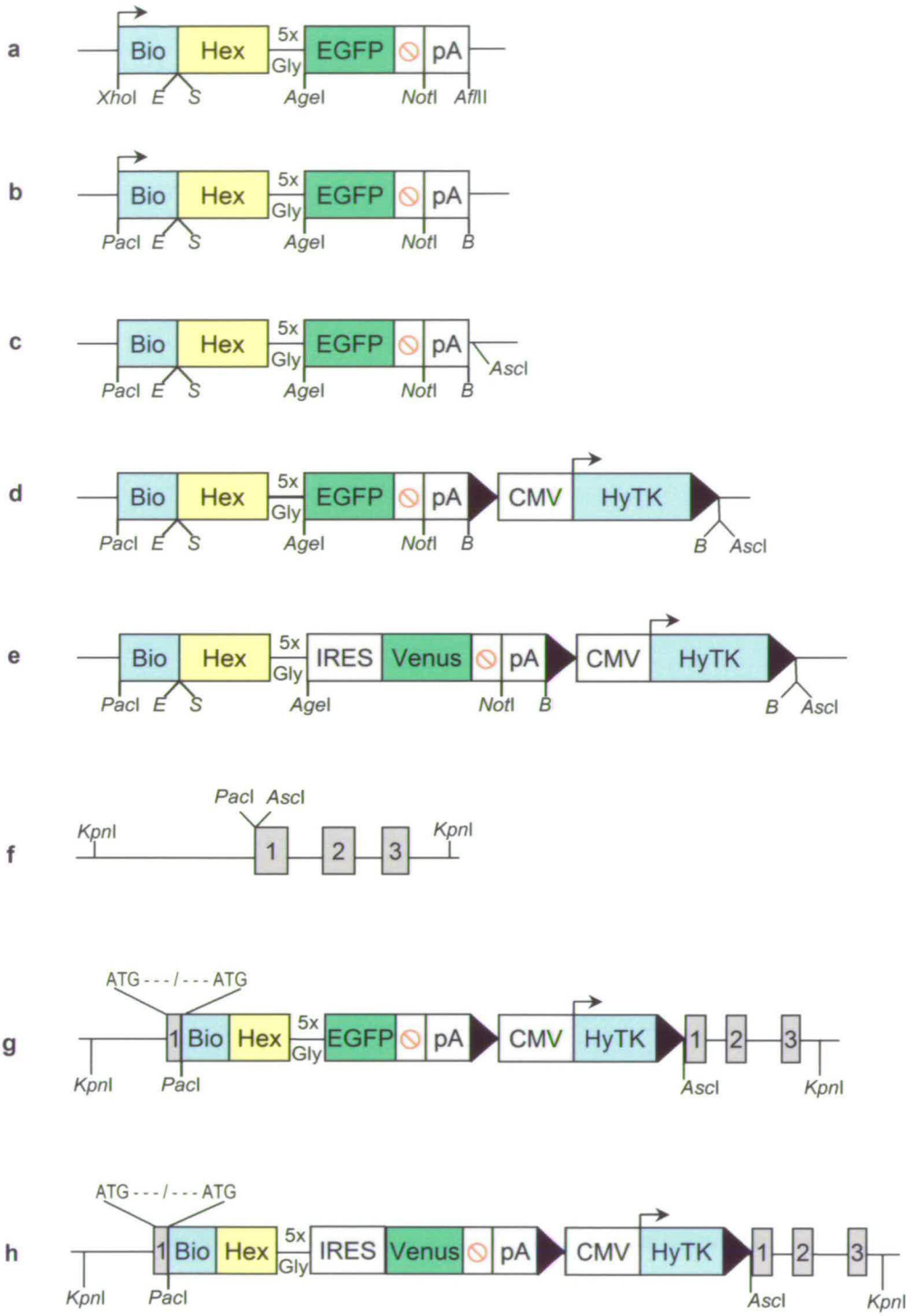
Several cloning steps were undertaken to introduce BioHex cDNA together with a selection cassette into a targeting vector containing homologous sequences to the *Hex* locus, pT1Pa, as detailed in figure 3.7. Two different targeting vectors were produced containing either BioHexGFP (pT1 BHG) or BioHex-IRES-Venus (pT1 BHIV) cDNA ligated into *Pac*I and *Asc*I restriction sites positioned 27nt downstream of the ATG sequence of exon 1. Sequencing was performed to confirm that the ATG of the Biotin tag in each targeting vector was in frame with the upstream ATG (Data not shown). The selection cassette used consisted of LoxP sites flanking a CMV promoter upstream of a cDNA coding for a fusion protein conferring hygromycin resistance and gancyclovir sensitivity (HyTK). This would allow positive selection of clones containing the targeting vectors, and subsequent selection against clones still containing the cassette following Cre recombinase addition.

The second targeting vector, pT1 BHIV, was designed to include a fluorescent reporter in conjunction with an IRES sequence downstream of the BioHex cDNA. An IRES Venus cassette was used from the gene trap vector pGTIV2 which has been shown to act as an efficient reporter of expression in undifferentiated and differentiated cultures (Tanaka and Stanford, personal communication). The sensitive nature of this cassette is due to the presence of an artificial sequence composed of tandem repeats of a 9nt *gtx* segment which has been shown to confer enhanced IRES activity (Chappell et al., 2000) in addition to the YFP variant, Venus, which is reported to be brighter than GFP (Nagai et al., 2002).

3.8 Expression and Function of BioHex IRES Venus in Hek293T cells.

Fig. 3.7 Cloning scheme for targeting BioHex to the *Hex* locus.

A series of cloning steps was performed in order to introduce BioHex cDNA into a targeting vector containing homologous regions to the *Hex* locus (**f**). Oligonucleotide linker sequences were introduced into the *Xho*I and *Afl*II sites of pBioHexGFP (**a**) in order to create *Pac*I and *BamH*I sites (**b**). The *Pac*I, *Bam*HI fragment of vector **b** was cloned into the respective sited of pNEB193 (**c**). A *Bam*HI fragment containing a floxed CMV HygroTk cassette was introduced into the *Bam*HI site of vector **c** (**d**). A variant of vector **d** was produced by swapping the EGFP sequence with an *Age*I, *Not*I cassette containing an IRES upstream of cDNA coding for the EYFP variant, Venus (**e**). The *Pac*I, *Asc*I fragments of vectors **d** and **e** were cloned into the respective sites of the targeting vector (**f**) to give pT1 BHG (**g**) and pT1 BHIV (**h**), respectively. *B* = *BamH*I, *E* = *EcoR*I, *S* = *Sal*I


a
Bio
Hex
5x
Gly
EGFP
pA
XhoI E S
AgeI
NotI AflII
b
Bio
Hex
5x
Gly
EGFP
pA
PacI E S
AgeI
NotI B
c
Bio
Hex
5x
Gly
EGFP
pA
AscI
PacI E S
AgeI
NotI B
d
Bio
Hex
5x
Gly
EGFP
pA
CMV
HyTK
PacI E S
AgeI
NotI B
B AscI
e
Bio
Hex
5x
Gly
IRES
Venus
pA
CMV
HyTK
PacI E S
AgeI
NotI B
B AscI
f
PacI AscI
KpnI
1
2
3
KpnI
g
ATG - - - / - - - ATG
1
Bio
Hex
5x
Gly
EGFP
pA
CMV
HyTK
1
2
3
KpnI
PacI
AscI
KpnI
h
ATG - - - / - - - ATG
1
Bio
Hex
5x
Gly
IRES
Venus
pA
CMV
HyTK
1
2
3
KpnI
PacI
AscI
KpnI

The BioHex in the context of the targeting vector has additional coding sequence derived from the first exon of Hex and the insertion of the tag. This sequence is depicted in Figure X and could alter Hex activity. Before targeting was initiated I checked that a Hex protein with this sequence could be expressed and function normally. To this end I first introduced BHIV cDNA into an expression vector with a CMV promoter (Fig. 3.8a). Then, an oligonucleotide linker coding for the first 27nt of *Hex* exon 1 was ligated into the *Pac*I site between the CMV promoter and BioHex sequence to produce pBHIV (Fig. 3.8b).

Transient transfection experiments of pBHIV in HEK293T cells were carried out to monitor expression of the IRES Venus reporter by fluorescence microscopy (Fig. 3.8c) and function of Hex by employing the TopFlash assay (Fig. 3.8d). Venus expression was easily detectable after 48 hours and Hex function appeared almost identical to that of wild-type Hex.

In order to test the detection and immunoprecipitation of BioHex by streptavidin, HEK293T cells were transiently transfected with pBHIV and cultured for 48 hours. Equal quantities of cell lysate from transfected cells were subjected to western analysis with streptavidin-HRP or immunoprecipitation with streptavidin coated magnetic beads (Dynabeads). A band of the expected size (42 kD) for the BioHex protein was observed in the input lane (Fig. 3.8e(i)). Background bands could be accounted for by the fact that there are other endogenously biotinylated proteins present in mammalian cells (de Boer et al., 2003). The BioHex protein in this context appeared to be efficiently precipitated with Dynabeads (bound fraction) with no BioHex protein evident in the unbound fraction (Fig. 3.8(ii)).

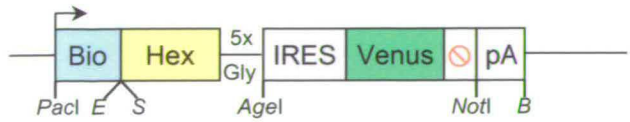
Discussion

This chapter has described the generation of various constructs containing BioHex cDNA in conjunction with a fluorescent reporter. Expression analyses in HEK293 cells were used to monitor the integrity of Hex function and viability of the reporter

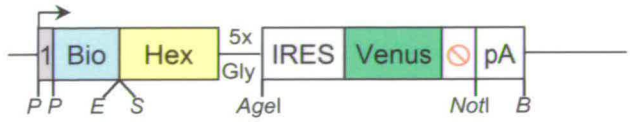
Fig. 3.8 Expression and function of BioHex IRES Venus in HEK293 cells.

In order to test the expression and function of the BioHex IRES Venus element in transient transfection experiments with a CMV promoter, the *Pac*I, *Not*I fragment of 'vector e' was cloned into the respective sites of 'vector b' (Fig. 3.7b,e) to give pbHIV (**a**). An oligonucleotide sequence coding for the first 27nt of *Hex* exon1 with *Pac*I ends was cloned into the *Pac*I site of pbHIV to give pBHIV (**b**). The vector pBHIV was used in transient transfection experiments to check the expression and function of BioHex IRES Venus. **c**, Following transfection into HEK293 cells, expression of pBHIV was monitored after 48 hours by fluorescence microscopy. Scale bars represent 50µm. **d**, The Topflash assay was performed with the indicated constructs (100ng of each) in HEK293T cells. In the presence of stabilised β-catenin, pBHIV and wild-type Hex (pmHex) comparably increase the activity of TOPflash beyond that obtained from an empty vector with a CMV promoter (pCS2). **e**, Following transient transfection of pBHIV into HEK293 cells, equal quantities of cell lysate were used for western analysis (**i**) or for precipitation with streptavidin beads and western analysis (**ii**). I = input, B = bound, U = unbound. *B* = *BamH*I, *E* = *EcoR*I, *S* = *Sal*I

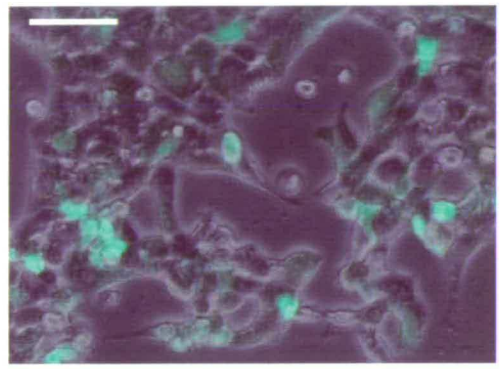
a


Bio
Hex
5x
Gly
IRES
Venus
pA
PacI E S
AgeI
NotI B

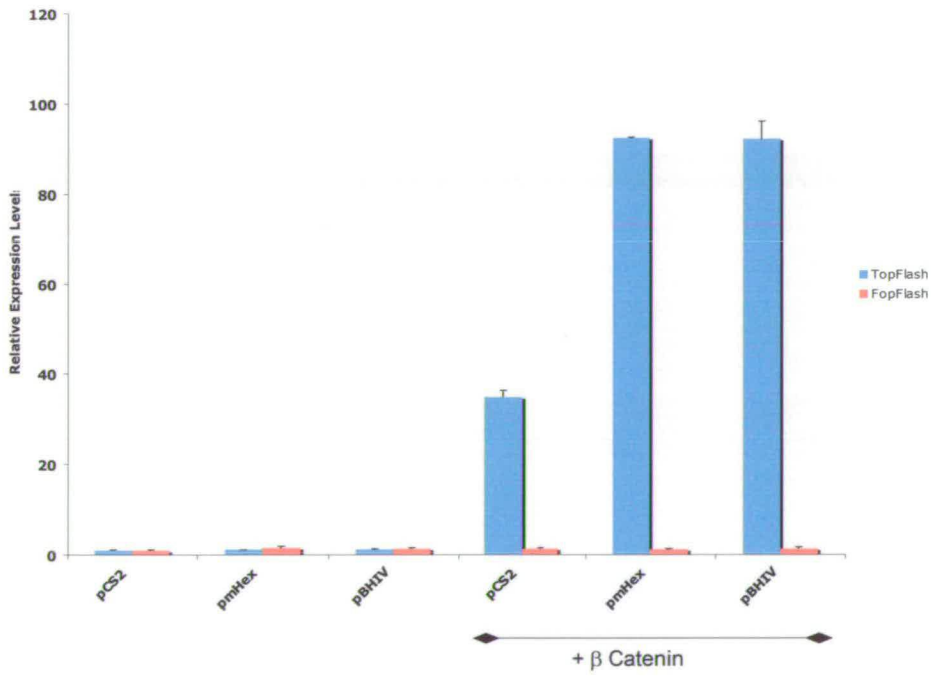
b


1
Bio
Hex
5x
Gly
IRES
Venus
pA
P P E S
AgeI
NotI B

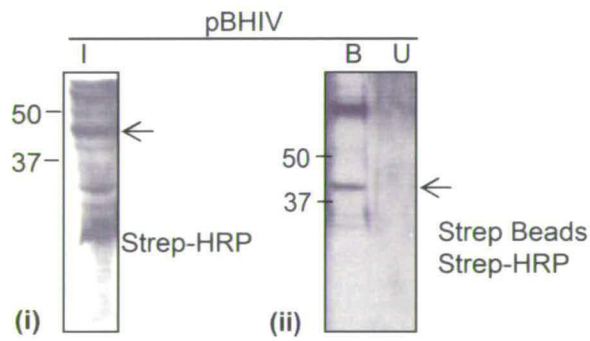
c



d


120
100
80
60
40
20
0
Relative Expression Level
TopFlash
FopFlash
pCS2
pmHex
pBHIV
pCS2
pmHex
pBHIV
+ β Catenin

e


pBHIV
I
B U
50
37
Strep-HRP
(i)
50
37
Strep Beads
Strep-HRP
(ii)

for each construct. The following chapters will describe the use of these constructs to monitor physiological expression (chapter 4) or for studies of the consequence of stable overexpression (chapter 5). Clearly the original design of these tools was intended for more extensive biochemistry although the focus of this thesis has centred on key observations made using the targeted line described in figure 4.1. In the future, either I or someone else in the lab will continue with the biochemistry and this chapter is designed to not only introduce the reagents used in Chapter 4 and 5, but provide a complete catalogue of all the tools I constructed, even those that do not feature prominently in this thesis.

Chapter 4

Physiological Expression of Hex in mouse ES cells

The expression of BioHex in conjunction with a fluorescent reporter from the *Hex* locus would provide the means to generate physiological levels of biotinylated protein for *in vivo* localisation and identification of direct targets and protein partners. This approach would ensure that wild-type levels of *Hex* were expressed and therefore provide confidence in the interpretation of results from downstream experiments as being truly representative of function and location. Expression of the triple fusion BioHexGFP cDNA from the *Hex* locus would provide the additional ability to look at sub-cellular localisation at various stages of early mouse development, providing further clues to the role of this transcription factor. Such knowledge might help generate an understanding of why it is the lack of *Hex* expression in the ADE and not extraembryonic tissues that gives rise to the observed phenotypes seen in the null embryo (Martinez Barbera et al., 2000). For biochemical analysis, a screen for downstream targets and partners could be performed directly from the *Hex* expression domains of mouse embryos following the isolation of Bio-tagged Hex proteins.

Additionally, such a cell line would prove a useful substrate for an *in vitro* differentiation protocol developed in our laboratory which can direct ES cells to form a population representing the ADE (Morrison et al. in preparation).

Identification of targets and protein partners from these methods in which BioHex is expressed at natural levels would reduce the chances of artefacts due to overexpression and/or misexpression, thus providing an insight into its function *in vivo*.

Two targeting vectors with homologous arms to the *Hex* locus were generated containing BioHex cDNA, pT1 BHG and pT1 BHIV (Fig. 3.7g,h). pT1 BHG was based on a triple fusion cDNA of BioHexGFP and the second, pT1 BHIV, contained an IRES Venus sequence downstream of BioHex. Thus, each targeting vector would

introduce BioHex into the *Hex* locus together with a fluorescent reporter expressed either as a fusion with BioHex or bicistronically from an IRES sequence.

This chapter will report on the successful use of pT1 BHIV to target BioHex to the *Hex* locus and the subsequent analyses of the resultant BHIV cell line. I will show that Venus reporter expression from the targeted allele reflects true transcriptional activity from the *Hex* locus and the existence of a subpopulation of Venus positive self-renewing ES cells. Manipulation of FGF signalling and *Nanog* expression together with gene expression analysis suggest that this population of cells may represent pre-fated precursors of the primitive endoderm. While both Venus positive and negative cells can interconvert and both are positive for ES cell markers Oct3/4 and SSEA1, the former have a restricted ability for clonal growth and embryo contribution.

4.1 Electroporation of pT1 BHIV into R26BirA cells.

pT1 BHIV was linearized with *Fsp*I and electroporated into R26birA ES cells which were then cultured in media containing 200μg/ml of hygromycin for 14 days. Surviving clones were expanded to isolate genomic DNA for Southern analysis with various probes in order to check for correct targeting to the *Hex* locus. Figure 4.1 depicts the strategy for targeting pT1 BHIV to the *Hex* locus and the resultant Southern analyses from hygromycin resistant clones. Two external probes revealed successful integration in the locus as defined by the presence of targeted bands of the expected sizes (Fig. 4.1b,d). Use of an internal probe which recognizes the Venus sequence showed the presence of only 1 integration event with pT1 BHIV (Fig. 4.1c). These results show the successful targeting of pT1BHIV to the *Hex* locus with no additional integration events into other loci.

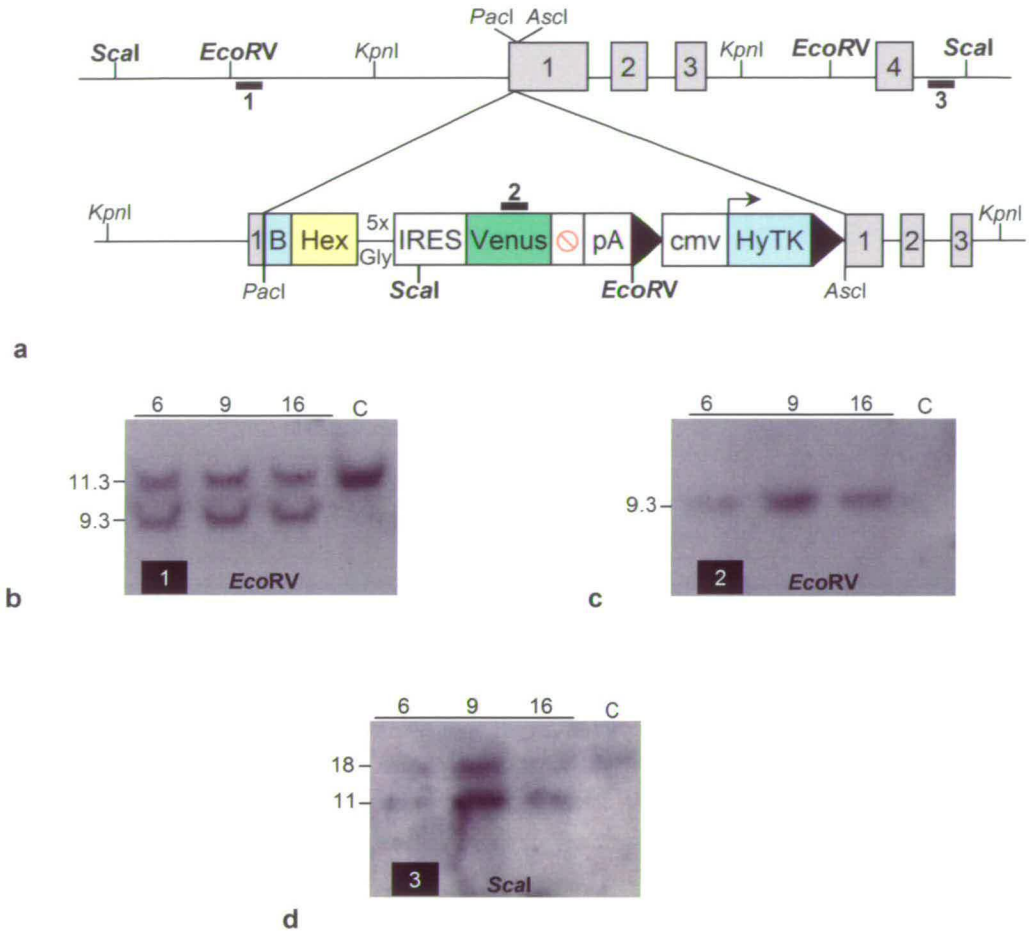


Fig. 4.1 Targeting of BioHex IRES Venus to the *Hex* locus.

a, Schematic representation of the gene targeting strategy to introduce BioHex IRES Venus (BHIV) to the *Hex* locus. pT1BHIV was designed such that the ATG of BHIV cDNA was downstream and in frame with the ATG of exon 1. A CMV driven HyTK cassette was introduced to allow for positive and negative selection. Following electroporation into R26 BirA cells and selection in hygromycin (200µg/ml), surviving clones were expanded and isolated genomic DNA was subjected to Southern analyses. Numbered bars represent the probes used for Southern analysis. **b**, Genomic DNA digested with *EcoR*V was hybridised with probe 1 revealing WT (11.3 kb) or targeted (9.3 kb) bands. **c**, Genomic DNA digested with *EcoR*V and hybridised to probe 2 shows the expected size (9.3 kb) for a single integration event only into the *Hex* locus. **d**, Genomic DNA digested with *Sca*I was hybridised with probe 3 revealing WT (17.8 kb) or targeted bands (11.5 kb). C = digested genomic DNA from unelectroporated R26 BirA cells.

4.2 Selection cassette removal and karyotypic analysis of BHIV clones.

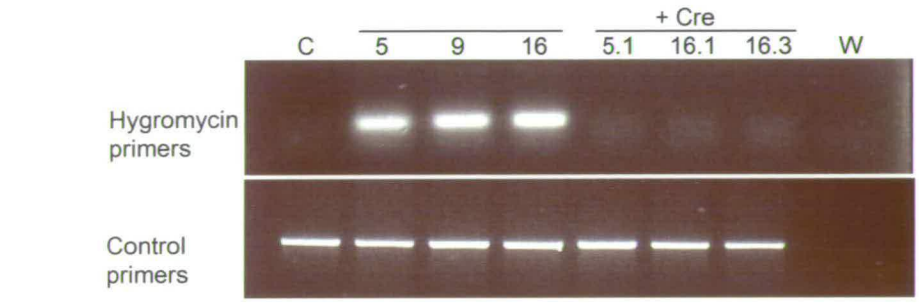
Three positive clones showing a correct targeting event were expanded and electroporated with a CAG driven Cre plasmid in order to remove the floxed selection cassette. Following two days in culture, cells were plated at clonal density and cultured with gancyclovir for 14 days. Gancyclovir resistant ($Ganc^R$) clones were expanded for genomic DNA extraction and PCR analysis to confirm removal of the selection cassette (Fig.4.2a). Chromosome spreads were prepared from $Ganc^R$ clones to check for karyotypic integrity (Fig. 4.2b). Forty chromosomes were seen for each of the three clones expanded.

4.3 Sequencing of BHIV clones.

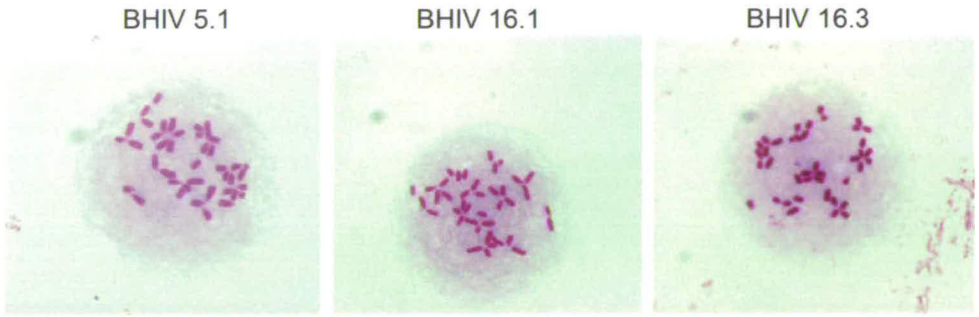
To check the integrity of the reading frame between the ATG of exon 1 of *Hex* and the ATG at the beginning of the BioHex cDNA, genomic DNA from BHIV clones was subjected to PCR and sequence analysis. Primers were chosen to amplify this region specifically from the targeted allele. Figure 4.3 shows partial resultant sequence from isolated PCR products. An intact ORF was present between the endogenous ATG of exon 1 and the beginning of the biotinylation tag sequence running into the downstream Hex cDNA.

4.4 Production and analysis of chimeric mice with BHIV clones.

ES cells from BHIV clones 5.1 and 16.1 were used to generate chimeric mouse embryos by aggregation with F1 C57Bl6 host morulae. Following transfer of aggregates into pseudo-pregnant mice, embryos were isolated at E9.5 for analysis by fluorescence microscopy. Approximately 50% of recovered embryos from each clone displayed green fluorescence indicative of contribution. The pattern of Venus expression in embryos resembled that of Hex seen from previous studies in embryos



a



b

Fig. 4.2 Removal of selection cassette and karyotypic analysis of BHIV clones.

Three BHIV clones were electroporated with a Cre expressing plasmid in order remove the CMV driven HyTK selection cassette from the targeted *Hex* locus. Following selection in gancyclovir surviving ($Ganc^R$) clones were expanded and genomic DNA was isolated for PCR analysis. **a**, Primers specific for the hygromycin resistance gene were used to confirm the removal of the selection cassette from $Ganc^R$ clones . A control PCR using primers specific to the *Hex* promoter region indicates genomic DNA from all samples can be amplified. **b**, Chromosome spreads were prepared from semi-confluent cultures of $Ganc^R$ BHIV clones for karyotype analysis. Forty chromosomes were observed for each clone. C = control parental genomic DNA, W = water.

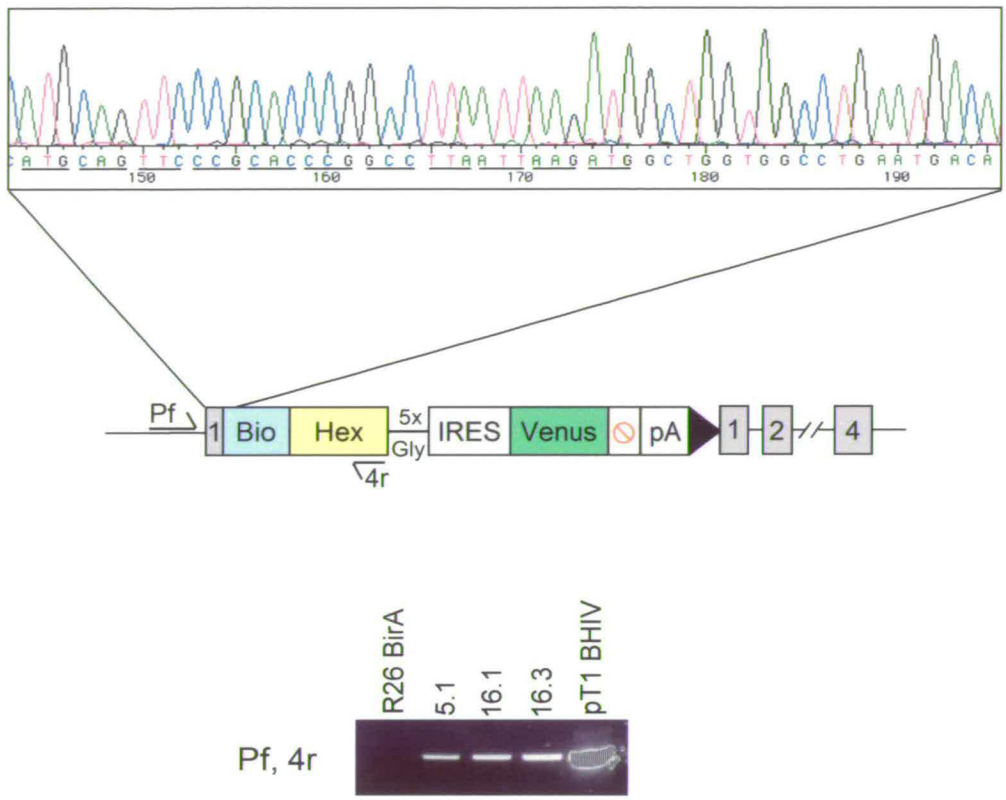


Fig. 4.3 Sequencing of BHIV clones.

PCR was performed on genomic DNA from three BHIV clones with the indicated primers, Pf and 4r. Products of the expected size for amplification from the targeted allele were obtained for each of the clones tested. DNA from unelectroporated R26BirA cells and pT1 BHIV was included as negative and positive controls, respectively. The resultant products were purified and sequenced. The schematic diagram shows sequence obtained spanning the ATG region of *Hex* confirming an intact ORF continuing into the BHIV cDNA.

at the same developmental stage by *in situ* hybridisation (Thomas et al., 1998) or by means of a GFP-transgenic BAC reporter (Rodriguez et al., 2001) (Fig. 4.4).

4.5 *In Vitro* differentiation of BHIV clones.

Previous work in our laboratory has produced a reporter cell line in which a variant of DsRed, Redstar, was knocked into the *Hex* locus. This Hex-Redstareporter line (HRS) has been used to establish *in vitro* differentiation protocols that generate 20-30% of cells expressing the Redstar protein. It has been shown that this population expresses genes of the ADE and does not express markers of neuroectoderm or those that are unique to mesoderm (Morisson et al., in press). BHIV clones were subjected to the 7 day protocol outlined in figure 4.5a in which ES cells are allowed to differentiate in non-adherent cultures in the presence of activin A, an agonist of TGF-β/Nodal signalling. HRS and parental R26BirA cells were included in the experiment as positive and negative controls, respectively. Resultant EBs, grown in the presence or absence of activin A, were dissociated and subjected to FACs analysis to measure the emergence of fluorescent cells. As expected for the HRS line, 22% cells were positive for red fluorescence, an effect that was activin specific. Each of the two BHIV clones showed high percentages of green fluoresce even in the absence of activin, however these percentages were 10-15% greater in the presence of activin (Fig. 4.5b).

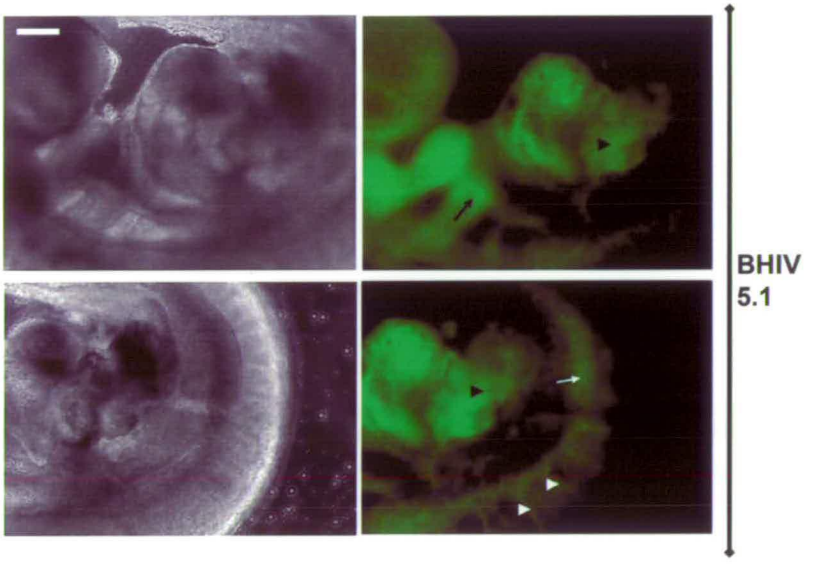
The high background fluorescence observed with the BHIV clones in the absence of activin was surprising due to the lack of redstar positive cells seen in the HRS line under the same conditions. This difference could be explained by the choice of fluorescent reporter used in the BHIV line. The use of Venus has been reported to have similar high background fluorescence in conjunction with a Sox17 reporter construct in mouse embryos (Sherwood et al., 2007). Additionally, the IRES element positioned between BioHex and Venus is known to be extremely efficient at driving translation bicistronically due to the presence of repeated gtx sequences

Fig. 4.4 Chimera analysis of BHIV cells

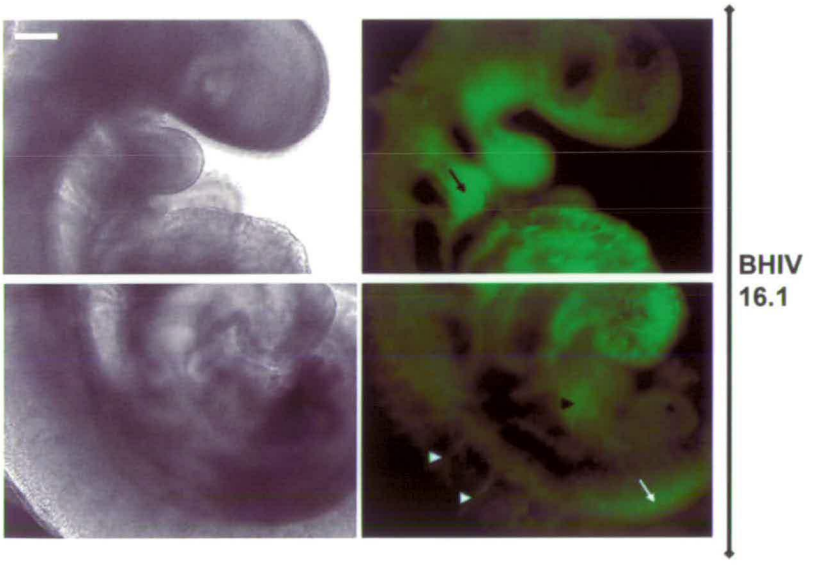
ES cells from BHIV clones 5.1 and 16.1 were aggregated with F1 host morulae for chimera analysis. Following transfer into psuedo-pregnant mice, embryos were obtained at E9.5 and observed with fluorescence microscopy. **a**, Whole mount in situ hybridisation analysis at E9.5 showing *Hex* expression in the thyroid (black arrowhead), liver primordia (white arrowhead), intersomitic vessels (white arrows) and roof of the dorsal aorta (black arrow) (Thomas et.al. 1998). **b**, Transgenic expression of GFP under the control of the *Hex* promoter of a simliar stage embryo reflects that seen for *Hex* in **a** (Rodriguez et. al. 2001). **c**, BHIV chimera produced from clone 5.1 showing expression of Venus in the thyroid (black arrow), intersomitic vessels (white arrowheads), the dorsal aorta region (white arrow) and liver primordium (black arrowhead). **d**, BHIV chimera produced from clone 16.1 showing similar regions of Venus expression.

a (Thomas et al, 1998)

b (Rodriguez et al, 2001)



c



d

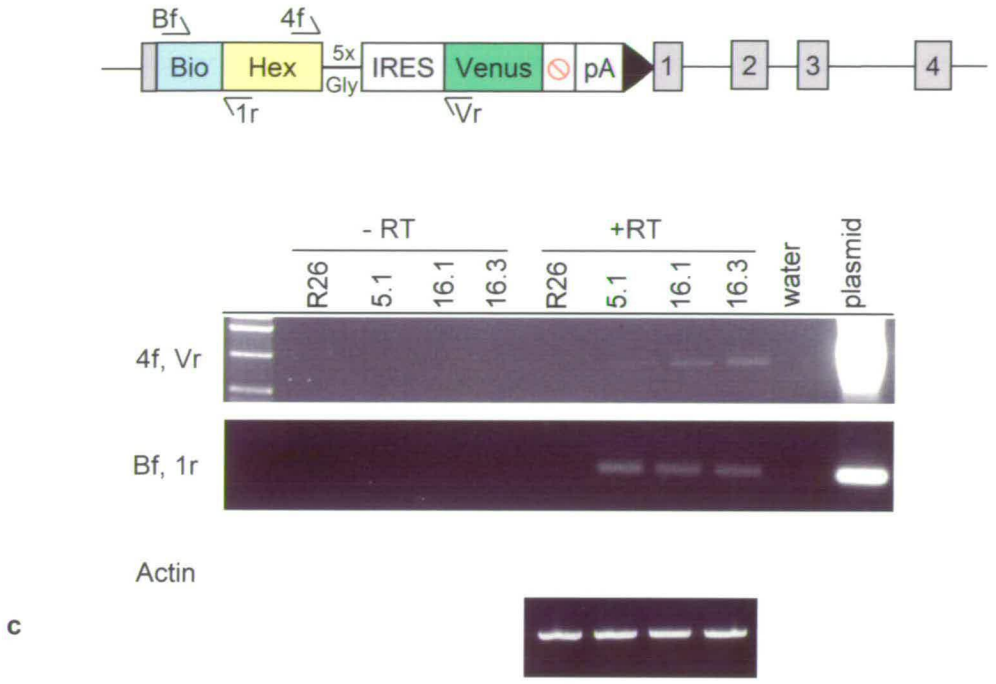
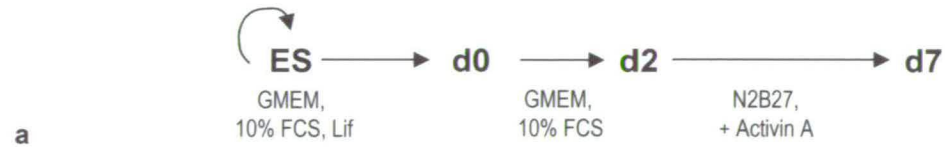
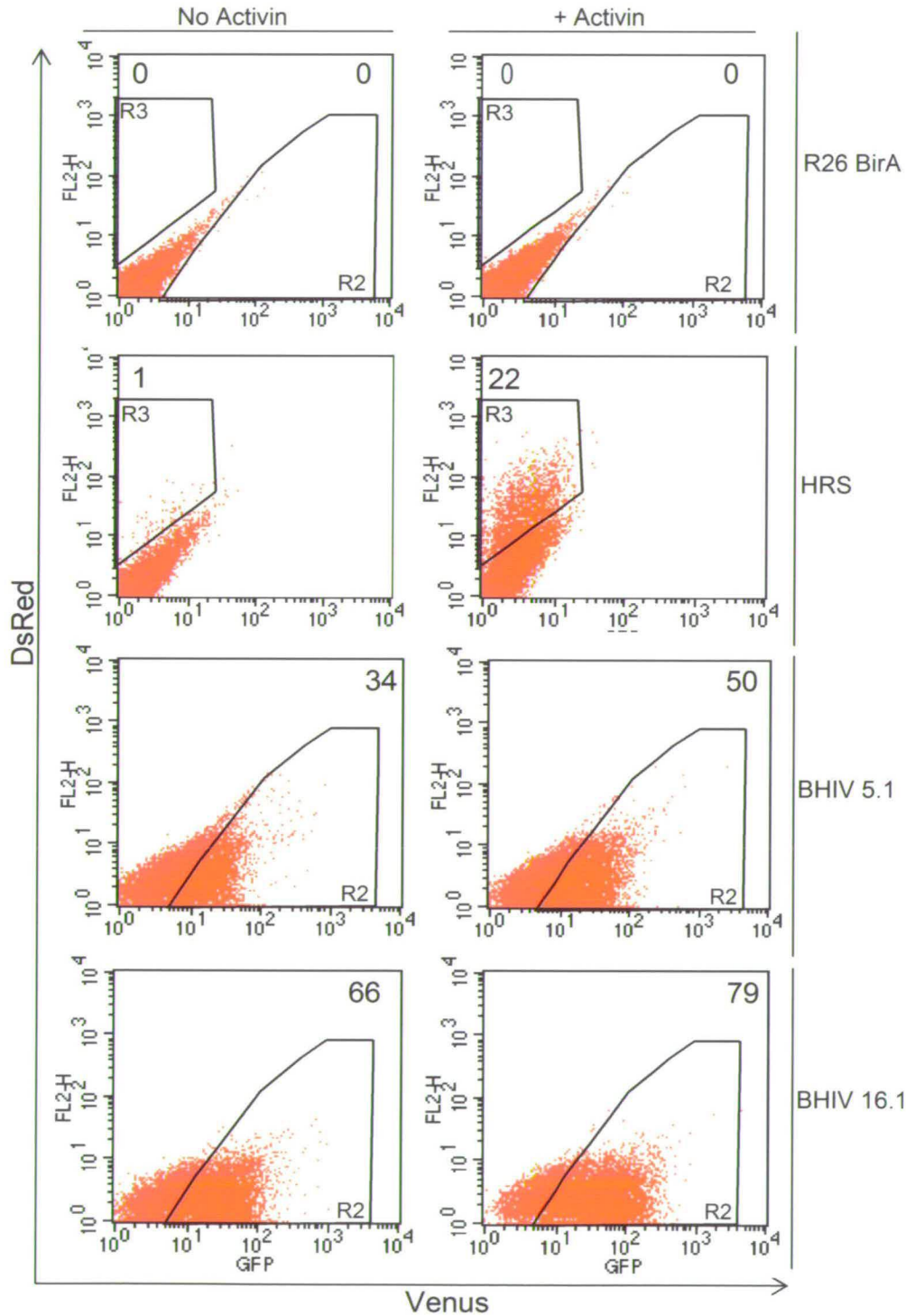


Fig. 4.5 *In vitro* differentiation of BHIV cells.

BHIV clones were cultured during *in vitro* differentiation to generate embryoid bodies (EBs) using a protocol shown to promote the appearance of cells expressing genes indicative of definitive endoderm (Morrison et al., in preparation). The Hex Redstar (HRS) reporter line and control R26BirA cells were cultured as positive and negative controls, respectively. **a**, schematic representation of the differentiation protocol. Each line was cultured in the presence (+) or absence of activin. **b**, At day 7, EBs were dissociated and analysed by FACs to measure the presence of red or green fluorescent cells. The Y axis represents the FL2 channel which was set to measure DsRed fluorescence in order to monitor *Hex* expression from the HRS line. The X axis represents the FL3 channel set to measure Venus fluorescence from the BHIV line. **c**, RNA was collected from 3 BHIV clones cultured with activin for cDNA synthesis and PCR analysis with the indicated primers to detect the presence of BioHex mRNA.


ES
d0
d2
d7
GMEM,
10% FCS, Lif
GMEM,
10% FCS
N2B27,
+ Activin A

a


No Activin
+ Activin
R26 BirA
HRS
BHIV 5.1
BHIV 16.1
0
0
0
0
1
22
34
50
66
79
R3
R2
FL2-H
DsRed
GFP
Venus

b

(Chappell et al., 2000). Despite this observation, the presence of activin always yielded 10-15% more Venus positive cells during this differentiation procedure.

To confirm the expression of BHIV mRNA, rtPCR was performed on RNA from activin treated samples. Primers specific to different regions of BHIV successfully detected expression of this cDNA in three different clones (Fig. 4.5c).

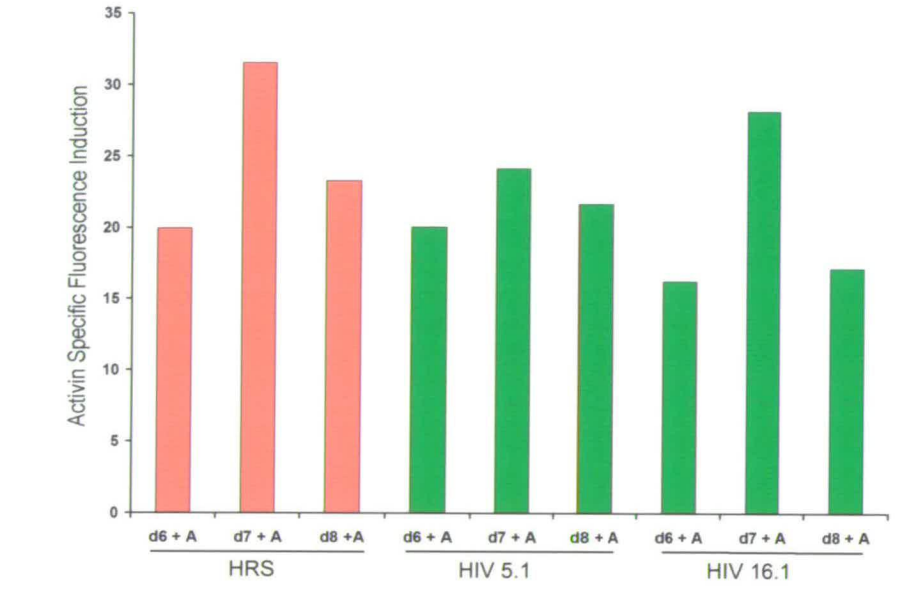
4.6 Time course of activin specific fluorescence induction and definitive endoderm marker expression.

A time course experiment was undertaken using the *in vitro* differentiation protocol to examine in greater detail the emergence of fluorescent cells and definitive endoderm marker expression in the BHIV line. EBs grown in the presence or absence of activin were collected at days 6, 7 and 8 for FACs and qPCR analysis. Figure 4.6a shows the additional percentage of fluorescent cells obtained with activin compared to the same samples untreated. For the HRS line and 2 BHIV clones, the peak of fluorescence induction using this protocol was seen at day 7. Quantitative PCR analysis with primers specific to Hex and Cerberus showed a similar pattern of expression among BHIV clones, the HRS line and the parental R26BirA cells (Fig. 4.6b). Expression levels of both genes were seen to be at a similar level to undifferentiated ES cells at each time point for samples cultured without activin. In the presence of activin, Hex and Cerberus mRNA levels were highest at day 6 in all cell lines. The difference in time for highest fluorescence induction and *Hex* gene expression can be explained by the lag in time for translation and folding of the fluorescent proteins. These results indicate that, in the presence of activin BHIV clones are able to differentiate to form a population of cells expressing anterior endoderm markers and that the increase in Venus expression during this time reflects this.

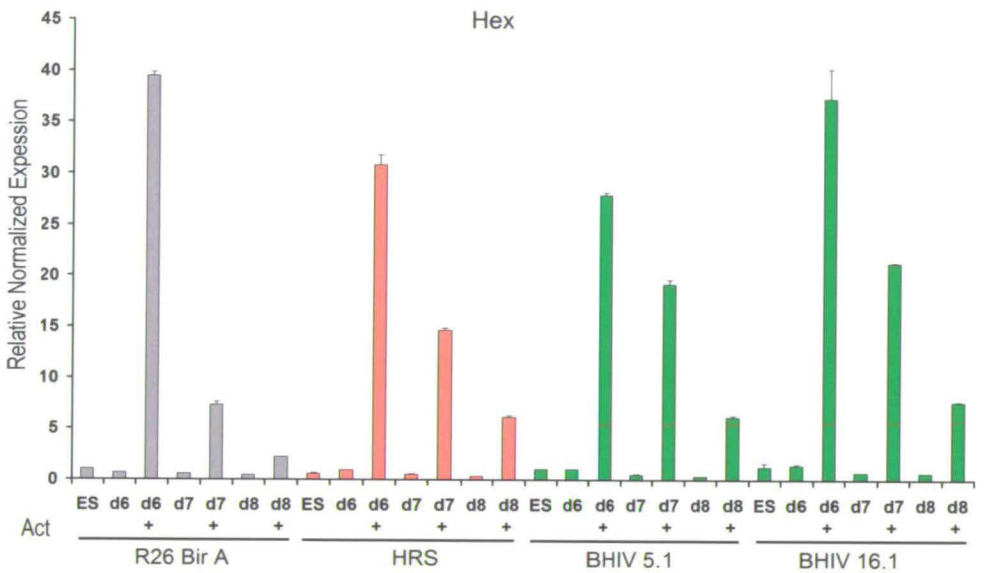
Despite the successful expected physiological expression levels of BioHex cDNA from the *Hex* locus in this ADE cell-type, detection of the BioHex protein proved to

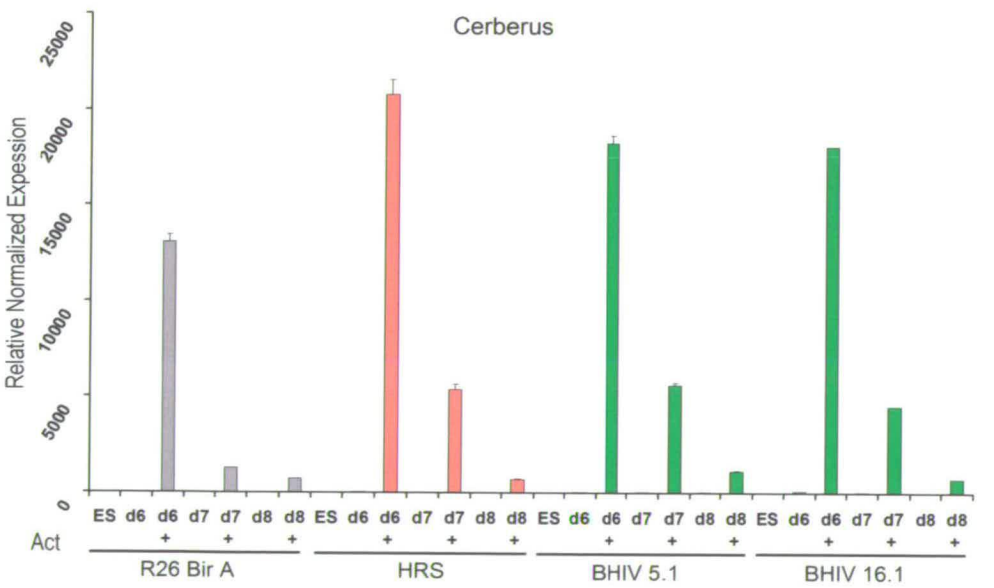
Fig. 4.6 Time course of activin specific fluorescence induction and definitive endoderm marker expression.

In vitro differentiation of 2 BHIV clones in parallel with the HRS reporter line and the R26 BirA parental line was carried out in the presence (+A) or absence of activin. EBs were collected at days 6, 7 and 8 and fluorescence was determined by FACs analysis. **a**, The graph displays the additional percentage of fluorescent cells obtained in samples treated with activin compared to the same samples untreated. **b**, EBs were collected at days 6, 7 and 8 and RNA was isolated for cDNA synthesis. Quantitative PCR using the UPL system was carried out to measure the expression levels of the definitive endoderm markers *Hex* and *Cerberus* at each time point alongside undifferentiated cells (ES) for each cell line. Hex and Cerberus levels were normalised to TBP levels for each sample. Normalised levels are related to the undifferentiated R26 BirA sample for each PCR.


Activin Specific Fluorescence Induction
35
30
25
20
15
10
5
0
d6 + A
d7 + A
d8 +A
d6 + A
d7 + A
d8 + A
d6 + A
d7 + A
d8 + A
HRS
HIV 5.1
HIV 16.1

a


Hex
Relative Normalized Expession
45
40
35
30
25
20
15
10
5
0
ES d6 d6 + d7 d7 + d8 d8 +
Act
R26 Bir A
HRS
BHIV 5.1
BHIV 16.1


Cerberus
Relative Normalized Expession
25000
20000
15000
10000
5000
0
ES d6 d6 + d7 d7 + d8 d8 +
Act
R26 Bir A
HRS
BHIV 5.1
BHIV 16.1

b

be problematic. One possible reason for this is the naturally occurring modest transcriptional activity at this genomic region. Despite the 50 fold induction of *Hex* from ES cell levels to those of day 7 embryoid bodies in the presence of activin, the absolute copy numbers were much less than those obtained for the gene *Cerberus* during this procedure. Additionally, a higher turn-over rate for the BioHex protein than that for Venus could have lessened the ability for detection further still. It is possible that with improved differentiation procedures in the future, more efficient production of anterior endoderm cells in culture will provide the opportunity to detect and utilise the BioHex protein expressed from the *Hex* locus in order to undertake target and partner screening.

4.7 Expression of Venus in undifferentiated cultures

Observation of BHIV clones under self-renewing conditions revealed detectable Venus fluorescence by microscopy in some but not all cells. FACs analysis of cultures confirmed this green fluorescence and staining with an anti-SSEA1 antibody demonstrated that most of the Venus positive and negative cells were undifferentiated ES cells (Fig. 4.7). This result was somewhat surprising as Hex is not expressed in the embryo until the formation of the primitive endoderm at E4.5 (Chazaud et al., 2006; Mesnard et al., 2006; Thomas et al., 1998; Torres-Padilla et al., 2007).

4.8 BHIV expression reflects endogenous noise from the Hex locus.

The detectable expression of Venus in ES cells of BHIV clones raised the concern of possible misexpression since Hex is expressed minimally under self-renewing conditions (Kubo et al., 2004; Kunath et al., 2005). Additionally, the HRS reporter line does not display red fluorescence when cultured similarly. To address whether this effect was due to either aberrant expression of BHIV cDNA from the targeted allele or the highly sensitive nature of IRES Venus as a reporter, qPCR was

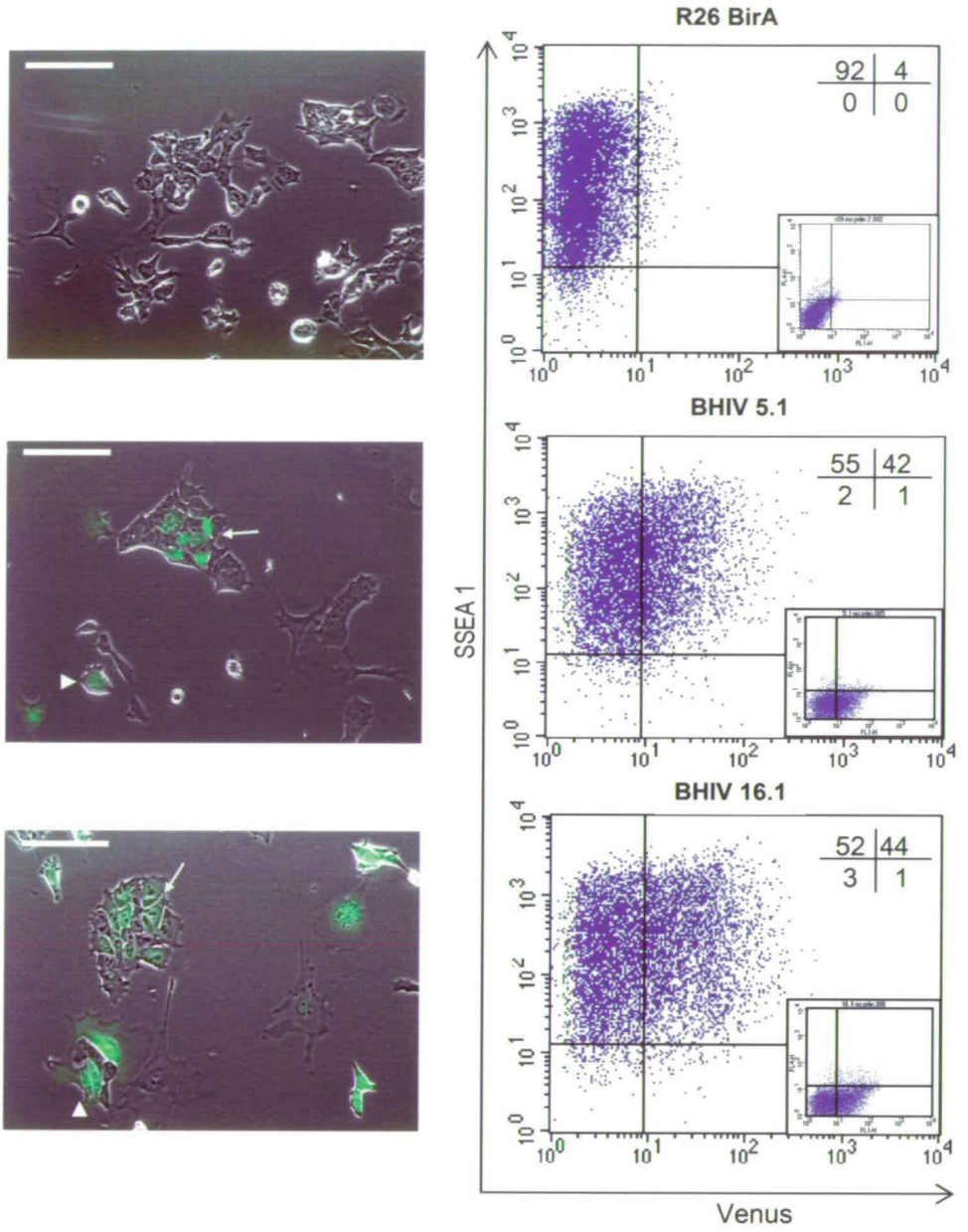


Fig. 4.7 Expression of Venus in a subpopulation of SSEA 1 positive BHIV cells under self-renewing conditions.

R26 BirA parental cells and BHIV clones maintained under self-renewing conditions were observed by fluorescence microscopy. Venus expression could be observed in a sub-population of ES cells from the BHIV clones. FACs analysis for each line is shown in which cells were stained for the presence of the cell surface protein SSEA-1. Percentages for each quadrant are shown. Scale bars represent 25μm. Control plots for SSEA-1 staining are shown as insets.

undertaken to measure transcript levels of BHIV mRNA. Primers were chosen to amplify either total Hex mRNA (endogenous Hex and BHIV mRNA) or BHIV mRNA only. Figure 4.8 shows the results of the qPCR in which the variation of total Hex mRNA among BHIV clones is closely reflected by the BHIV mRNA levels. Additionally, BHIV levels never exceeded those for total Hex in the parental R26BirA line or wild-type Cgr8 ES cells. These results would indicate that the detectable expression of Venus seen under self-renewing conditions is not as a result of misexpression from the targeted allele but rather it is due to the highly sensitive nature of the IRES Venus sequence as a reporter.

4.9 Venus expression co-localizes with Oct4 but not with Nanog.

To verify the presence of Venus in undifferentiated cultures, BHIV clones were stained with antibodies specific to the pluripotency markers, Oct4 or Nanog. While Oct4 appeared to be expressed in all ES cells from BHIV clones and the parental R26BirA line, Nanog expression was more heterogeneous (Fig. 4.9a,bc,d). This observation of Nanog heterogeneity in ES cells is in agreement with other studies (Chambers et al., 2007; Singh et al., 2007). Venus expression from BHIV clones could be detected in Oct4 positive cells, however, it tended to be mutually exclusive with Nanog expression (Fig. 4.9c,d). Interestingly, some ES cells showed neither Venus nor Nanog expression, suggesting the possibility of multiple subpopulations existing under self-renewing conditions.

4.10 Gene expression analysis of Venus positive and negative subpopulations.

In order to examine the differences existing between Venus positive and negative cells, BHIV clones were subjected to flow cytometry to segregate the subpopulations. Cells were stained with an anti-SSEA1 antibody so that any spontaneous differentiated cells, which occur naturally in culture, were

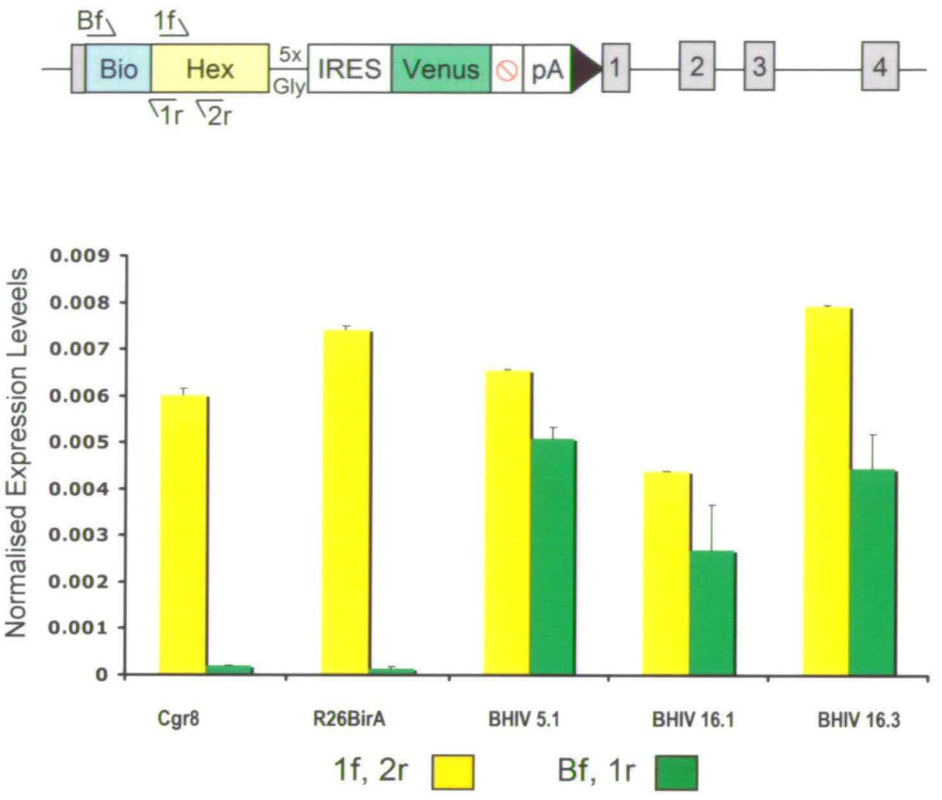
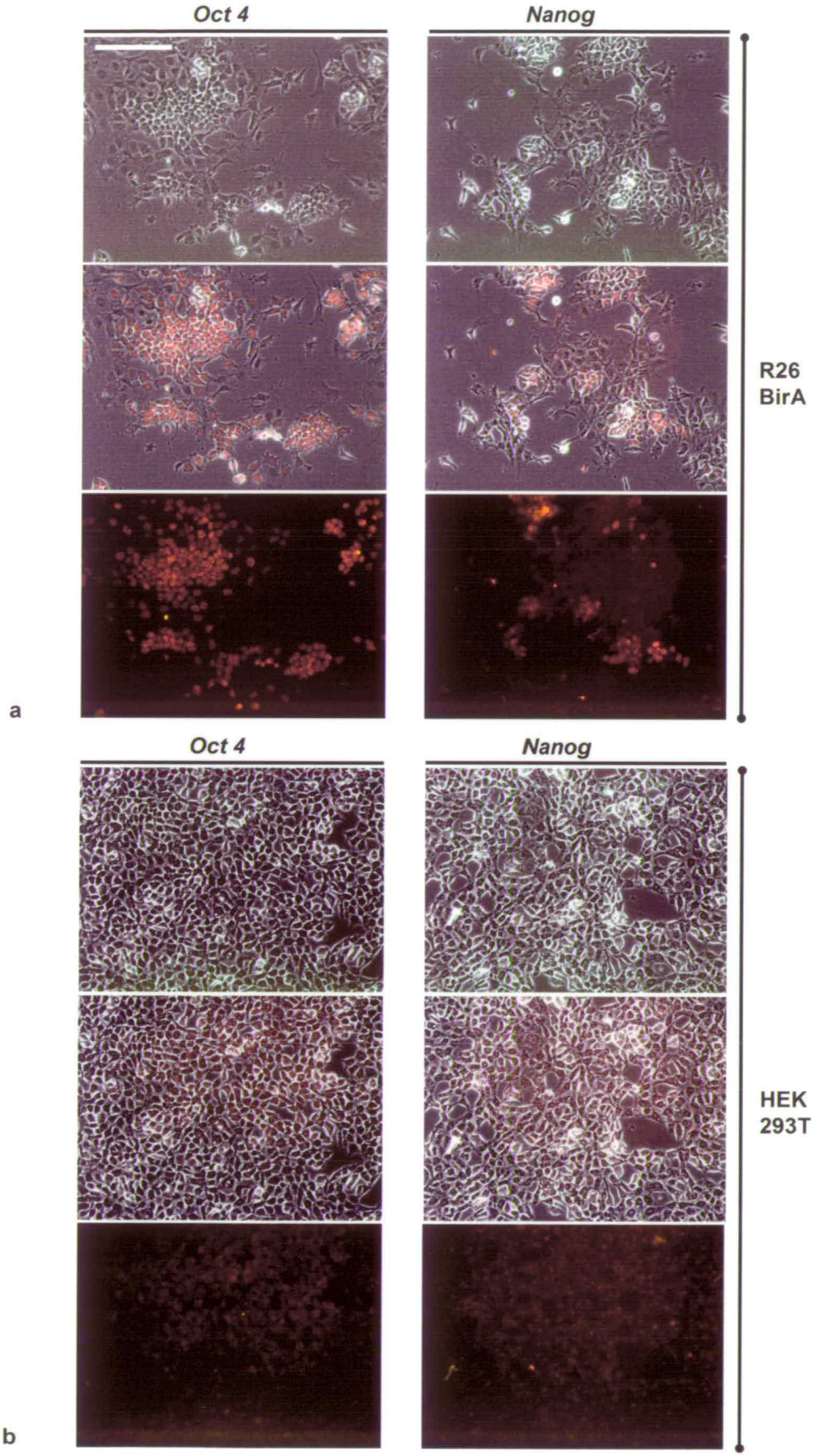


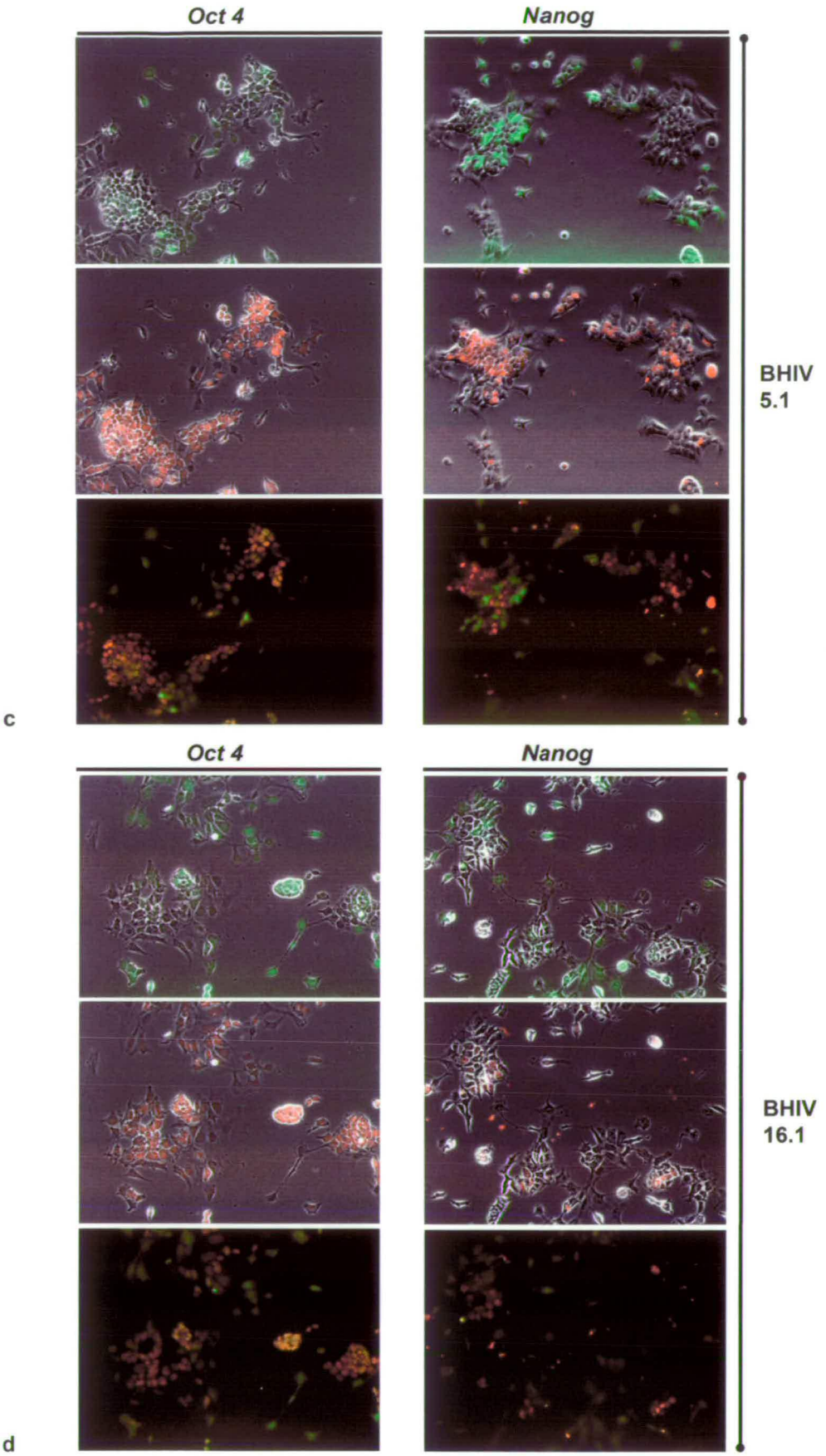
Fig. 4.8 Venus expression reflects endogenous noise at the *Hex* locus among different BHIV clones.

RNA and cDNA was prepared from self-renewing cultures of three BHIV clones, parental R26 BirA cells and Cgr8 cells. qPCR analysis was carried out using Syber Green to monitor levels of mRNA derived from both targeted and untargeted alleles of *Hex* (1f, 2r) or targeted allele only (Bf, 1r). The schematic diagram depicts the different primers used. Values for each primer set used were normalised to the actin value obtained for each sample.

Fig. 4.9 Immunocytochemistry of BHIV clones under self renewing conditions.

Two BHIV clones and R26BirA parental cells cultured under self-renewing conditions were fixed and immunostained to observe the presence of Oct4 and Nanog proteins. Primary antibodies specific to Oct4 and Nanog were detected using Alexa 568 conjugated secondary antidodies. Fixed HEK293 cells were immunostained in parallel as a non-ES cell control. **a**, R26BirA cells. **b**, HEK293 cells. **c**, BHIV 5.1. **d**, BHIV 16.1. Green/Bright field, Red/Bright field and mixed Green/Red field images are shown for each cell line and each antibody. Scale bar represents 50μm.


Oct 4
Nanog
R26
BirA
a
Oct 4
Nanog
HEK
293T
b


Oct 4
Nanog
BHIV
5.1
c
Oct 4
Nanog
BHIV
16.1
d

eliminatedfrom the sorting procedure (Fig.4.10a). RNA was isolated from each fraction and following cDNA synthesis, qPCR was carried out using the Roche UPL system. Comparison of Hex mRNA levels between the two subpopulations showed an average of 60% less in the Venus negative fractions than in the positive fractions from two different clones. These differences were seen to be greater still with another marker of primitive endoderm, *Gata6*, segregating with the Venus positive fraction (Fig. 4.10b).

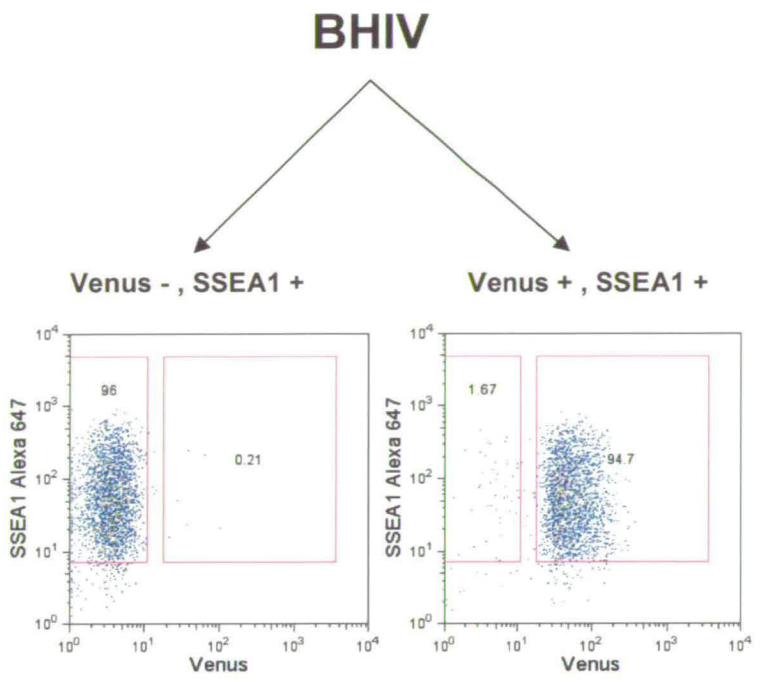
Markers for pluripotency were next examined by qPCR on the same samples. Expression of *Nanog*, *Klf4* and *Rex1* was seen to be highest in the Venus negative fractions, averaging at least 50% more than in the positive fractions. Levels of *Oct3/4*, on the other hand were on average equivalent between the two populations (Fig. 4.10c). These data suggest that Venus positive ES cells in BHIV clones represent a subpopulation pertaining to primitive endoderm in character and possibly a precursor of this lineage.

4.11 Microarray Analysis of BHIV subpopulations

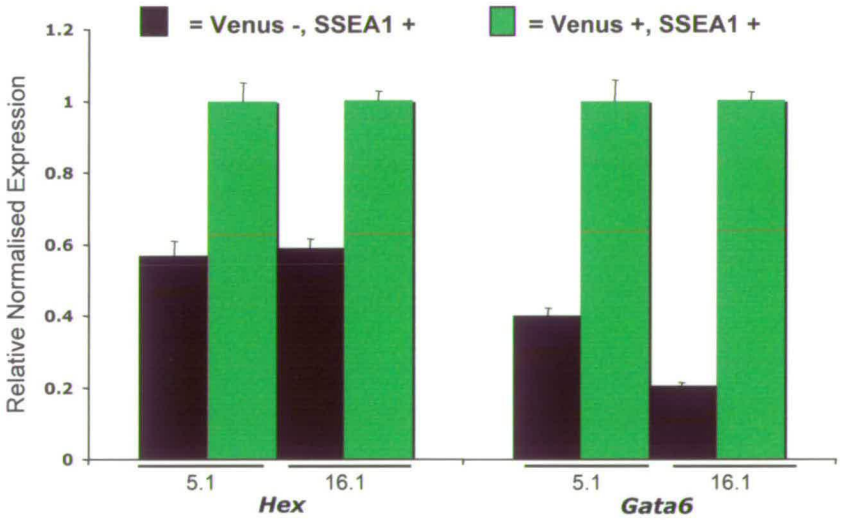
Two different BHIV clones (5.1 and 16.1) growing under standard self-renewing conditions were subjected to flow cytometry to separate four different fractions based on Venus (V) and SSEA1 (S) expression. RNA was isolated from the following fractions of both clones: V-,S+; V+,S+; V-,S-; V+,S- and samples were prepared in duplicate (a total of four replicates from each fraction, two technical and two biological). In collaboration with Minoru Ko and Alexei Sharov at the NIH in Baltimore, samples were hybridised to the NIA Mouse 44K Microarray v2.1 (Carter et al., 2005). Pairwise comparisons were performed using standard statistical conditions (FDR<0.05, > 1.5-fold expression levels) to unveil genes upregulated or downregulated between the Venus positive and negative cells of the SSEA1+ fraction (Fig. 4.11a). As can be seen, very few genes have expression level differences between the V- and V+ subpopulations of the SSEA1+ fraction (ES cells)

Fig. 4.10 Gene expression analysis of BHIV subpopulations.

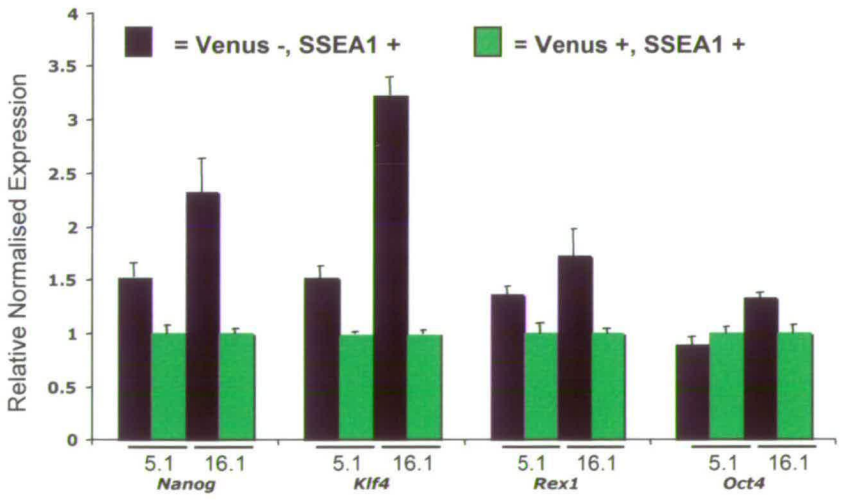
BHIV clones cultured under self renewing conditions were subjected to flow cytometry to separate venus positive and negative subpopulations. **a**, BHIV clones were labelled with an anti-SSEA 1 antibody in order to select SSEA 1 positive cells from the subpopulations. Purity check percentages are shown for both Venus negative, SSEA 1 positive and Venus positive, SSEA 1 positive fractions collected for RNA extraction and cDNA synthesis. **b**, qPCR analysis was performed with the UPL system to compare transcript levels of the primitive endoderm markers, *Hex* and *Gata6*, in the separated subpopulations derived from 2 BHIV clones. **c**, qPCR was carried out on the same samples to compare transcript levels of the pluripotency markers, *Nanog*, *Klf4*, *Rex1* and *Oct4*. Venus positive fractions are represented as green bars and venus negative, black bars. Transcript levels were normalised to the TBP value obtained for each sample. Normalised values are related to that obtained for the venus positive fraction derived from each clone.


BHIV
Venus - , SSEA1 +
Venus + , SSEA1 +
SSEA1 Alexa 647
Venus
96
0.21
1.67
94.7

a


= Venus -, SSEA1 +
= Venus +, SSEA1 +
Relative Normalised Expression
5.1
16.1
Hex
Gata6

b


= Venus -, SSEA1 +
= Venus +, SSEA1 +
Relative Normalised Expression
5.1
16.1
Nanog
Klf4
Rex1
Oct4

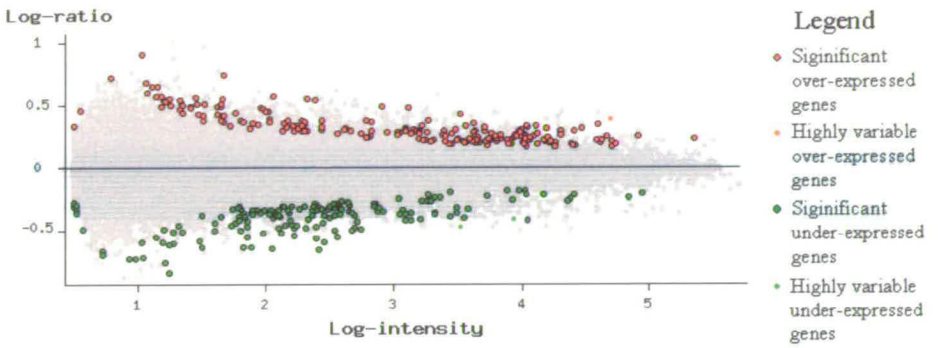
c

Fig. 4.11 Microarray analysis of BHIV clones

BHIV clones cultured under self renewing conditions were subjected to flow cytometry to separate four populations based on Venus (V) and SSEA1 (S) positivity. RNA was isolated from the following fractions: V-,S+; V+,S+; V-,S-; V+,S- and hybridised to a Universal mouse expression array, carried out at the laboratory of Genetics, NIH, Baltimore in collaboration with M. Ko and colleagues. Pair-wise comparisons (FDR<0.05, > 1.5-fold expression levels) were performed between **a**, Venus positive and negative cells in the SSEA1+ fraction and **b**, Venus positive and negative cells in the SSEA1- fraction. Plots are shown comparing mean log intensity values for each gene probe among the four RNA samples. Error bars represent standard deviation between technical (duplicate RNA samples) and biological (two different clones) replicates for each fraction (a total of four replicates). Plots are arranged in groups representing: **c**, pluripotency genes; **d**, primitive endoderm genes; **e**, mesoderm genes; **f**, neurectoderm genes; **g**, neural inhibition genes; and **h**, ICM/Epiblast genes.

Log-ratio chart for HIV_V+_S+ vs. HIV_V-_S+

200 genes overexpressed; 160 genes underexpressed
Click on genes to get info

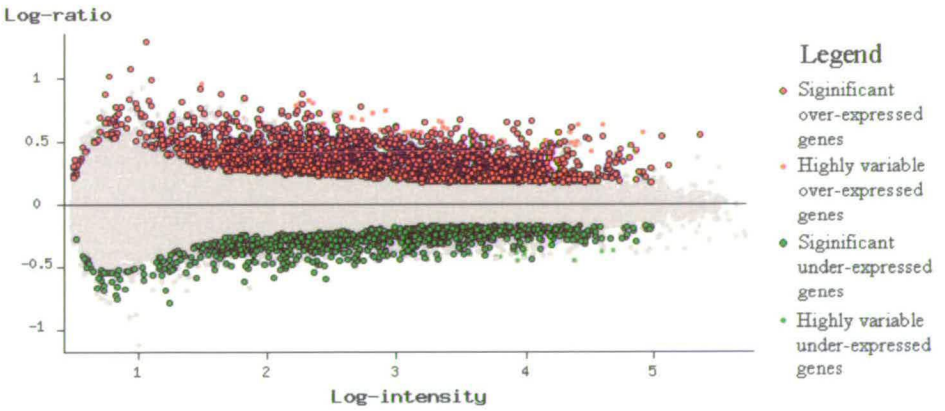


FDR=0.05 and **fold-change threshold**=1.5

a

Log-ratio chart for HIV_V+_S- vs. HIV_V-_S-

1994 genes overexpressed; 1000 genes underexpressed
Click on genes to get info

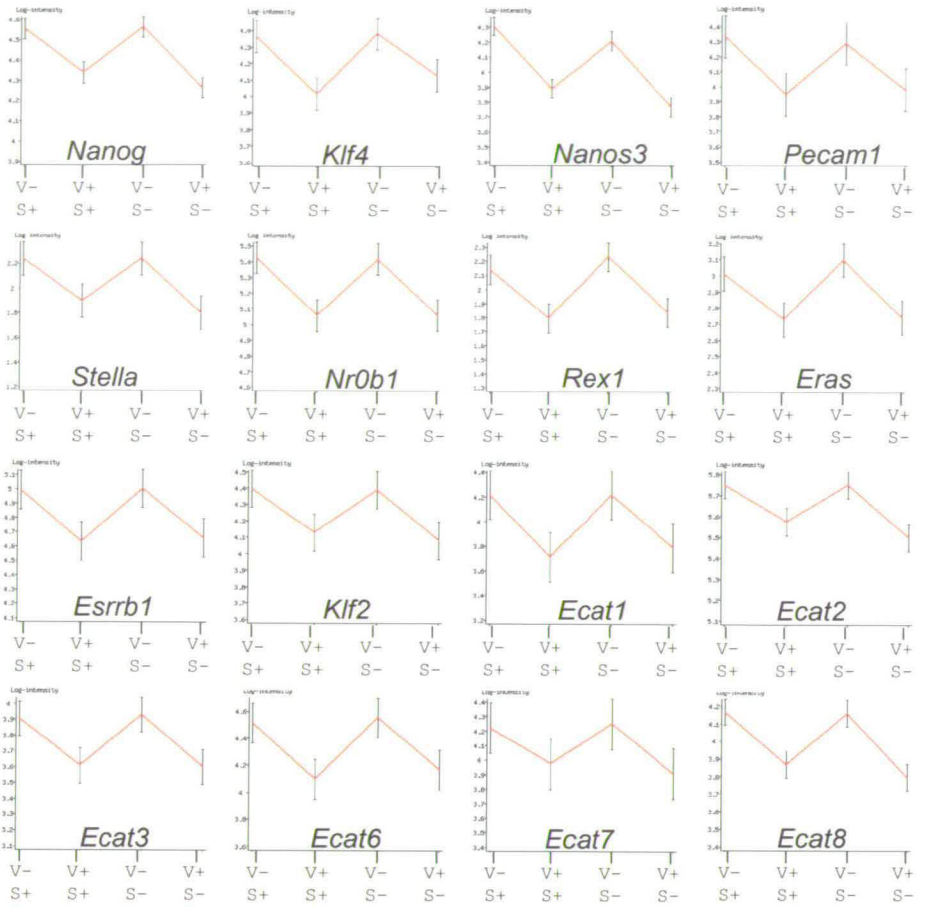


FDR=0.05 and **fold-change threshold**=1.5

Log-ratio vs. mean log-intensity chart for HIV_V+_S- vs. HIV_V-_S-

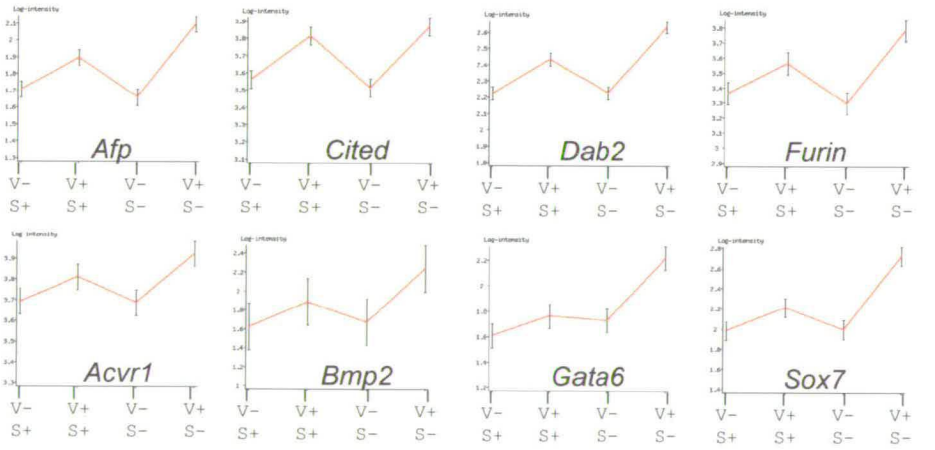
b

Pluripotency Genes



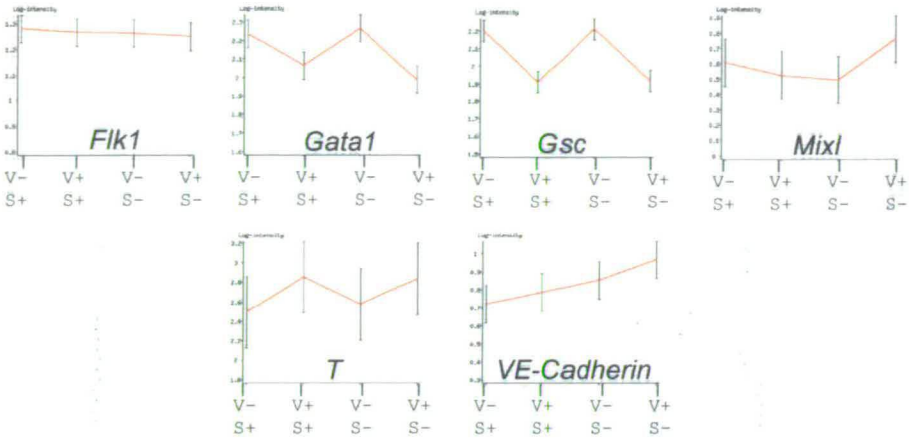
c

Primitive Endoderm Genes



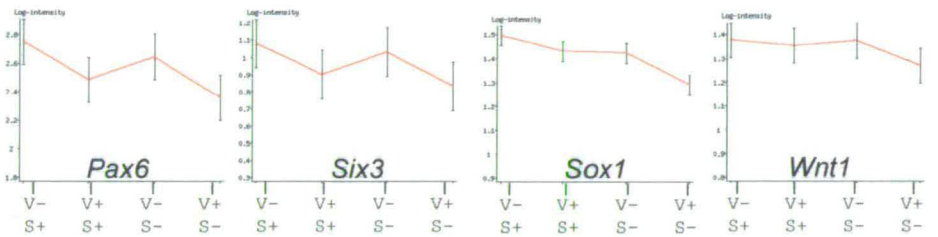
d

Mesoderm Genes



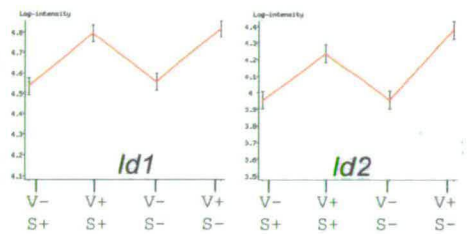
e

Neurectoderm Genes



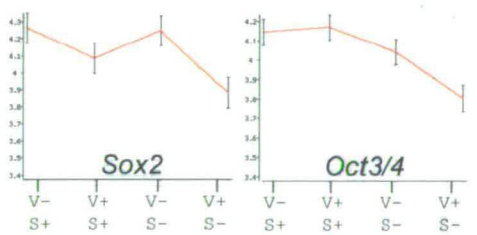
f

Neural Inhibition Genes



g

ICM//Epiblast Genes



h

with only 200 genes more highly expressed in the V+ cells and 160 more in the V- cells. As expected, the SSEA1 negative fraction showed greater differences between V- and V+ cells (Fig. 4.11b). Based on the expression of endoderm associated genes, this appears to reflect the fact that V+, S- cells are further differentiated than the V+, S+ fraction and this would be consistent with the notion that SSEA is a marker of an ICM like state. However, the exact cell typed defined by this marker is not always clear (see below).

Histogram plots were created for selected genes representing tissues of different developmental origin. Log intensity values are shown of each gene for each of the four populations. Error bars represent the standard deviation between the four replicates for each of the four samples. Analysis of 16 different 'superpluripotency' and gonad related genes all showed a strikingly consistent pattern of underexpression in V+ cells (Fig. 4.11c). This result is in agreement with the immunohistochemical and quantitative PCR results shown above in which *Hex* expression is mutually exclusive of *Nanog* and other pluripotency genes. As SSEA1 negative cells are thought to represent spontaneously differentiated cells in culture, it was surprising that the expression values of these pluripotency genes appeared similar between the V-,S+ and V-,S- fractions. This suggest that the V-,S- fraction maybe more 'pluripotent' in nature, while on the other hand, V+,S- cells maybe considered differentiated further still by comparison. Additionally, SSEA1 and *Nanog* have both been shown to display heterogeneous expression pattern in ES cells which do not necessarily overlap (Singh et al., 2007).

To affirm whether Venus positive cells pertain a primitive endoderm character as seen above with quantitative PCR, eight primitive endoderm genes were analysed. Again a consistent pattern can be seen whereby all these genes are expressed at higher levels in V+ cells (Fig. 4.11d). Differences are seen to be greater in the SSEA1 negative fraction, with the highest values for all genes belonging to the V+,S- subpopulation. Thus, while Hex positive ES cells (V+,S+) express slightly higher levels of primitive endoderm genes, the Hex positive spontaneously differentiated population (V+,S-) may be further differentiated toward this lineage.

Genes representative of other lineages were analysed also. No consistent pattern could be discerned among genes associated with mesoderm, haematopoietic or endothelial origin (Fig. 4.11e). This would suggest that Venus positive cells are not differentiating toward (or tending to differentiate toward) these lineages. While genes of the neurectoderm lineage were not significantly different between the fractions, slightly less expression appears consistently in the Venus positive cells, particularly the V+,S- subpopulation (Fig. 4.11f). Interestingly, two genes associated with the inhibition of the neurectoderm lineage , *Id1* and *Id2*, show consistent higher expression in Venus positive cells (Fig. 4.11g). These data would suggest that Venus positive cells have no tendency towards this lineage either and may in fact harbour characteristics inhibitory to its existence. Finally, analysis of Oct3/4 and Sox2 show that there is little if no difference of expression between the V- and V+ cells of the SSEA1+ fraction (Fig. 4.11h). As suggested above, the V+,S- subpopulation appears to be the most differentiated, while the V-,S- subpopulation may still have pluripotent character.

Taken together the microarray analysis demonstrates that Venus positive cells, are specifically enriched for genes of the primitive endoderm lineage while they underexpress pluripotency genes. This data is in agreement with the possibility that a 'pre-primitive endoderm' subpopulation exists among *Oct3/4* and SSEA1 positive ES cells.

Additionally, gene ontology analysis performed by A. Sharov revealed that Venus positive cells were enriched in genes downstream of both Nodal/Tgf-b and components of Wnt signalling (Table 4.1).

4.12 A Nanog binding site exists in a highly conserved region of *Hex* intron 1

TFGB/Nodal related genes
Anxa3, Tagln, Krt18, Krt19, Krt8, Gsn, Vim, Capn2, Anxa1, Cav1, Prss23, Mmp14, Hmga2, Egfr, Ghr, Id2, Rhoc, Acta1, Acta2, Igf2, Wnt9a, Foxc1, Fosl2, Inhba
Wnt related genes
Wnt9a, Wnt7a, Wnt2b, Wnt4, Wisp1, Dkk1

Table 4.1 Gene ontology of Venus positive ES cells.

Gene ontology analysis of the microarray experiment described in section 4.11 was performed by A. Sharov at the NIH. Examination of Venus positive ES cells showed an enrichment of genes activated by TGFB/Nodal signalling as well as genes related to the Wnt pathway.

It has been shown previously that the mosaic expression patterns of *Nanog* and *Gata6* in the ICM of E3.5 blastocysts are mutually exclusive of one another (Chazaud et al., 2006). Additionally, an attempt to derive *Nanog* null mice lead to embryos consisting of disorganised extraembryonic tissue at the expense of epiblast (Mitsui et al., 2003) while in an *in vitro* study, constitutive overexpression of *Nanog* repressed *Gata6* and other markers of primitive endoderm (Hamazaki et al., 2004). It was therefore interesting to ask whether *Nanog* may also be controlling the expression of *Hex*.

Recent analysis of Oct3/4, Sox2 and Nanog binding regions in the Human genome identified a binding site in the first intron of the *HEX* gene occupied by all three transcription factors (Boyer et al., 2005). This region lies within an area of intron 1 that is also bound by Ets and GATA factors important for haematopoietic specification (Donaldson et al., 2005). Therefore, to address whether binding to this region is important in an evolutionary conserved context., analysis of the Human *HEX* gene was performed using the Blast-like Alignment Tool (Genome Bioinformatics, UCSC. http://genome.ucsc.edu/). This binding region positioned at +1098 from the transcription start site was highly conserved among all mammals, chick and Xenopus (Fig. 4.12).

4.13 Overexpression of Nanog reduces Venus expression in BHIV cells

To examine whether the Venus expression seen in BHIV cells might be regulated by Nanog, a CAG driven Nanog IRES Puro cDNA was introduced into BHIV clone 16.1 by electroporation. Alternatively, cells were electroporated with a CAG driven stuffer sequence upstream of IRES Puro to control for the electroporation and selection processes. Clones derived from each construct following selection in 2μg/ml of puromycin were isolated for further analyses.

Initial studies of the role of Nanog have shown that its overexpression in ES cells resulted in an ability to maintain the undifferentiated state in the absence of LIF

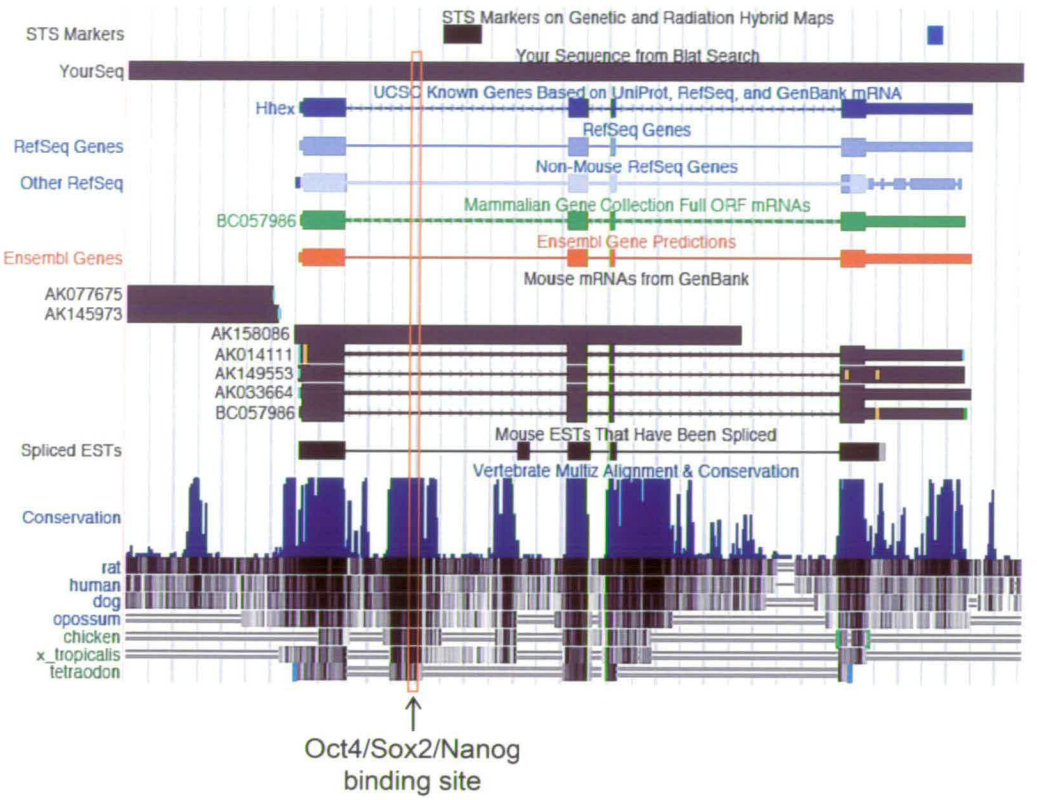


Fig. 4.12 Conservation of Oct4/Sox2 and Nanog binding site in intron 1 of the human *HEX* gene

Based on the study of combinatorial genomic site occupancy by Oct4/Sox2/Nanog using genome wide ChIP analysis, a region positioned at +1098 from the transcription start site of the human HHEX gene was identified (Boyer et al. 2005). Using the Blast-Like Alignment Tool (Blat) (Genome Bioinformatics, UCSC. http://genome.ucsc.edu/), this site was seen to be highly conserved among many species.

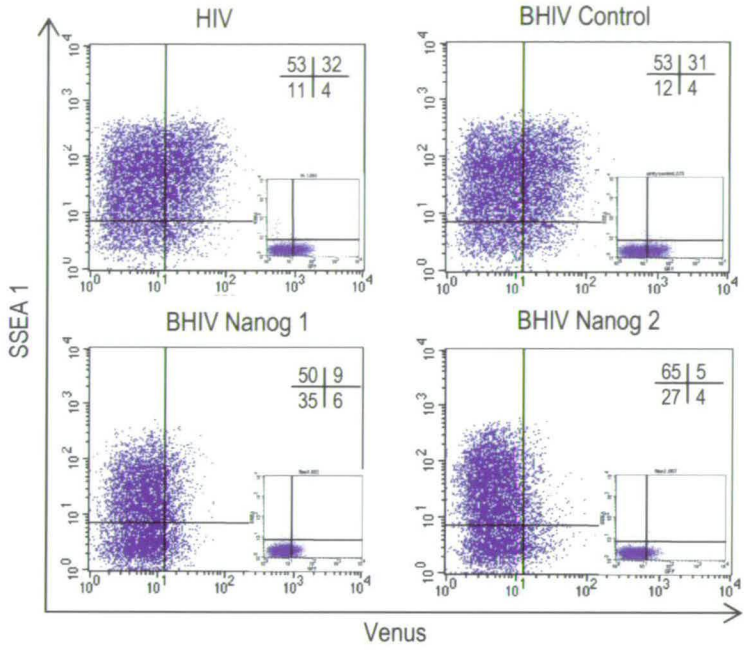
(Chambers et al., 2003; Mitsui et al., 2003). Therefore, to verify overexpression in BHIV cells, clonal transfectants derived from Nanog or Control transgenes were subjected to culturing under selection and in the absence of LIF for 7 days. Control clones showed obvious signs of differentiation while those containing the Nanog transgene remained morphologically similar to ES cells (Fig. 4.13a). Western analysis on the same clones cultured under selection in the presence of LIF confirmed increased Nanog protein levels in those harbouring the transgene (Fig. 4.13b).

Measurement of Venus levels by FACs in Nanog overexpressing BHIV clones revealed an average decrease of fluorescent cells from 35 to 12% when compared to parental untransfected BHIV cells or those expressing the control transgene (Fig. 4.13c). This data suggests that Nanog represses Hex expression in ES cell cultures potentially through similar mechanisms used to reduce other primitive endoderm markers. The following experiments address this possibility.

4.14 Manipulation of FGF signalling alters the levels of Venus Expression

Previous analyses have demonstrated that FGF signalling through the Grb2/Mek pathway is necessary for primitive endoderm specification. Grb2 null embryos fail to form primitive endoderm concomitant with an upregulation of Nanog expression in the ICM (Chazaud et al., 2006). Additionally, *in vitro* studies have shown that ES cells cultured with the FGFR inhibitor SU5402 lead to a similar dual effect of Nanog upregulation and a reduced ability to form primitive endoderm (Hamazaki et al., 2006).

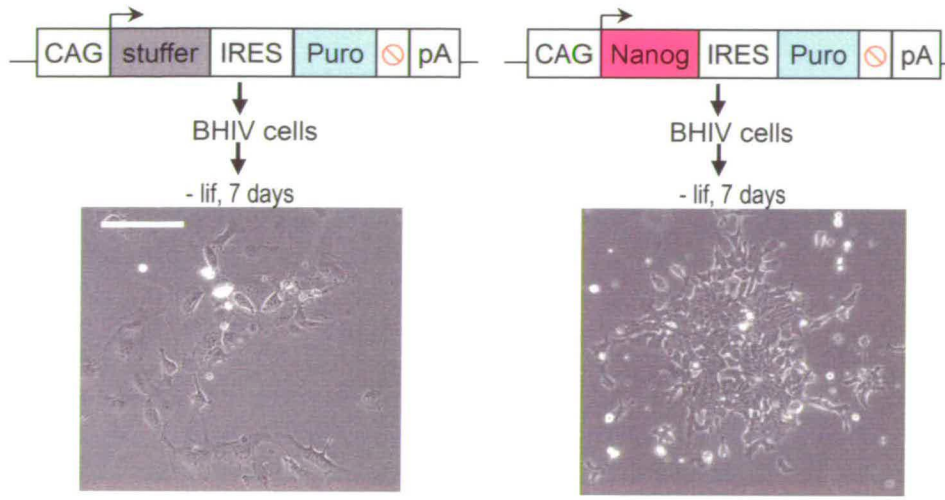
To test whether a block to FGF signalling could lead to a reduction in Hex, expression of Venus was monitored in two different BHIV clones that were cultured for 48 hours under self-renewing conditions in the presence of the FGFR inhibitor PD173074 (Mohammadi et al., 1998) or vehicle (DMSO). In addition, a Sox1-GFP



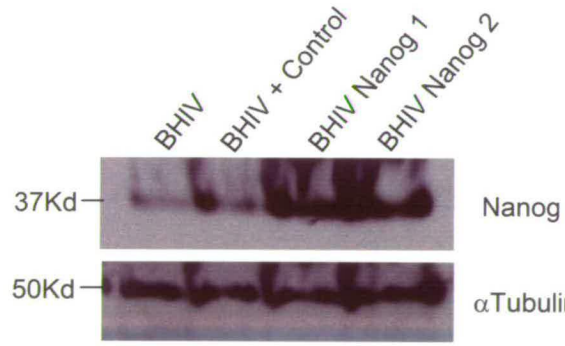
d

Fig. 4.13 Venus expression is reduced in Nanog overexpressing BHIV cells.

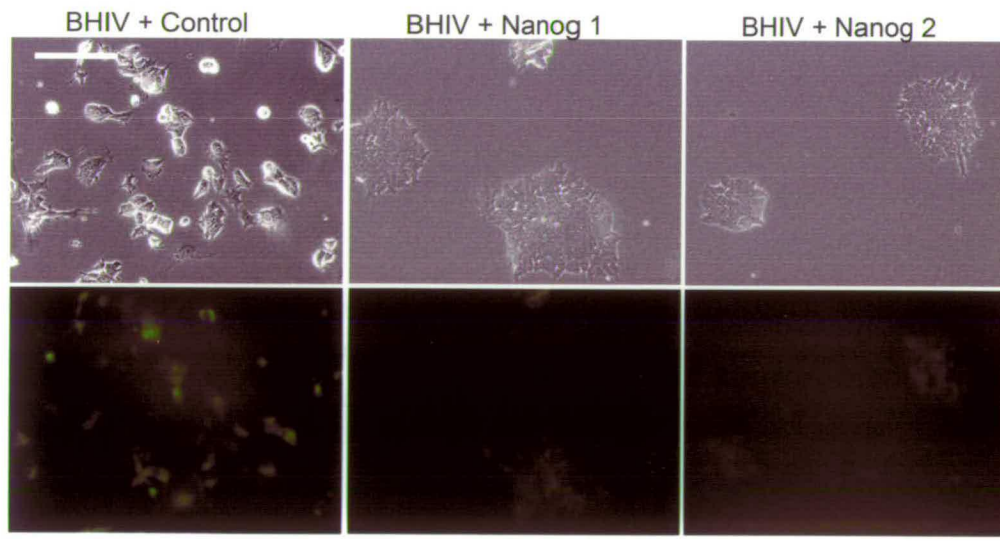
BHIV cells overexpressing Nanog were derived by electroporation of a vector containing the Nanog cDNA downstream of a CAG promoter and selection in puromycin (2µg/ml). Control clones were derived by electroporation of a similar construct but with stuffer DNA in place of the Nanog cDNA. **a**, Puromycin resistant clones derived from Control and Nanog vectors were cultured for 7 days in the absence of LIF. ES cells were still evident in the Nanog overexpressing clones. **b**, Lysates were obtained from control and nanog clones together with the BHIV parental cells grown under self-renewing conditions. Western analysis was performed on the lysates with antibodies specific to nanog or α-tubulin. **c**, Venus expression in Nanog overexpressing, control and parental BHIV cells cultured under self-renewing conditions was monitored by fluorescence microscopy and **d**, FACs analysis. Green and bright field images are shown for each cell line plated at clonal density. For FACs analysis, ES cells were labelled with an anti SSEA 1 antibody. Percentages for each quadrant are shown. Control plots for SSEA 1 staining are shown as insets. Scale bars represent 50µm.


CAG
stuffer
IRES
Puro
pA
BHIV cells
- lif, 7 days
CAG
Nanog
IRES
Puro
pA
BHIV cells
- lif, 7 days

a


BHIV
BHIV + Control
BHIV Nanog 1
BHIV Nanog 2
37Kd
Nanog
50Kd
αTubulin

b


BHIV + Control
BHIV + Nanog 1
BHIV + Nanog 2

c

knock-in line, marking early neural fate decisions, was cultured similarly as a non primitive endoderm reporter control (Ying et al., 2003). Quantitative PCR analysis of isolated RNA from PD173074 treated clones revealed an expected increase of Nanog expression with a slight decrease in Gata6 levels compared to untreated samples (Fig.4.14a). The percentages of Venus expressing cells in BHIV clones was reduced by 10-15% in the presence of PD173074 while fluorescence from the Sox1-GFP reporter line was unchanged (Fig. 4.14b). This result indicates that blockage of FGF signalling leads to a reduction in numbers of a Hex expressing subpopulation while favouring a more pluripotent phenotype reflected by the increase of Nanog mRNA levels.

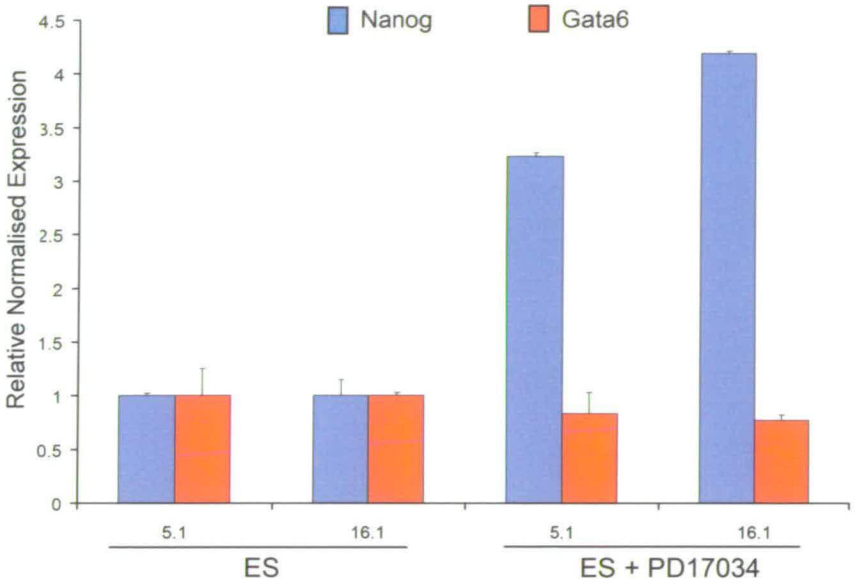
In a complimentary manner, FGF signalling through the Grb2/Mek pathway can be potentiated by the phosphatase inhibitor sodium vanadate leading to increased primitive endoderm formation in aggregate cultures as a consequence of Nanog repression (Hamazaki et al., 2006). BHIV clones were allowed to aggregate for 48 hours on gelatin-free dishes in the presence of LIF and 50μM of sodium vanadate or LIF alone. Measurement of RNA levels by qPCR showed a reduction of Nanog and a concomitant increase of Gata6 in BHIV clones cultured with sodium vanadate (Fig. 4.14c). Analysis by FACs of the same clones revealed an average increase of 25% in the number of cells displaying Venus fluorescence compared to untreated aggregated samples. Sodium vanadate did not increase the number of fluorescent cells from the Sox1-GFP reporter line (Fig 4.14d).

These data confirm the previous findings that potentiation of the FGF pathway leads to a decrease in *Nanog* resulting in the upregulation of the primitive endoderm marker *Gata6* and that the increase in Venus positive cells in BHIV clones reflects this.

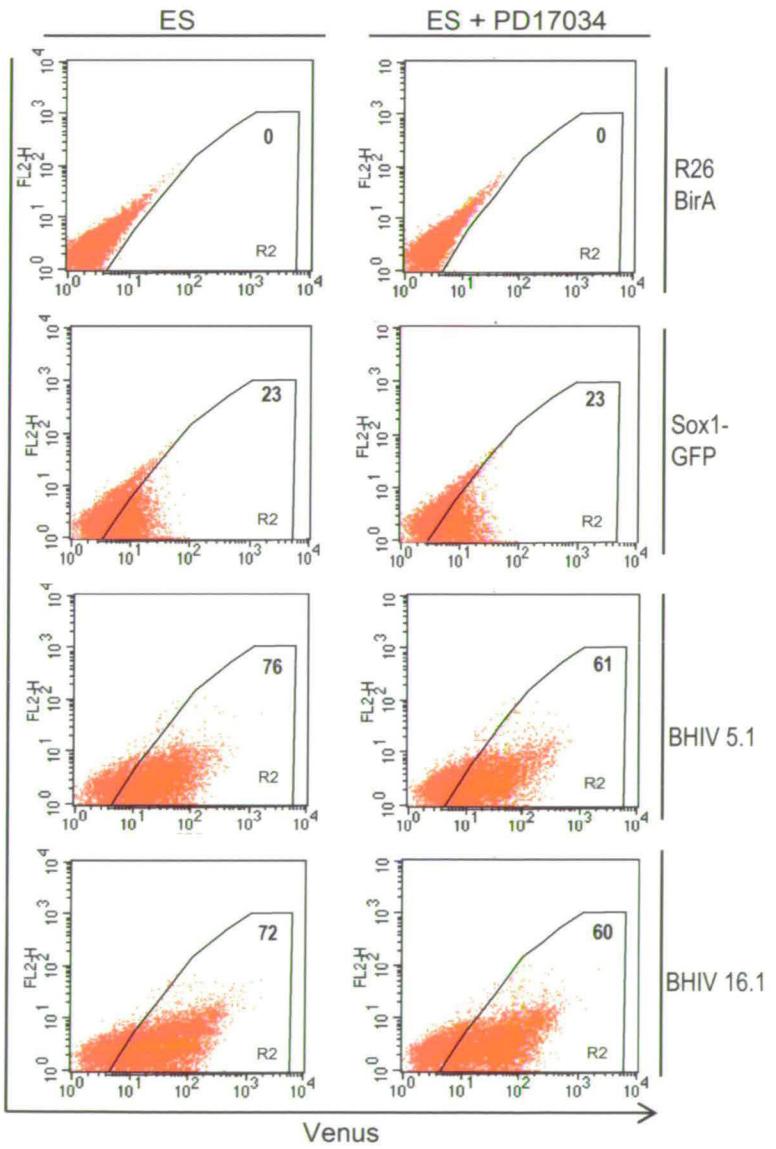
4.15 Reversibility of Venus positive and negative subpopulations

Fig. 4.14 Manipulation of FGF signalling alters levels of Venus expression.

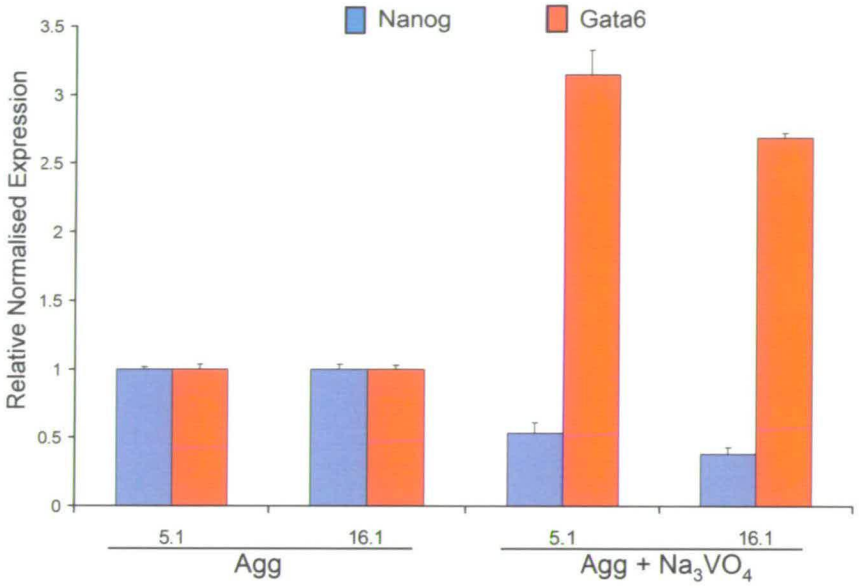
BHIV clones and R26BirA parental cells were cultured under self-renewing conditions in the presence or absence of the FGF receptor inhibitor PD173074 (10nM) for 48 hours. A Sox1-GFP cell line was included in the experiment in order to monitor expression of early neural fate. **a**, Treated and untreated BHIV clones were harvested for RNA extraction and cDNA synthesis. qPCR analysis using the UPL system was carried out to compare the transcript levels of Nanog and Gata6. Transcript levels were normalised to the TBP value obtained for each sample. Normalised values are related to the untreated sample for each clone. **b**, FACs analysis of untreated and treated cultures of R26BirA, Sox1-GFP, BHIV 5.1 and BHIV 16.1 cells are shown. **c**, BHIV clones and R26BirA parental cells were cultured on non-gelatin coated plates, allowing formation of ES cell aggregates in the presence (Agg + Na_3VO_4) or absence (Agg) of 50 μM sodium vanadate. Again, a Sox1-GFP cell line was included as a non primitive endoderm reporter. **c**, Treated and untreated BHIV clones were harvested for RNA extraction and cDNA synthesis. qPCR analysis using the UPL system was carried out to compare the transcript levels of Nanog and Gata6. Transcript levels were normalised to the TBP value obtained for each sample. Normalised values are related to the untreated sample for each clone. **d**, FACs analysis of untreated and treated cultures of R26BirA, Sox1-GFP, BHIV 5.1 and BHIV 16.1 cells are shown.


Nanog
Gata6
Relative Normalised Expression
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
5.1
16.1
5.1
16.1
ES
ES + PD17034

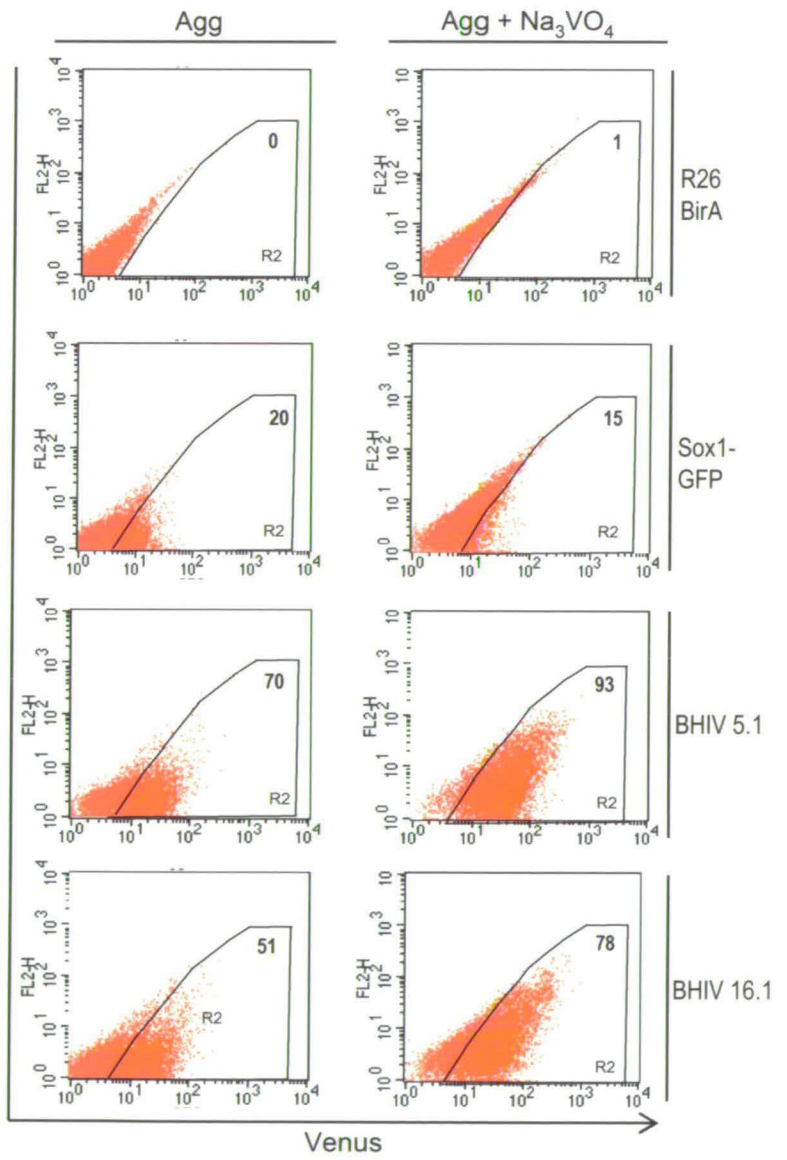
a


ES
ES + PD17034
FL2-H
R2
0
23
76
72
61
60
R26 BirA
Sox1-GFP
BHIV 5.1
BHIV 16.1
Venus

b


Nanog
Gata6
Relative Normalised Expression
3.5
3
2.5
2
1.5
1
0.5
0
5.1
16.1
5.1
16.1
Agg
Agg + Na3VO4

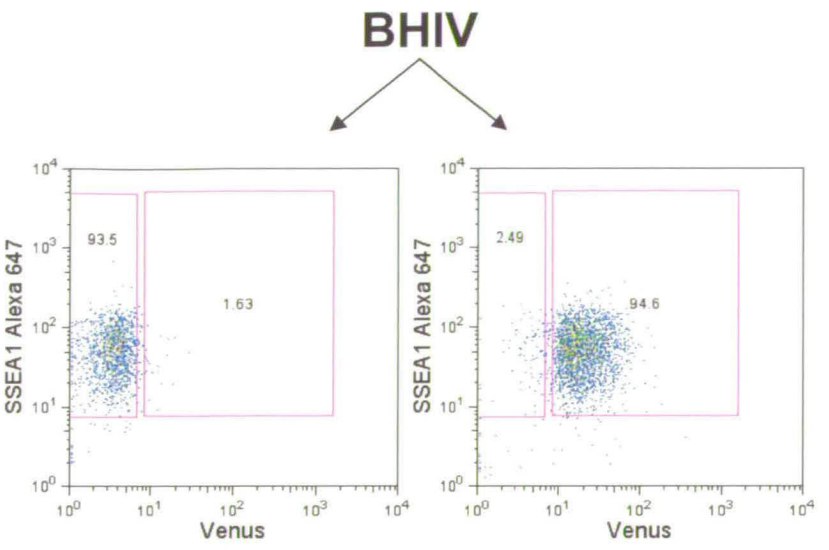
c


Agg
Agg + Na3VO4
R26 BirA
Sox1-GFP
BHIV 5.1
BHIV 16.1
0
1
20
15
70
93
51
78
R2
FL2-H
Venus

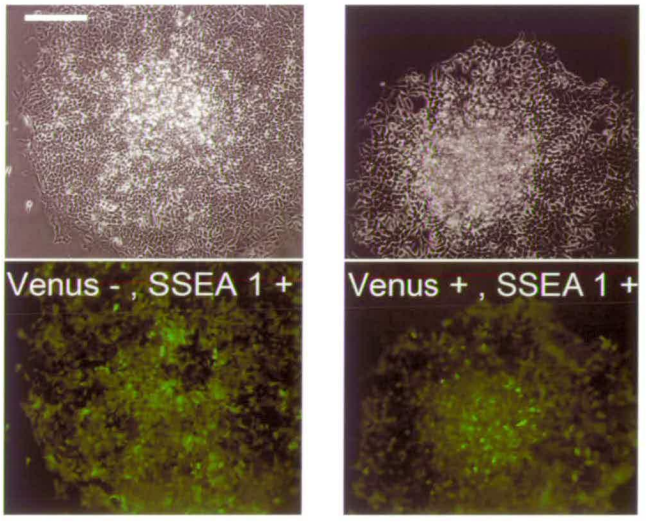
d

Fig. 4.15 Reversibility of Venus positive and negative populations.

BHIV cells cultured under self renewing conditions were subjected to flow cytometry to separate venus positive and negative subpopulations. **a**, BHIV 16.1 was labelled with an anti-SSEA 1 antibody in order to select SSEA 1 positive cells from the subpopulations. Purity check percentages are shown for both Venus negative, SSEA 1 positive and Venus positive, SSEA 1 positive fractions obtained for clonal density and single cell plating. **b**, Representative clones produced from each fraction plated at clonal density (100 cells/ml). Scale bar represents 50μm. **c**, Numbers of clones produced from clonal density plating. **d**, FACs analysis of clones produced from each fraction plated at single cell density. Representative data is shown for clones derived from venus positive and negative fractions together with unsorted BHIV 16.1 cells and R26BirA parental cells. **e**, Following culture under self-renewing conditions for 24 hours, cells from each fraction were stained for SSEA 1 and subjected to FACs analysis. Control plots for SSEA 1 staining are shown as insets.



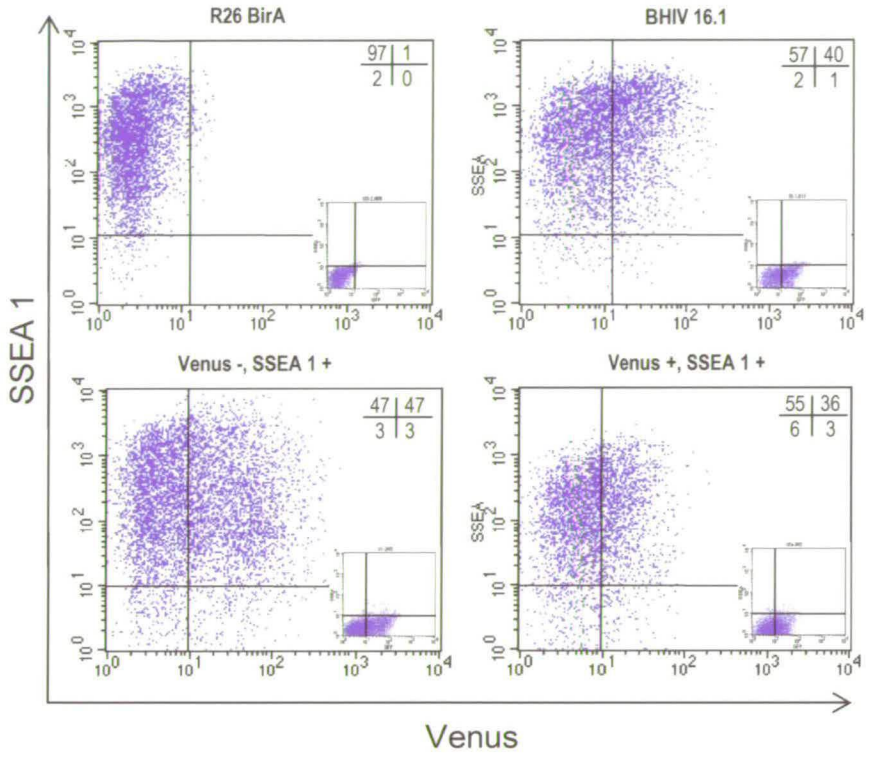
a



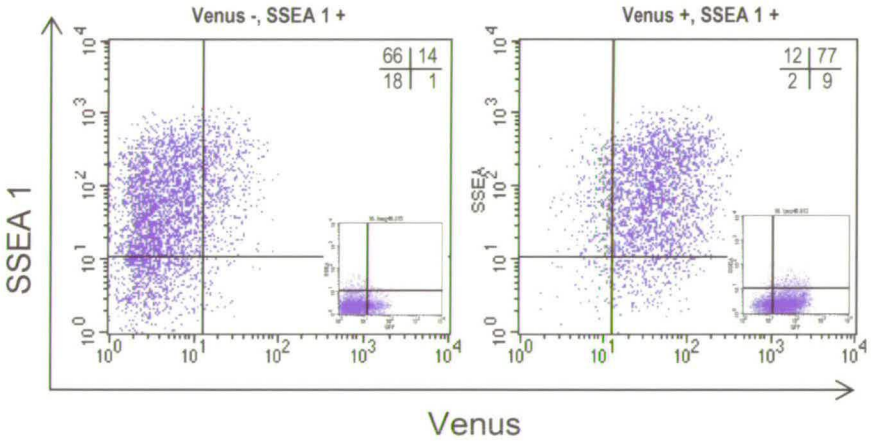
b

	No. of clones obtained	% of clones fluorescent by microscopy
Venus -, SSEA 1 +	90	100
Venus +, SSEA 1 +	21	100

c


R26 BirA
BHIV 16.1
Venus -, SSEA 1 +
Venus +, SSEA 1 +
97 | 1
2 | 0
57 | 40
2 | 1
47 | 47
3 | 3
55 | 36
6 | 3
SSEA 1
Venus

d


Venus -, SSEA 1 +
Venus +, SSEA 1 +
66 | 14
18 | 1
12 | 77
2 | 9
SSEA 1
Venus

e

To examine whether Venus positive and negative subpopulations were restricted to generating cells of their own kind *de novo*, each fraction was purified by flow cytometry and grown in culture (Fig. 4.15a). SSEA 1 positive cells of each purified fraction were plated at clonal density and at a single cell level. Clonal density plating of each population of sorted cells produced colonies with heterogeneous Venus expression, indicating reversibility (Fig 4.15b). Over three times more clones from the Venus negative cells than positive cells were produced suggesting an enhanced ability to permit clonal growth in the former subpopulation (Fig. 4.15c). Additionally, clones derived from single cell plating of the sorted fractions showed similar proportion of fluorescence as unsorted parental cells after 10 days growth (Fig. 4.15d). The ability of single BHIV cells to generate both subpopulations confirms clonality of the line, ruling out the possibility of two different cell types existing together as a mixed clone.

FACs analysis of Venus negative cells plated after flow cytometry separation revealed the emergence of Venus positive cells within 24 hours. Conversely, Venus positive cells cultured for the same duration lead to a similar proportion of cells that interconverted. This result demonstrates that oscillations in *Hex* expression occur rapidly (Fig. 4.15e).

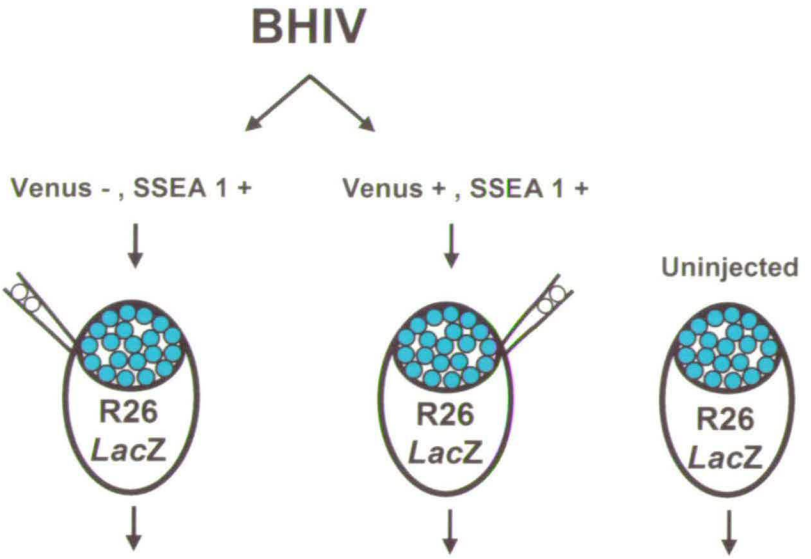
4.16 Fetal contribution potential of Venus positive and negative cells

To test the ability of each subpopulation to contribute to the embryo, each was separated by flow cytometry and SSEA1 positive cells of each fraction were injected into ROSA26 lacZ blastocysts. In this experiment the ES cell contribution should be negative for β-galactosidase contribution (Fig. 4.16a). Analysis of chimeric embryos revealed Venus fluorescence in those generated from both V+ and V- cells (Fig. 4.16b,c). Following staining for lacZ expression embryos were observed for contribution. The level of contribution appeared much greater in embryos generated from the Venus negative cells (Fig. 4.16d). No difference in contribution to specific lineages could be seen in transverse cryostat sections (Fig. 4.16e). Analysis of all the

Fig. 4.16 Chimera and contribution potential analysis of venus positive and negative subpopulations.

BHIV cells cultured under self renewing conditions were subjected to flow cytometry to separate venus positive and negative subpopulations. Cells were stained for SSEA-1 prior to sorting to select the ES cells from each population **a**, Schematic diagram showing the strategy to generate chimeric mice by injection of the sorted populations using R26LacZ host blastocysts. Sorted cells were injected 1 hour after flow cytometry. Images are shown of representative chimeras produced from each population, **b**, bright field and **c**, green filter images showing venus expression. **d**, X-gal staining of all embryos produced from each sorted population was carried out to judge the contribution potential of injected cells. **e**, Transverse cryostat sections of X-gal stained embryos indicated in **d**. **f**, Total number of embryos analysed for each population indicating the percentage which showed contribution from injected cells (presence of white areas). Scale bars in d = 600µm.

a

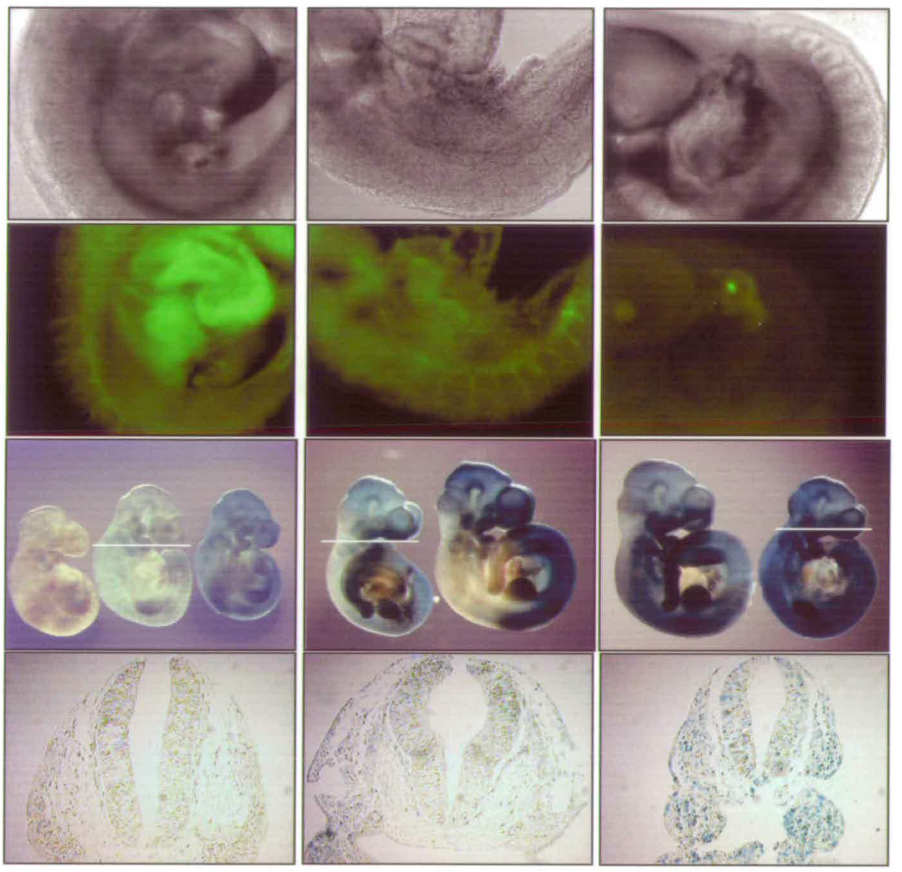


b

c

d

e



f

	Venus +, SSEA 1 +	Venus -, SSEA 1 +
n	19	24
No Contribution	74%	26%
Contribution	29%	71%

chimeric embryos showed that the Venus positive cells contributed to only 29% of injected blastocsyts (n=19) compared with 71% in the case of the Venus negative cells (n=24) (Fig. 4.16f). Interestingly, the reduced ability of the V+S+ population to generate ES cell colonies *in vitro* (see Fig. 4.15c) appeared to be reflected in the reduced likelihood of chimera contribution by this population *in vivo*.

While these results may have shed some light on the contribution potential of each fraction to definitive tissues of the embryo, the experimental design is not ideal due to the requirement for a 'negative result' to prove a positive point. Lack of β-galactosidase may either indicate positive contribution of injected cells or simply a failure of the staining procedure to faithfully penetrate all cells of the embryo equally. Thus, to address this issue, the BHIV cell line is currently being modified to constitutively express β-galactosidase such that the injected cells will stain blue in a wild-type host. The staining of β-galactosidase would be independent of Venus expression and therefore contribution from each fraction to the embryo may be more confidently judged.

While the above experiment can give an idea as to the pluripotent potential of Venus positive and negative cells, analysis of embryos at an earlier time point may shine light on the reasons for apparent reduced embryonic contribution of the latter cell type. Given that a component of murine ES cell cultures can contribute to the primitive endoderm lineage upon injection into host blastocysts (Beddington and Robertson, 1989), it might be hypothesised that Venus positive self-renewing ES cells may represent this component. To test this hypothesis it will be necessary to examine the location of injected Venus positive and negative ES cells by β-galactosidase expression in day 6.5 embryos. If Venus positive cells truly represent that component which can contribute to the primitive endoderm, they should be located in the parietal endoderm. Additionally it might be expected that the Venus negative cells would very rarely if ever at all be found there. This work is currently underway.

4.17 Discussion

In this chapter I have described the generation of the highly sensitive BHIV *Hex* reporter cell line that maintains wild type levels of *Hex* while giving a highly sensitive read out from the locus. The use of a translationally amplified IRES Venus cassette allowed the identification of cells expressing *Hex* and thus facilitating downstream analyses. It has been demonstrated previously that this particular IRES, which contains reiterated *get* sequences, acts as a translational amplifier for downstream genes, and behaves up to 60 times more efficiently than conventional ECMV IRESs (Chappell et al., 2000). Additionally, gene trap vectors containing this novel IRES sequence driving translation of Venus act as sensitively as those with *LacZ* (Tanaka et al., 2008). Quantitative PCR analysis of BHIV clones demonstrated that expression from the transgenic allele was reflecting normal *Hex* transcriptional activity, thus suggesting that observed 'background' Venus expression was due to the sensitivity of the IRES Venus reporter rather than its misexpression from the targeted locus. Extensive Southern analyses ruled out the possibility of incorrect or additional integration events of the transgene.

Generation of mouse chimeras with BHIV cells revealed strong expression of Venus in regions previously reported to express *Hex*, namely, the liver, the thyroid area, the dorsal aorta and the intersomitic vessels (Rodriguez et al., 2001; Thomas et al., 1998). That an increase of *Hex* expression could be reflected by an increase of Venus was revealed by a cell culture protocol developed in our lab to promote the *in vitro* differentiation of ES cells to an ADE fate (Morrison et al., in press). Induced expression of definitive endoderm markers with Activin during this protocol coincided with an increase of fluorescence in BHIV clones to a level similar to that obtained by a second *Hex* reporter cell line (HRS). Together these data confirmed that Venus expression in the BHIV line faithfully reflected transcriptional activity from the *Hex* locus.

The observed heterogeneous expression of Venus in undifferentiated ES cell cultures suggested the existence of multiple ES cell subpopulations potentially differing in

character from one another. The presence of SSEA1 and Oct3/4 together with morphological examination confirmed that the majority of the self-renewing Venus positive and negative cells were indeed ES cell in nature. Such heterogeneity among ES cells has been reported by other groups using reporter cell lines for the pluripotency genes *Nanog* and *Rex1* (Chambers et al., 2007; Singh et al., 2007; Toyooka et al., 2008). These studies demonstrated that undifferentiated ES cells are in fluctuation between different states of either super-pluripotency or an increased propensity to differentiate and that these states may represent an *in vitro* equivalent to the transition from ICM to epiblast/primitive ectoderm in the mouse embryo. It was interesting then that some ES cells appeared to be enriched for *Hex*, a gene whose earliest domain of expression has been found to be the primitive endoderm of 4.5 peri-implantation embryos (Chazaud et al., 2006; Mesnard et al., 2006; Thomas et al., 1998; Torres-Padilla et al., 2007). Due to the developmental origin of ES cells and their contribution potential when injected back into blastocysts, it seemed plausible that the Venus positive cells may constitute a population existing under self-renewing conditions that is predisposed to become primitive endoderm. ES cells are derived from the ICM of implanting blastocysts, at a time when the developmental choice of some ICM cells is to become primitive endoderm. Indeed, when either early ICM cells or ES cells are injected into blastocysts, they can contribute, albeit infrequently, to the primitive endoderm lineage (Beddington and Robertson, 1989; Gardner, 1985). Additionally, ES cells can readily differentiate into primitive endoderm *in vitro* (Keller, 2005; Niwa et al., 2000). So, could Venus positive cells constitute yet another subpopulation of ES cells existing alongside those previously observed with *Nanog* and *Rex1*?

To examine the relationship of the ES cell heterogeneity observed here with that of previous reports, *Nanog* expression was monitored in BHIV clones. Interestingly, *Hex* and *Nanog* expression appeared mutually exclusive of one another. This relationship appeared similar to that observed in ICM cells of E3.5 embryos between *Nanog* and another primitive marker, *Gata6*, (Chazaud et al., 2006). Interestingly, some ES cells appeared negative for either Venus or *Nanog* expression suggesting the existence of a third population. It is tempting to speculate that these cells may be

representatives of the early primitive ectoderm, previously reported to exist among ES cells (Toyooka et al., 2008). Given the early developmental choices of an ICM cell in the embryo, i.e. to become the epiblast/primitive ectoderm or the primitive endoderm, it would not be surprising that populations with such tendencies might exist together with an earlier ICM-like state enriched for *Nanog* among self-renewing ES cell cultures.

Gene expression analysis between Venus positive and negative ES cells revealed an enrichment of primitive endoderm and pluripotency genes in these populations, respectively. Pluripotency genes analysed were those that have previously been reported to be associated with *Nanog* and *Rex1* (Chambers et al., 2007; Mitsui et al., 2003; Toyooka et al., 2008). Genes of other lineages such as mesoderm, haematopoietic or neurectoderm showed no pattern of enrichment for either cell type. Of note, Venus positive non-ES cells (i.e. spontaneously differentiated cells, negative for SSEA1) showed an even greater enrichment for primitive endoderm gene expression, suggesting that these cells were further differentiated toward this particular lineage. Importantly, levels of *Oct3/4* and *Sox2* were not much different between Venus positive and negative cells of the SSEA1 positive fraction, underpinning the fact that both these subpopulations are pluripotent in nature. Interestingly, previous work in our lab examining gene expression changes following the misexpression of Hex from the ROSA26 locus, suggested that two of the pluripotency markers enriched in Venus negative cells of this study are targets of *Hex*, namely *Nr0b1* and *Klf2* (Zamparini et al., 2006).

Additionally, gene ontology analysis of the microarray data show that Venus positive ES cells are enriched for genes whose expression lies downstream of Nodal/TGF signalling as well as an enrichment of genes involved in the Wnt pathway. It is interesting that these pathways which are important for establishing A-P polarity of the pre-gastrulation embryo may have an additional role at an earlier stage in determining transcriptional noise fluctuations among ES cells that could dictate the decision between self-renewal and immediate early differentiation toward primitive endoderm (see chapter 6 for further discussion).

Previous work on *Nanog* in ES cells and mouse embryos has pointed to role for this transcription factor in repressing the emergence of the primitive endoderm (Chambers et al., 2003; Chazaud et al., 2006; Hamazaki et al., 2006; Hamazaki et al., 2004; Mitsui et al., 2003). Increased expression of *Nanog* either by forced expression of a Nanog transgene or inhibition of FGF signalling, resulted in a decrease of the Hex positive population. Conversely, inhibition of *Nanog* expression by stimulation of FGF signalling by treatment with sodium vanadate lead to an increase of Hex expressing cells concomitant with upregulation of *Gata6*. While the agent used in this particular experiment acts as a general phosphatase inhibitor and might therefore have additional effects on other pathways, it has been shown previously that such treatment specifically potentiates the MEK/ERK branch of the FGF pathway (Hamazaki et al., 2006).

As with *Nanog* and *Rex1* reporter cell lines, subpopulations reported here also showed the ability to interconvert to one another following clonal and single cell plating. Interestingly, the dynamics of this interconversion were fast, whereby Venus positive ES cells were evident just 24 hours after plating purified Venus negative cells.

Nanog negative ES cells display a reduced capacity for clonal growth and *Rex1* negative cells have a reduced ability to contribute to embryonic tissues thus suggesting differing developmental capabilities of ES cells (Chambers et al., 2007; Toyooka et al., 2008). Examination of populations in the BHIV line also showed differing functional potential whereby the Venus positive ES cells showed a reduced capacity for clonal growth in culture and a reduced likelihood to contribute to embryonic lineages. The primitive endoderm character of these ES cells may mean that they are not at an ideal neutral ground state to provide optimal clonal growth or be poised to differentiate into other lineages. It is important to note however that Venus positive ES cells do have these capabilities, just at reduced level. Thus, some of these cells may have the capability to return from their transient primitive endoderm direction back to a state similar to that of Venus negative cells. It will be

interesting to examine whether Venus positive ES cells are fated to the primitive endoderm lineage in early embryos. Work is currently underway to address this issue.

In the next chapter I discuss the consequences of overexpressing Hex in ES and other cell types.

Chapter 5

Overexpression of Hex in ES cells

As *Hex* is both an early marker of A-P axis formation and also a potential marker of the hemangioblast a number of studies have probed its function by gain of function experiments. Previous reports on the overexpression of *Hex* in various systems have shed some insights into a number of potential activities. Most molecular data suggests that Hex can act as a transcriptional repressor (Brickman et al., 2000; Pellizzari et al., 2000; Tanaka et al., 1999; Zamparini et al., 2006) and a number of studies have indicated that *Hex* maybe an inhibitor of differentiation by the repression of lineage specific promoters (Brickman et al., 2000; Zamparini et al., 2006). Thus, overexpression of *Hex* in *Xenopus* can affect the developmental potential of different germ layers and cell lineages. Inhibition of dorsal mesoderm formation has been demonstrated by the suppression of organiser genes such as *Goosecoid* and *Chordin* (Brickman et al., 2000; Zamparini et al., 2006). *In vitro* analysis in ES cell derived differentiation cultures has also indicated a role as a suppressor of mesoderm derivatives in EB cultures (Kubo et al., 2005). Indeed, studies in these and other systems have shown that *Hex* may play a similar role later in haematopoietic differentiation. *Hex* is expressed in and is an important regulator of mesoderm derived haematopoietic and endothelial lineages. While overexpression of *Hex* in *Xenopus* and zebrafish suggests a positive role for vascular-endothelial lineage development (Liao et al., 2000; Newman et al., 1997), studies in various endothelial and haematopoietic cell lines were indicative of a negative effect on the proliferation or maturation of these lineages (George et al., 2003; Jayaraman et al., 2000; Mack et al., 2002; Nakagawa et al., 2003; Ying et al., 2002). This inhibitory effect on proliferation was also observed in cell types representative of vascular-endothelium and early haematopoietic progenitors in suspension ES cell derived suspension cultures overexpressing *Hex*, while neurectodermal differentiation was unaffected (Kubo et al., 2005). As with the above examples, this maybe due to the repression of either cell or cell cycle specific promoters. However, at least with respect to aspects of cell cycle regulation, the effect of *Hex* may not be transcriptional as it can inhibit proliferation in promonocytic cell lines via an interaction with the eIF4e component of the ribosome (Topisirovic et al., 2003).

Forced expression of *Hex* in *Xenopus* has also been shown to play a positive role in promoting anterior endoderm (Brickman et al., 2000; Jones et al., 1999; Zamparini et al., 2006; Zorn et al., 1999). While these studies have suggested a mechanism for promoting anterior character to endoderm through direct interaction of Nodal and Wnt signalling components, it is unknown whether these would be direct targets of *Hex* in endoderm of the mouse embryo or in endoderm derived from differentiation of murine ES cells.

Consistent with a role in promoting anterior endoderm specification, loss of function analysis in *Xenopus* by injection of morpholino sequences specific to *XHex* lead to anterior truncations (Jones et al., 1999) and a null mutant in mouse has shown that *Hex* is required in the anterior definitive endoderm for correct anterior development of the forebrain and endoderm derived organs such as the thyroid and liver (Martinez Barbera et al., 2000). Similarly as mentioned above in zebrafish, ectopic expression of *Hex* can lead to the induction endothelial precursors, and loss of function models in zebrafish and mouse display defects in endothelial specification. In the endoderm the regulation of *Hex* regulates lineage specification through transcriptional regulation of signalling pathway components. How much of the ability of *Hex* to promote lineage specification is accounted for by this and how extensive Hex dependent transcriptional networks relay remain to be shown. Moreover, whether *Hex* can influence gene expression in these contexts is unknown.

Thus, a number of lines of evidence suggest that *Hex* can both act to promote and inhibit lineage specification. The evidence presented in chapter 4 of this thesis suggests that *Hex* is a very early marker of endoderm specification. In this chapter I use the reagents described in chapter 3 to address whether Hex misexpression has any impact on ES cell culture under self renewing conditions. While there is no evidence for this activity in *Hex* null ES cells (Guo et al., 2003; Kubo et al., 2005), the heterogeneous expression patterns described in Chapter 4 are very provocative. Additionally, the overexpression in ES cells of other transcription factors important in primitive endoderm formation lead to a change in morphology to resemble this

lineage (Fujikura et al., 2002). In this chapter I show that I could not observe any such effect, but rather that misexpression of *Hex* at high levels induces transcription dependent apoptosis in ES cell culture and that low level expression has no apparent phenotype. These studies have also allowed me to make a number of interesting observations about ES cell selection strategies and the levels of gene expression they can be used to achieve.

5.1 Generation of stable lines that stably over express Hex

I attempted to use two different selection strategies to generate ES cells that constitutively express *Hex*. In all experiments the expression levels obtained by these strategies were determined with a control vector which expressed only GFP (Chapter 3, Figures 3.3b and 3.4b).

5.1a Electroporation of hygromycin constructs into R26birA ES cells

To generate stable clones, the three hygromycin constructs (see Fig. 3.3) were linearised with *Fsp*I and electroporated in R26birA cells. Following selection for 12 days in 200μg/ml of hygromycin, clones generated from each construct were counted and visualised for GFP fluorescence. Table 5.1 shows the average number of clones obtained from two independent electroporation experiments and the total number subsequently expanded and analysed for GFP fluorescence by FACs analysis. In each of the two electroporations equal numbers of clones were obtained from the three constructs, however, none could be seen to be GFP positive using fluorescence microscopy (data not shown). Twenty clones generated from each construct were further analysed by FACs for the presence of GFP but again none of these were positive (Fig. 5.1a). The lack of detectable fluorescence by microscopy or FACs may be explained by the stringency of the hygromycin selection. If low expression of hygromycin phosphotransferase is required for antibiotic resistance, then its use to

	Average No. Clones Obtained	% Fluorescent By microscopy	Total No. Analysed by FACS (numbers positive for GFP)
pBGiH	190	0	20 (0)
pBHGiH	200	0	20 (0)
pBHiGiH	170	0	20 (0)

Table 5.1a Summary of stable clones produced from 100 µg of pBGiH, pBHGiH or pBHiGiH in R26birA cells following selection in hygromycin (200µg/ml). The numbers relate to two independent electroporation experiments.

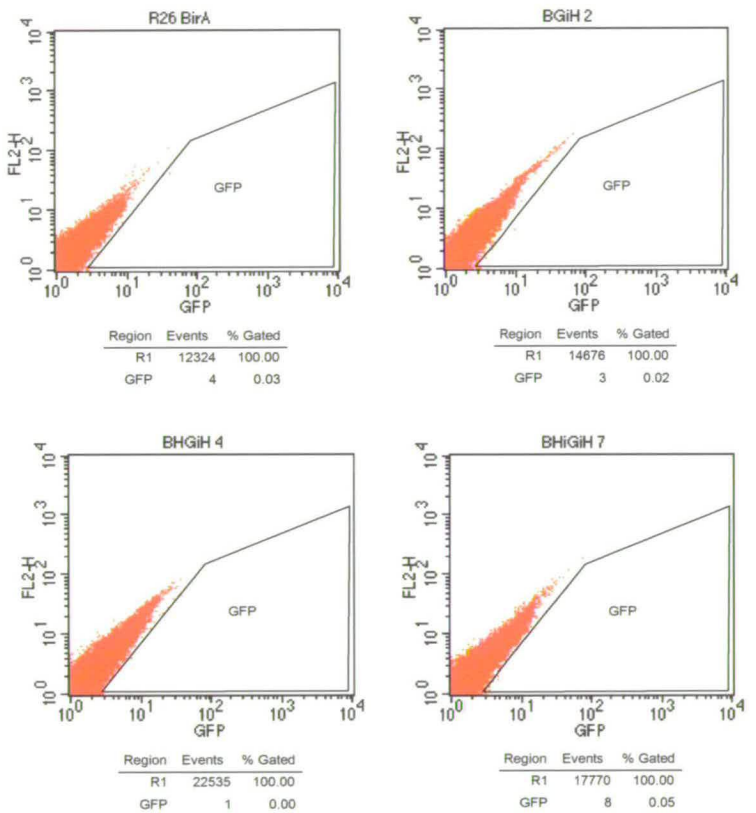


Fig. 5.1a Fluorescence analysis of hygromycin resistant stable clones.

Clones produced from electroporation of pBGiH, pBHGiH and pBHiGiH in R26birA cells were analysed by FACs for GFP expression. Representative data of clones produced from each construct are shown including unelectroporated R26birA cells.

select for high expression would not be efficient. Thus, obtaining high level expressers might require extensive screening of large numbers of colonies.

Previous work in our lab using a similar system of CAG driven *Hex* under hygro selection has shown that transcript levels were similar to those obtained from expression of Hex from the ROSA26 locus (Zamparini et al., 2006). Each of these previously generated lines had no apparent phenotype. Expression analysis of Hex from the ROSA26 locus showed no upregulation of primitive endoderm genes. It was decided therfore to attempt a higher level of expression using a different selection system.

To overcome this problem, the use of a more stringent selection system was considered to favour higher and therefore detectable levels of GFP which should reflect higher levels of Hex. Selection with puromycin is known to have advantages over other drugs such as G418 and hygromycin: it acts quickly to kill non-transfected cells within 48 hours reducing the problem of colony overgrowth during selection, it is active at low concentrations (1-10μg) and it is inexpensive (de la Luna and Ortin, 1992). Thus, subcloning of the Bio-tagged variants of the hygromycin constructs into a similar vector harbouring puromycin resistance was undertaken in an attempt to derive stable clones with higher expression (Fig. 3.4).

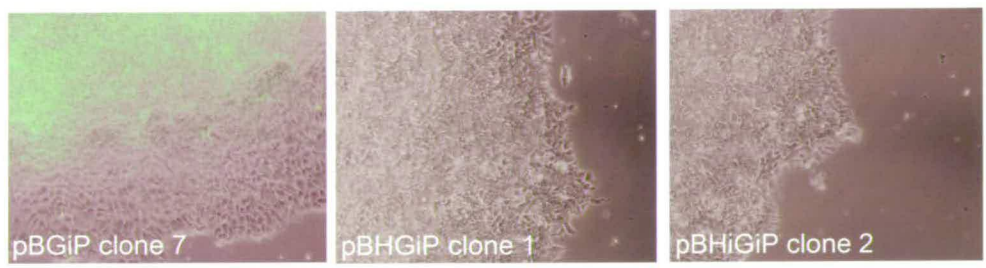
5.1b Electroporation of puromycin constructs into R26birA ES cells

The three puromycin constructs (see Fig. 3.4) were linearised with *Sca*I and electroporated into R26birA cells. Following 12 days in selection with 2μg/ml of puromycin, clones generated from each construct were counted and visualised for GFP fluorescence. Table 5.2 shows the average number of clones obtained from two independent electroporation experiments and the total number subsequently expanded and analysed for GFP fluorescence by FACs analysis. While each electroporation generated many clones from the control construct pBGiP, very few were obtained from the two Hex containing constructs, pBHGiP and pBHiGiP. These

	Average No. Clones Obtained	% Fluorescent By microscopy	Total No. Analysed by FACS (numbers positive for GFP)
pBGiP	220	50	7 (6)
pBHGiP	4.5	0	5 (0)
pBHiGiP	2	0	2 (0)

Table 5.1b Summary of stable clones produced from 100 µg of pBGiP, pBHGiP or pBHiGiP in R26birA cells following selection in puromycin (2µg/ml). The numbers relate to two independent electroporation experiments.

(i)



(ii)

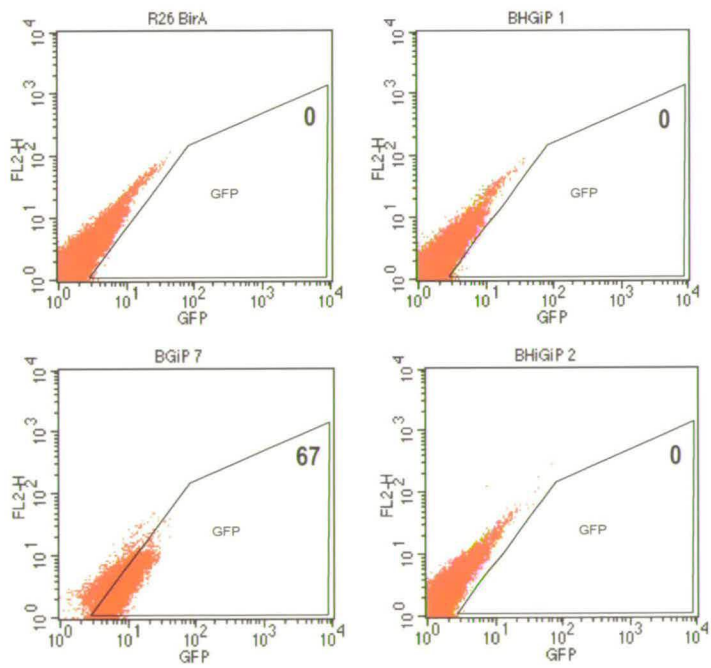


Fig. 5.1b Fluorescence analysis of puromycin resistant stable clones.

(i), Clones produced from electroporation of pBGiP, pBHGiP and pBHiGiP in R26birA cells were analysed by fluorescence microscopy for GFP expression. Representative photos of clones produced from each construct are shown. **(ii)**, FACs analysis of the same clones together with unelectroporated R26birA cells. Percentages positive for GFP are shown.

results differ from those obtained with the Hygro selection and suggested that if Puro does indeed drive high levels of *Hex* expression, this might in someway be detrimental to ES cell colony growth.

To compare the levels of expression in the puro selected clones to those obtained with hygro, GFP was examined by both flow cytometry and fluorescence microscopy. Of the clones transfected with the control vector, pBGiP, at least fifty percent of the colonies obtained were fluorescent whereas none of those transfected with a control vector under hygro selection were found to GFP positive (Compare Table 5.1a with Table 5.1b and Fig 5.1b(i)). However, of the clones derived from transfection with either pBHGIP or pBHiGiH, none were seen to fluoresce. Seven of the Hex negative control clones along with five and two clones generated from pBHGiP and BHiGiP, respectively, were further expanded and analysed by FACs. Six of the seven control clones were positive for GFP to differing degrees, however, none of the Hex containing clones were positive (Fig. 5.1b(ii)).

5.1c Western Analysis of hygromycin and puromycin resistant clones

In order to further compare differences between hygro and puro selected clones, western analysis was undertaken using an anti-GFP antibody on a selection of these lines. Two pBHGiH and two pBHGiP clones together with three of the Hex negative pBGiP clones were chosen to reflect the range of GFP expression as seen by microscopy and FACs analysis. Figure 5.3 shows the results of this western blot. A band of the expected size for BioGFP was seen in each of the three pBGiP lanes. The intensities of these bands reflected the percentages of GFP in these clones as detected by FACs. No band could be detected in the two pBHGiP lanes, however, a very faint band of the expected (64 kD) size for BioHexGFP was seen in both pBHGiH lanes.

Taken together the data obtained from the hygromycin and puromycin resistant clones suggest that only low levels of Hex expression are tolerated by ES cells. The absence of detectable GFP fluorescence in clones derived from all four BioHex

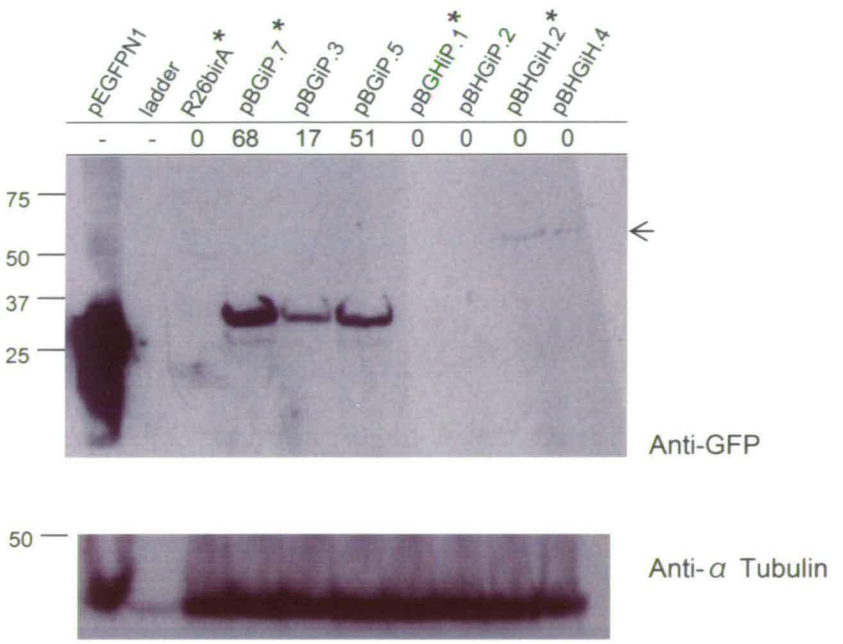


Fig. 5.1c Western Analysis of hygromycin and puromycin resistant clones.

Two pBHGiH clones, two pBHGiP clones and three Hex negative pBGiP clones were lysed and equivalent amounts for each were subjected to SDS-PAGE. Lysates from Un-electroporated R26birA cells and HEK293 cells transiently transfected with pEGFP-N1 were loaded as negative and positive controls for GFP, respectively. Western analysis was performed with an anti-GFP followed by an anti-α Tubulin antibody . Numbers under each sample indicate the GFP percentage for each clone obtained by FACs analysis. Arrow indicates the expected size for BioHexGFP fusion protein. Arrow indicates the position of BioHexGFP protein.

containing constructs appears to reflect this. To examine the possibility that this is an ES cell specific effect, we attempted misexpression coupled to puro selection in other cell types.

5.2 Electroporation of puromycin constructs in other cell types

To establish stable clones expressing BioHex, different cell types were chosen to represent different tissues during embryonic developmental stages downstream of that represented by ES cells, namely the ICM. These were Xen cells and P19 embryonal carcinoma (EC) cells which are derived from the extraembryonic endoderm, and the epiblast of the mouse embryo, respectively (Kunath et al., 2005; Riego et al., 1995; van der Heyden and Defize, 2003). Additionally, the HepG2 cell line derived from a human hepatocarcinoma was chosen due to high expression of Hex in the developing liver primordium (Knowles et al., 1980; Thomas et al., 1998).

Due to the higher efficiency observed with puromycin selection, the three puromycin constructs were electroporated into these three cell types. Table 5.2 summarizes the total numbers of colonies produced in each cell line from each construct following selection in puromycin. The numbers of clones obtained from these electroporations are extremely similar to what was obtained in ES cells. Again, the overexpression of Hex appears to select against the growth in these cell types also. None of the few clones derived from the two BioHex constructs could be seen to fluoresce, however, clones produced in these cell-types from the Hex-negative control construct, pBGiP, could be seen to fluoresce (Fig. 5.2 and data not shown).

Taken together these data suggest that expression of Hex in ES, Xen, P19 EC or HepG2 cell lines results in a block to cell growth.

	IM8A1, Xen (% fluorescent)	P19, EC (% fluorescent)	HepG2 (% fluorescent)
pBGiP	89 (75)	140 (50)	210 (50)
pBHGiP	2 (0)	1 (0)	1 (0)
pBHiGiP	2 (0)	4 (0)	0 (0)

Table 5.2 Production of Stable clones expressing pBGiP, pBHGiP and pBHiGiP in IM8A1 Xen, P19 EC and HepG2 cells. Numbers reflect the total number of clones produced for each construct after selection in puromycin (2µg/ml). Numbers in parentheses indicate the percentage of clones seen to fluoresce by microscopy.

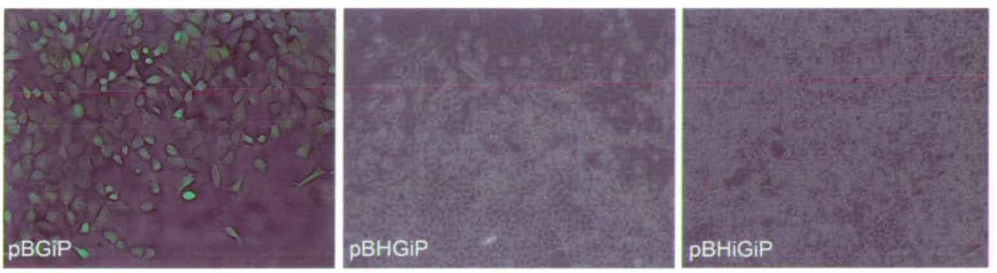


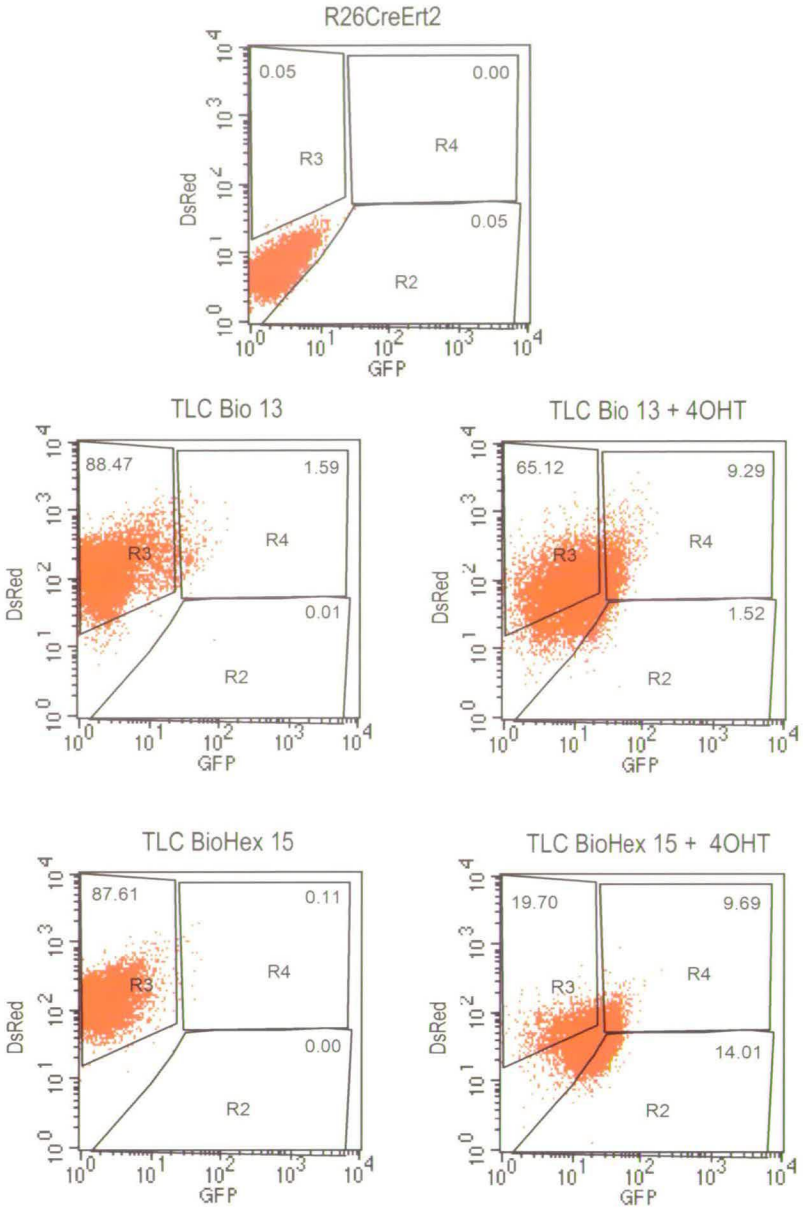
Fig. 5.2 Expression of puromycin resistant stable clones in IM8A1 Xen cells.

Clones produced from electroporation of pBGiP, pBHGiP and pBHiGiP in IM8A1 Xen cells were analysed by fluorescence microscopy. Representative clones of pBGiP, pBHGiP and BHiGiP are shown.

5.3 Inducible expression of BioHex

A number of transcription factors can be expressed in ES cells using an IRES puro selection cassette (Chambers et al., 2003; Mountford et al., 1994). As a tet inducible system for *Hex* expression in ES cells had been previously established (Kubo et al., 2005), an inducible system was chosen to address the failure to obtain high levels. This system would also be useful for downstream biochemical applications. I chose to use an inducible system that coupled a fluorescent reporter to Puro selection, so that I could derive cells that would be predicted to be high expressers. Induction would then be accomplished by a Cre-mediated recombination in the construct to bring the *Hex* cDNA under the promoter that had previously been driving Puro. Stable clones were derived using the pTLC BioHex construct (see Chapter 3.6) in an ES cell line constitutively expressing Cre recombinase as a fusion with a modified ligand binding domain of the estrogen receptor (CreERt2), from the ROSA26 locus (RC cells). Upon the addition of 4-hydroxytamoxifen (4OHT), CreERt2 translocates into the nucleus and can then excise the upstream selection cassette that is flanked by loxP site, thus allowing expression of the downstream BioHex IRES GFP cDNA.

pTLC BioHex or the control construct, pTLC Bio were electroporated with into RC cells together with a CAG driven BirA ligase expression construct harbouring a blastocidin resistance gene (pBirA). Colonies were derived following selection in 2μg/ml of puromycin and 5μg/ml of blastocidin for 10 days. Table 5.3 summarises the total number of clones derived from pTLC Bio and pTLC BioHex, the percentages of colonies seen to display DsRed fluorescence by microscopy, and the subsequent numbers analysed by FACs. Three clones derived from each construct, which displayed strong DsRed fluorescence were chosen for 4OHT treatment. Figure 5.3a shows a fluorescence change from red to green following 4OHT treatment of TLC Bio (control) and TLC BioHex clones. Bright field images show an unhealthy, rounded-up phenotype only in the induced BioHex. The absence of this phenotype in the 4OHT treated control suggests a BioHex specific effect. FACs analysis of the same samples confirmed the switch from DsRed to GFP expression (Fig. 5.3b).



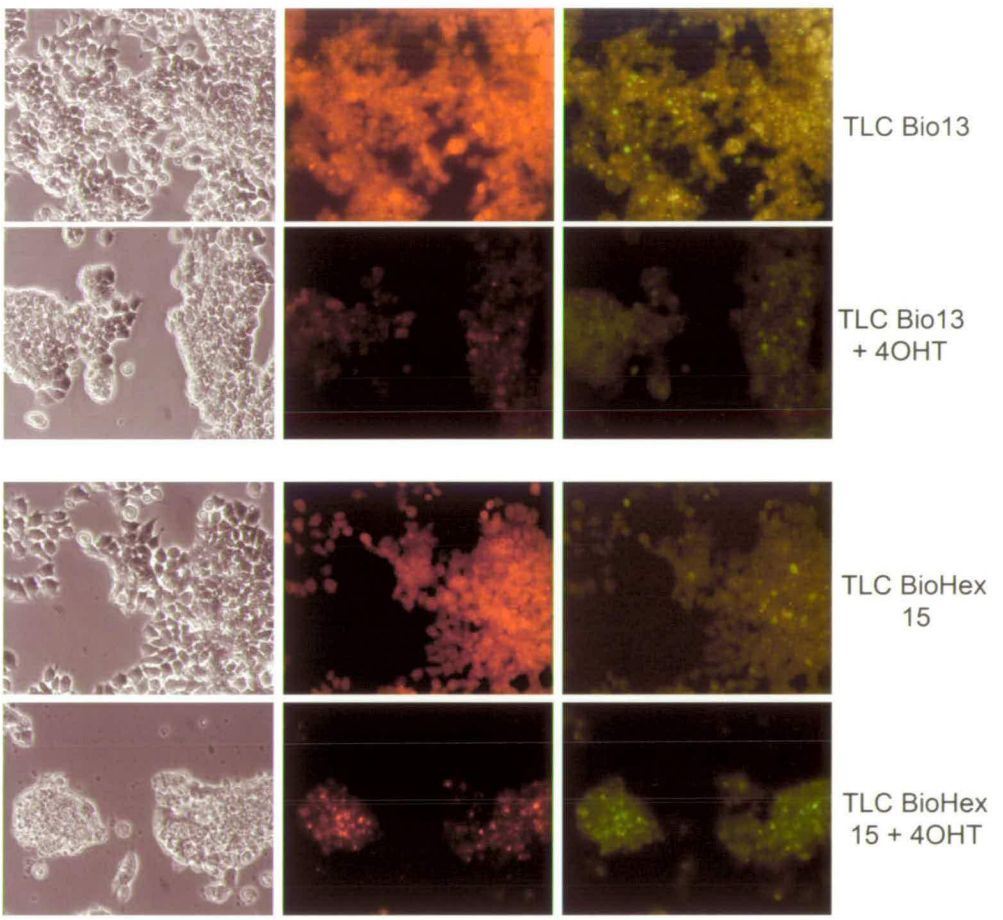
b

Fig. 5.3 Fluorescence analysis of TLC clones incubated with 4OHT.

TLC clones were incubated for 12 hours alone or with 500nM 4OHT. Following the removal of 4OHT clones were cultured for a further 36 hours before analysis by fluorescence microscopy and FACs. **a**, Red, Green and bright field images shown for representative TLC Bio and TLC BioHex clones cultured alone or with 4OHT. **b**, FACs analysis of the same clones. The plot for unelectroporated R26CreERT2 cells is shown.

	Average No. Clones Obtained	% Fluorescent By microscopy	Total No. Analysed by FACS (numbers positive for DsRed)
pTLC Bio	80	75	6 (5)
pTLC BioHex	90	75	6 (6)

Table 5.3 Summary of stable clones produced from pTLC Bio and pTLC BioHex in R26CreErt2 cells following selection in puromycin (2µg/ml).



a

5.4 Expression of BioHex mRNA and protein following 4OHT induction.

The presence of GFP in 4OHT treated clones suggests the successful induction of the BioHex IRES-GFP transcript due to the removal of the upstream DsRed IRES-Puro selection cassette. To check the integrity at the 5 prime end of BioHex IRES-GFP mRNA, rtPCR analysis was performed on isolated RNA from 4OHT treated clones over a time course of three days (Fig. 5.4a). Bands of increasing intensity correlated with incubation time following 4OHT treatment as seen in a representative pTLC BioHex clone (Fig. 5.4b).

Lysates were prepared from untreated and 4OHT treated TLC Bio and TLC BioHex clones and subjected to SDS-PAGE. Western analysis using streptavidin-HRP showed the presence of a band of the predicted size for BioHex protein in the induced TLC BioHex sample (Fig. 5.4c). This result also indicated the successful biotinylation of the BioHex protein from the stable expression of the BirA ligase enzyme.

Taken together these data suggest that the observed phenotype following 4OHT treatment of TLC BioHex clones is due to induction of BioHex protein.

5.5 Induced expression of BioHex leads to cell death and caspase-3 truncation.

The unhealthy phenotype associated with induced overexpression of BioHex protein reflected the possibility of cell toxicity due to this transcription factor. There is also precedent in the literature that Hex expression can induce apoptosis and this would seem a reasonable explanation for the phenotypes observed (Topisirovic et al., 2003). Indeed, when cultured up to 96 hours after 4OHT treatment, few if any surviving cells could be seen. To test whether ectopic expression of Hex induced apoptosis I examined Caspase3 expression following Hex induction and compared this to a

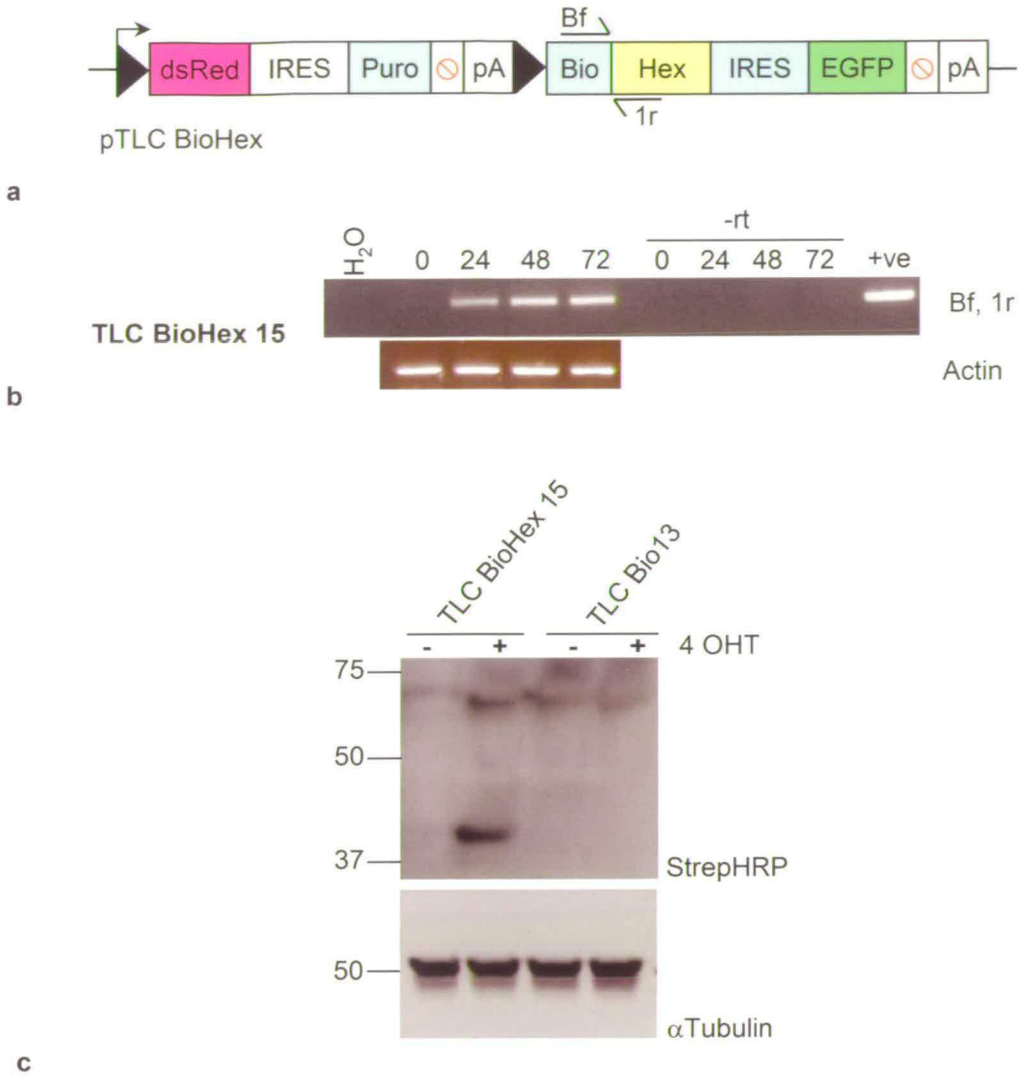


Fig. 5.4 Inducible expression of BioHex mRNA and protein in TLC BioHex clones.

TLC BioHex clones were incubated alone or for various times with 500nM 4OHT. RNA was extracted at the end of each time point followed by cDNA synthesis. **a**, Schematic representation of the primers used for the detection of BioHex mRNA. **b**, rtPCR analysis was performed on cDNA synthesized from RNA extracted at the indicated timepoints (shown in hours) from a representative TLC BioHex clone. Primers Bf and 1r were used for the detection of BioHex mRNA. Actin primers were used as a control for the presence of cDNA. **c**, Western analysis on lysate from the 48 hour time point following 4OHT treatment on the same pTLC BioHex clone and a conntrol pTLC Bio clone. Lysates were subjected to SDS-PAGE and western analysis was performed using Strepavidin-HRP or an anti-αtubulin antibody. The predicted size of the BioHex protein is 42kD.

known inducer of apopotosis, namely cyclohexamide (Borner et al., 1995). TLC BioHex and control TLC Bio clones were treated with either 4OHT or 1 mM cycloheximide (CHX). Figure 5.5a shows the morphology of TLC BioHex and TLC Bio clones cultured alone or treated with either 4OHT or CHX. While incubation of control lines with 4OHT had no effect, similar treatment of *Hex* inducible lines lead to a morphology resembling that induced by CHX suggesting that they are indeed undergoing apoptosis. Importantly, treatment of cultures with either CHX or 4-OHT leads to the induction of a faster migrating fragment of Caspase-3, indicative of apoptosis (Fernandes-Alnemri et al., 1994). Figure 5.5b shows a western blot of clones incubated alone, with 4OHT or with CHX and probed with an Anti-Caspase3 antibody. In both the CHX treated and 4-OHT Hex inducible cultures the truncated 19kD and 17kD caspase-3 species are clearly present and are absent in all uninduced, or samples not containing the inducible transgene. This finding suggests that ectopic expression of *Hex* induces apoptosis in ES cell culture.

In addition to its role as a transcription factor, Hex inhibits proliferation and transformation by interfering with eIf4E-dependent mRNA transport of cyclin D1 (Topisirovic et al., 2003). Thus one explanation for Hex's ability to induce this apoptotic phenotype may have to do with acute interference of the cell cycle which could trigger a cascade leading to apoptosis. Alternatively, high levels of Hex protein may lead to exaggerated alteration of gene expression levels of its targets or off-targets to an intolerable extent.

5.6 Generation and functional analysis of pTLC BioHex mutants.

To examine whether Hex induced apoptosis was due to either cell cycle interference through interaction with eIF4e or as a consequence of aberrant transcription regulation through DNA binding, mutations in the Hex sequence of the pTLC BioHex construct were generated by site-directed mutagenesis (Fig. 5.6a). Substitution of residues 23 and 24 of Hex from leucine-leucine to alanine-alanine has been shown to negate binding to eIF4e (Topisirovic et al., 2003). Glutamine and

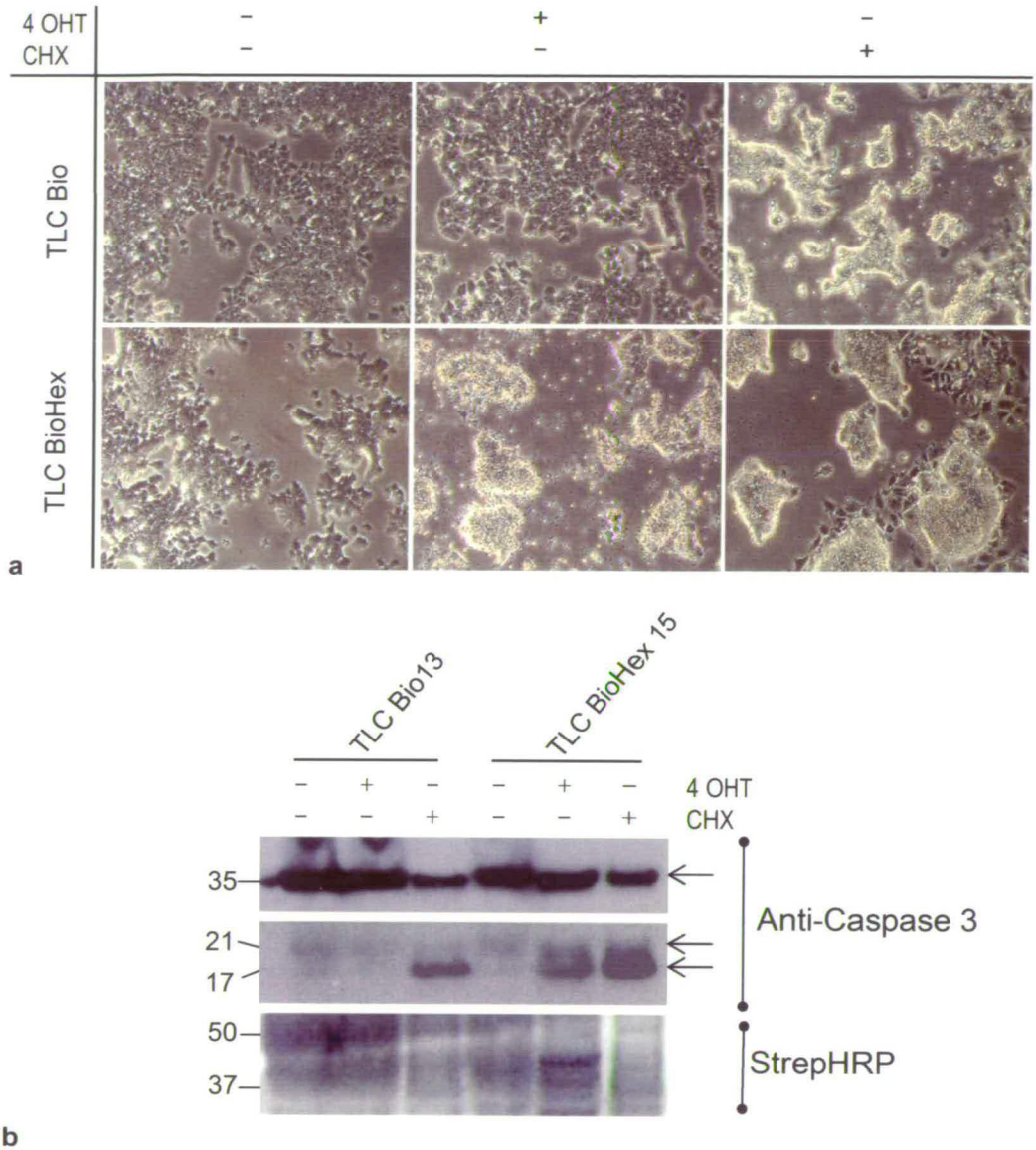
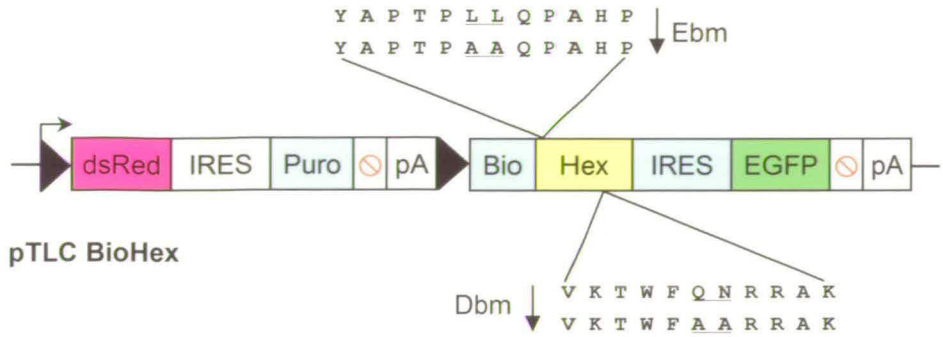
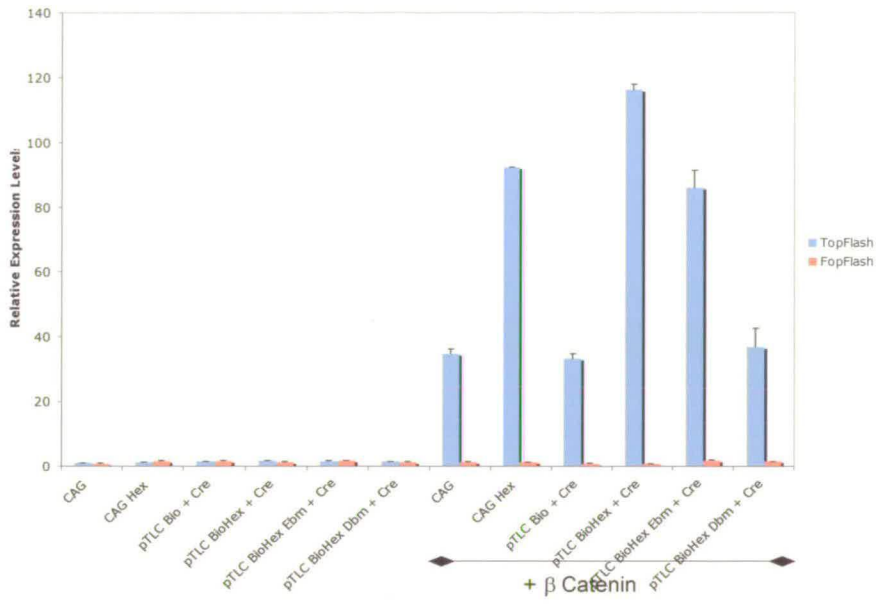


Fig. 5.5 Induced expression of BioHex leads to cell death and caspase-3 truncation.

TLC Bio and TLC BioHex clones were incubated alone, with 500nM 4OHT or 1mM CHX for 12 hours. Following the removal of 4OHT or CHX, clones were cultured for a further 36 hours before analysis. A, Bright field images of representative TLC Bio and TLC BioHex clones. **b**, Lysates prepared from treated and untreated clones were subjected to SDS-PAGE and western analysis was performed using an anti-caspase 3 antibody or streptavidin-HRP. Arrows indicate positions of full length (35 kD) and truncated (21 + 17 kD) caspase 3 proteins. Analysis with steptavidin-HRP reveals the presence of the BioHex protein in the TLC BioHex clone incubated with 4OHT.



a



b

Fig. 5.6 Generation and function of BioHex mutants in pTLC BioHex.

Site directed mutagenesis was performed on the construct pTLC BioHex in order to generate an eIF4e binding mutant (Ebm) or a DNA binding mutant (Dbm). **a**, Schematic representation of the amino acid changes made to generate either pTLC BioHex Ebm or pTLC BioHex Dbm. **b**, The Topflash assay was performed with the indicated constructs (100ng of each) in HEK293T cells. In the presence of stabilised β-catenin, pTLC BioHex + Cre, pTLC BioHex Ebm + Cre and wild-type Hex (CAG-Hex) comparably increase the activity of TOPflash beyond that obtained from an empty vector with a CAG promoter (CAG). The control vector pTLC Bio + Cre and pTLC BioHex Dbm + Cre do not amplify this signal.

asparagine residues at positions 50 and 51 of the Hex homeodomain and other homeodomain transcription factors are responsible for binding to its core recognition sequence (Crompton et al., 1992; Gehring et al., 1994). Thus, substitution of these two hydrophilic residues to hydrophobic alanines would be predicted to interfere with DNA binding.

Before establishing stable clones with the mutant forms of the pTLC BioHex construct, they were subjected to the TopFlash assay to test for Hex function (Fig. 5.6b). The eIf4e binding mutant (Ebm) lead to a similar amplification of β-catenin induced transcription as unmutated BioHex. However, this function of Hex was not seen with the DNA binding mutant (Dbm) suggesting the successful interruption of transcription regulation through DNA binding.

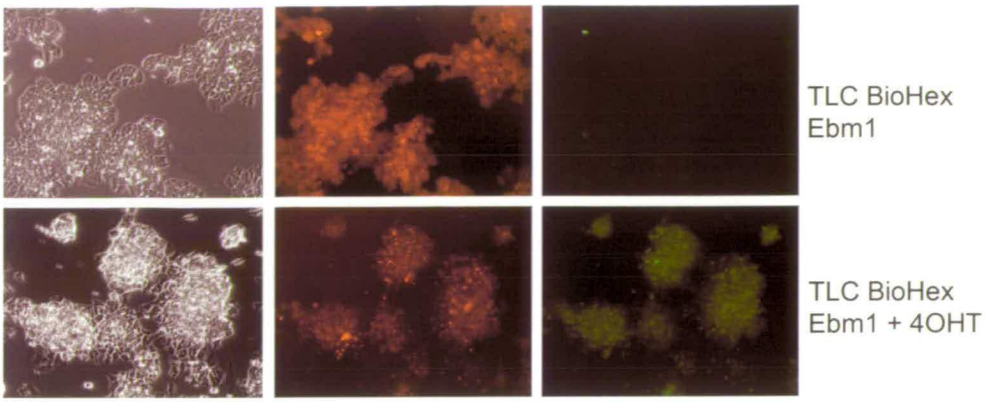
5.7 Hex induced apoptosis is due to DNA binding.

The Ebm or Dbm Hex mutant constructs were electroporated into RC cells together with pBirA and selection was carried as described above. Equal numbers of clones were produced for each construct with the majority of colonies displaying strong red fluorescence. Ebm and Dbm clones were expanded and exposed to 4OHT. After 48 hours of culturing, Ebm clones displayed a shift from red to green fluorescence and the morphology resembled the cell death phenotype seen with wild-type pTLC BioHex clones (Fig. 5.7a). Dbm clones on the other hand did not show signs of cell death concomitant with the fluorescence change (Fig. 5.7b).

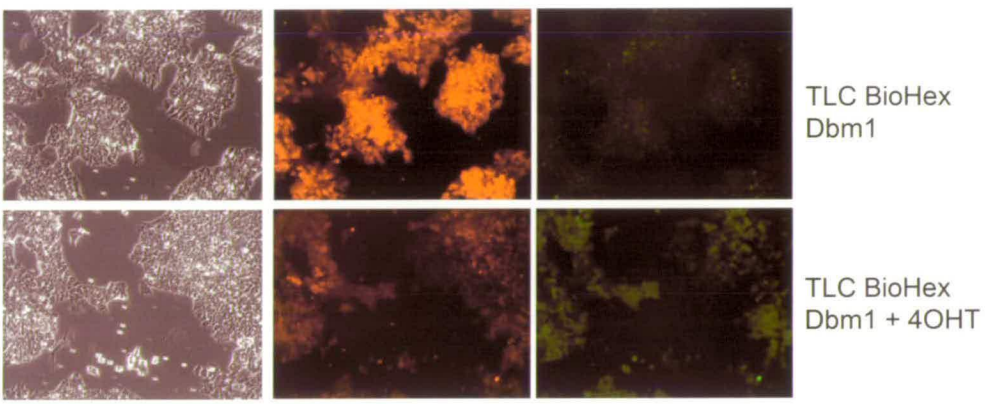
Western analysis was performed on lysates from 4OHT treated wild type TLC BioHex and mutant clones to confirm the expression of BioHex proteins. In each case a band corresponding to the expected size for a full length BioHex protein was observed (Fig. 5.7c). Use of an anti-Caspase3 antibody revealed the presence of truncated Caspase3 proteins in the 4OHT treated TLC BioHex and TLC BioHexEbm samples. This marker for apoptosis was not observed in the 4OHT treated Dbm sample which is in agreement with the lack of cell death phenotype from this mutant.

Fig. 5.7 Apoptosis due to induced BioHex is a result of DNA binding.

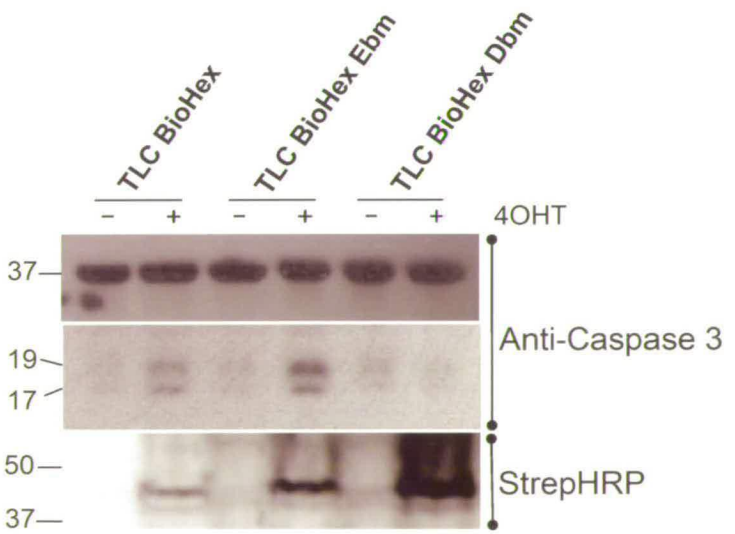
TLC BioHex mutant clones were incubated for 12 hours alone or with 500nM 4OHT. Following the removal of 4OHT clones were cultured for a further 36 hours before analysis by fluorescence microscopy. **a**, Representative clone derived from pTLC BioHex Ebm showing a switch from red to green fluorescence and cell death upon addition of 4OHT. **b**, Representative clone derived from pTLC BioHex Dbm showing a switch from red to green fluorescence upon induction by 4OHT, however, cell death is not observed. **c**,TLC BioHex and TLC BioHex mutant clones were incubated for 12 hours alone or with 500nM 4OHT. Following the removal of 4OHT clones were cultured for a further 36 hours before analysis by western blotting. The induction of BioHex and BioHex mutant proteins by 4OHT is shown by analysis with streptavidin HRP. A concomitant appearance of truncated caspase 3 proteins (21 and 17kD) is seen with induction of BioHex and BioHex Ebm but not with BioHex Dbm. Truncated and full length (35kD) caspase 3 proteins were detected with an anti-caspase 3 antibody.


TLC BioHex
Ebm1
TLC BioHex
Ebm1 + 4OHT

a


TLC BioHex
Dbm1
TLC BioHex
Dbm1 + 4OHT

b


TLC BioHex
TLC BioHex Ebm
TLC BioHex Dbm
– + – + – +
4OHT
37
19
17
Anti-Caspase 3
50
37
StrepHRP

c

5.8 Discussion

This chapter has described attempts to overexpress BioHex proteins in ES and other cell types. While small changes of Hex expression appear to mark alterations in early lineage choices as seen in chapter 4, these levels of Hex expression produce no overt phenotype in ES cells. However, when *Hex* is expressed at high levels it is not possible to establish cell lines and this appears to be due to the onset of apoptosis upon its induction. I have also shown that this observed cell death is due to the ability of *Hex* to bind DNA. This would suggest that the observed apoptosis is due either to misregulation of transcription by *Hex* or by steric hindrance due to the over-occupation at promoter sequences utilised by other transcription factors.

It has been shown that overexpression of Hex in a promonocytic cell line leads to G1 arrest by interfering with the transport of cyclin D mRNA by eIf4e (Topisirovic et al., 2003). Although such interference with the cell cycle could lead to the apoptosis seen with BioHex in this study, overexpression of a mutant of Hex unable to bind to eIF4e but with normal transcriptional activity did not present this phenotype.

While various studies have shown that *Hex* function can be reliant on the N-terminal region for repressive activity (Guiral et al., 2001; Tanaka et al., 1999) and the C-terminal domain for activation of transcription (Kasamatsu et al., 2004), the affects of *Hex* overexpression *Xenopus*, ES cell cultures, haematopoietic cells and dermal fibroblasts are disrupted by mutation of the homeodomain (Brickman et al., 2000; Guiral et al., 2001; Obinata et al., 2002).

Screening for targets of *Hex* in an ES cells system lead to the identification of *Nodal* which is known to be important for the maintenance of human and mouse ES cell cultures (Brickman et al., 2000; James et al., 2005; Ogawa et al., 2007). Therefore, high levels of Hex could lead to an intolerable degree of *Nodal* repression resulting in cell death. However, it was not shown here whether *Hex* induced apoptosis was due to the misregulation of its targets or through promiscuous binding to non-specific sequences as a consequence of the high levels. Due to the similarity of recognition

sequences between different homoedomain proteins, at increased levels, *Hex* may bind to and influence transcription at sites occupied by other HOX transcription factors, particularly of the Antennopedia class, to which *Hex* is closely related (Crompton et al., 1992; Gehring et al., 1994). It has been previously reported that *Hex* can bind to and alter transcription of simple promoter sequences containing only the TATA box sequence *in vitro* (Guiral et al., 2001) and thus, at high levels, Hex may compete with the TATA box-binding protein to repress transcription, an effect seen with *Engrailed* (Ohkuma et al., 1990). Thus, at high levels, Hex may occupy and therefore influence promoter sequences globally in addition to altering expression of its natural targets.

Another possible mechanism of induced apoptosis by Hex overexpression maybe through transcriptional squelching. Hex can impart transcriptional repression via its proline rich domain, at least in part, by interaction with co-repressors like Tle1 (Swingler et al., 2004). Inappropriate sequestering of such binding partners due to abnormally high concentrations of Hex may thus prove deleterious to a cell's survival. However, this would have to depend on DNA binding or an intact recognition helix and as I have shown, a point mutation in this region abolishes the effect.

In conclusion this work has demonstrated that ES cells expressing high levels of Hex undergo apoptosis. The inducible expression of *Hex* may prove useful for future biochemical analysis of BioHex protein following differentiation procedures which may generate cell-types allowing such high levels. Additionally, the ES cell lines with lower and therefore tolerable overexpression levels of *Hex* as seen with those under hygro selection could provide further insights into the work seen in Chapter 4. Therefore it would be interesting to see if the level of *Nanog* or the proportion of cells expressing *Nanog* is lowered in the line overexpressing Hex under hygro selection.

Chapter 6

Final Discussion

The work I have described in this thesis has uncovered some key findings into the significance of expression of the homeodomain transcription factor *Hex*. The use of a sensitive reporter line has uncovered a novel domain for its expression in ES cell cultures and has raised important questions into the heterogeneous nature of apparently uncommitted ES cells. While ES cells remain pluripotent in appropriate culturing conditions, they also harbour the ability to differentiate into many lineages *in vitro* and *in vivo*. Heterogeneity described here complements the previous findings by others that multiple subpopulations exist as Oct3/4 positive ES cells that differ in their gene expression patterns and functional abilities (Chambers et al., 2007; Singh et al., 2007; Toyooka et al., 2008). Such fluctuations between the ranges of multilineage potential states have recently been reported in the haematopoietic system (Huang et al., 2007). In this work a considerable number of lineage specific genes were found to be promiscuously expressed in bipotential precursors, such that this metastable state is always primed for differentiation to either direction. The chance upregulation of opposing lineage specific transcription factors in different cells of bipotent origin may then be translated into a mutual inhibition and self-reinforcing program that ensures progression of each population to a terminally differentiated stable state (Chang et al., 2008; Huang et al., 2007).

The existence of a pluripotent ground state for ES cells appear to be made possible by transcription factor duels between *Oct3/4* and *Cdx2* in determining a choice between ICM or trophectoderm (placental), while expression of *Nanog* appears to buffer cells against endogenous FGF/Erk signalling that leads to a primitive endoderm fate (Chazaud et al., 2006; Niwa et al., 2005). Thus, the uncommitted state is the result of many gene regulatory networks that cancel each other out such that the breaks are put on these differentiation-prone cells.

Stochastic events leading to determined cell fates are also seen in the ICM and this is particularly evident in a self–enhancement lateral inhibition (SELI) mechanism as seen by the expression of the downstream antagonist of Nodal signalling, namely, *Lefty1* (Dietrich and Hiiragi, 2007; Takaoka et al., 2007). This mechanism proposes that *Lefty1* expression fluctuates in cells of the ICM such that the one or two cells

that happen to be expressing slightly higher levels are able to maintain this signal while the others lose it. This stochastic choice is then converted into asymmetric localisation within the primitive endoderm that is eventually translated as an expression domain just anterior to the DVE at E5.5 following cavitation. Therefore, heterogeneous expression of *Lefty1* may later determine the A-P axis of the mouse embryo. It is of particular interest then that another primitive endoderm marker that is also reported to be a downstream antagonist of *Nodal*, namely *Hex*, should display a heterogenous pattern of expression in ES cells as observed here (Rodriguez et al., 2001; Zamparini et al., 2006). Microarray analysis in this thesis shows an enrichment of primitive endoderm and genes downstream of TGFβ/Nodal signalling in the *Hex* positive population. It could be that *Nodal*, which is required for pluripotency in mouse ES cells may also dictate transcriptional noise events in downstream primitive endoderm genes, priming them for their later and eventual asymmetric expression (Ogawa et al., 2007). In the ICM at E3.5, stimulation of the ERK branch of the FGF pathway is required for the downregulation of *Nanog* and thus allowing the emergence of the primitive endoderm, as seen with *Gata6* (Chazaud et al., 2006). It appears that Hex is also regulated by *Nanog*, as seen in this thesis, while *Lefty1* has been reported to activate Erk pathway which would lead to suppresion of *Nanog* (Hamazaki et al., 2006; Ulloa et al., 2001). Thus, an additional or complemetary mechanism of primitive endoderm specification may also exist by virtue of *Nodal* signalling, downstream components of which may reinforce this fate by suppression of *Nanog*. Finally, it is interesting that a heterogeneous expression pattern for *Gata6* also appears to exist in undifferentiated ES cell cultures as judged by *LacZ* expression from this locus (Koutsourakis et al., 1999).

It will be interesting to see whether the pattern of Venus expression seen in ES cells is reflected in the ICM at E3.5 in mice that will eventually be derived from the BHIV line.

While the Hex negative subpopulation appeared to be enriched for super-pluripotency and gonad related genes, components of Wnt signalling appeared to associate with Venus positive fraction. This finding was curious since stimulation of

the Wnt pathway, through Gsk3β inhibiton, promotes self-renewal of mouse and human ES cells (Sato et al., 2004). However, Hex has been previously reported to stimulate canonical Wnt signalling in ES cells and Xenopus indirectly by interfereing with an inhibitor of this pathway, namely *Tle4* (Zamparini et al., 2006). Previous reports on Wnt stimulation alone either by use of Gsk3β inhibitors or the constituitive expression of the *Wnt1* inhibits neural differentiation (Aubert et al., 2002; Ying et al., 2008). It was therefore interesting that two inhibitors of the neurectoderm lineage, *Id1* and *Id2* (Ying et al., 2003), were also enriched in the Venus positive population. Additionally, the Wnt antagonist, *Dkkl*, is present in this population. This secreted factor lies downstream of Wnt signalling and is reported to act as a guidance cue in the movement of the AVE and therefore mediates A-P axis polarization in the pre-gastrula mouse embryo (Kimura-Yoshida et al., 2005). It will be intersting to monitor the effects of either Wnt stimulation or antagonism on Venus expression of the BHIV line.

Previous work in our lab has revealed that Hex misexpression from the ROSA26 locus lead to the downregulation of to of the pluripotency markers *Nr0b1* and *Klf2* (Zamparini et al., 2006). Interesetingly, these same pluripotency markers were identified together with many others in the Venus negative fraction supporting the idea that they could potentially be targets. Thus, it will be valuable to examine whether *Hex* has a role in the suppression other such pluripotency markers. The generation of a constituitive low level Hex expression cell line (see Chapter 5.1a) will help to address such question for future work. Additionally, upon further optimisation, the cell lines reported in this thesis will become useful substrates for biochemical anlaysis and the identification of protein partners targets of *Hex*.

Agius, E., Oelgeschlager, M., Wessely, O., Kemp, C. and De Robertis, E. M. (2000). Endodermal Nodal-related signals and mesoderm induction in Xenopus. *Development* **127**, 1173-83.

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. *Mol Cell Biol* **17**, 6321-9.

Ang, S. L., Jin, O., Rhinn, M., Daigle, N., Stevenson, L. and Rossant, J. (1996). A targeted mouse Otx2 mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain. *Development* **122**, 243-52.

Arceci, R. J., King, A. A., Simon, M. C., Orkin, S. H. and Wilson, D. B. (1993). Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol Cell Biol* **13**, 2235-46.

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**, 1240-5.

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**, 126-40.

Beddington, R. S. (1994). Induction of a second neural axis by the mouse node. *Development* **120**, 613-20.

Beddington, R. S. and Robertson, E. J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733-7.

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

Bedford, F. K., Ashworth, A., Enver, T. and Wiedemann, L. M. (1993). HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human. *Nucleic Acids Res* **21**, 1245-9.

Belo, J. A., Bouwmeester, T., Leyns, L., Kertesz, N., Gallo, M., Follettie, M. and De Robertis, E. M. (1997). Cerberus-like is a secreted factor with neutralizing activity expressed in the anterior primitive endoderm of the mouse gastrula. *Mech Dev* **68**, 45-57.

Bogue, C. W., Ganea, G. R., Sturm, E., Ianucci, R. and Jacobs, H. C. (2000). Hex expression suggests a role in the development and function of organs derived from foregut endoderm. *Dev Dyn* **219**, 84-9.

Borner, M. M., Myers, C. E., Sartor, O., Sei, Y., Toko, T., Trepel, J. B. and Schneider, E. (1995). Drug-induced apoptosis is not necessarily dependent on macromolecular synthesis or proliferation in the p53-negative human prostate cancer cell line PC-3. *Cancer Res* **55**, 2122-8.

Bort, R., Martinez-Barbera, J. P., Beddington, R. S. and Zaret, K. S. (2004). Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* **131**, 797-806.

Bouwmeester, T., Kim, S., Sasai, Y., Lu, B. and De Robertis, E. M. (1996). Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. *Nature* **382**, 595-601.

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**, 947-56.

Brantjes, H., Roose, J., van De Wetering, M. and Clevers, H. (2001). All Tcf HMG box transcription factors interact with Groucho-related co-repressors. *Nucleic Acids Res* **29**, 1410-9.

Brennan, J., Lu, C. C., Norris, D. P., Rodriguez, T. A., Beddington, R. S. and Robertson, E. J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**, 965-9.

Brickman, J. M., Jones, C. M., Clements, M., Smith, J. C. and Beddington, R. S. (2000). Hex is a transcriptional repressor that contributes to anterior identity and suppresses Spemann organiser function. *Development* **127**, 2303-15.

Brons, I. G., Smithers, L. E., Trotter, M. W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A. et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191-5.

Cadigan, K. M. and Nusse, R. (1997). Wnt signaling: a common theme in animal development. *Genes Dev* **11**, 3286-305.

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**, 643-55.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L. and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-4.

Chambers, I. and Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**, 7150-60.

Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544-7.

Chappell, S. A., Edelman, G. M. and Mauro, V. P. (2000). A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. *Proc Natl Acad Sci U S A* **97**, 1536-41.

Chazaud, C. and Rossant, J. (2006). Disruption of early proximodistal patterning and AVE formation in Apc mutants. *Development* **133**, 3379-87.

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**, 615-24.

Chickarmane, V., Troein, C., Nuber, U. A., Sauro, H. M. and Peterson, C. (2006). Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol* **2**, e123.

Conlon, F. L., Lyons, K. M., Takaesu, N., Barth, K. S., Kispert, A., Herrmann, B. and Robertson, E. J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**, 1919-28.

Crompton, M. R., Bartlett, T. J., MacGregor, A. D., Manfioletti, G., Buratti, E., Giancotti, V. and Goodwin, G. H. (1992). Identification of a novel vertebrate homeobox gene expressed in haematopoietic cells. *Nucleic Acids Res* **20**, 5661-7.

Cui, L., Johkura, K., Yue, F., Ogiwara, N., Okouchi, Y., Asanuma, K. and Sasaki, K. (2004). Spatial distribution and initial changes of SSEA-1 and other cell adhesion-related molecules on mouse embryonic stem cells before and during differentiation. *J Histochem Cytochem* **52**, 1447-57.

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**, 7480-5.

de la Luna, S. and Ortin, J. (1992). pac gene as efficient dominant marker and reporter gene in mammalian cells. *Methods Enzymol* **216**, 376-85.

Dietrich, J. E. and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. *Development* **134**, 4219-31.

Ding, J., Yang, L., Yan, Y. T., Chen, A., Desai, N., Wynshaw-Boris, A. and Shen, M. M. (1998). Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo. *Nature* **395**, 702-7.

Donaldson, I. J., Chapman, M., Kinston, S., Landry, J. R., Knezevic, K., Piltz, S., Buckley, N., Green, A. R. and Gottgens, B. (2005). Genome-wide identification of cis-regulatory sequences controlling blood and endothelial development. *Hum Mol Genet* **14**, 595-601.

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**, 477-82.

Dufort, D., Schwartz, L., Harpal, K. and Rossant, J. (1998). The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis. *Development* **125**, 3015-25.

Dunn, N. R., Koonce, C. H., Anderson, D. C., Islam, A., Bikoff, E. K. and Robertson, E. J. (2005). Mice exclusively expressing the short isoform of Smad2 develop normally and are viable and fertile. *Genes Dev* **19**, 152-63.

Dunn, N. R., Vincent, S. D., Oxburgh, L., Robertson, E. J. and Bikoff, E. K. (2004). Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. *Development* **131**, 1717-28.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6.

Feldman, B., Gates, M. A., Egan, E. S., Dougan, S. T., Rennebeck, G., Sirotkin, H. I., Schier, A. F. and Talbot, W. S. (1998). Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* **395**, 181-5.

Fernandes-Alnemri, T., Litwack, G. and Alnemri, E. S. (1994). CPP32, a novel human apoptotic protein with homology to Caenorhabditis elegans cell death protein Ced-3 and mammalian interleukin-1 beta-converting enzyme. *J Biol Chem* **269**, 30761-4.

Finley, K. R., Tennessen, J. and Shawlot, W. (2003). The mouse secreted frizzled-related protein 5 gene is expressed in the anterior visceral endoderm and foregut endoderm during early post-implantation development. *Gene Expr Patterns* **3**, 681-4.

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

Gardner, R. L. (1982). Investigation of cell lineage and differentiation in the extraembryonic endoderm of the mouse embryo. *J Embryol Exp Morphol* **68**, 175-98.

Gardner, R. L. (1985a). Clonal analysis of early mammalian development. *Philos Trans R Soc Lond B Biol Sci* **312**, 163-78.

Gardner, R. L. (1985b). Regeneration of endoderm from primitive ectoderm in the mouse embryo: fact or artifact? *J Embryol Exp Morphol* **88**, 303-26.

Gardner, R. L. and Rossant, J. (1979). Investigation of the fate of 4-5 day post-coitum mouse inner cell mass cells by blastocyst injection. *J Embryol Exp Morphol* **52**, 141-52.

Gehring, W. J. (1996). The master control gene for morphogenesis and evolution of the eye. *Genes Cells* **1**, 11-5.

Gehring, W. J., Affolter, M. and Burglin, T. (1994). Homeodomain proteins. *Annu Rev Biochem* **63**, 487-526.

George, A., Morse, H. C., 3rd and Justice, M. J. (2003). The homeobox gene Hex induces T-cell-derived lymphomas when overexpressed in hematopoietic precursor cells. *Oncogene* **22**, 6764-73.

Glinka, A., Wu, W., Delius, H., Monaghan, A. P., Blumenstock, C. and Niehrs, C. (1998). Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. *Nature* **391**, 357-62.

Gu, Z., Nomura, M., Simpson, B. B., Lei, H., Feijen, A., van den Eijnden-van Raaij, J., Donahoe, P. K. and Li, E. (1998). The type I activin receptor ActRIB is required for egg cylinder organization and gastrulation in the mouse. *Genes Dev* **12**, 844-57.

Guiral, M., Bess, K., Goodwin, G. and Jayaraman, P. S. (2001). PRH represses transcription in hematopoietic cells by at least two independent mechanisms. *J Biol Chem* **276**, 2961-70.

Guo, Y., Chan, R., Ramsey, H., Li, W., Xie, X., Shelley, W. C., Martinez-Barbera, J. P., Bort, B., Zaret, K., Yoder, M. et al. (2003). The homeoprotein Hex is required for hemangioblast differentiation. *Blood.*

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* **26**, 7539-49.

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**, 5681-6.

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**, 187-98.

Hay, D. C., Sutherland, L., Clark, J. and Burdon, T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells* **22**, 225-35.

Heyer, J., Escalante-Alcalde, D., Lia, M., Boettinger, E., Edelmann, W., Stewart, C. L. and Kucherlapati, R. (1999). Postgastrulation Smad2-deficient

embryos show defects in embryo turning and anterior morphogenesis. *Proc Natl Acad Sci U S A* **96**, 12595-600.

Ho, C. Y., Houart, C., Wilson, S. W. and Stainier, D. Y. (1999). A role for the extraembryonic yolk syncytial layer in patterning the zebrafish embryo suggested by properties of the hex gene. *Curr Biol* **9**, 1131-4.

Hobbs, S., Jitrapakdee, S. and Wallace, J. C. (1998). Development of a bicistronic vector driven by the human polypeptide chain elongation factor 1alpha promoter for creation of stable mammalian cell lines that express very high levels of recombinant proteins. *Biochem Biophys Res Commun* **252**, 368-72.

Hogan, B. L. and Tilly, R. (1981). Cell interactions and endoderm differentiation in cultured mouse embryos. *J Embryol Exp Morphol* **62**, 379-94.

Huang, S., Guo, Y. P., May, G. and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev Biol* **305**, 695-713.

Huelsken, J. and Behrens, J. (2002). The Wnt signalling pathway. *J Cell Sci* **115**, 3977-8.

Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C. and Birchmeier, W. (2000). Requirement for beta-catenin in anterior-posterior axis formation in mice. *J Cell Biol* **148**, 567-78.

James, D., Levine, A. J., Besser, D. and Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132**, 1273-82.

Jayaraman, P. S., Frampton, J. and Goodwin, G. (2000). The homeodomain protein PRH influences the differentiation of haematopoietic cells. *Leuk Res* **24**, 1023-31.

Jones, C. M., Broadbent, J., Thomas, P. Q., Smith, J. C. and Beddington, R. S. (1999). An anterior signalling centre in Xenopus revealed by the homeobox gene XHex. *Curr Biol* **9**, 946-54.

Jones, C. M., Kuehn, M. R., Hogan, B. L., Smith, J. C. and Wright, C. V. (1995). Nodal-related signals induce axial mesoderm and dorsalize mesoderm during gastrulation. *Development* **121**, 3651-62.

Kasamatsu, S., Sato, A., Yamamoto, T., Keng, V. W., Yoshida, H., Yamazaki, Y., Shimoda, M., Miyazaki, J. and Noguchi, T. (2004). Identification of the transactivating region of the homeodomain protein, hex. *J Biochem (Tokyo)* **135**, 217-23.

Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* **19**, 1129-55.

Keng, V. W., Yagi, H., Ikawa, M., Nagano, T., Myint, Z., Yamada, K., Tanaka, T., Sato, A., Muramatsu, I., Okabe, M. et al. (2000). Homeobox gene Hex is essential for onset of mouse embryonic liver development and differentiation of the monocyte lineage. *Biochem Biophys Res Commun* **276**, 1155-61.

Kim, J., Chu, J., Shen, X., Wang, J. and Orkin, S. H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049-61.

Kimura-Yoshida, C., Nakano, H., Okamura, D., Nakao, K., Yonemura, S., Belo, J. A., Aizawa, S., Matsui, Y. and Matsuo, I. (2005). Canonical Wnt signaling and its antagonist regulate anterior-posterior axis polarization by guiding cell migration in mouse visceral endoderm. *Dev Cell* **9**, 639-50.

Kimura, C., Yoshinaga, K., Tian, E., Suzuki, M., Aizawa, S. and Matsuo, I. (2000). Visceral endoderm mediates forebrain development by suppressing posteriorizing signals. *Dev Biol* **225**, 304-21.

Knowles, B. B., Howe, C. C. and Aden, D. P. (1980). Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science* **209**, 497-9.

Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R. and Grosveld, F. (1999). The transcription factor GATA6 is essential for early extraembryonic development. *Development* **126**, 723-32.

Kubo, A., Chen, V., Kennedy, M., Zahradka, E., Daley, G. Q. and Keller, G. (2005). The homeobox gene HEX regulates proliferation and differentiation of hemangioblasts and endothelial cells during ES cell differentiation. *Blood* **105**, 4590-7.

Kubo, A., Shinozaki, K., Shannon, J. M., Kouskoff, V., Kennedy, M., Woo, S., Fehling, H. J. and Keller, G. (2004). Development of definitive endoderm from embryonic stem cells in culture. *Development* **131**, 1651-62.

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**, 1649-61.

Kunath, T., Saba-El-Leil, M. K., Almousailleakh, M., Wray, J., Meloche, S. and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895-902.

Levin, M., Johnson, R. L., Stern, C. D., Kuehn, M. and Tabin, C. (1995). A molecular pathway determining left-right asymmetry in chick embryogenesis. *Cell* **82**, 803-14.

Lewis, S. L. and Tam, P. P. (2006). Definitive endoderm of the mouse embryo: formation, cell fates, and morphogenetic function. *Dev Dyn* **235**, 2315-29.

Liao, W., Ho, C. Y., Yan, Y. L., Postlethwait, J. and Stainier, D. Y. (2000). Hhex and scl function in parallel to regulate early endothelial and blood differentiation in zebrafish. *Development* **127**, 4303-13.

Liu, P., Wakamiya, M., Shea, M. J., Albrecht, U., Behringer, R. R. and Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. *Nat Genet* **22**, 361-5.

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* **38**, 431-40.

Lowe, L. A., Yamada, S. and Kuehn, M. R. (2001). Genetic dissection of nodal function in patterning the mouse embryo. *Development* **128**, 1831-43.

Lu, C. C. and Robertson, E. J. (2004). Multiple roles for Nodal in the epiblast of the mouse embryo in the establishment of anterior-posterior patterning. *Dev Biol* **273**, 149-59.

Lu, C. C., Robertson, E. J. and Brennan, J. (2004). The mouse frizzled 8 receptor is expressed in anterior organizer tissues. *Gene Expr Patterns* **4**, 569-72.

Mack, D. L., Leibowitz, D. S., Cooper, S., Ramsey, H., Broxmeyer, H. E. and Hromas, R. (2002). Down-regulation of the myeloid homeobox protein Hex is essential for normal T-cell development. *Immunology* **107**, 444-51.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R. et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55-70.

Manfioletti, G., Gattei, V., Buratti, E., Rustighi, A., De Iuliis, A., Aldinucci, D., Goodwin, G. H. and Pinto, A. (1995). Differential expression of a novel proline-rich homeobox gene (Prh) in human hematolymphopoietic cells. *Blood* **85**, 1237-45.

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 U S A* **78**, 7634-8.

Martinez-Barbera, J. P. and Beddington, R. S. (2001). Getting your head around Hex and Hesx1: forebrain formation in mouse. *Int J Dev Biol* **45**, 327-36.

Martinez Barbera, J. P., Clements, M., Thomas, P., Rodriguez, T., Meloy, D., Kioussis, D. and Beddington, R. S. (2000). The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. *Development* **127**, 2433-45.

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**, 4261-9.

Mesnard, D., Guzman-Ayala, M. and Constam, D. B. (2006). Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. *Development* **133**, 2497-505.

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**, 631-42.

Mohammadi, M., Froum, S., Hamby, J. M., Schroeder, M. C., Panek, R. L., Lu, G. H., Eliseenkova, A. V., Green, D., Schlessinger, J. and Hubbard, S. R. (1998). Crystal structure of an angiogenesis inhibitor bound to the FGF receptor tyrosine kinase domain. *Embo J* **17**, 5896-904.

Morrison, G. M. and Brickman, J. M. (2006). Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. *Development* **133**, 2011-22.

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 U S A* **91**, 4303-7.

Mukhopadhyay, M., Shtrom, S., Rodriguez-Esteban, C., Chen, L., Tsukui, T., Gomer, L., Dorward, D. W., Glinka, A., Grinberg, A., Huang, S. P. et al. (2001). Dickkopf1 is required for embryonic head induction and limb morphogenesis in the mouse. *Dev Cell* **1**, 423-34.

Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K. and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* **20**, 87-90.

Nakagawa, T., Abe, M., Yamazaki, T., Miyashita, H., Niwa, H., Kokubun, S. and Sato, Y. (2003). HEX acts as a negative regulator of angiogenesis by modulating the expression of angiogenesis-related gene in endothelial cells in vitro. *Arterioscler Thromb Vasc Biol* **23**, 231-7.

Newman, C. S., Chia, F. and Krieg, P. A. (1997). The XHex homeobox gene is expressed during development of the vascular endothelium: overexpression leads to an increase in vascular endothelial cell number. *Mech Dev* **66**, 83-93.

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**, 379-91.

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**, 372-6.

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

Niwa, H., Yamamura, K. and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193-9.

Norris, D. P., Brennan, J., Bikoff, E. K. and Robertson, E. J. (2002). The Foxh1-dependent autoregulatory enhancer controls the level of Nodal signals in the mouse embryo. *Development* **129**, 3455-68.

Nusse, R. and Varmus, H. E. (1992). Wnt genes. *Cell* **69**, 1073-87.

Obinata, A., Akimoto, Y., Omoto, Y. and Hirano, H. (2002). Expression of Hex homeobox gene during skin development: Increase in epidermal cell proliferation by transfecting the Hex to the dermis. *Dev Growth Differ* **44**, 281-92.

Ogawa, K., Saito, A., Matsui, H., Suzuki, H., Ohtsuka, S., Shimosato, D., Morishita, Y., Watabe, T., Niwa, H. and Miyazono, K. (2007). Activin-Nodal signaling is involved in propagation of mouse embryonic stem cells. *J Cell Sci* **120**, 55-65.

Ohkuma, Y., Horikoshi, M., Roeder, R. G. and Desplan, C. (1990). Engrailed, a homeodomain protein, can repress in vitro transcription by competition with the TATA box-binding protein transcription factor IID. *Proc Natl Acad Sci U S A* **87**, 2289-93.

Okita, K., Ichisaka, T. and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-7.

Osada, S. I. and Wright, C. V. (1999). Xenopus nodal-related signaling is essential for mesendodermal patterning during early embryogenesis. *Development* **126**, 3229-40.

Park, I. H., Zhao, R., West, J. A., Yabuuchi, A., Huo, H., Ince, T. A., Lerou, P. H., Lensch, M. W. and Daley, G. Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141-6.

Pellizzari, L., D'Elia, A., Rustighi, A., Manfioletti, G., Tell, G. and Damante, G. (2000). Expression and function of the homeodomain-containing protein Hex in thyroid cells. *Nucleic Acids Res* **28**, 2503-11.

Perea-Gomez, A., Lawson, K. A., Rhinn, M., Zakin, L., Brulet, P., Mazan, S. and Ang, S. L. (2001a). Otx2 is required for visceral endoderm movement and for the restriction of posterior signals in the epiblast of the mouse embryo. *Development* **128**, 753-65.

Perea-Gomez, A., Rhinn, M. and Ang, S. L. (2001b). Role of the anterior visceral endoderm in restricting posterior signals in the mouse embryo. *Int J Dev Biol* **45**, 311-20.

Perea-Gomez, A., Shawlot, W., Sasaki, H., Behringer, R. R. and Ang, S. (1999). HNF3beta and Lim1 interact in the visceral endoderm to regulate primitive streak formation and anterior-posterior polarity in the mouse embryo. *Development* **126**, 4499-511.

Perea-Gomez, A., Vella, F. D., Shawlot, W., Oulad-Abdelghani, M., Chazaud, C., Meno, C., Pfister, V., Chen, L., Robertson, E., Hamada, H. et al. (2002). Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. *Dev Cell* **3**, 745-56.

Popperl, H., Schmidt, C., Wilson, V., Hume, C. R., Dodd, J., Krumlauf, R. and Beddington, R. S. (1997). Misexpression of Cwnt8C in the mouse induces an ectopic embryonic axis and causes a truncation of the anterior neuroectoderm. *Development* **124**, 2997-3005.

Rebagliati, M. R., Toyama, R., Fricke, C., Haffter, P. and Dawid, I. B. (1998). Zebrafish nodal-related genes are implicated in axial patterning and establishing left-right asymmetry. *Dev Biol* **199**, 261-72.

Riego, E., Perez, A., Martinez, R., Castro, F. O., Lleonart, R. and de la Fuente, J. (1995). Differential constitutive expression of interferon genes in early mouse embryos. *Mol Reprod Dev* **41**, 157-66.

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**, 2354-66.

Rodriguez, T. A., Casey, E. S., Harland, R. M., Smith, J. C. and Beddington, R. S. (2001). Distinct enhancer elements control Hex expression during gastrulation and early organogenesis. *Dev Biol* **234**, 304-16.

Rosenquist, T. A. and Martin, G. R. (1995). Visceral endoderm-1 (VE-1): an antigen marker that distinguishes anterior from posterior embryonic visceral endoderm in the early post-implantation mouse embryo. *Mech Dev* **49**, 117-21.

Rossant, J. (2008). Stem cells and early lineage development. *Cell* **132**, 527-31.

Ruzinova, M. B. and Benezra, R. (2003). Id proteins in development, cell cycle and cancer. *Trends Cell Biol* **13**, 410-8.

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.

Schier, A. F. and Shen, M. M. (2000). Nodal signalling in vertebrate development. *Nature* **403**, 385-9.

Shaner, N. C., Steinbach, P. A. and Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. *Nat Methods* **2**, 905-9.

Shawlot, W., Wakamiya, M., Kwan, K. M., Kania, A., Jessell, T. M. and Behringer, R. R. (1999). Lim1 is required in both primitive streak-derived tissues and visceral endoderm for head formation in the mouse. *Development* **126**, 4925-32.

Sherwood, R. I., Jitianu, C., Cleaver, O., Shaywitz, D. A., Lamenzo, J. O., Chen, A. E., Golub, T. R. and Melton, D. A. (2007). Prospective isolation and global gene expression analysis of definitive and visceral endoderm. *Dev Biol* **304**, 541-55.

Singh, A. M., Hamazaki, T., Hankowski, K. E. and Terada, N. (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* **25**, 2534-42.

Smith, A. G. (2001). Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* **17**, 435-62.

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**, 688-90.

Smith, W. C., McKendry, R., Ribisi, S., Jr. and Harland, R. M. (1995). A nodal-related gene defines a physical and functional domain within the Spemann organizer. *Cell* **82**, 37-46.

Smithers, L. E. and Jones, C. M. (2002). Xhex-expressing endodermal tissues are essential for anterior patterning in Xenopus. *Mech Dev* **119**, 191-200.

Swingler, T. E., Bess, K. L., Yao, J., Stifani, S. and Jayaraman, P. S. (2004). The proline-rich homeodomain protein recruits members of the Groucho/Transducin-like enhancer of split protein family to co-repress transcription in hematopoietic cells. *J Biol Chem* **279**, 34938-47.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-72.

Takaoka, K., Yamamoto, M. and Hamada, H. (2007). Origin of body axes in the mouse embryo. *Curr Opin Genet Dev* **17**, 344-50.

Tam, P. P. and Steiner, K. A. (1999). Anterior patterning by synergistic activity of the early gastrula organizer and the anterior germ layer tissues of the mouse embryo. *Development* **126**, 5171-9.

Tanaka, T., Inazu, T., Yamada, K., Myint, Z., Keng, V. W., Inoue, Y., Taniguchi, N. and Noguchi, T. (1999). cDNA cloning and expression of rat homeobox gene, Hex, and functional characterization of the protein. *Biochem J* **339 (Pt 1)**, 111-7.

Tanaka, T. S., Davey, R. E., Lan, Q., Zandstra, P. W. and Stanford, W. L. (2008). Development of a gene-trap vector with a highly sensitive fluorescent protein reporter system for expression profiling. *Genesis* **46**, 347-56.

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. and McKay, R. D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196-9.

Thomas, P. and Beddington, R. (1996). Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Curr Biol* **6**, 1487-96.

Thomas, P. Q., Brown, A. and Beddington, R. S. (1998). Hex: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors. *Development* **125**, 85-94.

Topisirovic, I., Culjkovic, B., Cohen, N., Perez, J. M., Skrabanek, L. and Borden, K. L. (2003). The proline-rich homeodomain protein, PRH, is a tissue-specific inhibitor of eIF4E-dependent cyclin D1 mRNA transport and growth. *Embo J* **22**, 689-703.

Torres-Padilla, M. E., Richardson, L., Kolasinska, P., Meilhac, S. M., Luetke-Eversloh, M. V. and Zernicka-Goetz, M. (2007). The anterior visceral endoderm of the mouse embryo is established from both preimplantation precursor cells and by de novo gene expression after implantation. *Dev Biol* **309**, 97-112.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909-18.

Tremblay, K. D., Hoodless, P. A., Bikoff, E. K. and Robertson, E. J. (2000). Formation of the definitive endoderm in mouse is a Smad2-dependent process. *Development* **127**, 3079-90.

Ulloa, L., Creemers, J. W., Roy, S., Liu, S., Mason, J. and Tabibzadeh, S. (2001). Lefty proteins exhibit unique processing and activate the MAPK pathway. *J Biol Chem* **276**, 21387-96.

Vallier, L., Alexander, M. and Pedersen, R. A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* **118**, 4495-509.

van der Heyden, M. A. and Defize, L. H. (2003). Twenty one years of P19 cells: what an embryonal carcinoma cell line taught us about cardiomyocyte differentiation. *Cardiovasc Res* **58**, 292-302.

Vincent, S. D., Dunn, N. R., Hayashi, S., Norris, D. P. and Robertson, E. J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes Dev* **17**, 1646-62.

Waldrip, W. R., Bikoff, E. K., Hoodless, P. A., Wrana, J. L. and Robertson, E. J. (1998). Smad2 signaling in extraembryonic tissues determines anterior-posterior polarity of the early mouse embryo. *Cell* **92**, 797-808.

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**, 364-8.

Wang, Z. and Jaenisch, R. (2004). At most three ES cells contribute to the somatic lineages of chimeric mice and of mice produced by ES-tetraploid complementation. *Dev Biol* **275**, 192-201.

Xu, R. H., Peck, R. M., Li, D. S., Feng, X., Ludwig, T. and Thomson, J. A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Methods* **2**, 185-90.

Yamamoto, M., Saijoh, Y., Perea-Gomez, A., Shawlot, W., Behringer, R. R., Ang, S. L., Hamada, H. and Meno, C. (2004). Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature* **428**, 387-92.

Yatskievych, T. A., Pascoe, S. and Antin, P. B. (1999). Expression of the homebox gene Hex during early stages of chick embryo development. *Mech Dev* **80**, 107-9.

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**, 281-92.

Ying, Q. L., Nichols, J., Evans, E. P. and Smith, A. G. (2002). Changing potency by spontaneous fusion. *Nature* **416**, 545-8.

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-23.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R. et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-20.

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**, 3789-94.

Zamparini, A. L., Watts, T., Gardner, C. E., Tomlinson, S. R., Johnston, G. I. and Brickman, J. M. (2006). Hex acts with beta-catenin to regulate anteroposterior patterning via a Groucho-related co-repressor and Nodal. *Development* **133**, 3709-22.

Zhang, W., Yatskievych, T. A., Baker, R. K. and Antin, P. B. (2004). Regulation of Hex gene expression and initial stages of avian hepatogenesis by Bmp and Fgf signaling. *Dev Biol* **268**, 312-26.

Zhang, W., Yatskievych, T. A., Cao, X. and Antin, P. B. (2002). Regulation of Hex gene expression by a Smads-dependent signaling pathway. *J Biol Chem* **277**, 45435-41.

Zorn, A. M., Butler, K. and Gurdon, J. B. (1999). Anterior endomesoderm specification in Xenopus by Wnt/beta-catenin and TGF-beta signalling pathways. *Dev Biol* **209**, 282-97.