

Directed Differentiation of Human Embryonic Stem Cells to Insulin Producing Cells by Regulating GATA Genes Cascade



Independent Learning Project Final Report

Supervisors

Dr Kuldip Sidhu

Professor Bernie Tuch

Faculty of Medicine

University of New South Wales, Sydney Australia

March, 2007

ABSTRACT

Human embryonic stem cells (hESC) are pluripotent cells that may serve as an attractive option for the generation of a renewable source of pancreatic β cells for the treatment of Type 1 diabetes. Previous studies have indicated the differentiation hESCs *in vitro* first into definitive endoderm to form pancreatic β cell progenitors, and then secrete insulin in response to high glucose concentrations. This project has focused on the directed differentiation of hESC to definitive endoderm via genetic manipulation of hESCs to overexpress *GATA4*, a transcription factor known to play a pivotal role in endodermal development. The DNA sequence of human *GATA4* gene was isolated from day 5 Activin A treated hESC and inserted into the pcDNA4/HisMax TOPO vector by cloning. A cell line of transgenic zeocin-resistant human fetal fibroblasts (HFF) was successfully created and shown to serve as a feeder layer for the antibiotic selection of the transfected hESC. pcDNA4/*GATA4* recombinant plasmid was transfected into hESCs by nucleofection. Positive clones were selected and characterised by RT-PCR for gene expression of *GATA4* and definitive endoderm markers. Results indicated that wild-type hESC expressed a range of markers from definitive endoderm and also a marker of pluripotency. Unexpectedly, semi-quantitative RT-PCR indicated that transfected hESC expressed lower levels of *GATA4* compared with wild-type hESC, while little change in expression of definitive endoderm markers was observed. Further investigation of protein expression and repeated hESC transfections would be essential to confirming the results obtained. However the approaches described here in modifying and monitoring hESC gene expression provide a useful model to better understand the molecular details of the pathway from hESC to definitive endoderm and insulin-producing cell.

INTRODUCTION

Type 1 diabetes is a chronic disease characterized by autoimmune destruction of insulin-producing β cells, located in islets of the pancreas. The loss of β cells results in insulin deficiency and high blood glucose levels. Type 1 Diabetes affects more than 140,000 Australians, and its incidence is increasing both in Australia and worldwide ¹. Current treatment involves daily insulin injections, routine monitoring of blood glucose levels and strict diet control however this is not ideal as it does not prevent long term complications such as kidney failure, diabetic retinopathy and neuropathy ².

Several approaches have been attempted to reverse the disease process for type 1 diabetes, such as replacing the β cells via whole organ pancreas transplants and islet transplants³. Despite the improvement of transplant techniques⁴, the severe shortage of islets from cadaver or live donors remains a major factor in restricting the viability of this option. Thus, interest has focused on developing renewable sources of insulin-producing cells appropriate for transplant.

Human embryonic stem cells (hESC), derived from the inner cell mass (ICM) of 5-7 day old blastocysts may provide an attractive option for the generation of functional pancreatic β cells. HESC are pluripotent cells which possess unique properties of unlimited self-renewal and proliferation when propagated *in vitro*. Such pluripotent cells have the capacity to develop into cell types representing the three embryonic germ layers under both *in vitro* and *in vivo* conditions ⁵.

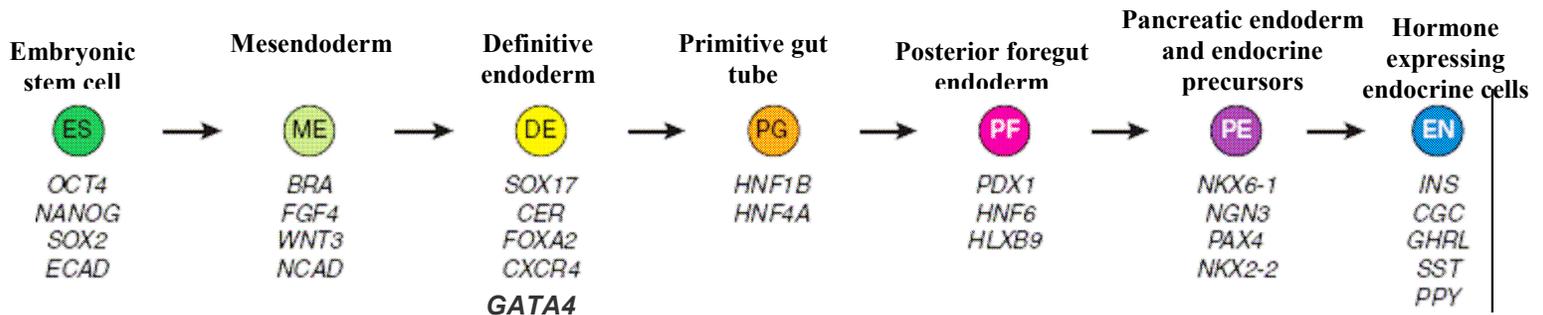
During early embryonic development *in vivo*, the trilaminar germ disc is formed by gastrulation and consists of three germ layers: endoderm, ectoderm, and mesoderm ⁶. The definitive endoderm is the inner most germ layer and gives rise to the epithelial lining of the primitive gut and its derivatives including the entire gastrointestinal tract, the thyroid, thymus, respiratory tract, liver and pancreas ⁷.

Although the identification reliable set of gene markers of definitive endoderm remains under scrutiny, several common components of the molecular pathways have been

established in vertebrates. These involve transcription factors of nodal pathways such as the TGF- β superfamily, the GATA and Forkhead-domain families, and high mobility group DNA-binding SOX proteins⁷⁻⁹. Genetic markers used to identify each stage of β cell derivation are listed in **Fig. 1**.

Members of the GATA family of zinc finger transcription factors have been shown to play critical roles in definitive endoderm formation¹⁰. GATA transcription factors are implicated in the activation of Sox transcription factors and are also known to operate in the later stages of endoderm development (Shivdasani, 2002). *GATA4* transcription factor is expressed in hESC (Segev, 2005). Studies *in vitro* have found that the overexpression of *GATA4* and *GATA6* by transfection into mouse embryonic stem cells induced morphological changes and the transcription of extra-embryonic endodermal genes^{11, 12}.

Figure 1 Pathway of β cell differentiation from hESC, adapted from D'Amour et al.¹³



The molecular details of the differentiation pathway from stem cell to insulin-secreting β cell are however, not completely understood, and remain under intense scrutiny in current research. Various studies focusing on mice and humans have demonstrated the ability of embryonic stem cells (ESC) to differentiate into insulin-secreting cells^{1, 2, 14-17}. There are generally two approaches to achieve this: the optimization of culture conditions in the cell culture dish; or genetic manipulation of the embryonic stem cell to express known transcription factors of endodermal and β cell nature¹⁸.

In mouse embryonic stem cells (ESC), there have been reports of directed differentiation into insulin-producing cells via the ectodermal pathway by using a step-wise protocol of

various culture conditions¹⁹. The conclusions obtained from these experiments have been controversial because it was shown that immunostaining for the insulin protein in culture was misleading since insulin from the surrounding culture media adheres to the cells. Rather than producing insulin, the cells may have instead been concentrating it from the culture media of this protocol²⁰.

D'Amour et al. reported the use of Activin A in combination with low serum concentrations to efficiently differentiate hESC to definitive endoderm¹³ and later insulin-secreting cells²¹. The cultures produced consisted of up to 80% definitive endoderm cells expressing gene markers including *Sox17* and *FoxA2*. Similarly, Shi et al. achieved differentiation of mouse ESC to insulin-secreting cells by using a combination of Activin A and retinoic acid¹⁵. Such research has indicated that the embryonic conditions inducing differentiation need to be recreated *in vitro*, following the natural development of β cells from definitive endoderm²².

In studies of genetic manipulation, the transfection and subsequent overexpression of Pax4 into mouse ESCs was found to significantly promote the development of islet-like spheroid structures that produced insulin in response to glucose^{1, 23}. Similarly, overexpression of Pdx1 and FoxA2 in hESCs led to the earlier expression of pancreatic markers in transfected cells, with further differentiation achieved *in vivo* with the formation of teratomas²⁴.

Several methods of transfection have been developed for gene transfer into mouse and human ESC, including electroporation, liposome-based and viral methods with varying rates of efficiency²⁵. Nucleofection is a more recently developed non-viral method in which delivery of plasmid DNA is achieved directly into the cell nucleus, resulting in enhanced gene expression, with 20% greater efficiency rates than electroporation in hESC²⁶. The enhanced transfection efficiency obtained with nucleofection lends it use as an effective tool to study the effect of ectopic expression of transcription factors such as *GATA4*.

The major aim of the present study is to directly differentiate hESC to endoderm by overexpressing *GATA4* after stable transfection by nucleofection and selection. It is hypothesised that the overexpression of *GATA4* in hESC may promote their differentiation to endoderm and thus to insulin-producing cells under *in vitro* conditions.

MATERIALS AND METHODS

hESC Culture

Preparation of HFF feeder plates

The Human fetal fibroblast (HFF) cell line, HFF01 was an in-house line established from fetal dermal tissue derived from the therapeutic termination of an early second trimester pregnancy. Thawing of frozen HFF01 and preparation of feeder plates were carried out as per standard procedure²⁷. Briefly, frozen HFF at passage 5 were thawed, washed and grown in T75 tissue culture flasks (Greiner Bio-one, Sydney Australia) with 20mL FDMEM (**Supplementary Fig. 1**) at 37°C in CO₂ incubator (Forma Scientific, Marietta, USA) with 5% CO₂ / 95% O₂ for 7 days to reach confluency. The cells were harvested with 0.25% Trypsin/1mM EDTA (Invitrogen, Victoria Australia), gamma irradiated (45Gyr) before being seeded onto gelatin-coated six well plates (Cellstar) at 1.5 x 10⁵ cells/mL, as feeder layers.

Thawing and subculturing hESC

Colonies from the in-house hESC line, Endeavour-1 (E1), were thawed at 37°C and diluted with 10mL Knock Out Serum Replacement Media (KOSR) (**Supplementary Fig. 2**) to remove remaining DMSO and cultured on six-well plates containing HFF01 feeder layers. hESC were incubated at 37°C in CO₂ incubator with 5% CO₂ / 95% O₂ and media changed daily.

Upon reaching 80% confluency within 7 days, hESC colonies were passaged by washing twice with pre-warmed PBS (without Ca²⁺ and Mg²⁺) (Invitrogen) and then incubated with 0.05% Trypsin/1mM EDTA for 2 minutes. KOSR was used to neutralise trypsin. The colonies were manually detached from the wells using a plastic scrapper and transferred to fresh feeder plates. Differentiating colonies if any were removed by dissection before sub-culturing. Excess hESC colonies were cryopreserved by a slow freezing procedure as previously described²⁷.

Vector Construction

Isolation of GATA4 cDNA

Human cDNA of *GATA4*, a 1328 base-pair sequence (NM002052) was isolated from Activin A (Day 5) treated hESCs by RT-PCR with Platinum® *Taq* DNA Polymerase

High Fidelity (Invitrogen, Australia). As a taq designed for cloning, this added a poly-A tail at the ends of the PCR product which would base-pair with the T-overhangs at the vector cloning site. Human *GATA4*-specific primers were used with the forward primer containing the start codon (5' atgtatcagagcttgccat 3'), and reverse primer containing the stop codon (5' ttacgcagtgattatgtccc 3')

The *GATA4* primers were optimised for Polymerase Chain Reaction (PCR) Amplification with the conditions:

97 °C for 2 min;
97°C for 15s
60.3°C for 45s
72°C for 90s; } 40 cycles
72°C for 5 min

MgSO₄ concentration = 1mM

The amplified PCR product was resolved on 1% agarose gel containing ethidium bromide, by electrophoresis to visualize the band. After confirming the presence and correct size of the band, the DNA was extracted and purified from the gel using the Wizard SV Gel and PCR Clean-Up System (Promega).

Ligation of the vector and bacterial transformation

The resulting *GATA4* cDNA was sequentially ligated to the pcDNA4/HisMax TOPO TA vector (mammalian expression vector, Invitrogen) (**Fig. 2**) using the kit provided. Briefly, the provided salt solution was added to the linearised vector (1µl) and PCR product (2µg) and incubated at room temperature for 5 minutes.

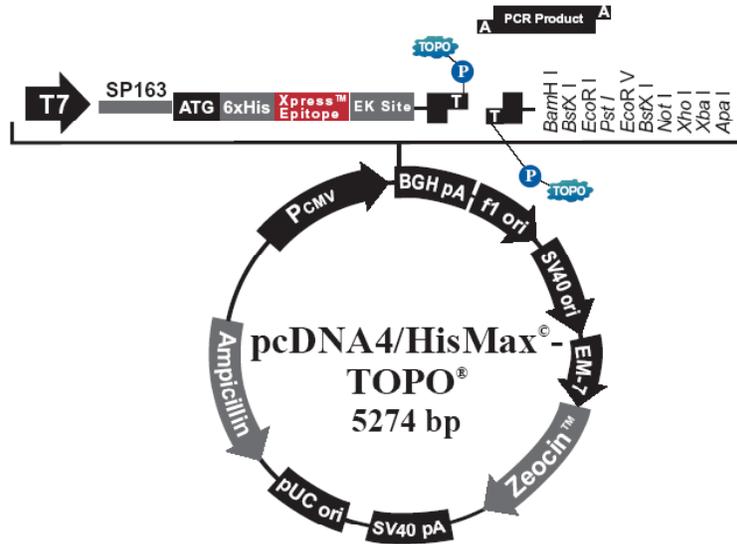


Figure 2 pcDNA4/HisMax-TOPO vector map and cloning site.
(<http://www.invitrogen.com>)

For the transformation of the plasmid into chemically competent *E. coli*, the mixture was incubated with *E. coli* (JM109) cells for 30 min on ice. The cells were heat shocked at 42°C for 30s and then incubated with LB media at 37°C for 1hr. Bacteria were then centrifuged at 1000 rpm and resuspended in a small amount of LB media before plating onto agar plates containing ampicillin (50µg/mL) for overnight incubation at 37°C (**Fig. 3**).

Several positive colonies were selected and further propagated on agar plates overnight at 37°C, and then in LB media with ampicillin (50µg/mL). The bacteria were harvested and the DNA was isolated using Wizard Plus SV Minipreps DNA Purification System (Promega). This was performed according to the protocols provided in the kit.

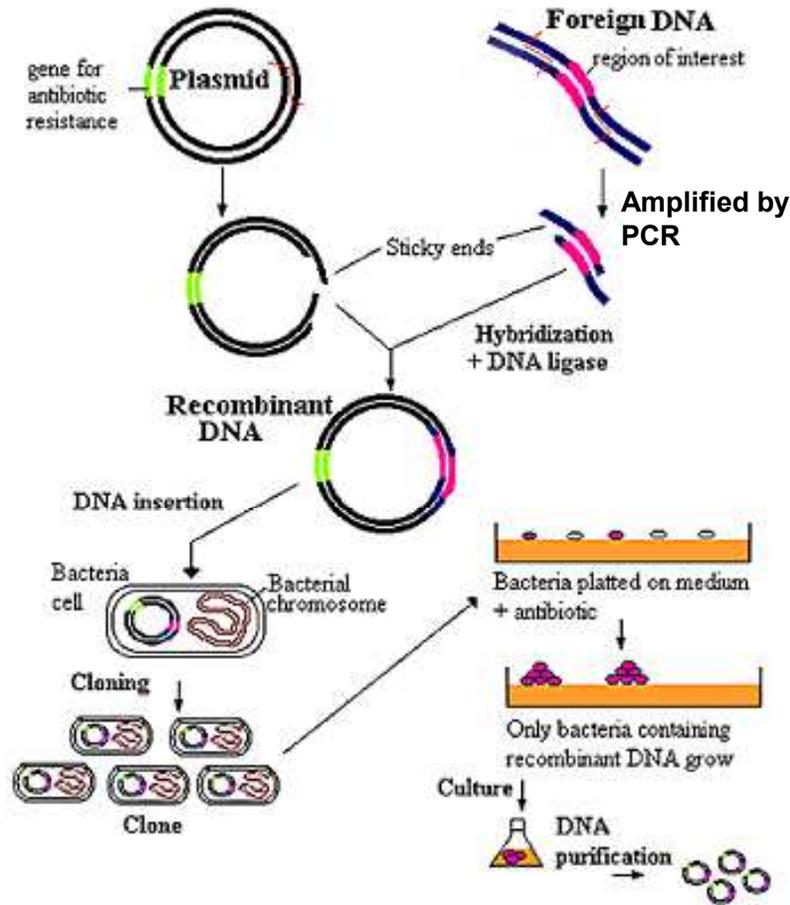


Figure 3 Cloning DNA into a vector. The plasmid carrying genes for antibiotic resistance is linearised and the gene of interest amplified by PCR. The plasmid and the PCR product are mixed with DNA ligase, which ligates the two pieces resulting in recombinant DNA. The resulting plasmid is allowed to transform a bacterial culture, which is then exposed to antibiotics. Only the plasmid and thus antibiotic resistance bearing bacteria survive the selection process and form colonies containing the recombinant DNA.

(Modified from source: <http://www.accessexcellence.org/RC/VL/GG/plasmid.html>)

Restriction enzyme digestion

The resulting vector DNAs were subjected to restriction enzyme digestion with *Xba*I and *Hind*III (**Fig. 4**) to check for presence of the insert. The protocol used was as follows:

Protocol: Restriction enzyme digestion

Plasmid DNA	5 μ g	} 10 μ l
<i>Xba</i> I (20 U/ μ l)	1 μ l	
<i>Hind</i> III (20 U/ μ l)	1 μ l	
<i>Xba</i> I and <i>Hind</i> III buffer (E)	2 μ l	
Milli-Q water	1 μ l	

For undigested plasmid DNA, 5 μ g of the DNA was mixed with 5 μ l of Milli-Q water. The undigested and digested mixtures were incubated at 37°C for 1h and 30 min. The digests were mixed with 1 μ l of 6X loading dye (Promega) and loaded onto a 1% agarose gel containing ethidium bromide for electrophoresis at 90V for 40 min. Cut plasmids were run against uncut plasmids on the gel and band sizes were compared to check for the *GATA4* insert.

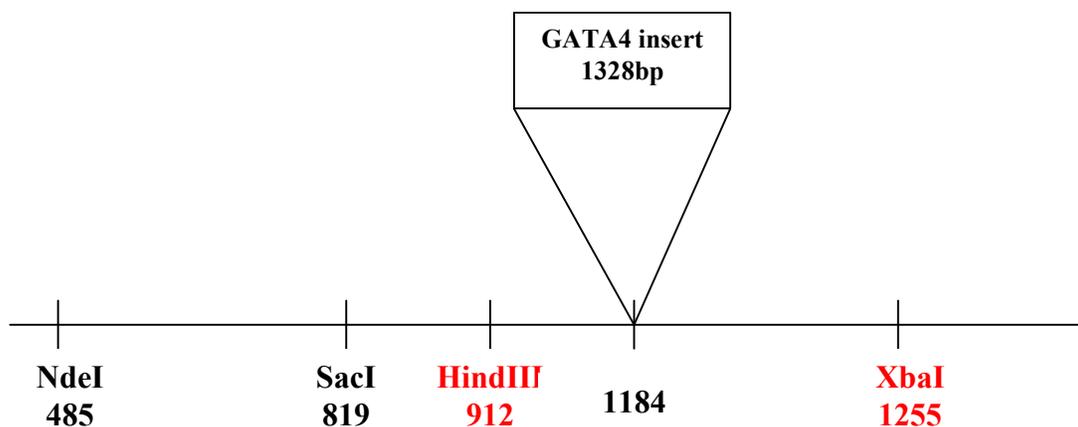


Figure 4 pcDNA4/HisMax- TOPO vector (5274bp) restriction enzyme map

Plasmid DNA containing the *GATA4* gene were sent for DNA sequence analysis at the Sydney University Prince Alfred Macromolecular Analysis Centre (SUPAMAC) to confirm the correct *GATA4* sequence and orientation.

hESC Transfection and Differentiation

The parent and clone 1 cell lines of E1-hESC cultured on HFF01 feeder layers were transfected by nucleofection according to the manufacturer (Amaxa Biosystem) and optimised to achieve 40% transfection efficiency. hESC at 70-80% confluency were harvested by washing twice with PBS (without Ca²⁺ and Mg²⁺) and incubating with collagenase (1mg/mL; Gibco) at 37°C for 7 minutes to remove the HFF. Collagenase was removed and hESC were treated with Acutase (Chemicon) for 25 minutes at 37°C to break the clumps into single cells. Following resuspension in KO-SR, aliquots of 2 million cells per sample were pelleted by centrifugation at 500 rpm for 5 minutes and resuspended in Solution V. Cells were nucleofected with either the eGFP-N1 plasmid as the positive control, or the pcDNA4-*GATA4* plasmid (5ug), using nucleofector program B-16.

Transfected hESCs recovered for 2-3 days before commencing antibiotic selection with the addition of zeocin at 50mg/mL. These hESCs were cultured and selected on either zeocin-resistant transgenic HFF feeder layers, or on Matrigel Basement Membrane Matrix (BD Biosciences) coated plates in a feeder-free system.

Zeocin-resistant transgenic HFFs as feeder layer

HFF10 at passage 3-5 was thawed as per standard procedure (Sidhu, 2005) and cultured in T75 flasks for 6-7 days in FDMEM (**Supplementary Fig. 1**). Upon reaching 80% confluency, HFF10 cells were transfected with the mammalian expression vector, pTracer CMV2 (Invitrogen) (**Supplementary Fig. 3**) containing a CMV promoter and GFP-Zeocin fused selection marker. The plasmid was recovered from frozen glycerol stock of plasmid-containing *E. coli* obtained from another member in the lab. *E. coli* were cultured in LB media containing ampicillin (50µg/ml) overnight at 37°C with shaking at

200 rpm, and plasmid DNA was extracted based on the Maxipreps DNA purification system (Promega).

The HFF10 cells were transfected with the resulting pTracer plasmid DNA (5 µg) using a nucleofector (Amaxa Biosystem), optimised to achieve 40% transfection efficiency. The nucleofection protocol was carried out according to the manufacturer's recommendations, using Solution V and program U-20.

Transfected HFFs were selected and propagated in FDMEM containing zeocin (50mg/mL) and passaged according to standard procedure²⁷. Feeder plates of transgenic zeocin-resistant HFFs were prepared according to standard procedure as described above.

Feeder-free system

Six well plates were coated with Matrigel basement membrane matrix (BD Biosciences) as previously described (Xu et al, 2001). Briefly, Matrigel was aliquoted to a 1:20 dilution in cold KO-DMEM to coat the plates. The surface of six-well plates (Cellstar) was overlaid with the diluted Matrigel for 1 hour at room temperature. Excess liquid was aspirated and the plates washed once with KO-DMEM before use.

The hESC cultured on Matrigel plates required conditioned medium. HFF-conditioned medium was prepared 24 hours before use on γ -irradiated (45 Gyr) HFF01 p5 cells with KO-SR plus 4ng/mL bFGF. An additional 8ng/mL bFGF was added before use, and media was changed daily.

Reverse Transcription-Polymerase Chain Reaction (RT-PCR) Analysis

Total RNA was extracted from transfected and non-transfected E1-hESCs using RNeasy extraction kit (Qiagen, Australia). The RNA was treated on column with 80µl Dnase 1 incubation mix (Qiagen) to remove any genomic DNA. The resulting RNA was then quantified by absorbance at 260nm using a WPA Lightwave spectrometer (Linton, Cambridge, UK). cDNA synthesis was carried out using Superscript III Kit (Invitrogen) from 200ng of RNA template, with incubation at 50°C for 50 min and 85°C for 5 min. RnaseH (2U/µl) was added to remove excess RNA templates.

Polymerase chain reaction (PCR) amplification reactions for *GATA4* were performed with Platinum Taq DNA Polymerase (Invitrogen) under the following conditions: 95°C for 5 min; 35 cycles of 95°C for 30s, 59°C for 30s, 72°C for 45s; 72°C for 5 min.

PCR reactions for endodermal markers *Sox17* & *FoxA2*, mesendoderm marker *GSC* and pluripotent marker, *Nanog* were similarly performed. Primer sequences, PCR conditions and the size of the final products are described in **Table 1**. PCR products were assessed by gel electrophoresis on 1.2% agarose ethidium bromide-stained gels.

Table 1 Primers used for PCR, conditions and their expected products

Gene	Primers (5'-3')		PCR conditions	Size (bp)
<i>β-actin</i>	Forward	ACGGCATCGTCACCAACT	97°C for 2min; 35 cycles of 96°C for 30s, 59°C for 30s, 72°C for 30s; 72°C for 5min	583
	Reverse	AGGAAGGAAGGCTGGAAGAG		
<i>Nanog</i>	Forward	CCTATGCCTGTGATTTGTGG	97°C for 2min; 35 cycles of 96°C for 30s, 58°C for 30s, 72°C for 30s; 72°C for 5min	208
	Reverse	CTGGGACCTTGTCTTCCTTT		
<i>GATA4</i>	Forward	CTAGACCGTGGGTTTTGCAT	96°C for 2min; 35 cycles of 96°C for 30s, 60°C for 30s, 72°C for 30s; 72°C for 5min	395
	Reverse	AGCGGGAAGAGGGATTTTA		
<i>FoxA2</i>	Forward	AGCGGTGAAGATGGAAGG	96°C for 2min; 35 cycles of 96°C for 30s, 59°C for 30s, 72°C for 30s; 72°C for 5min	582
	Reverse	CGGTAGAAGGGGAAGAGGT		
<i>Sox17</i>	Forward	AGCGCCCTTCACGTGTACTA	96°C for 2min; 35 cycles of 96°C for 30s, 59°C for 30s, 72°C for 30s; 72°C for 5min	246
	Reverse	CTTGACACGAAGTGCAGAT		
<i>GSC</i>	Forward	TCTCAACCAGCTGCACTGTC	96°C for 2min; 35 cycles of 96°C for 30s, 59°C for 30s, 72°C for 30s; 72°C for 5min	383
	Reverse	TCGTCTGTCTGTGCAAFTCC		

RESULTS

hESC culture

E1-hESC parent and clone 1 lines were propagated on a feeder layer of γ -irradiated human fetal fibroblast (HFF01) (**Fig. 5**). The undifferentiated state of the hESC was confirmed by a typical morphology, consisting of compactly arranged cells with high nucleoplasmic ratios and by RT-PCR expression of a pluripotent marker gene, *Nanog*.

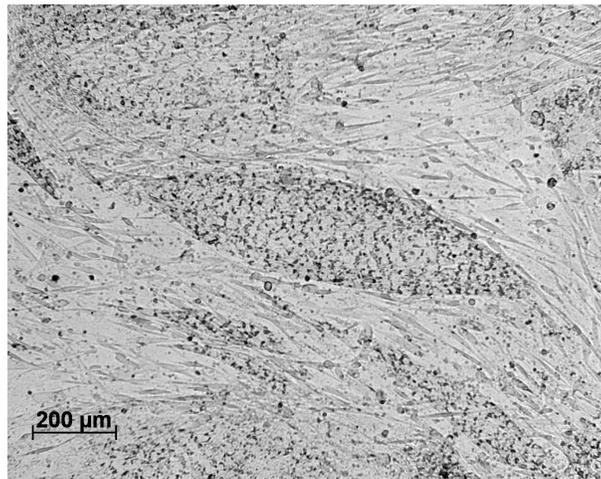
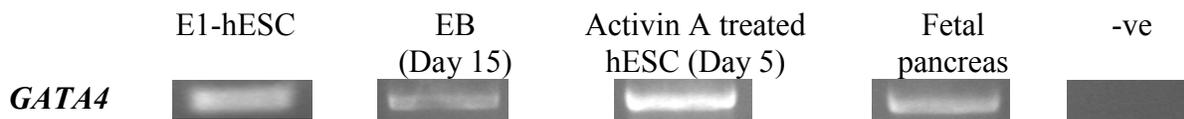


Figure 5 Typical morphology of undifferentiated E1 parent line hESC colonies (passage 46) grown on HFF feeder layer; at 10x magnification.

Vector Construction

Various RNA samples from different tissues and cells were tested for the presence of the *GATA4* RNA. *GATA4* was detected in various positive samples such as fetal pancreas, EB and Activin A treated (Day 5) hESCs, confirmed by RT-PCR. Endogenous expression of *GATA4* was also observed in Endeavour-1 hESC. **Fig. 6** shows RT-PCR results of *GATA4* positive controls.

Figure 6 *GATA4* expression in E1-hESC, Embryoid Bodies (Day 15), Activin A treated hESC (Day 5) and fetal pancreas. HFF01 cDNA was the negative control.



The Activin A treated hESC sample showed the greatest band intensity, and was chosen for the isolation of the full human *GATA4* DNA sequence for cloning. The PCR product was extracted from the gel for cloning into the pcDNA4 vector. **Fig. 7** shows the bands from which the PCR product were extracted and purified

Following ligation of the PCR product to the linearised pcDNA4 vector, and cloning with chemically competent *E. Coli*, 8 positive colonies were selected and propagated. From restriction enzyme digestion conducted on the resulting vector DNA, 2 out of the 8 colonies analysed showed the expected band sizes indicating the presence of the insert (**Fig. 8**).

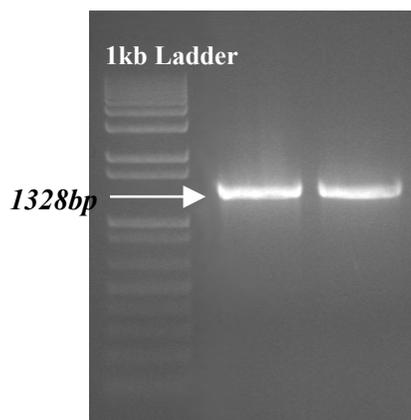


Figure 7 *GATA4* PCR product

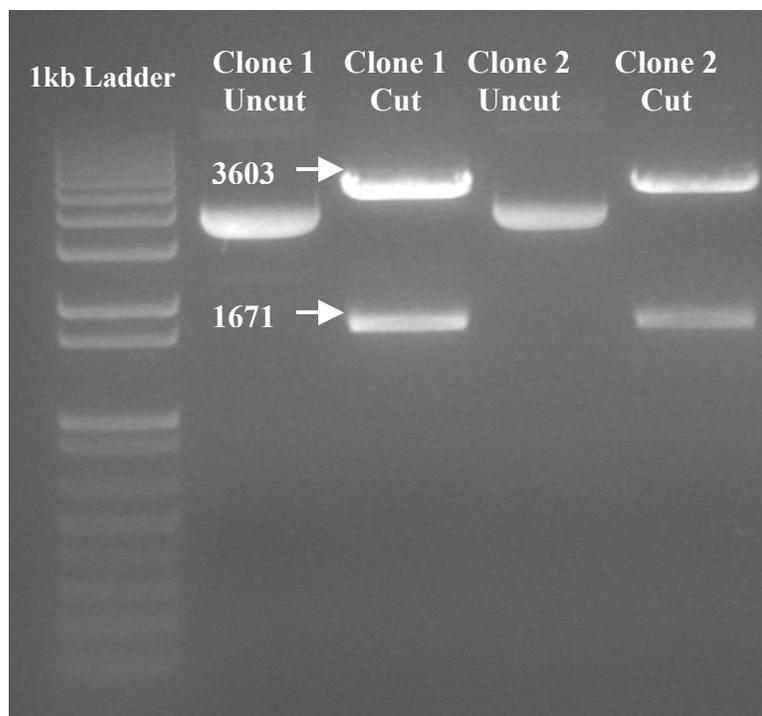


Figure 8 Restriction enzyme digestion. HindIII and XbaI restriction enzymes were used to cut the vector near the *GATA4* sequence insert. Cut vectors were run against un-cut vectors. Clones 1 and 2 demonstrated the expected band sizes (3603bp and 1671bp) after digestion, and were DNA sequenced.

Zeocin resistant transgenic HFFs

From a dose curve analysis using non-transfected HFF10, a concentration of 50mg/mL zeocin was selected as it killed the cells at approximately 14-17 days. In order to create a line of zeocin-resistant transgenic human fetal fibroblasts, HFF10 cells at passage 3-5 were initially nucleofected with the pIRESbleo3 vector (mammalian expression vector, Clontech). However repeated transfections ($n = 5$) with this vector proved unsuccessful as all transfected cells appeared to eventually die off during and after the selection period of 14-17 days, with no apparent cell proliferation as observed with a light microscope. As a positive transfection control, the HFF10 cells were transfected with an eGFP vector provided with the nucleofection kit (Amaxa). Fluorescent HFF were detected at Day 2 using a fluorescent inverted microscope, indicating the successful transfection of the cells. This shows that the nucleofection program settings and technique followed according to the manufacturer were appropriate. However the pIRESbleo3 vector was not suitable for transfection into this particular HFF cell line.

Subsequently a different vector pTracer CMV2, also conferring resistance to zeocin was transfected into HFF10 cells. It was found that positive clones survived and proliferated gradually in the presence of zeocin (50mg/mL), and non-transfected cells did not survive zeocin selection ($n = 2$). After the selection period of 14-17 days, growth of several small circular-shaped colonies was observed with gradual outward expansion and increased cell density within the centre of colonies. Morphology of cells (**Fig. 10**) appeared typical to that of normal fibroblasts, however observed growth rate was noticeably slower, particularly after subculture. The transgenic cells took much longer than the parent HFF10 cells to reach confluency even when media was changed every 2-3 days.

This delay in propagating the transgenic HFF restricted the production of an adequate amount of feeder layers on which to grow the prospective hESCs transfected with the pcDNA4-*GATA4* plasmid for selection. Therefore, an alternative method to culture the hESCs was devised, i.e. a feeder-free system using Matrigel Basement Membrane Matrix (BD Biosciences).

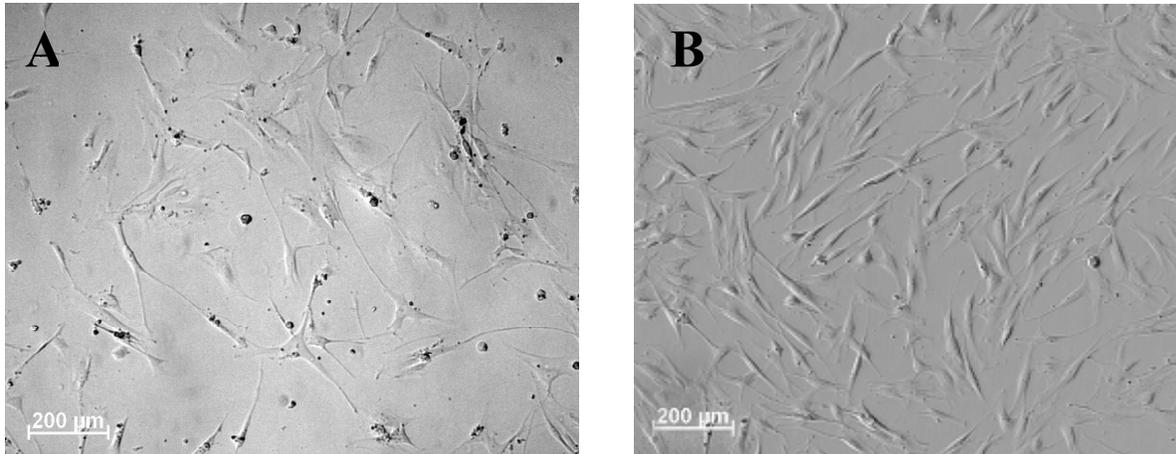


Figure 10 Morphology of transfected HFF10 p5 growing in a T75 flask (A), was typical of parent HFF10 cells (B).

Feeder-free hESC culture

In the present study, hESCs could only be maintained as pluripotent, undifferentiated colonies for the first couple of passages using the feeder-free system with Matrigel. After two passages, the morphology of the hESC colonies changed, with the appearance of mesenchymal-like cells around the periphery when observed under the microscope.

Transfected hESCs were therefore cultured on Matrigel only during the selection period of 14-17 days, without subculture as to minimise the disturbance of recovering transfected cells, and thereafter propagated on HFF feeder layers as per standard procedure²⁷.

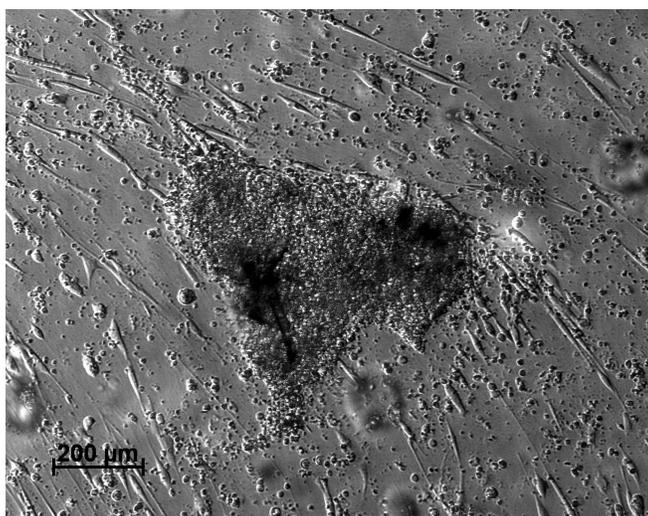
Stable Transfection of hESCs and Characterisation

The major aim of the project was to overexpress the transcription factor, *GATA4* in hESCs by nucleofection with the constructed pcDNA4-*GATA4* plasmid. E1-hESCs cultured on HFF01 feeder layers were transfected as single cells with the pcDNA4-*GATA4* plasmid. After transfection they were seeded onto either zeocin-resistant HFF feeder layers, or Matrigel coated plates in a feeder-free system, for antibiotic selection of stable *GATA4* expression.

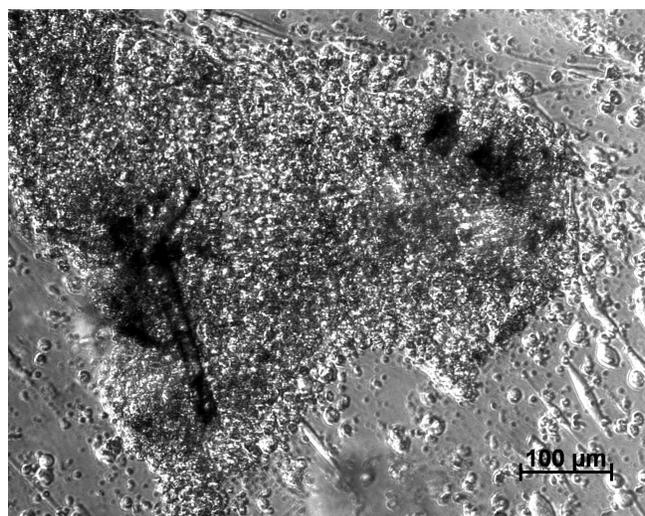
During the selection period of 14-17 days many cells died off and detached from the base of the wells. The positive clones that survived zeocin selection, were observed as single cells or small clumps adhering to the base of the wells, however these hESCs did not appear in the early stages to proliferate well and form colonies.

It was not until the first passage of hESC onto HFF feeder layer that a few small colonies began to appear at day 27. The morphology of these early colonies appeared typical to that of the parent non-transfected E1-hESC: the cells had high nucleus to cytoplasm ratios, prominent nucleoli and formed tightly compacted multicellular colonies (**Fig. 11**). However, even in these later stages there remained a noticeable difference in the growth characteristics of transfected hESCs compared with non-transfected hESC which proliferated at a much faster rate.

Figure 11 Morphology of transfected E1-hESC after subculture on to HFF feeder layer (Day 27).



10 X



20 X

Several colonies of positive clones from each well were pooled and RNA extraction was performed at two time points: day 20 and 37. The RNAs were reversely transcribed into cDNA before PCR amplification. RNA analysis was conducted to determine the

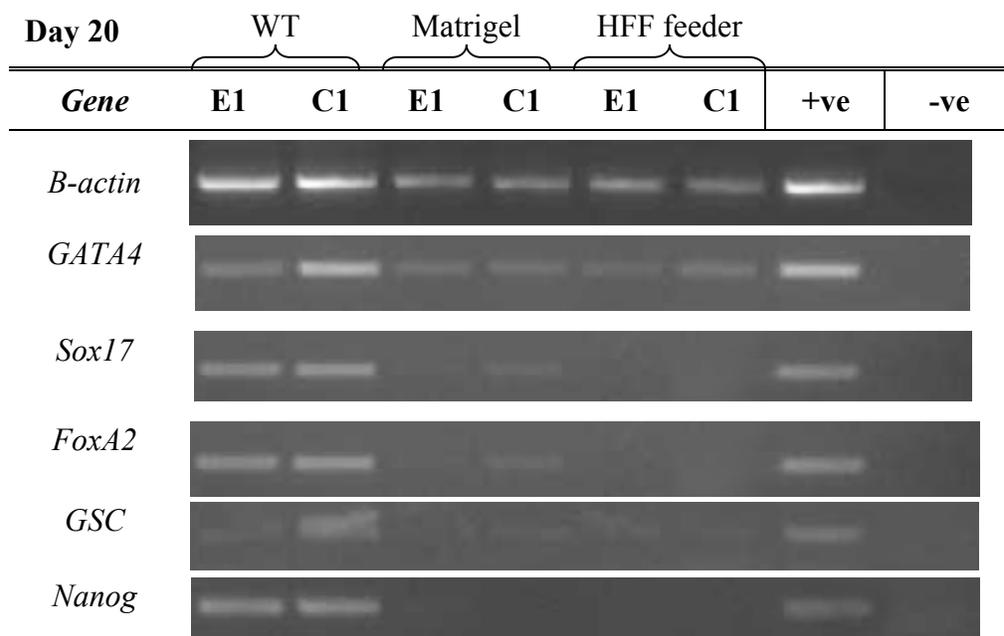
expression of *GATA4* in transfected and non-transfected cells. The expression profiles were analysed semi-quantitatively by ratio-metric analysis using β actin expression as a control house keeping gene.

As shown in **Fig. 12**, there was strong expression of *GATA4* in WT hESC, as this cell line expressed endogenous *GATA4*. However, after day 20, in the positive clones the expression of *GATA4* was surprisingly weak. Similarly, there was only weak expression of *GATA4* detected after day 37 in positive clones compared with high expression levels in WT cells, which confirmed the earlier finding.

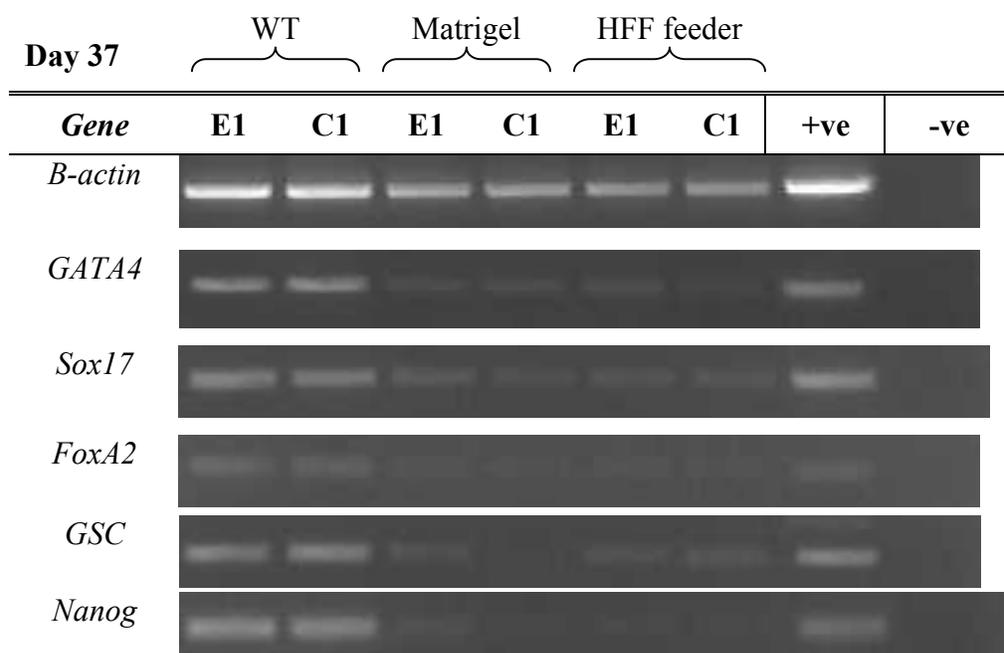
Further, RT-PCR was conducted to determine whether transfected and non-transfected hESC expressed markers representative of different lineages. The expression of *Nanog* was observed in WT cells, indicative of their pluripotency. However no expression was observed in the transfected cells at both time points, suggesting that these cells had lost their pluripotent state and were differentiating.

WT cells expressed definitive endoderm markers, *Sox17* and *FoxA2* and some faint expression was evident for day 20 and day 37 transfected hESC indicating some spontaneous differentiation. However the expression levels were relatively weaker compared with WT cell expression. Very faint or no bands were observed for *GSC* expression, suggesting no mesendoderm lineage formation in either the WT or the transfected cells (Fig. 12A and 12B). From t-test analyses in each group, there was no significant difference in expression levels between transfected cells growing on HFF feeder layer, and those growing on Matrigel ($p > 0.05$). T-test statistical analyses were also conducted to determine any significant differences in expression levels between E1 parent and clone 1 lines and these were marked with an asterisk. The only significant difference observed was in *Nanog* Day 37 cells, where transfected E1 parent line showed significantly ($p < 0.05$) more expression of *Nanog* compared with the clone 1 line.

Figure 12A: RT-PCR gene expressions in wild-type and transfected hESC at two time points (Day 20 and 37). *β actin* was used to assess the quality of the cDNA. Each sample was screened for the expression of GATA4, definitive endodermal markers Sox17 and FoxA2, mesendoderm marker Goosecoid (GSC) and pluripotent marker, Nanog. Activin-A treated (day 1) hESC was used as the positive control for *β actin*, GATA4, Sox17, FoxA3, GSC. Wild-type E1-hESC were used as the positive control for Nanog. HFF cDNA was used for the negative controls.



E1 = Endeavour-1 hESC parent line; C1 = Endeavour-1 hESC clone 1 line



E1 = Endeavour-1 hESC parent line; C1 = Endeavour-1 hESC clone 1 line

Figure 12B Semi-quantitative RT-PCR. All data sets are normalised to β -actin and charted in relative expression to WT (non-transfected) E1-hESC. The 2 columns represent E1-hESC parent line, and clone 1 line respectively. For GATA4 Day 20, $n=2$, and data is expressed as mean + standard error mean (SEM). For the rest of the data sets, $n=1$. In each group, t-test statistical analysis was performed to determine any significant differences between expression in parent and clone 1 E1-hESC. $*p < 0.05$, represents a significant difference.

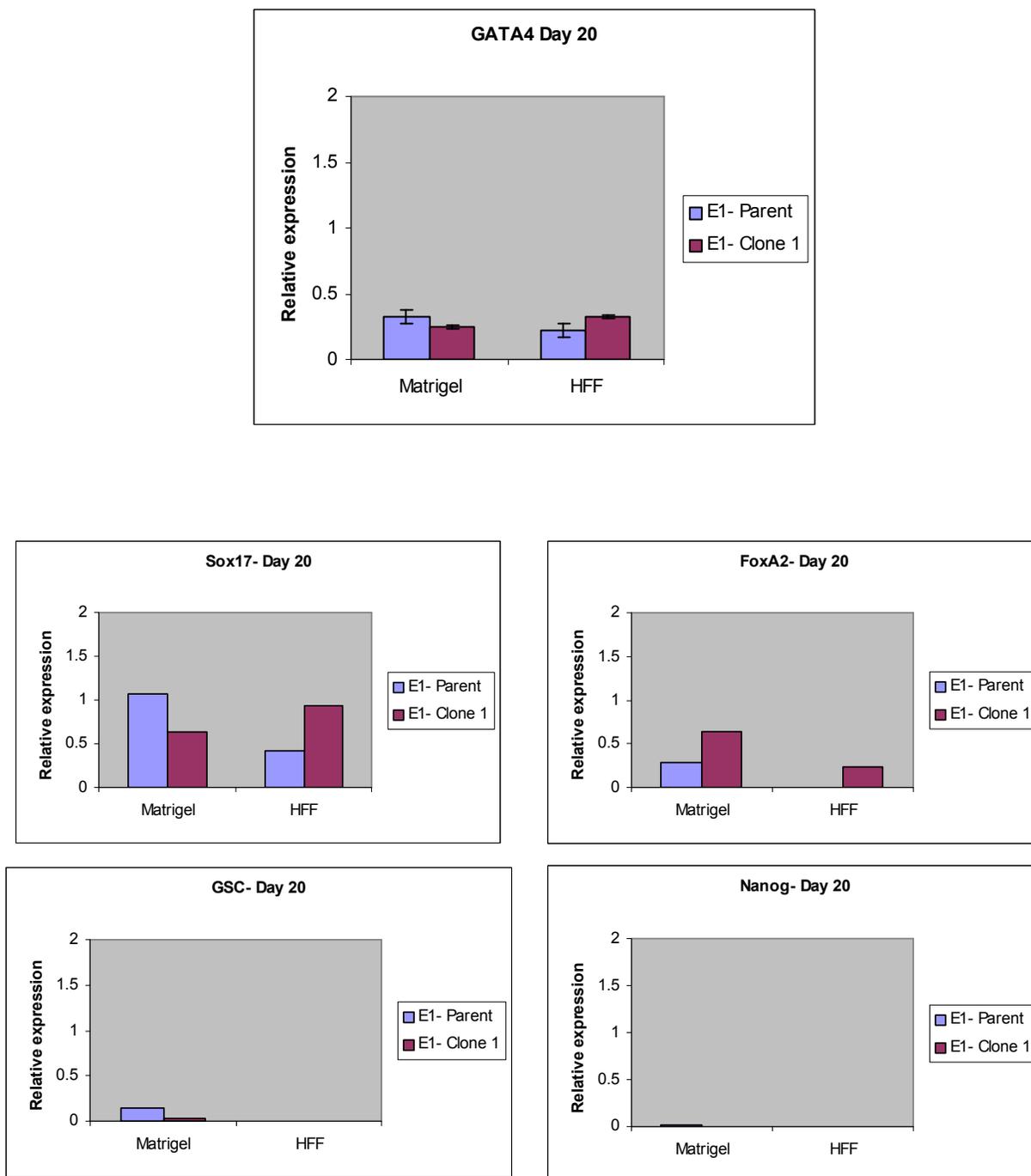
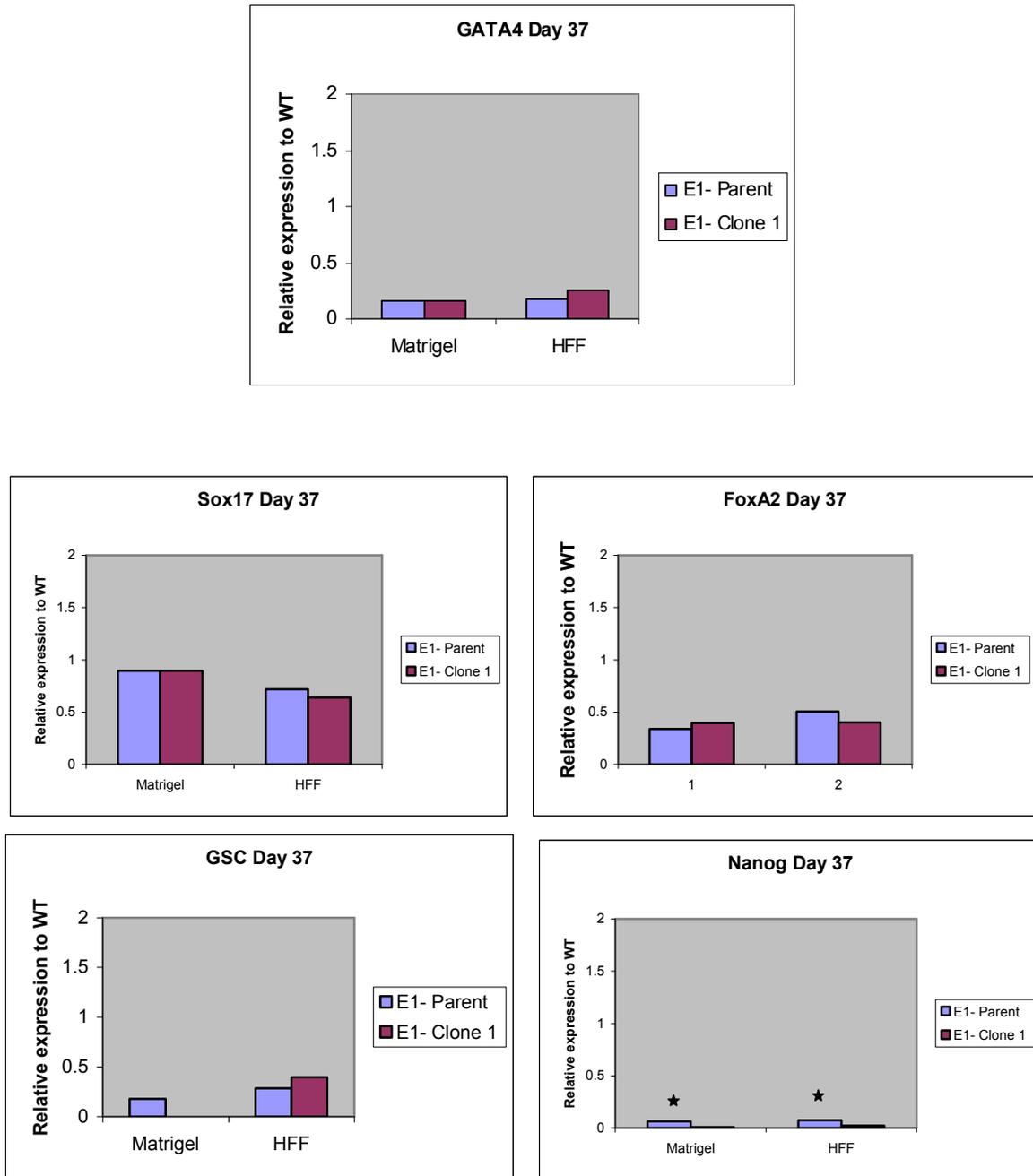


Figure 12B (cont'd)



DISCUSSION

Although hESC possess enormous potential in regenerative cellular therapies, many aspects regarding directed differentiation into specific cell types remain uncertain. Elucidating the molecular details of the pathway from hESC to insulin-producing β cells could facilitate the development of a renewable source of insulin-producing cells suitable for use in the treatment of Type-1 diabetes.

The present study focused on demonstrating the overexpression of *GATA4* in hESC as a method of genetic manipulation to directly differentiate hESC towards definitive endoderm, an essential pathway for β cell derivation. In order to investigate ectopic expression of *GATA4* in hESC, an expression vector was first constructed with the human *GATA4* gene sequence insert by cloning, and a cell line of zeocin-resistant HFF was created to serve as a feeder layer on which transfected hESC could be selected. The stable transfection of hESC with the pcDNA4-*GATA4* plasmid could be achieved and preliminary observations for morphology and gene expression were made. Further experiments are required to support these observations, and will provide a greater understanding of the effect of *GATA4* and other transcription factors on hESC differentiation to definitive endoderm and insulin producing cells.

It has been established that *GATA4* transcription factor, along with other genes such as *GATA6*, *FoxA2* and *Sox17*, function to regulate mesoderm and endoderm formation during embryogenesis²⁸. In the present study, *GATA4* sequence was isolated based on Activin A treated hESC. This was in concordance with studies conducted by D'Amour et al.^{13, 21} demonstrating the differentiation of hESCs into definitive endoderm and expression of gene markers including *Sox17*, *FoxA2* and *GSC*.

Notably, wild-type E1-hESC grown in standard culture conditions in the present study exhibited endogenous expression of *GATA4* and other endodermal markers including *Sox17* and *FoxA2*, as well as expression of *Nanog*, a marker of pluripotency. This suggests that a sub-population of the hESC were differentiating towards definitive

endoderm. A possible explanation could be that some hESC lines are more prone than others to undergo spontaneous differentiation in culture²⁹.

The *GATA4* gene was successfully cloned into a vector for transfection into the hESC, and DNA sequencing revealed the presence of two point mutations when compared with the human *GATA4* gene sequence obtained from NCBI (NM002052). In each mutation, there was a substitution of a single base nucleotide with another nucleotide (**Fig. 9**). Unexpectedly, both mutations were exactly the same in the two clones (G1 and G2) that were analysed. This suggests the possibility of the source being the positive sample from which the gene sequence was deciphered (Activin A Day 5 treated hESC), rather than mutation events occurring during the cloning process. Errors occurring during PCR amplification was less likely as a taq designed for cloning was used; Platinum® *Taq* DNA Polymerase High Fidelity (Invitrogen) which contained a proof-reading (3'→5' exonuclease activity) enzyme *Pyrococcus* species *GB-D* polymerase.

In the plasmid clone (G1) that was used for transfection, one point mutation at 1220bp was translated to a mutation in the protein, where a P was substituted with a Q. The mutation at 1138bp did not translate to a change in the protein and was therefore a silent mutation. The extent to which the protein function was affected could be an issue of consideration in the analysis of gene expression results. Protein expression assays of the transfected cells was not carried out in the present study due to time restraints, though further investigations would be beneficial for better understanding the effects at the protein level.

A line of zeocin-resistant HFF was created for use as a feeder layer on which the transfected hESC could be selected by zeocin-resistance. As such, the HFF were stably transfected with a vector conferring zeocin resistance. Stable expression can be influenced by two factors: the transfection method used and the vector being integrated into the cell³⁰. It was observed that HFF transfected with the pIRESbleo3 vector did not survive or proliferate after transfection. The positive control with an eGFP vector demonstrated an effective transfection method, as green fluorescent cells were detected a

few days after transfection. A feasible explanation would therefore implicate the vector, which contains an IRES (Internal Ribosome Entry Site) in front of the zeocin-resistance gene. Most likely, the IRES site was unsuccessful with the HFF cell line³¹ and failed to provide zeocin-resistance to the transfected cells.

The subsequent stable transfection of the HFFs with the pTracer plasmid (Invitrogen) was successful in producing positive zeocin-resistant clones which were pooled and propagated in culture. Creating a line of zeocin-resistant HFF shown to be capable of supporting the growth and selection of transfected hESC was a key achievement in this study. Such cells would also have diverse applications in future investigations.

A major restricting factor in the current study was the lack of time and this was hampered by the slow growth rate of the transgenic zeocin-resistant fibroblasts, which took much longer than normal HFF to propagate. This prompted the search for an alternative culture method for selection of transfected hESC before there was sufficient numbers of HFF to create feeder plates. A feeder-free system with Matrigel did not support the undifferentiated growth of hESC for more than a couple of passages, with the appearance of differentiated cells around the periphery of colonies. These findings are consistent with Ullmann et al. who described the presence of mesenchymal-like differentiated cells associated with feeder-free culture³². The flaws in the present study's protocol could involve the conditioned media used, since it has been shown that there is substantial variation between different human feeder types in their capacity to support the self-renewal of hESC³³. Further investigation to devise an optimal feeder-free protocol for hESC would be desirable.

Compared with widely used cell lines, hESCs are known to be difficult to transfect with general methods such as microinjection, electroporation and liposome-based reagents, yielding low transduction efficiencies³⁴. Research has shown that nucleofection of hESCs yielded a three-fold greater transfection efficiency than electroporation³⁵, but was also associated with a high rate of cell death. The nature and characteristics of hESC vary from one cell line to the next³⁶, rendering some more susceptible to cell damage and unintended toxic effects of transfection. The damaging effect on the cells of

nucleoporation was experienced in the present study, reflected by the difficulties that arose during zeocin selection process. HESC and even HFF that were nucleofected took a prolonged period of time (3-5 weeks) to recover and begin to proliferate in culture.

It could be speculated, however, that the presence of zeocin itself in the media for selection of the positive clones may have had an undesirable effect on the cells and their rate of growth. A study by Trastoy et al found that zeocin, added in the culture medium of human ovarian cells stably expressing the *Sh ble* resistance gene, was responsible for the formation of DNA strand breaks in the recombinant cells³⁷. This suggests that the zeocin is not completely detoxified and is still able to cleave DNA, despite the stable expression of the *Sh ble* gene in the recombinant clones.

Following transfection of E1-hESC with pcDNA4-*GATA4* recombinant plasmid, gene expression of *GATA4* was analysed by RT-PCR at two time points (day 20 & day 37) in transfected hESC. Levels of *GATA4* expression were compared with expression levels in non-transfected WT E1-hESC by semi-quantitative analysis. Surprisingly, results obtained from both time points clearly showed weaker *GATA4* expression in the transfected hESC, at around one-fifth the amount expressed in WT cells. This observation could suggest the cells had not been successfully transfected with the vector. However, the observed survival and proliferation of some colonies in the presence of zeocin suggests that these cells were zeocin-resistant and thus contained the vector. It is possible that errors that occurred during the integration of the *GATA4* DNA into the target cell's chromosome would alter the normal mechanism of gene expression and cascade³⁰. The findings here were not consistent with previous studies in mouse ESCs demonstrating induction of endogenous *GATA4* expression in transfected cells¹¹. Repetition of the transfection experiments conducted would be required to confirm the integrity of the results produced in the present study and further investigations of protein expression via immunohistochemistry would prove beneficial.

Investigation of functional *GATA4* protein activity was examined by assaying levels of gene expression in markers of definitive endoderm. After both time points, the expression of *Sox17* and *FoxA2* in transfected hESC were similar to, or less than the levels of expression in WT cells. Hence the transfected hESC did not appear to be differentiating

toward definitive endoderm lineage. Of interest, strong expression of *Nanog* was detected in WT hESC but none or very faint expression was observed in transfected hESC, which was indicative of their loss of pluripotency. RT-PCR with *Sox17*, *FoxA2*, *GSC* and *Nanog* on the samples could not be repeated due to time restrictions, hence repeated analyses would improve reliability of the results.

Further investigation involving mRNA analyses and immunohistochemistry with markers of the three embryonic lineages, would also provide a more comprehensive depiction as to which lineage the transfected hESC were differentiating.

In summary, although the main aim of reaching definitive endoderm was not achieved, the results obtained here describe the approaches involved in working towards a model of ectopic gene expression in hESC. As hESC are capable of undergoing some of the differentiation characteristic of early human development, these approaches for monitoring and modifying hESC gene expression provide a useful model for human functional genomics. This in turn, can help to reveal the molecular pathways of differentiation from hESC to definitive endoderm and β cell, and facilitate the development of a regenerative source of insulin-producing cells for use in cellular replacement therapies.

REFERENCES

1. Blyszczuk P, Asbrand C, Rozzo A, et al. Embryonic stem cells differentiate into insulin-producing cells without selection of nestin-expressing cells. *International Journal of Developmental Biology*. 2004;48:1095-1104.
2. Hardikar AA, Lees, J. G., Sidhu, K. S., Colvin, E., Tuch, B. E. Stem-Cell Therapy for Diabetes Cure: How Close Are We? *Current Stem Cell Research & Therapy*. 2006.
3. Fujikawa T, Oh, S., Pi, L., Hatch, H., Shupe, T., Peterson, B. Teratoma Formation Leads to Failure of Treatment for Type I Diabetes Using Embryonic Stem Cell-Derived Insulin-Producing Cells. *American Journal of Pathology*. 2005;166.
4. Shapiro AM, Lakey, J.R., Ryan, E.A., Korbitt, G.S., Toth, E., Warnock, G.L., Kneteman, N.M., Rajotte, R.V. Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoidfree immunosuppressive regimen. *New England Journal of Medicine*. 2000;343:230-238.
5. Thomson JA, Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science*. 1998;282:1145-1147.
6. Larsen W. *Human Embryology*: Churchill Livingstone; 2001.
7. Tam P, Kanai-Azuma, M., Kanai, Y. Early endoderm development in vertebrates: lineage differentiation and morphogenetic function. *Current Opinion in Genetics & Development*. 2003;13:393-400.
8. Ober E, Field, H., Stainier D. From endoderm formation to liver and pancreas development in zebrafish. *Mech Development*. 2003;120:5-18.
9. Stainier D. A glimpse into the molecular entrails of endoderm formation. *Genes & Development*. 2002;16:893-907.
10. Molkenkin JD, Lin, Q., Duncan, S.A., and Olson, E.N. Requirement of the transcriptional factor GATA-4 for heart tube formation and ventral morphogenesis. *Genes & Development*. 1997;11:1061-1072.
11. Fujikura J, Yamato, E., Yonemura, S., Kiminori, H., Masui, S., Nakao, K., Miyazaki, J., Niwa, H. Differentiation of embryonic stem cells is induced by GATA factors. *Genes & Development*. 2002;16:784-789.
12. Futaki S, Hayashi, Y., Emoto, T., Weber, C. N., Sekiguchi, K. Sox7 Plays Crucial Roles in Parietal Endoderm Differentiation in F9 Embryonal Carcinoma Cells through Regulating Gata-4 and Gata-6 Expression. *Molecular and Cellular Biology*. 2004;24:10492-10503.
13. D'Amour K, Agulnick, A. D., Eliazer, S., Kelly, O. G., Kroon, E., Baetge, E. E. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nature Biotechnology*. 2005.
14. Segev H, Fishman, B., Ziskind, A., Shulman, M., Itskovitz-eldor, J. Differentiation of Human Embryonic Stem Cells into Insulin-Producing Clusters. *Stem Cells*. 2004;22:265-274.
15. Shi Y, Hou, L., Tang, F., Jiang, W., Wang, P., Ding, M., Deng, H. Inducing Embryonic Stem Cells to Differentiate into Pancreatic β Cells by a Novel Three-

- Step Approach with Activin A and All-Trans Retinoic Acid. *Stem Cells*. 2005;23:656-662.
16. Xu X. KB, Forgianni A., Jing P., Jacobson L., Browning V., Treff N., Odorico J. Endoderm and pancreatic islet lineage differentiation from human embryonic stem cells. *Cloning & Stem Cells*. 2006;8:96-107.
 17. Baharvand H, Jafary, H., Massumi, M., Ashtiani, S. K. Generation of insulin-secreting cells from human embryonic stem cells. *Development Growth & Differentiation*. 2006;48:323-332.
 18. Menendez P, Wang, L., Bhatia, M. Genetic Manipulation of Human Embryonic Stem Cells: A System to Study Early Human Development and Potential Therapeutic Applications. *Current Gene Therapy*. 2005;5:375-385.
 19. Lumelsky N, Blondel, O., Laeng, P. et al. Differentiation of embryonic stem cells to insulin-secreting structures similar to pancreatic islets. *Science*. 2001;292:1389-1394.
 20. Conley BJ, Young, J. C., Trounson, A. O., Mollard, R. Review- Derivation, propagation and differentiation of human embryonic stem cells. *The International Journal of Biochemistry & Cell Biology*. 2004;36:555-567.
 21. D'Amour KAB, A.G. Eliazer, S. et al. . Production of pancreatic hormone – expressing endocrine cells from human embryonic stem cells. *Nature Biotechnology*. 2006.
 22. Stainier D. No Stem Cell Is an Islet (Yet). *New England Journal of Medicine*. 2006;354.
 23. Blyszczuk P, et al. Expression of Pax4 in embryonic stem cells promotes differentiation of nestin-positive progenitor and insulin-producing cells. *Proc Natl Acad Sci USA*. 2003;100:998.
 24. Lavon N, Yanuka, O., Benvenisty, N. The Effect of Overexpression of Pdx1 and Foxa2 on the Differentiation of Human Embryonic Stem Cells into Pancreatic Cells. *Stem Cells*. 2006;24:1923-1930.
 25. Ryan EA, Lakey, J.R., Rajotte, R.V., Korbitt, G.S., Kin, T., Imes, S., et al. Clinical outcomes and insulin secretion after islet transplantation with the Edmonton protocol. *Diabetes*. 2001;50:710-719.
 26. Eiges R, Schuldiner, M., Drukker, M., Yanuka, O., Itskovitz-Eldor, J., & Benvenisty, N. Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. *Current Biology*. 2001;11:514-518.
 27. Sidhu KS. Growing Human Embryonic Stem Cells: A Class Manual. *Diabetes Transplant Unit, Sydney*; 2005:pp1-45.
 28. Kubo AS, K. Shannon, J.M. et al. . Development of definitive endoderm from embryonic stem cells in culture. . *Development* 2004;131:1651-1662.
 29. Bodnar MSM, J.J. Rodriguez, R.T. et al. . Propagation and maintenance of undifferentiated human embryonic stem cells *Stem Cells Dev*. 2004;13:243-253.
 30. Murnane JP, Yezzi, M. J., Young, B. R. Recombination events during integration of transfected DNA into normal human cells. *Nucleic Acids Res*. 1990;18:2733-2738.
 31. Mizuguchi H, Xu, Z., Ishii-Watabe, A., Uchida, E., Hayakawa, T. IRES-Dependent Second Gene Expression Is Significantly Lower Than Cap-Dependent

- First Gene Expression in a Bicistronic Vector. *Molecular Therapy*. 2000;1:376-382.
32. Ullmann Uea. Epithelial-mesenchymal transition process in human embryonic stem cells cultured in feeder-free conditions. . *Molecular Human Reproduction* 2007;13:21-32.
 33. Richards M TS, Fong CY, Biswas A, Chan WK, Bongso A. . Comparative evaluation of various human feeders for prolonged undifferentiated growth of human embryonic stem cells. *Stem Cells*. 2003;21:546-556.
 34. Ren C, et al. Establishment of Human Embryonic Stem Cell Line Stably Expressing Epstein-Barr Virus-Encoded Nuclear Antigen. *Acta Biochimica et Biophysica Sinica*. 2005;37:68–73.
 35. Lakshimipathy U, Pelacho, B., Sudo, K., Linehan, J. L., Coucouvanis, E., Kaufman, D. S., Verfaillie, C. Efficient transfection of embryonic and adult stem cells. *Stem Cells*. 2004;22:531-543.
 36. Mikkola M, Olsson, C., Palgi, J., et al. . Distinct differentiation characteristics of individual human embryonic stem cell lines. *Dev Biol*. 2006;6.
 37. Trastoy M, Defais, M., Larminat, F. Resistance to the anitbiotic zeocin by stable expression of the SH BLE Gene does not fully suppress zeocin-induced DNA cleavage in human cells. *Mutagenesis*. 2005;2:111-114.

SUPPLEMENTARY DATA: TISSUE CULTURE MEDIA*Figure 1: Preparation of Human Fetal Fibroblast Media (FDMEM)*

To make 500mL:

DMEM – high glucose (GIBCO BRL, Invitrogen)	442.5mL
Fetal Bovine Serum (FBS) (GIBCO BRL, Invitrogen)	50mL
200mM (100x) L-glutamine (GIBCO BRL, Invitrogen)	5mL
5000U/mL Penicillin/Streptomycin (GIBCO BRL, Invitrogen)	2.5mL

1. Aliquot all reagents into a T75 tissue culture flask
2. Remove 57.5mL DMEM from 500mL bottle
2. Filter mixed reagents through a 0.22 µm filter (Millipore) directly into medium
3. Store medium in the dark at 4°C

Note: Medium has shelf life of 4 weeks at 4°C

Figure 2: Preparation of Knock Out Serum Replacement (KOSR) Medium (20%)

To make 500mL:

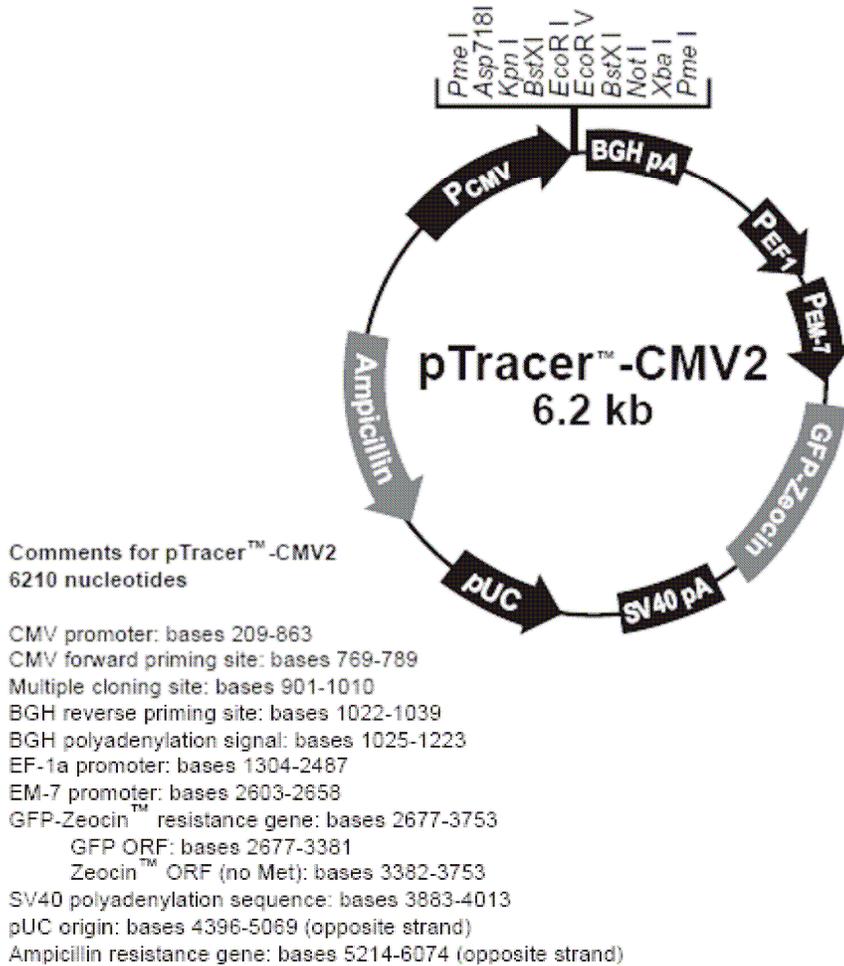
Knock Out DMEM– high glucose (KODMEM) (GIBCO BRL, Invitrogen)	380.6mL
Knock Out Serum Replacer (GIBCO BRL, Invitrogen)	100mL
10mM (100x) Non essential amino acids (GIBCO BRL, Invitrogen)	5mL
100x Insulin-Transferrin Selenium (GIBCO BRL, Invitrogen)	5mL
200mM (100x) L-glutamine (GIBCO BRL, Invitrogen)	5mL
5000U/mL Penicillin/Streptomycin (GIBCO BRL, Invitrogen)	2.5mL
55mM (1000x) buffered 2-mercaptoethanol (GIBCO BRL, Invitrogen)	0.9mL
Human basic fibroblast growth factor (bFGF) (Invitrogen)	1mL

1. Aliquot all reagents into a T75 tissue culture flask
2. Remove 119.4mL KODMEM from 500mL bottle
3. Filter mixed reagents through a 0.22 µm filter (Millipore) directly into medium
4. Store medium in the dark at 4°C
5. bFGF is added to a final concentration of 4ng/mL

Note: Medium has shelf life of 4 weeks at 4°C

SUPPLEMENTARY DATA: VECTOR CONSTRUCTION

Figure 3: pTracer – CMV2 vector map



(<http://www.invitrogen.com>)

Figure 4: Complete DNA sequence of *GATA4* (NM002502) (top row) aligned with Clone 1 sequence (second row)

```

Matches():1326
Mismatches(#):3
Caps():0
Unattempted(.):0

      *      *      *      *      *      *      *      *      *      *
1  ATGTATCAGAGCTTGGCCATGGCCGCCAACACAGGGCCGCCCGGCTGCCTACGAGGCGGGCGGCCCGCCGCTTCATGCACGGCCGGGCGCCGCT 100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1  ATGTATCAGAGCTTGGCCATGGCCGCCAACACAGGGCCGCCCGGCTGCCTACGAGGCGGGCGGCCCGCCGCTTCATGCACGGCCGGGCGCCGCT 100
      *      *      *      *      *      *      *      *      *      *

101 CCTCGCCACTCTACGTGCCACACCCGGGTGCCCTCCTCCGTGCTGGGCTGCTCCTACTCCAGGGCGGAGGGCGGGCTCTGCGTCCGGAGGGCGCTC 200
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
101 CCTCGCCACTCTACGTGCCACACCCGGGTGCCCTCCTCCGTGCTGGGCTGCTCCTACTCCAGGGCGGAGGGCGGGCTCTGCGTCCGGAGGGCGCTC 200
      *      *      *      *      *      *      *      *      *      *

201 GGGCGGCAGCTCCGGTGGGCGCGCTCTGGTGGGGGCCCGGGACCCAGCAGGGCAGCCGGGATGGAGCCAGGGGGAGCCGAGCGGCTTACACC 300
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
201 GGGCGGCAGCTCCGGTGGGCGCGCTCTGGTGGGGGCCCGGGACCCAGCAGGGCAGCCGGGATGGAGCCAGGGGGAGCCGAGCGGCTTACACC 300
      *      *      *      *      *      *      *      *      *      *

301 CCGCCCGCGGTGTCCGCGCGCTTCTCCTTCCGGGGACACCGGGTCCCTGGCGGCCCGCGCCGCTGCGCGGGCCGGGAAGCTGCGGCTACAGCA 400
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
301 CCGCCCGCGGTGTCCGCGCGCTTCTCCTTCCGGGGACACCGGGTCCCTGGCGGCCCGCGCCGCTGCGCGGGCCGGGAAGCTGCGGCTACAGCA 400
      *      *      *      *      *      *      *      *      *      *

401 GTGGCGGCAGCGCGGGTGTGGGCTGGCGGGCCGGAGCAGTACGGCGCGCGGGCTTGGCGGGCTCCTACTCCAGCCCTACCCGGCTTACATGGC 500
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
401 GTGGCGGCAGCGCGGGTGTGGGCTGGCGGGCCGGAGCAGTACGGCGCGCGGGCTTGGCGGGCTCCTACTCCAGCCCTACCCGGCTTACATGGC 500
      *      *      *      *      *      *      *      *      *      *

501 CGACGTGGCGCGCTCTGGGCGCAGCCGCGCGCTCCGCGGGCCCTTCGACAGCCCGGTCTGCACAGCCTGCCGGCGGGGCAACCCGGCGCGC 600
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
501 CGACGTGGCGCGCTCTGGGCGCAGCCGCGCGCTCCGCGGGCCCTTCGACAGCCCGGTCTGCACAGCCTGCCGGCGGGGCAACCCGGCGCGC 600
      *      *      *      *      *      *      *      *      *      *

601 CGACACCCCAATCTCGATATCTTTGACGACTTCTCAGAAGCCAGAGACTGTCTCAACTGTGGGCTATCTCCACCCCGCTCTGGAGGCGAGATGGGACGG 700
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
601 CGACACCCCAATCTCGATATCTTTGACGACTTCTCAGAAGCCAGAGACTGTCTCAACTGTGGGCTATCTCCACCCCGCTCTGGAGGCGAGATGGGACGG 700
      *      *      *      *      *      *      *      *      *      *

701 GTCACATCTGTGCAACGGCTGCGGCTCTACCACAAGATCAACGGCATCAACCGCGGCTCATCAAGCCTCAGCGCGGCTGTCCGCGCTCCCGCGGACT 800
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
701 GTCACATCTGTGCAACGGCTGCGGCTCTACCACAAGATCAACGGCATCAACCGCGGCTCATCAAGCCTCAGCGCGGCTGTCCGCGCTCCCGCGGACT 800
      *      *      *      *      *      *      *      *      *      *

801 GGGCTCTCCTGTGCAACTGCCAGACCACCACACCGCTGTGGCGGCCCAATCGGGAGGGGAGCCTGTGTGCAATGCCTGCGGCTCTACATGAAG 900
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
801 GGGCTCTCCTGTGCAACTGCCAGACCACCACCACCGCTGTGGCGGCCCAATCGGGAGGGGAGCCTGTGTGCAATGCCTGCGGCTCTACATGAAG 900
      *      *      *      *      *      *      *      *      *      *

901 CTCACGGGCTCCCGAGGCTTCCCAATCGGAAAGAGGGGATCCAAACCAGAAAACGAAAGCCCAAGAACTGAATAAATCTAAGACACCAGCAGCTC 1000
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
901 CTCACGGGCTCCCGAGGCTTCCCAATCGGAAAGAGGGGATCCAAACCAGAAAACGAAAGCCCAAGAACTGAATAAATCTAAGACACCAGCAGCTC 1000
      *      *      *      *      *      *      *      *      *      *

1001 CTTACGGCACTGAGAGCCTTCTCCCGCCAGCGTCTCCAGCAACTCCAGCAACGCCACCACCAGCAGCAGGAGATGCGTCCCATCAAGACCGA 1100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1001 CTTACGGCACTGAGAGCCTTCTCCCGCCAGCGTCTCCAGCAACTCCAGCAACGCCACCACCAGCAGCAGGAGATGCGTCCCATCAAGACCGA 1100
      *      *      *      *      *      *      *      *      *      *

1101 GCCTGGCTGTCTACTACGGGCACAGCAGCTCCGTGTCACAGAGCTTCTCAGTCAAGTGGATGTCTGGCCATGGGCGCTCCATCCACCCCTGTCTC 1200
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1101 GCCTGGCTGTCTACTACGGGCACAGCAGCTCCGTGTCACAGAGCTTCTCAGTCAAGTGGATGTCTGGCCATGGGCGCTCCATCCACCCCTGTCTC 1200
      *      *      *      *      *      *      *      *      *      *

1201 TCGGCCCTGAAGCTCTCCCAACAAGGCTATGCGTCTCCCGTCAGCCAGTCTCCACAGACCAGCTCCAAGCAGGACTCTTGGAAACAGCCTGTCTTGGCCG 1300
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1201 TCGGCCCTGAAGCTCTCCCAACAAGGCTATGCGTCTCCCGTCAGCCAGTCTCCACAGACCAGCTCCAAGCAGGACTCTTGGAAACAGCCTGTCTTGGCCG 1300
      *      *      *      *      *      *      *      *      *      *

1301 ACAGTCACGGGGACATAATCACTGCGTAA 1329
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1301 ACAGTCACGGGGACATAATCACTGCGTAA 1329
      *      *

```

Figure 5: Complete DNA sequence of *GATA4* (NM002502) (top row) aligned with Clone 2 sequence (second row)

```

1 ATGTATCAGAGCTTGGCCATGGCCGCAACCACGGGGCCGCCCCCGGTGCTTACGAGGGGGGGCCCCGGCGCCTTCATGCACGGCCGGGGCCGCGCT 100
1 ATGTATCAGAGCTTGGCCATGGCCGCAACCACGGGGCCGCCCCCGGTGCTTACGAGGGGGGGCCCCGGCGCCTTCATGCACGGCCGGGGCCGCGCT 100
101 CCTCGCCAGTCTACGTGCCACACCGCGGGTGCCTCCTCCGTGCTGGGCTGTCTTACCTCCAGGGGGAGGGCGGGGCTCTGCGTCCGGAGGGCGCTC 200
101 CCTCGCCAGTCTACGTGCCACACCGCGGGTGCCTCCTCCGTGCTGGGCTGTCTTACCTCCAGGGGGAGGGCGGGGCTCTGCGTCCGGAGGGCGCTC 200
201 GGGCGGCAGTCCGGTGGGGCCGCTCTGCTGCGGGCCCGGACCCAGCAGGGCAGCCCGGGATGGAGCCAGCGGGAGCCAGCGAGCCGCTTACACC 300
201 GGGCGGCAGTCCGGTGGGGCCGCTCTGCTGCGGGCCCGGACCCAGCAGGGCAGCCCGGGATGGAGCCAGCGGGAGCCAGCGAGCCGCTTACACC 300
301 CGCCGCGCGTGTGCGCGCGCTTCTCCTTCCCGGGACACCGGGTCCCTGGGGCCCGCGCCCGCGCTGCCCGGGCCGGGAAGCTGCGGCCTACAGCA 400
301 CGCCGCGCGTGTGCGCGCGCTTCTCCTTCCCGGGACACCGGGTCCCTGGGGCCCGCGCCCGCGCTGCCCGGGCCGGGAAGCTGCGGCCTACAGCA 400
401 GTGGCGCGGAGCGGGGCTGCGGGCTGGCGGGCCGCGAGCAGTACGGGGCCCGGGCTTCCGGGGCTCCTACTCCAGCCCTACCGGGCTTACATGGC 500
401 GTGGCGCGGAGCGGGGCTGCGGGCTGGCGGGCCGCGAGCAGTACGGGGCCCGGGCTTCCGGGGCTCCTACTCCAGCCCTACCGGGCTTACATGGC 500
501 CGACGTGGCGCGTCTGCGGGCCAGCCCGCCCGGCTTCCCGCGCCCTTCGACAGCCGGTCTGACAGCCCTGCCCGCCGGGCCAACCCGGCCGCG 600
501 CGACGTGGCGCGTCTGCGGGCCAGCCCGCCCGGCTTCCCGCGCCCTTCGACAGCCGGTCTGACAGCCCTGCCCGCCGGGCCAACCCGGCCGCG 600
601 CGACACCCCAATCTCGATATGTTTGACGACTTCTCAGAAAGCAGAGAGTGTGTAAGTGTGGGCTATGTCCACCCCGCTCGGAGCCGAGATGGGACGG 700
601 CGACACCCCAATCTCGATATGTTTGACGACTTCTCAGAAAGCAGAGAGTGTGTAAGTGTGGGCTATGTCCACCCCGCTCGGAGCCGAGATGGGACGG 700
701 GTCACTATCTGTGCAACCGCTGCGGCCCTTACCACAAGATGAACGGCATCAACCGCGGCTCATCAAGCCTCAGCGCCGGCTGTCCGCTCCCGCCGAGT 800
701 GTCACTATCTGTGCAACCGCTGCGGCCCTTACCACAAGATGAACGGCATCAACCGCGGCTCATCAAGCCTCAGCGCCGGCTGTCCGCTCCCGCCGAGT 800
801 GGGCCTCTCCTGTGCAACTGCCAGACCACCACCACCGCTGTGGCCGCAATCGCGAGGGCGAGCCTGTGTGCAATGCCTGCGCCCTTACATCAAG 900
801 GGGCCTCTCCTGTGCAACTGCCAGACCACCACCACCGCTGTGGCCGCAATCGCGAGGGCGAGCCTGTGTGCAATGCCTGCGCCCTTACATCAAG 900
901 CTCACGGGTCCTCCAGGCTCTTTCGAATGCGCAAAAGAGGGATCCAAACAGAAAACGGAAAGCCAAAGCACTGAATAAATCTAAACACACAGCAGCTC 1000
901 CTCACGGGTCCTCCAGGCTCTTTCGAATGCGCAAAAGAGGGATCCAAACAGAAAACGGAAAGCCAAAGCACTGAATAAATCTAAACACACAGCAGCTC 1000
1001 CTTCAGGCAGTGAGAGCCTTCTCCCGCCAGCGGTGCTTCCAGCAACTCCAGCAACGCCACCACCAGCAGCAGCGAGGAGATGCGTCCCATCAAGACGGA 1100
1001 CTTCAGGCAGTGAGAGCCTTCTCCCGCCAGCGGTGCTTCCAGCAACTCCAGCAACGCCACCACCAGCAGCAGCGAGGAGATGCGTCCCATCAAGACGGA 1100
1101 GCCTGGCCTGTCTACTACTACGGGCACAGCAGCTCCGTGTCACAGAGTCTCAGTCAAGTGGATGTCTGGCCATGGGCCCTCCATCCACCTGTCTC 1200
1101 GCCTGGCCTGTCTACTACTACGGGCACAGCAGCTCCATGTCACAGAGTCTCAGTCAAGTGGATGTCTGGCCATGGGCCCTCCATCCACCTGTCTC 1200
1201 TCGGCCCTGAAGCTCTCCCAAGGCTATGCGTCTCCCGTACGCCAGTCTCCACAGACAGCTCCAAAGCAGGACTCTTGGAAACAGCCTGGTCTTGGCCC 1300
1201 TCGGCCCTGAAGCTCTCCCAAGGCTATGCGTCTCCCGTACGCCAGTCTCCACAGACAGCTCCAAAGCAGGACTCTTGGAAACAGCCTGGTCTTGGCCC 1300
1301 ACAGTCACGGGGACATAATCACTCGGTA 1329

```