Development of human embryonic stem cell technologies : genetic manipulation and stem cell selection

Tan, Shen Mynn

2008

Tan, S. M. (2008). Development of human embryonic stem cell technologies : genetic manipulation and stem cell selection. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/38902>

<https://doi.org/10.32657/10356/38902>

DEVELOPMENT OF HUMAN EMBRYONIC STEM CELL TECHNOLOGIES: GENETIC MANIPULATION AND STEM CELL SELECTION

Tan Shen Mynn

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

2008

PREFACE

In 2003, human embryonic stem cell (hESC) research was just emerging from its infancy, with the first cell lines isolated barely five years ago. As groups all over the world jump-started the field, many protocols were developed to isolate, culture, characterize, differentiate and manipulate these pluripotent cells. Four years later, we know a lot more about these cells, but are only marginally closer to bringing the promise of cell therapy from bench to bedside. This work is dedicated to all adventurous, forward-thinking and tireless hESC researchers.

This thesis is the final work of my Ph.D. study in the School of Biological Sciences, Nanyang Technological University (NTU), Singapore. It is a record of my four year experience with hESC research, spanning from laboratory and cell culture setup to culmination of the two projects contained in this thesis, from 2003 to 2007. This study has been financed by both NTU and A*STAR.

The thesis consists of three main parts. The first describes my experience in optimizing human embryonic stem cell culture, characterization and transfection. The following two parts describe in detail the two main projects undertaken during my stay in NTU. In the first project, I demonstrate the utility of genetic modification systems in hESCs and the successful transfer of the λ integrase site-specific recombination system to these cells. In the second project, I establish the importance and sensitivity of UTF1 as a pluripotency

marker, and show that a *UTF1*-driven antibiotic resistance gene insert in hESCs can be used to select for homogeneously pluripotent hESC cultures.

This thesis consists of work done exclusively by myself, save for three parts: construction of plasmid *pTZ-UTF1-EGFP-M1mut* (C.-T. Lee), help with the Southern blot analysis (R.J. Zhou), and construction of hESC line EPN14 (K. Neef). Both these projects have been published separately in peer-reviewed journals and presented in international conferences:

Shen Mynn Tan, Siew Tein Wang, Hannes Hentze and Peter Dröge. 2007. A *UTF1*-based selection system for stable homogeneously pluripotent human embryonic stem cell cultures, *Nucleic Acids Research* 2007 35(18):e118.

Shen Mynn Tan and Peter Dröge. 2005. Comparative analysis of sequence-specific DNA recombination systems in human embryonic stem cells, *Stem Cells* 23:868-873.

Shen Mynn Tan and Peter Dröge. 2006. The pluripotency marker *UTF1* and maintenance of undifferentiated human embryonic stem cells, poster presentation at the 4th *International Society for Stem Cell Research Annual Meeting, Toronto*.

Shen Mynn Tan and Peter Dröge. 2005. Comparative analysis of sequence-specific DNA recombination in human embryonic stem cells, poster presentation at *Keystone Symposia: Stem Cells, Senescence and Cancer, Singapore*.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, A/P Peter Dröge for believing in me and giving me the opportunity to initiate my human embryonic stem cell adventure in NTU. His tremendous drive, intense results-oriented focus and excellent scientific stewardship have been a great source of inspiration throughout the four years, and will be far into the future.

I would also like to express my gratitude to my lab mates: Bao, Dennis, Ray, Li Ou, Feng Shu, Sabrina, Aygul, Helena, Cammy and in particular Klaus, for their engaging intellectual banter, wonderful company and friendship over these past years.

Finally I would like to thank my family who has supported me all these years, my band of brothers, who kept me sane throughout the entire course of my research, and especially Yvonne, my muse, confidant and soul mate.

ABSTRACT

Human embryonic stem cells, or hESCs, are pluripotent cells isolated from blastocyst-stage preimplantation embryos that have great potential in basic research, drug development and cell replacement therapy. However, advances in many facets of hESC research have to be attained before these goals can be reached. We describe here the development of two technologies that will aid in this regard: genetic manipulation and stem cell selection.

Controlled manipulation of hESC genomes through site-specific DNA recombination (SSR) systems may play a significant role in future hESC applications. However, very little is known about the functionality of these systems in hESCs. We show here that three SSR systems: mutant phage λ integrase, phage P1 Cre recombinase, and mutant $\gamma\delta$ resolvase displayed distinct activities on episomal recombination substrates. In addition, we demonstrate successful λ integrase-mediated integrative intramolecular recombination in both HeLa and hESC integrated genomic substrates. These proof-of-concept experiments should allow future improvements to the utility and efficiency of SSR technology, which can be used to excise resistance genes in transgenic hESCs or even for gene targeting and gene delivery in hESCs.

An understanding of the regulatory mechanisms responsible for pluripotency in hESCs is critical for both the maintenance of undifferentiated cells and the ability to direct differentiation. One such important pluripotency marker, undifferentiated transcription

factor 1 (UTF1), was initially identified in mouse embryonic stem cells, and is also expressed in human embryonic and adult stem cells. *UTF1* transcription ceases at the onset of differentiation, which clearly distinguishes it from less sensitive pluripotency markers, such as *Oct4*, *Sox2* or *Nanog*. We present here two transgenic “ZUN” hESC lines, each harboring a copy of the *UTF1* promoter/enhancer driving a resistance gene. ZUN cultures under selection pressure consist of highly homogeneous hESC populations, with a larger proportion of Oct4 and Sox2 positive cells. While ZUN cultures, like parental hESC cultures, retained the capacity to differentiate into progenitors of all three germ layers, they exhibited an increased refractoriness to various differentiation cues. Together with the small size of only 2.4 kb for the entire *UTF1* transgenic cassette, these features render our selection system a powerful novel tool for many stem cell applications and cell reprogramming strategies.

As additions to the hESC research toolkit, the development of these two systems represents significant advances in our ability to manipulate hESCs. Precise and efficient genomic manipulation by SSR systems can replace inefficient homologous recombination and unspecific viral vector technologies; the expansion of pluripotent ZUN cultures to bioreactor quantities can improve the yield for cell-replacement therapy. Further optimization and improvement to both of these technologies will aid to unlock the potential of hESC research, and can potentially be transferred to other related fields of research.

LIST OF ABBREVIATIONS

AP	Alkaline Phosphatase
att	Int Attachment Site
bFGF	Basic Fibroblast Growth Factor
BMP	Bone Morphogenetic Protein
BSA	Bovine Serum Albumin
ChIP	Chromatin Immunoprecipitation
CHO	Chinese Hamster Ovary
CMV	Cyto Megalo Virus
dFBS	Defined Fetal Bovine Serum
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DOPA	Dihydroxyphenylalanine
EC	Embryonic Carcinoma
EDTA	Ethylenediaminetetraacetic Acid
EGFP	Enhanced Green Fluorescent Protein
ESC	Embryonic Stem Cell
EST	Expressed Sequence Tag
FACS	Fluorescence-activated Cell Sorting
FBS	Fetal Bovine Serum
FSC	Forward Scatter
GPI	Glycosylphosphatidylinositol
GTP	Guanosine Triphosphate
H&E	Hematoxylin and Eosin (staining)
hESC	Human Embryonic Stem Cells
HLA	Human Leucocyte Antigen
HSV	Herpes Simplex Virus
ICM	Inner Cell Mass
IHF	Integration Host Factor
Int	Phage λ -encoded Integrase
ITS	Insulin-Transferrin-Selenium

MEF	Mouse Embryonic Fibroblasts
mESC	Mouse Embryonic Stem Cells
MoMLV	Moloney Murine Leukemia Virus
MPSS	Massively Parallel Signature Sequencing
NEAA	Non-Essential Amino Acids
Neo	Neomycin Antibiotic Resistance
NLS	Nuclear Localization Signal
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
POU	Pit-Oct-Unc
qRT-PCR	Quantitative Real-Time PCR
RA	Retinoic Acid
rAAV	Recombinant Adeno-associated Virus
RE	Restriction Enzyme
RNA	Ribonucleic Acid
RNAi	RNA Interference
RT-PCR	Reverse Transcription PCR
SAGE	Serial Analysis of Gene Expression
SCID	Severe Combined Immunodeficiency Disease
scIHF	Single Chain IHF
SCNT	Somatic Cell Nuclear Transfer
siNT	Non-targeting siRNA
siRNA	Short Interfering RNA
SR	Serum Replacement
SSC	Side Scatter
SSEA	Stage Specific Embryonic Antigens
SSR	Site-Specific Recombination
TH	Tyrosine Hydroxylase
tk	Thymidine Kinase
TRA	Tumor Recognition Antigens
UTF1	Undifferentiated Transcription Factor 1
XIS	Excisionase
ZUN	<i>pTZ-UTF1-Neo</i> -transfected Stable hESC Line

CONTENTS

Preface	i
Acknowledgements	iii
Abstract	iv
List of abbreviations	vi
I. INTRODUCTION	1
I.1 Human embryonic stem cells	1
I.1.1 Overview	1
I.1.2 Derivation	1
I.1.3 Properties	3
I.1.4 Differentiation	4
I.1.5 Potential	7
I.2 Genetic manipulation of human embryonic stem cells	13
I.2.1 Transfection	14
I.2.2 Genomic integration	16
I.2.3 Serine and tyrosine recombinases	18
I.2.4 λ integrase	19

I.3	Pluripotency of human embryonic stem cells	22
I.3.1	Standards of pluripotency	22
I.3.2	Pluripotency marker <i>UTF1</i>	24
I.3.3	Spontaneous differentiation	27
I.3.4	Maintenance of pluripotency	28
II.	STATEMENT OF AIMS	32
III.	MATERIALS AND METHODS	34
III.1	Molecular biology	34
III.1.1	DNA manipulation	34
III.1.2	Bacterial work	35
III.1.3	Construction of plasmids	37
III.2	Cell culture	50
III.2.1	Routine cell line maintenance	50
III.2.2	Derivation of mouse embryonic fibroblasts	51
III.2.3	Human embryonic stem cell maintenance	52
III.2.4	Differentiation protocols	56

III.3	Transfection	59
III.3.1	Electroporation	59
III.3.2	Reagent-based protocols	60
III.3.3	Flow cytometry	64
III.3.4	Stable cell line generation	65
III.4	Characterization of cell lines	68
III.4.1	Pluripotency assays	68
III.4.2	Transgene validation	70
III.4.3	Karyotyping	73
III.4.4	Proliferation assay	73
III.4.5	Modulation of selection pressure	74
III.4.6	Transcript knockdown	74
III.4.7	Transcript analysis	74
IV.	RESULTS	78
IV.1	Characterization of human embryonic stem cells	78
IV.1.1	Morphology	78
IV.1.2	Mouse embryonic fibroblast quantification	79
IV.1.3	Pluripotency and karyotyping	80

IV.2	Transfection of human embryonic stem cells	84
IV.2.1	Transfection efficiency	84
IV.2.2	Generation of stable cell lines	88
IV.3	Site-specific recombination	89
IV.3.1	Episomal site specific recombination in hES2	89
IV.3.2	λ integrase-mediated genomic site-specific recombination	94
IV.4	Characterization of <i>UTF1</i>	100
IV.4.1	Analysis of transcript expression in various cell lines	100
IV.4.2	Transcript knockdown	103
IV.4.3	Transcript expression after RA/DMSO treatment	105
IV.5	Characterization of ZUN cell lines	108
IV.5.1	Genomic characterization	108
IV.5.2	Pluripotency and proliferation	111
IV.5.3	Modulation of selection pressure	116
IV.6	Comparison of ZUN differentiation	120
IV.6.1	Spontaneous differentiation	120
IV.6.2	Directed differentiation	123
IV.6.3	Long-term differentiation	130

V.	DISCUSSION	135
V.1	Human embryonic stem cell culture	135
V.2	Site-specific recombination in human embryonic stem cells	138
V.3	Characterization of <i>UTF1</i>	145
V.4	Characterization and differentiation of ZUN cell lines	147
VI.	CONCLUSION	151
VII.	REFERENCES	152
VIII.	APPENDICES	180
Appendix I	DNA sequences for constructed plasmids	180
Appendix II	Sequencing of <i>pTZ-UTF1-EGFP-M1mut</i>	196
Appendix III	Validation gels for qRT-PCR primers	197
Appendix IV	Amplification curves for qRT-PCR results	199
Appendix V	Calculations for qRT-PCR results	203
Appendix VI	List of publications	214

I. INTRODUCTION

I.1 Human embryonic stem cells

I.1.1 Overview

Human embryonic stem cells, or hESCs, are derived from the inner cell mass of the blastocyst-stage preimplantation embryo. These cells are immortal and retain the potential to differentiate into cells of all three germ layers, and are thus pluripotent in culture over many cell generations (Thomson *et al.*, 1998; Reubinoff *et al.*, 2000). This property allows research into many aspects of human development from investigations into molecular mechanisms of differentiation, to organogenesis and even fetal maturation. Eventually, it is hoped that hESCs can live up to clinical promises of cellular transplantation therapy for many degenerative disorders and diseases.

I.1.2 Derivation

In less than a decade, hundreds of hESC lines have been derived and characterized since the first report from the Thomson laboratory (Amit and Itskovitz-Eldor, 2006), although only 22 of these are currently registered with the National Institutes of Health and available to researchers utilizing public funds in the United States (<http://escr.nih.gov>). The main source of these lines was normal surplus embryos from *in vitro* fertilization clinics that were voluntarily donated. To isolate hESCs, these embryos were first cultured to the blastocyst stage in specialized medium and the zona pelliucida was digested with pronase. To remove the trophoblastic cells, the difference in the permeability of inner cell masses (ICMs) and trophoblastic cells to certain antibodies was exploited: these cells are killed when the blastocysts are exposed to antiserum alone and then transferred to guinea

pig complement (Solter and Knowles, 1975). Thus the ICMs were isolated from human blastocysts via exposure to anti-human rabbit serum antibody and followed by exposure to guinea pig complement. These ICMs were subsequently plated onto inactivated mouse embryonic fibroblasts (MEFs) (Thomson *et al.*, 1998). ICM-derived outgrowths were selected for uniformity and undifferentiated morphology, and continually passaged onto fresh MEF layers via mechanical dissociation, dispase treatment (Thomson *et al.*, 1998; Reubinoff *et al.*, 2000), or trypsin digestion (Cowan *et al.*, 2004) (**Fig. I.1**). It is important to note that the hESCs derived in this fashion have been exposed to many potentially contaminating non-human “xeno” agents, from rabbit antiserum during blastocyst immunosurgery to poorly defined media components and MEFs in routine culture conditions. This is widely regarded as unacceptable for potential clinical-grade hESC treatments, as these have to be xeno-free to safeguard patients from cross-species diseases and contaminants. In addition, most hESC lines consist of a heterogeneous cell population derived from an inner cell mass, as opposed to a clonal, homogeneous cell line that would be more predictable in downstream applications.

Recently, various studies have addressed these issues. Laser-assisted blastocyst dissection was used successfully in the mouse model to isolate embryonic stem cells (ESCs) (Tanaka *et al.*, 2006). Culture media for undifferentiated hESC growth have been refined to include only defined components with or without MEFs (Lu *et al.*, 2006; Ludwig *et al.*, 2005) and hESCs have also been derived without feeder cells (Klimanskaya *et al.*, 2005). Human fibroblasts of foreskin (Amit *et al.*, 2003), adult marrow (Cheng *et al.*, 2003) and autologous spontaneously-differentiated hESC (Stojkovic *et al.*, 2005) origin have also

been used successfully to culture pluripotent hESCs. As a means to address the heterogeneity of current lines, as well as to sidestep certain ethical concerns, hESCs have recently been derived from a single blastomere using a single-cell biopsy similar to that used in preimplantation genetic diagnosis (Klimanskaya *et al.*, 2006).

Figure I.1. Derivation of hESCs. Inner cell masses were extracted via immunosurgery from donated, frozen-thawed blastocysts and cultured on inactivated MEFs with human embryonic stem cell medium.

I.1.3 Properties

The properties and general features of hESCs are essentially similar to that of mouse ESCs (mESCs). It is vital that they possess all of the following characteristics before being considered *bona fide* embryonic stem cells (Pera *et al.*, 2000):

- Originate from a pluripotent cell population
- Maintain a normal karyotype and be stably diploid *in vitro*
- Immortal and can be cultured indefinitely in the primitive embryonic state

- Ability to differentiate spontaneously into cell types of all three germ layers in either:
 - Teratomas after grafting in immune-deficient mice
 - Suspended embryoid body cultures *in vitro*
 - Under appropriate culture conditions *in vitro*
- Ability to contribute to any cell type when allowed to colonize a host blastocyst

The above criteria are easily and routinely tested during hESC characterization, save for the last one, due to ethical opposition against human chimerical experiments. However, hESCs have been reported to contribute to the mouse ICM niche in hESCs aggregated with blastomere stage mouse embryos. Interestingly, some of these embryonic chimeras implant and develop in the uterus of pseudopregnant foster mice (James *et al.*, 2006).

I.1.4 Differentiation

Although hESCs retain the potential to differentiate into all cell types of the endoderm, mesoderm and ectoderm, harnessing this capability to achieve differentiation to a specific cell type of choice is no simple task. Thus far, hESCs have been successfully coaxed to differentiate into cardiomyocytes, hematopoietic progenitor cells, leucocytes, endothelial cells, neurons, glia, insulin-producing cells, and even germ cells (Hoffman and Carpenter, 2005). However, the efficiency of the various protocols varies, along with the utility of the differentiated cells. Essentially, success in this field depends on the knowledge of mechanisms controlling lineage decisions during embryogenesis, and subsequent recreation of these conditions *in vitro* (Spagnoli and Hemmati-Brivanlou, 2006).

Figure I.2. Differentiation of hESCs. Cell types from all three germ layers can potentially be differentiated from hESC cultures. Differentiation protocols include co-culture with somatic cells, culture on surface matrices, along with the addition of differentiation factors into culture media or embryoid body induction, applied either sequentially or concurrently.

This direction of differentiation can be achieved in three main ways: modification of defined media components and matrices, for example the addition of cell-permeable transcription factors (Heng *et al.*, 2004), enforced or selected expression of desired marker genes, or co-culture with differentiated somatic cells or their conditioned media (Heng *et al.*, 2004). All three approaches, separately or sequentially, have been used with varying degrees of success; alternative protocols also include an embryoid body-forming intermediate step (Fig. I.2). Target cell types of particular interest to researchers are those with the potential for straightforward cell-based therapy, where the restoration of loss of

function requires only a single cell population to be transplanted. The best examples of these cell types are: cardiomyocytes to improve myocardial dysfunction, dopaminergic neurons to replenish those lost to Parkinson's disease, and insulin-producing β -cells for the treatment of type 1 diabetes mellitus.

Cardiomyocyte differentiation received early attention from Geron Corporation, where efficiencies of up to 70% were reported by using spontaneous differentiation and Percoll density centrifugation, with embryoid body formation as an intermediate step. These cells showed the ability to spontaneously contract, and expressed GATA4, Nkx2.5, MEF2, cardiac α -myosin heavy chain and atrial natriuretic factor, among other cardiomyocyte markers (Xu *et al.*, 2002). Cells differentiated using this protocol were recently characterized and their expression pattern analyzed in detail using qualitative PCR, thus validating the cardiomyocytic phenotype observed (Segev *et al.*, 2005). Alternatively, a novel approach could be to utilize these hESC-derived cardiomyocytes to reprogram autologous stem cells to express cardiomyogenic function, thus reducing the risk to teratoma formation in patients (Heng *et al.*, 2004; Heng *et al.*, 2005).

Dopaminergic neurons have also been successfully generated with various protocols. Park and colleagues managed to relieve symptomatic motor behavior in a rat model of Parkinson's disease by the transplantation of hESCs with randomly inserted cDNAs coding for two important genes: tyrosine hydroxylase (TH), the enzyme responsible for catalyzing the conversion of L-tyrosine to dihydroxyphenylalanine (DOPA), a precursor to dopamine, and guanosine triphosphate (GTP) cyclohydrolase I, which is responsible

for the hydrolysis of GTP to form 7,8-dihydroneopterin 3'-triphosphate (Park *et al.*, 2003). In another study, co-culture with stromal feeder cells and sequential application of defined patterning molecules resulted in progressive expression of key factors, dopamine release and the presence of action potentials and TH-positive synaptic terminals (Perrier *et al.*, 2004). Chemically-defined human-derived media additives and substrata were also used successfully to differentiate dopaminergic neurons expressing related markers and characteristics in culture (Lacovitti *et al.*, 2006).

Much research effort has been put into the differentiation of β islet or islet-like cells for the treatment of diabetes. The over-expression of important pancreatic marker genes *Pdx1* and *Foxa2* were not sufficient to direct the differentiation of insulin-producing cells *in vitro* (Lavon *et al.*, 2006). Co-transplantation of spontaneously differentiated pancreas progenitor hESCs with cells from the embryonic mouse pancreas was a more successful method, generating insulin-positive β -cell clusters (Brolén *et al.*, 2005). Most recently, Novocell Inc. has developed a multi-stage variable growth factor protocol that mimics *in vivo* pancreatic organogenesis. This successfully differentiated hESCs to insulin-expressing cells that release C-peptide in response to multiple secretory stimuli, but unfortunately only minimally to glucose (D'Amour *et al.*, 2006).

I.1.5 Potential

Cell-based therapy

One of the most far-reaching potentials of hESCs is in cell-based therapy. With the potential to differentiate into virtually any mature cell type in the human body, the

promise of treatment for a multitude of diseases seems closer than ever before. Recently, the large scale production of embryonic red blood cells from hESCs was achieved, paving the way to production of hESC-derived mature red blood cells as an alternative source of transfusion (Olivier *et al.*, 2006). In addition, the induction of neurogenic progenitors capable of generating neurons, astrocytes and oligodendrocytes was achieved via a simple protocol that exposed adherent hESCs to fibroblast growth factor-2 (Benzing *et al.*, 2006). In the future, tissue or even organ regeneration from hESCs may be possible.

The terminal differentiation of hESCs depends on two crucial development-based intermediate phenotypes: lineage-specific cells and progenitor cells. Thus, the possibility remains that the isolation of intermediate progenitor cells would be an alternative to somatic cell-based therapy. *In vivo* differentiation of either a specific or a small subset of multipotent stem cells could be used in this therapeutic context, for example hematopoietic repopulating cells for *in vivo* clinical reconstitution (Menendez, Wang and Bhatia, 2005).

However, despite the successes in directing hESC differentiation, there is much to improve upon as none of the studies discussed above have yet reached an efficacy or purity required for advancement to clinical trials. Therefore, the risk of potential teratogenicity from transplanted, undifferentiated cells is an important concern. One possible fail-safe mechanism against this is to insert ablative genes driven by pluripotency marker gene elements into hESCs. The feasibility of this has been

demonstrated in experiments utilizing the herpes simplex virus thymidine kinase (HSV-tk) gene, which rendered derivative hESCs sensitive to ganciclovir and successfully ablated tumors caused by these cells in mice (Schuldiner *et al.*, 2003).

Additional applications

Besides direct clinical applications, hESCs have great utility in the investigation of human development and disease, drug development and even cancer research. These cells have been used to study the specification of the trophoblast lineage and the formation of the human placenta by acting as surrogate cells of the preimplantation embryo, or as starting material for trophoblast lineage differentiation (Golos *et al.*, 2006). They have also been used to model Lesch-Nyhan disease, where the *HPRT1* gene was mutated via homologous recombination (Urbach *et al.*, 2004). In drug development, hESCs and their differentiated derivatives can be involved in all stages, from high throughput screening at the lead development stage, to absorption, distribution, metabolism and excretion (ADME) analysis at the candidate stage and even target validation and toxicology at pre-clinical stages (McNeish, 2004).

Additionally, there is a growing acknowledgement of the similarities between cancer and stem cells as most of the cell signaling pathways essential for normal development, such as Notch, Wnt or Hedgehog, are involved in the progression of both cell types (Menendez *et al.*, 2005). Indeed, the embryonic stem cell marker Oct4 has been recently discovered to be expressed in breast carcinoma, primary germ cell tumors, and even bladder cancer (Ezeh *et al.*, 2005; Jung *et al.*, 2006; Atlasi *et al.*, 2007). These discoveries lend weight to

the stem cell theory of carcinogenesis and tumor malignancy. Unlike primary tumor cells, hESCs are easily available and relatively simple to culture. Therefore, *in vitro* differentiation of hESCs can offer an unprecedented tool for studying cancer biology and emergence of transformation events (Menendez *et al.*, 2005); undifferentiated hESCs themselves can be used to elucidate components of parallel pathways that lead to both pluripotency and tumorigenesis.

From bench to bedside

The hESC journey from bench to bedside is fraught with many hurdles. Initially, cells must be derived and kept in xeno-free conditions to remove concerns about cross-species transfer of pathogens (Stojkovic *et al.*, 2004); these cells should be as homogeneous and as undifferentiated as possible in culture conditions. The second challenge is to establish pure cultures of specific cell types, either by an efficient differentiation protocol, or by the selection of differentiated cells via cell sorting or ablation, or more likely a combination of both. These cells then have to be tested for physiological functionality, for example stimulated dopamine or insulin release. Thereafter, efficacy *in vivo* has to be analyzed, absence of tumor formation, infectious diseases and xeno-proteins confirmed, karyotypic and genetic stability assured, and immune rejection dealt with before proceeding to human clinical trials (Odorico *et al.*, 2001). After transplantation into patients, two additional issues that should be considered are the tracking of transplanted cells, and the possibility to ablate these cells if necessary (Hentze *et al.*, 2007).

Undoubtedly, the main issue hESCs researchers have to solve before progressing to the clinic would be immunologic rejection of transplanted cells. Some possible solutions to this are: the development of a “universal” hESC line by blocking HLA class I expression or the creation of HLA-genotyped hESC line banks, with an estimated 150 different genotypic donors required to match at least one HLA-DR for 84% of all patients. Alternatively, the generation of hematopoietic chimerism via the transplantation of hematopoietic progenitors with the hESC genotype could be used to avoid the rejection of these cells in recipients (Drukker and Benvenisty, 2004). However, even with the proposed combination of these approaches as a solution, it is still likely that immunosuppressive drugs would be required to prevent immunologic rejection because of HLA mismatches and minor histocompatibility antigens (Cabrera *et al.*, 2006).

The solution to this problem might come from another approach altogether: the use of isogenic embryonic or lineage-restricted stem cells for cell-replacement therapy. Patient-specific hESCs can potentially be isolated from blastocysts which were generated via somatic cell nuclear transfer (SCNT) technology. Termed therapeutic cloning, this has been achieved in mice and cattle, and is thought to be possible in humans as well (Yang *et al.*, 2007). Alternatively, reprogrammed somatic cells from patients can also be used to derive embryonic-like pluripotent or multipotent stem cells. The possibility of this option was recently demonstrated with the generation of pluripotent stem cells from mouse fibroblasts via the introduction of four pluripotency-related genes (Takahashi and Yamanaka, 2006), or even Oct4 transient overexpression alone (Grinnell *et al.*, 2007). In addition to the sustained expression of such genes, reprogramming is thought to require

both alteration of the epigenetic state and remodeling of chromatin structure, along with other factors such as specific microRNA expression. Further research has to be conducted to assess the fate of these reprogrammed cells before this promising cell-based therapeutic approach becomes feasible (Alberio *et al.*, 2006).

All of these issues have to be overcome before hESC research makes the successful transition from bench to bedside. To fulfill the regenerative promise of hESC-based medicine, many new techniques and protocols have to be developed. We describe here the progress made in two of these technologies that will greatly aid the development of this field: genetic manipulation of hESCs via site-specific recombination (SSR) and pluripotency marker-driven antibiotic selection for undifferentiated stem cells.

I.2 Genetic manipulation of human embryonic stem cells

The introduction of foreign DNA into hESCs is an important technique that can be used to enhance the potential of hESCs in cell-based therapy. Transgenes inserted into safe non-disruptive and non-silenced genomic loci could prove useful in a variety of functions from the maintenance of pluripotency to the direction of differentiation. Transgenic hESCs have been generated with various transfection and genomic integration protocols. In fact, it is encouraging to note that hESCs randomly inserted with enhanced green fluorescent protein (*EGFP*) and neomycin antibiotic resistance (*Neo*) genes driven by separate elongation factor-1 α promoters have been reported to remain pluripotent whilst maintaining transgene expression (Liu *et al.*, 2004). Therefore, careful selection of both the transgene and site of genomic insertion should enable researchers to manipulate hESCs without impairment to its intrinsic properties such as pluripotency. The refinement of this technology would be of value to many aspects of hESC research.

To achieve gene insertion, two steps are required: Initially, intact transgenic DNA that is usually plasmid-based has to be transported to the cell nucleus. This is accomplished by three main transfection techniques: reagent-based transfection reagents, electroporation-related, or viral-based protocols. The subsequent integration of these transgenes into the hESC genome is then attained indiscriminately by random insertion or viral insertion. Alternatively, the precise insertion of transgenes into sequence-specific genomic loci can be achieved by homologous recombination or site-specific recombination.

I.2.1 Transfection

Reagent-based methods

Transfection reagents are a popular and straightforward method of introducing DNA into a cell. A variety of chemical transfection reagents have been developed and most fall into two categories: lipid or liposomal-based and polymer-based. These reagents optimize DNA complexation/condensation, membrane fusion, endosomal release or nuclear targeting, all of which are presumed barriers to gene delivery (Luo and Saltzman, 2000). Of these, only a select few have been shown to work in hESCs. A group in Israel was the first to compare commonly available transfection reagents: the cationic lipid Lipofectamine, the non-liposomal lipid Fugene and the polymer-based polyethylenimine Exgen (Eiges *et al.*, 2001). With optimized conditions, a transfection efficiency of less than 10% was achieved with the best candidate, Exgen, which was used to generate stable EGFP-expressing lines. Additionally, they also reported that hESCs do not survive well and cannot be efficiently transfected via electroporation. This is unlike their murine counterparts where electroporation is one of the more commonly used transfection protocols. More recently, Fugene has been used for the generation of EGFP-expressing hESCs (Liu *et al.*, 2004; Gerrard *et al.*, 2005).

Electroporation-related protocols

Electric-pulse mediated DNA transfer is a common and efficient transfection tool in many cell types. Until a recent publication which detailed the optimization of this method and achieved up to 14% efficiency, electroporation was perceived as too harsh for hESCs (Mohr *et al.*, 2006). Previously, only a small number of publications reported successful

electroporation and subsequent transgene integration, with very low transfection and integration efficiencies (Zwaka and Thomson, 2003; Costa *et al.*, 2005).

A recent variant of the electroporation technique using cell-specific nucleofector solution and customized pulse settings, termed nucleofection, has been used to effectively transfect hESCs. In two separate studies, it was reported to be at least two-fold more effective than conventional lipid-based reagents, and two to four-fold more efficient than Zwaka and Thompson's electroporation protocol (Lakshimipathy *et al.*, 2004; Siemen *et al.*, 2005). In addition, these studies also demonstrated that nucleofection could be used to generate transient and stable hESC lines that maintained the expression of key pluripotency markers.

Viral-based transduction

Viral-based transduction protocols are usually more efficient at DNA delivery into hESCs. Oncoretroviral vectors such as the modified Moloney murine leukemia virus (MoMLV) can be used to randomly integrate a cloned fragment into the chromosomes of dividing cells, with overall transduction efficiencies in hESCs reaching 20% (Trono, 2000; Koch *et al.*, 2006). Recombinant adeno-associated viruses (rAAV) have also been developed, and these have an important additional benefit of viral plasmid integration into specific sites within the host genome (Vasileva and Jessberger, 2005). Aside from oncoretroviral vectors and rAAVs, lentiviral vectors appear to be the vehicle of choice for many hESC researchers as these possess a broader host range and has the ability to infect even non-dividing cells (Gropp *et al.*, 2003). The development of these replication-

deficient, self-inactivating, pseudotyped lentiviruses used in gene therapy provides a way to introduce a gene into the hESC genome with efficiencies of up to 87% (Ma *et al.*, 2003). Both these studies showed high-level, sustained transgene expression in pluripotent hESCs that was maintained even in their differentiated progeny. This was similarly demonstrated in a third study which used a bicistronic, dual-promoter lentivirus vector that incorporated an antibiotic resistance gene (Ben-Dor *et al.*, 2006).

However, viral transductions do have one main disadvantage: the abundance of target sites in the human genome. This creates two problems, the first of which is the lack of control over the number of copies of integrated viral vector. Secondly, there is a very high risk of insertional mutagenesis, where the viral vector is integrated into a tumor-suppressor or oncogene, ultimately leading to malignant transformation of cells (Hacein-Bey-Abina *et al.*, 2003; Glover *et al.*, 2005).

1.2.2 Genomic integration

Random integration

Aside from the random genomic insertion accomplished from the many viral target sites, free plasmid DNA in the nucleus can also be integrated randomly into the human genome, albeit at a much lower efficiency. This process has been shown to occur via a mechanism similar to that used for repair of spontaneous or gamma ray-induced strand breaks, as evidenced by short genomic deletions at plasmid integration sites (Murnane *et al.*, 1990). Although the inherent shortfalls of random integration are similar to those of viral integration, the number of sites and transgenic cassettes inserted per cell can be as

low as one. It is therefore theoretically possible to screen, characterize and isolate a particular hESC clone that contains only one transgene, integrated into a permissive and yet non-disruptive genomic loci, that is expressed at an optimal level. However, this would require a large, inefficient expense of time and manpower.

Homologous recombination

The homologous recombination strategy seeks to overcome the problem of insertional mutagenesis in viral transduction and random integration. As opposed to random integration where DNA is spontaneously integrated into the host genome at an unspecific location, homologous recombination occurs with the assistance of endogenous gene repair mechanisms, replacing chromosomal DNA with identical or almost identical exogenous DNA (Yates and Daley, 2006; Zhou and Dröge, 2006). Both processes enlist the host cell's DNA double strand break repair mechanisms and are relatively inefficient, with homologous mechanisms being significantly less frequent than non-homologous ones (Johnson and Jasin, 2001). This approach is the only way to obtain isogenic, fully competent cells for cell therapy, and is routinely used in mESCs to generate disease models and transgenic animals. However, very few examples of homologous recombination in hESCs have been achieved thus far (Zwaka and Thomson, 2003; Urbach *et al.*, 2004; Zhou and Dröge, 2006).

Site-specific recombination

Loci-specific integration of transgenic cassettes into the human genome can also occur through site-specific DNA recombination (SSR) systems derived from prokaryotic cells.

These systems catalyze recombination reactions between short, conserved DNA sequences known as recombination sites, resulting in excision, integration or inversion reactions (Grindley *et al.*, 2006). Some of these systems, like the Cre recombinase and λ and Φ C31 integrases, have already been successfully developed into valuable tools for various applications in eukaryotic cells (Kolb, 2002), including analyzing gene function, genomic engineering and even gene therapy. Furthermore, the Cre, Flp, Φ C31, and λ integrase systems can achieve targeted insertion of foreign DNA into defined predetermined artificial or natural genome sequences (Kolb, 2002; Dröge, Christ and Lorbach, 2003; Lindenbaum *et al.*, 2004). The latter application is especially relevant to future gene therapy approaches with hESCs because it minimizes the risk of unwanted genome alterations due to random DNA insertions (Zhou and Dröge, 2006).

1.2.3 Serine and tyrosine recombinases

SSR systems fall into two large families based on amino acid sequence homology and catalytic residue: tyrosine integrases which includes λ integrase and Cre and Flp recombinases, and serine recombinases like Φ C31 integrase and $\gamma\delta$ resolvase (Groth and Calos, 2003). However, their recombination mechanisms are distinctly different. Tyrosine recombinases break and rejoin single strands in pairs via a Holiday junction intermediate; serine recombinases cut all strands in advance of strand exchange and re-ligation (Grindley *et al.*, 2006).

The recombination efficiencies of these different SSR systems also vary in mammalian cells. Notably, the phage P1 recombinase Cre is particularly efficient, and is frequently

used to splice out marker genes from genomes after gene targeting and for conditional mutagenesis in model organisms, particularly in the mouse (Branda and Dymecki, 2004). Most recently, mouse codon-optimized FLP (FLPo) and Φ C31 (Φ C31o) have also been reported to attain Cre-like recombination efficiencies (Raymond and Soriano, 2007). In hESCs, transduced cell-permeant Cre protein successfully catalyzed genomic recombination with an efficiency of almost 100%. These cells maintained pluripotency after transduction, and retained the ability to differentiate into all three germ layers (Nolden *et al.*, 2006). The fact that Cre recombinase expression does not have a detrimental effect on stem cell pluripotency is an important result as it is essential to prove not only the efficiency, but also the safety of such systems, especially in this context. These advances in SSR technology will enable highly controlled genetic modifications in hESCs, enabling safe, precise transgene integrations that can be used for many downstream applications.

1.2.4 λ integrase

The phage λ -encoded integrase (Int) is the prototype of the integrase family of SSRs. It executes both the integration and excision of the phage genome into and out of the *Escherichia coli* genome through recombination between pairs of attachment sites, termed *attB/attP* and *attL/attR*, respectively (Radhakrishna *et al.*, 1998). Each *att* site sequence is composed of two inverted core Int binding sites separated by a 7-bp overlap region that is identical in all *att* sites. In addition to these, all *att* sites except *attB* contain arm Int binding sites as well as binding sites for accessory proteins like heterodimeric integration host factor (IHF; required for both integration and excision) and excisionase

(XIS; required for excision only) (Christ and Dröge, 2002) (**Fig. I.3**). Both of these SSR pathways are more tightly controlled in its natural host than the corresponding reactions catalyzed by the related Cre and Flp recombinase (Landy, 1989, 1993). The additional appeal of the λ integrase system in comparison with other SSR systems is its ability to distinguish between two different recombination reactions via different pairs of *att* sites. This can negate the problem of managing the excision/integration reaction equilibrium faced by other systems, and effectively increase efficiencies of genomic insertion.

It has been demonstrated recently that Int-h/218, a mutant phage λ integrase (E174K/E218K), was able to catalyze SSR reactions in the absence of cofactors in human cells. Episomal intramolecular experiments utilizing Int-h/218 in HeLa cells revealed an efficiency of 20% for integrative recombination between *attB* and *attP*, and around 10% for excisive recombination between *attL* and *attR* (Lorbach *et al.*, 2000); episomal intermolecular tests also achieved similar efficiencies (Christ *et al.*, 2002). Int-h/218 was also able to delete a resistance marker flanked by either pair of *att* sites inserted into the permissive ROSA26 genomic locus of mESCs, albeit at low efficiencies (Christ and Dröge, 2002). In addition, an engineered single chain IHF, termed scIHF2, has recently been shown to aid integrative recombination by wild type Int in HeLa cells (Corona *et al.*, 2003).

The high fidelity, directionality and safety features of the λ integrase SSR system give it many advantages as a tool for genomic manipulation. Mutant Int-h/218, along with

scIHF2, could be used to improve recombination and genome targeting efficiencies in hESCs.

Figure I.3. Recombination schematic for λ-integrase. Integrative recombination employs target sites *attP* and *attB* and requires IHF, in addition to Int. Excisive recombination employs hybrid sites *attL* and *attR* and IHF and XIS as co-factors. The *attP* site is about 240bp long and contains five additional arm-type DNA binding sites for Int (○), three IHF binding sites (□), and two binding sites for XIS (△). The *attB* site is 21bp long and lacks any additional binding sites for Int and co-factors. (Adapted from Christ and Dröge, 2002).

I.3 Pluripotency of human embryonic stem cells

I.3.1 Standards of pluripotency

Pluripotency, or the potential for differentiation, is one of the defining features of hESCs. There are three different aspects of pluripotency that can be easily assessed: cell morphology, transcript status and protein marker expression. When cultured on MEFs, hESCs proliferate in tight, rounded colonies, and are morphologically quite distinct from other differentiated cell types, for example fibroblasts (**Fig. I.4**). In particular, hESCs should have a high nuclear to cytoplasm ratio, prominent nucleoli and be compact in size (Amit *et al.*, 2000).

Figure I.4. Morphology of hESCs. Tight colonies, prominent nucleoli and dense, compact cells distinguish hESCs from MEFs in co-culture conditions.

The pluripotency of hESCs in culture can be routinely assessed by mRNA transcript expression levels via either microarray, semi-quantitative reverse transcription PCR (RT-PCR) or quantitative real-time PCR (qRT-PCR) techniques. Other techniques such as

serial analysis of gene expression (SAGE) (Richards *et al.*, 2006) and massively parallel signature sequencing (MPSS) (Wei *et al.*, 2005) are also used to confirm pluripotency-related transcripts, as well as elucidate new “stemness” genes. The three important, well-studied, key regulatory genes that are highly expressed in undifferentiated hESCs and downregulated upon differentiation are POU transcription factor Oct4, high mobility group box protein Sox2 and homeobox transcription factor Nanog (Laslett *et al.*, 2003; Hoffman and Carpenter, 2005; Hirst *et al.*, 2006; Boyer *et al.*, 2006). In particular, Oct4 and Sox2 have been found to bind to adjacent domains in regulatory regions of *Oct4* and *Nanog* (Okumura-Nakanishi *et al.*, 2005; Kuroda *et al.*, 2005; Rodda *et al.*, 2005), as well as many other pluripotency-related genes such as *UTF1*, *Rex1* and *TDGF1* (Rao, 2004; Boyer *et al.*, 2005; Babaie *et al.*, 2006). This indicates that the concerted action of many transcription factors expressed at an optimal level is required to maintain the pluripotential state in hESCs. Some examples of recently identified pluripotency markers are SLAIN1 (Hirst *et al.*, 2006) and Sall4 (Zhang *et al.*, 2006).

The expression of certain cell-surface or intracellular proteins or markers can also be analyzed to gauge hESC pluripotency. Certain surface markers like the globoseries glycolipid stage specific embryonic antigens (SSEA) and the keratin sulphate-associated tumor recognition antigens (TRA) have been well characterized; pluripotent hESCs stain positively for TRA-1-60, TRA-1-81, SSEA-3 and SSEA-4, but negatively for SSEA-1 (Henderson *et al.*, 2002; Carpenter *et al.*, 2004). This profile is similar to human embryonic carcinoma and human ICM cells. In contrast, mESCs stain positively for SSEA-1, but negatively for the other 4 markers. Despite being associated closely with

hESC pluripotency, the function of these surface antigens presents a conundrum, with various roles in cell-cell adhesion or embryonic pathology suggested (Henderson *et al.*, 2002). In conjunction with surface marker analyses, expression levels of transcription factor Oct4 and alkaline phosphatase, a hydrolase enzyme responsible for catalyzing dephosphorylation, are commonly used to gauge the pluripotency of hESC cultures (Laslett *et al.*, 2003). Both these proteins have an exceptionally high expression level in ESCs, and are downregulated upon differentiation (Amit and Itskovitz-Eldor, 2006; Trounson, 2006).

I.3.2 Pluripotency marker *UTF1*

Function and characterization

Along with Oct4, Sox2 and Nanog, the transcriptional cofactor undifferentiated transcription factor 1 (UTF1) is regarded as one of the most important pluripotency markers defining the hESC niche (Bhattacharya *et al.*, 2004; Trounson, 2006). It is consistently profiled as a transcript present in many undifferentiated hESC lines (Wobus and Boheler, 2005), and was first discovered in mouse embryonic carcinoma (EC) cells and ESCs in 1998. UTF1 expression was rapidly extinguished upon induced differentiation in mouse EC cells and downregulated in the primitive endoderm of the mouse embryo during the primitive streak stage (Okuda *et al.*, 1998). This same group cloned and characterized the human UTF1 homolog, which is 87% similar to that of the mouse. Mutagenesis analyses revealed that UTF1 was essential for potentiation of ATF-2 activity via the leucine zipper motif of the former binding to the metal binding motif of the latter. UTF1 also binds to the TATA-binding protein containing complex, and as such

plays an important role in activating specific transcription (Fukushima *et al.*, 1998). Recently, they also showed that UTF1 might be involved in mESC proliferation (Nishimoto *et al.*, 2005).

The *UTF1* gene consists of a TATA-less promoter with four GC boxes, two exons, one intron, and an enhancer comprising of the PEA3 and En-like motifs, as well as octamer and sox binding sequences (Nishimoto *et al.*, 2001). Whilst the mutation of PEA3 and En-like sequences demonstrated little effect on expression, the synergistic binding of transcription factors Oct4 and Sox2 on the twin octamer sequence in the 3' enhancer region was essential for UTF1 expression (Nishimoto *et al.*, 1999). This has been verified recently *in vivo* in hESCs via chromatin immunoprecipitation (ChIP), showing four to six fold enrichment of Oct4 and Sox2 (but not Nanog) at the 3' enhancer region of *UTF1* (Boyer *et al.*, 2005). Sp1-like transcription factors that bind to GC boxes like those in the *UTF1* promoter are currently thought to be regulators of embryonic development in vertebrates (Zhao and Ming, 2005). These factors have also been shown to be involved in the transcriptional control of both *Oct4* and *Nanog* (Yang *et al.*, 2005; Wu and Yao, 2006). Along with key ES-associated genes like *Oct4*, *Nanog* and *Rex1*, *UTF1* switches its replication timing to later in S-phase in response to differentiation (Perry *et al.*, 2004). Thus the genomic profile of *UTF1* is consistent with its function as an undifferentiated transcription factor.

Furthermore, RNA interference (RNAi) of *Oct4* transcripts in mESCs has been shown to cause downregulation of *UTF1* transcripts (Sen *et al.*, 2004), and more than 90% of all

cells in undifferentiated mESC populations stain positively for both these markers (Hoof *et al.*, 2006). These two experiments further allude to the prevalence of UTF1 and Oct4 co-expression in undifferentiated ESCs. However the phenotypic effects of a constitutive UTF1 knockout have not been analyzed, although *UTF1*-knockout 129/SvEv mice have been generated by the Texas institute for genomic medicine (<http://www.tigm-knockouts.org>).

Expression profile

A GeneAtlas microarray analysis of *UTF1* shows an expression pattern similar to that of key embryonic transcription factors *Oct4* and *Nanog* (<http://symatlas.gnf.org>). Its expression levels are very low in differentiated human and mouse tissues, but high in mouse embryos and blastocysts (Su *et al.*, 2002; Su *et al.*, 2004). Additionally, UTF1 was the second most overrepresented gene when mouse ESC and somatic cell expressed sequence tag (EST) libraries were compared via digital differential display (Mitsui *et al.*, 2003).

Like any important pluripotency-related marker, UTF1 is expressed in both mouse and human ESCs (Ginis *et al.*, 2003; Brimble *et al.*, 2004; Sun *et al.*, 2006; Hoof *et al.*, 2006), and is downregulated upon differentiation. In fact, *UTF1* has been shown to be more sensitive to differentiation as it was downregulated sooner, and with a larger magnitude, than either *Oct4* or *Sox2* upon spontaneous differentiation via embryoid body formation of both mouse and human ESCs (Nishimoto *et al.*, 1999; Cai *et al.*, 2005; Wei *et al.*, 2005). In a comparison of hESCs cultured with MEFs in media containing either fetal

calf serum or defined serum replacement (SR), *UTF1* was found to be downregulated to a larger extent in serum medium compared to *Oct4* or *Sox2*, suggesting strongly that it is a good marker for early differentiation (Skottman *et al.*, 2005). This same trend was observed during the directed differentiation of mouse EC and ESCs through retinoic acid (RA) induction (Eifert *et al.*, 2006; Ivanova *et al.*, 2006).

Aside from ES and EC cells, UTF1 expression has been detected in other pluripotent cell types. Spermatogonial stem cells isolated recently from the adult mouse testis were shown to exhibit pluripotent ESC characteristics. These cells expressed Oct4, Nanog, Rex1 and UTF1, and also downregulated UTF1 sooner than other pluripotent markers upon embryoid body formation (Guan *et al.*, 2006). Pluripotent stem cells induced from mouse fibroblasts via the introduction of four factors, Oct4, Sox2, c-Myc and Klf4, expressed UTF1 as well. Interestingly, the expression level of UTF1 in these variably pluripotent cell lines coincided more closely with that of important ESC marker gene Oct4, compared to other markers like Cripto or Nanog (Takahashi and Yamanaka, 2006).

I.3.3 Spontaneous differentiation

Spontaneous differentiation of hESCs in laboratory conditions can occur with relative ease. Thus, most cultures will consist of a population of heterogeneous cells, and usually contain a subset of less pluripotent, and therefore, more differentiated cell types (Amit *et al.*, 2000; Eiges *et al.*, 2001; Stewart *et al.*, 2006). This is also true for mouse and medaka ESCs (McWhir *et al.*, 1996; Hong *et al.*, 1996). The heterogeneity of hESC cultures was demonstrated recently via the isolation of SSEA-3⁺ and SSEA-3⁻ populations, and the

subsequent comparison of Oct4 and Nanog transcript and protein levels within these two groups of cells (Stewart *et al.*, 2006). Interestingly, cultures seeded with either of the sorted populations could regenerate a mixture of both SSEA-3⁺ and SSEA-3⁻ hESCs. Although the two separate factions displayed similar transcript levels of *Oct4* and *Nanog*, only SSEA-3⁺ cells expressed SSEA-4, Oct4 and Nanog proteins. This suggested that a dynamic equilibrium was established between both SSEA-3⁺ and SSEA-3⁻ hESC compartments via translational control of these genes, and confirmed the heterogenic nature of hESCs in culture (Stewart *et al.*, 2006).

I.3.4 Maintenance of pluripotency

Culture conditions

Given that the presence of minor populations of differentiated cells can compromise the pluripotency of entire stem cell cultures (Mountford *et al.*, 1998), it is essential that the hESC lines are kept in laboratory culture conditions tailored to maintain pluripotency. Research in this area has progressed from MEF-layered surfaces and ill-defined culture media relying on MEF-secreted factors to defined extracellular matrices and media formulations (Hoffman and Carpenter, 2005). Two groups have also successfully cultured pluripotent hESCs by replacing MEFs and related factors with basic fibroblast growth factor (bFGF) alone (Xu *et al.*, 2005), or in combination with noggin, a bone morphogenetic protein (BMP) antagonist (Xu *et al.*, 2005). In addition, overexpression of Nanog in hESCs has been shown to enable sustained pluripotency for multiple passages upon the removal of MEFs (Darr *et al.*, 2006).

Cell sorting

Besides research focused upon keeping hESCs pluripotent, some groups have taken an alternative approach, which is to select for undifferentiated cells. Fluorescence-activated cell sorting (FACS) is a common way to group cells according to the different expression profiles of selected surface epitopes. Although hESCs express a unique set of surface markers at a high level, the percentage of the total population expressing SSEA4, TRA-1-60 and TRA-1-81 in culture at any one time is very variable, ranging from 75-99% (Rosler *et al.*, 2004; Carpenter *et al.*, 2004). Due to the low efficiency of generating single-cell clones for hESCs, successful sorting and clonal isolation of pluripotent cultures was only achieved recently (Forsyth *et al.*, 2006; Hewitt *et al.*, 2006; Stewart *et al.*, 2006).

This strategy can also be accomplished by means of a fluorescent transgene driven by the promoter and/or enhancer of a pluripotency marker. In hESCs, this was first demonstrated with an *EGFP* gene under the control of a murine *Rex1* promoter (Eiges *et al.*, 2001). Similar human *Oct4-EGFP* constructs have also been used to screen for undifferentiated hESCs via flow cytometry (Zwaka and Thomson, 2003; Gerrard *et al.*, 2005). A recent improvement in semi-automated FACS clonal recovery utilized a mouse *Oct4* promoter-driven *EGFP* construct to select for homogeneously pluripotent hESCs at physiological oxygen levels (2%) (Hewitt *et al.*, 2006).

A sorting strategy utilizing a combination of carefully selected markers or transgenes to improve the selection criteria should be more ideal for future experiments. However, the

main disadvantage is that the FACS procedure has to be repeated routinely to ensure a constant high fraction of pluripotent hESCs in culture.

Antibiotic selection

An alternative solution is to ablate differentiating cells present in hESC cultures by antibiotic selection. Antibiotic selection has the advantage of being less intrusive to hESCs; the expression of the neomycin resistance gene (*Neo*) and corresponding G418 antibiotic selection, along with EGFP expression, have no adverse effect on hESC pluripotency (Liu et al., 2004; Gerrard *et al.*, 2005). Thus far, only one hESC study for maintenance of pluripotency utilizing the twin markers of EGFP and neomycin resistance driven by an *Oct4* promoter has been reported. However, the ability to ablate or select against partially differentiated cells was not established in these studies (Zwaka and Thomson, 2003).

In the mouse, the *Oct4-Neo* system has been used to isolate ES and EC cell lines from murine embryos with a non-permissive genetic background via selective ablation of differentiating cells (McWhir *et al.*, 1996; Gallagher et al., 2003). In addition, the same strategy has been used in both mouse and medaka ESCs to eliminate differentiated progeny and sustain homogeneously pluripotent ESC cultures. (Mountford *et al.*, 1998; Hong et al., 2004).

These studies show that such a system would be useful if transferred into the hESC context. However the use of Oct4 as a marker does have its disadvantages. Aside from

being less sensitive to differentiation compared to UTF1, its expression level is also tightly controlled as an increase of less than two-fold will lead to differentiation into primitive endodermal and mesodermal lineages in mouse ESCs (Niwa *et al.*, 2000). This can mean that the *Oct4-Neo* system is potentially less sensitive to these differentiation pathways, and thus will not be able to efficiently ablate hESCs that are undergoing differentiation into these lineages.

II Statement of aims

Even with recent advances in derivation, culture and differentiation, the field of hESC research is still barely out of infancy. The development of more tools and techniques, and deeper understanding of hESC culture and characteristics are required to enable the successful transition from research laboratories to clinical trials. In this project, we endeavor to advance current scientific knowledge in this field and improve and expand the current hESC toolkit.

The acquisition of techniques and protocols required for hESC culture and characterization is therefore our initial goal. Subsequently, two key issues of hESC research are tackled: genetic manipulation and selection for pluripotency in culture.

The λ integrase SSR system is functional in human cells, flexible, easy to manipulate and accurate in its specificity (Christ, Corona and Dröge, 2002; Christ and Dröge, 2002; Dröge, Christ and Lorbach, 2003). Therefore for the purpose of precise genome manipulation, we intend to transfer this system into the hESC context. The development and optimization of this system would greatly facilitate transgene insertion experiments with a view to future clinical procedures and possibly stem cell gene therapy.

To address the problem of heterogeneous hESC populations in culture, we propose to develop a selective system based on the pluripotency gene, *UTF1*. UTF1 has been shown to be rapidly downregulated during stem cell differentiation (Okuda *et al.*, 1998), and has

been shown to be controlled by two key pluripotency markers Oct4 and Sox2 (Nishimoto *et al.*, 1999; Boyer *et al.*, 2005). In fact, it is more sensitive to differentiation and is downregulated sooner and by a larger magnitude than either Oct4 or Sox2 (Nishimoto *et al.*, 1999; Cai *et al.*, 2005; Wei *et al.*, 2005). Antibiotic resistance driven by the genetic elements of *UTF1* can thus potentially be employed to generate homogenously pluripotent hESC cultures. This could reduce the complexities of maintaining uniform undifferentiated hESCs in the laboratory and opens up the possibility of increasing the scale of such cultures to industrial quantities.

Therefore our three main aims can be summarized as follows:

- 1) Establish a hESC laboratory and optimize culture, characterization and manipulation protocols.
- 2) Develop the λ integrase SSR system for use in hESCs.
- 3) Develop a *UTF1*-based antibiotic selection system to select for pluripotent hESCs.

III. MATERIALS AND METHODS

III.1 Molecular biology

III.1.1 DNA manipulation

Restriction digests and ligation

Restriction digests were performed with restriction enzymes (REs) from NEB. In general, 1 μ g of DNA was incubated with 1 μ l of RE (10u/ μ l) in 20 μ l of buffered solution for 4 hours, unless indicated otherwise. DNA products were then visualized via agarose gel electrophoresis (Biorad) and extracted via Qiagen's gel extraction kit, if needed. DNA ligation reactions were performed at a 3:1 stoichiometric ratio (insert to backbone) with 1 μ l of T4 DNA Ligase (10u/ μ l) in 10 μ l of buffered solution (NEB) at room temperature overnight.

Polymerase Chain Reaction (PCR)

PCR reactions were performed using a thermal cycler (ABI), and *Pfu* or *Taq* DNA polymerases with relevant buffers from Promega and Qiagen respectively, according to manufacturer's protocol. Detailed protocols are listed below in individual sections.

In general, the programmed cycle was as follows:

Stage	Temperature (°C)	Time (min)	35 Cycles
Initial denaturation	94	10	
Denaturation	94	1	
Annealing	Variable	1	
Extension	72	Variable	
Final extension	72	10	

PCR reactions contained the following:

Component	Concentration	Volume
dNTPs	10mM	0.4μl
Taq/Pfu	5U/μl	0.5μl
Template DNA	Variable	Variable
Primers	20μMol each	1μl x2
Q Buffer	5X (Optional)	4μl
Buffer	10X	2μl
Water	Top up to	20μl

Agarose gel electrophoresis and DNA purification

DNA products and plasmids were separated via agarose gel electrophoresis (Biorad) and stained with ethidium bromide (Biorad) for 10min before visualization with Biorad’s gel documentation system. Desired DNA fragments were extracted and purified via Qiagen’s gel extraction kit. Qiagen’s PCR purification kit was also used to extract solitary DNA products from PCR reaction mixtures, which were subsequently used for ligation.

III.1.2 Bacterial work

Materials:

Type	Composition	Source
LB media	1% (w/v) bacto-tryptone 0.5% (w/v) yeast extract 1% (w/v) NaCl, pH 7.4	BD Biosciences BD Biosciences Fluka
LB agar plates	LB media 1.5% (w/v) bacto-agar	As above BD Biosciences
LB amp plates	LB agar plates 300μg/ml ampicillin	As above Sigma

Competent cell preparation

DH5α *supE44 ΔlacU169 (φ80 lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1 E. coli* cells (Hanahan, 1983) were streaked onto LB agar plates and cultured overnight at

37°C. A single colony was then picked and inoculated into 50ml LB media and incubated at 300rpm, 37°C until the OD₆₀₀ value reached between 0.4 to 0.6. The cultures were harvested via centrifugation at 4000g for 5min at 4°C. The resultant pellet was re-suspended in 30ml of cold 0.1M MgCl₂ (Sigma) and centrifuged again at 4000g for 5min at 4°C. With the supernatant discarded, the pellet was re-suspended in 20ml of 0.1M CaCl₂ (Sigma) and kept on ice for 30min. The cells were collected via centrifugation at 4000g for 5min at 4°C, re-suspended in 2ml of cold 0.1M CaCl₂/15% glycerol (Sigma) and aliquoted to 100μl/tube to be stored at -80°C.

Transformation of *E. coli* and plasmid DNA preparation

100ng of plasmid or 20μl of DNA ligation mix was added to 100μl of competent DH5α bacteria and incubated on ice for 30min. Heat shock was administered at 42°C for 2min, and the cells were placed on ice again for 5min. Recovery was performed by incubating at 230rpm, 37°C for 40min, after the addition of 900μl of LB media. Centrifugation of the tube thereafter collected the cell pellet, which was re-suspended in 100μl of LB media and plated out onto LB amp plates. The plates were then incubated overnight at 37°C and colonies were picked, inoculated and expanded in LB media with 300μg/ml ampicillin. Plasmid DNA was prepared from transformed *E. coli* cell pellets via the Qiagen miniprep kit for general molecular biology work, or the Qiagen endo-free maxi kit for cell culture work, performed according to manufacturer's protocols.

III.1.3 Construction of plasmids

pCH-RLNRLE

pCH-RLNRLE was employed as the substrate vector to gauge recombination efficiencies of both Cre and $\gamma\delta$ SSR systems, utilizing EGFP expression as readout. It was derived from *pCH-RLNRLZ* (Schwikardi and Dröge, 2000) via the replacement of the original *LacZ* gene with the coding region for *EGFP*. The *EGFP*+PolyA signal sequence was derived from *pCMVssEGFPbpa* (Christ *et al.*, 2002) via PCR, with a short linker (GGSGG) replacing the *EGFP* start codon (**Fig. III.1A**). This strategy enabled us to remove most of the *LacZ* coding region without resorting to complex cloning steps, and made use of endogenous KpnI and BamHI single-cut restriction sites in *pCH-RLNRLZ*.

The following primers were used to generate the PCR product: L1EGFP-F (CGGGGTACCGGGTGGAAGCGGCGGTGTGAGCAAGGGCGAGGA) and L1EGFP-R (CGCGGATCCGAGGCTAGAACTAGTGG), resulting in a KpnI/BamHI-flanked 1kb insert. The PCR reaction was conducted using *Pfu* polymerase with 50ng of template plasmid *pCMVssEGFPbpa*, at 55°C annealing temperature and a 3min extension step. Both the PCR product and vector *pCH-RLNRLZ* were then cleaved with KpnI and BamHI restriction enzymes and ligated (**Fig. III.1B**). The resultant clones were then tested via either KpnI or BsrGI RE digests, which should result in a linearized plasmid. *pCH-RLNRLE* was subsequently transformed, expanded and (**Fig. III.1C**). The region from the *LacZ* coding fragment to the polyA signal (1398-2379bp) was sequenced to confirm its authenticity. The *EGFP* insertion replaced a large portion of the *LacZ* gene and resulted in the expression of an N-terminal LacZ-Linker-EGFP fusion protein.

Figure III.1. *pCH-RLNRLE* construction strategy. (A) Parental plasmids. (B) *pCH-RLNRLZ* was digested with *KpnI* and *BamHI* to remove the *LacZ* fragment, which was replaced via ligation with a similarly-digested PCR insert comprising of a linker, *EGFP* gene and polyA signal, (C) thus creating substrate vector *pCH-RLNRLE* which was validated with either *KpnI* or *BsrGI* RE digests.

pCH-RLE

The recombined product vector *pCH-RLE* served as a positive control and was generated through transformation of *pCH-RLNRLE* into 294-Cre cells. This *Escherichia coli* strain expresses Cre constitutively at a low level (Aranda *et al.*, 2001), and is thus able to recombine substrate plasmid *pCH-RLNRLE*, resulting in the excision of the *Neo* gene between the two LoxP sites. This resultant *pCH-RLE* plasmid (**Fig. III.2A**) was then verified by digesting with NcoI and BamHI, which gives a 1.4kb and a 3.4kb fragment, and ClaI, which linearizes the plasmid (**Fig. III.2B**). In addition, EGFP-expressing cells were observed after successful transfection of *pCH-RLE* into HeLa cells via electroporation, demonstrating its utility as a positive control plasmid (data not shown).

Figure III.2. *pCH-RLE* generation. (A) *pCH-RLNRLE* was transformed into 294-Cre *E. coli* cells expressing Cre, generating *pCH-RLE*. (B) *pCH-RLE* checked via restriction digest with NcoI and BamHI (1.4kb and 3.4kb fragments), and ClaI (4.8kb linearized plasmid).

pBS-UTF1

The human *UTF1* gene was cloned into a *pBluescript* backbone to generate *pBS-UTF1*. Genomic DNA was extracted from HeLa cells using the DNeasy blood and tissue genomic DNA purification kit (Qiagen), and used as a template to amplify the human *UTF1* gene. The genomic PCR was performed in a 100 μ l reaction with 1 μ g of gDNA, 50pMol of primers UTF1-F (CCGGAATTCAGCGCCAGGACCGACCCCTTA) and UTF1-R2 (TGCTCTAGAGGGTTCAGCACTTCTCCTGCCTC) each, and *Pfu* DNA polymerase. PCR conditions were: 60°C annealing temperature, 6min extension time and 35 cycles. The resulting 2.8kb fragment (**Fig. III.3A**) consisting of the *UTF1* 5' promoter, 3' enhancer and coding regions (two exons and one intron) was extracted and cleaved using *EcoRI* and *XbaI* REs, and ligated to a similarly-digested *poIHF2GFP* (lab stock) to remove the promoter and coding regions of the *pBluescript*-based backbone (**Fig. III.3B**). The resultant colonies were screened via a *NotI* and *SacII* double digest, and correct clones yielded four fragments (3.2kb, 2.2kb, 772bp and 263bp), as shown in **Fig. III.3C**. This *pBS-UTF1* plasmid (**Fig. III.3D**) was transformed into competent *E. coli*, expanded and the full 2.8kb *UTF1* insert sequenced to confirm its identity.

Figure III.3. *pBS-UTF1* generation. (A) The UTF1 genomic fragment was amplified from HeLa cells via PCR. This fragment was digested with both EcoRI and XbaI, and ligated to a similarly-digested *pBluescript*-based backbone from (B) *poIHF2GFP*. (C) NotI and SacII restriction digest of a correct *pBS-UTF1* clone showing the four expected DNA fragments. (D) Map of *pBS-UTF1* showing the *UTF1* 5' promoter, coding regions, and 3' enhancer which consist of the OCT, SOX and M1 octamer sequences.

pBS-UTF1-EGFP

We next proceeded to replace the *UTF1* coding regions with that of the *EGFP* gene. This generated *pBS-UTF1-EGFP*, which has the *EGFP* gene driven by the *UTF1* 5' promoter and controlled by the *UTF1* 3' enhancer. KpnI and NcoI restriction sites were first

introduced between the end of the *UTF1* promoter and before the ATG start codon, by amplifying the promoter region of *pBS-UTF1* via PCR with UTF1-F and UTF1-MR3 (CGGGGTACCCCGGGGCTGGGGCGCGG) primers using *Pfu* polymerase with 50ng of template plasmid, at 60°C annealing temperature and a 1min extension step. This fragment was cleaved with EcoRI and NcoI, and ligated to a similarly-digested *pCMVssEGFPbpa* plasmid to replace the CMV promoter (**Fig. III.4A**). This plasmid, *pUTF1pro-EGFPbpa*, was digested with EcoRI and BsrGI to produce an (EcoRI)*UTF1promoter-EGFP*(BsrGI) sticky-ended 1kb fragment (**Fig. III.4B**, right). *pBS-UTF1* was used as a PCR template to amplify a segment of the *UTF1* 3' UTR region, whilst inserting a BsrGI restriction site near the end of the *UTF1* transcript using primers UTF1-MF2 (AGCTGTACAAGTGAGTCCCGGCTGCGGC) and UTF1-R2. This PCR reaction was conducted using *Pfu* polymerase with 50ng of template plasmid, at 60°C annealing temperature and 3min extension time. When digested, this produced a sticky-ended BsrGI/XbaI 1.4kb fragment (**Fig. III.4B**, middle). *pBS-UTF1* was then digested with EcoRI and XbaI, and this 3.6kb backbone fragment (**Fig. III.4B**, left) was ligated to the above two fragments. Three positive *pBS-UTF1-EGFP* clones were detected via restriction digest with NcoI, yielding the correctly-sized 4.7kb and 1.3kb DNA fragments and also an uncut 6kb band due to short incubation time (**Fig. III.4C**). One of these was selected, transformed and expanded. The region from the 5' *UTF1* promoter to the 3' *UTF1* enhancer (2251-4600bp) was sequenced to confirm its identity.

Figure III.4. *pBS-UTF1-EGFP* construction strategy. (A) Generation of interim plasmid *pUTF1pro-EGFPbpa* from a *pCMVssEGFPbpa* backbone (see Fig. III.1A). (B) Three fragment ligation of *pBS-UTF1* backbone, *UTF1* enhancer insert, and *UTF1* promoter with *EGFP* gene insert, resulting in (C) *pBS-UTF1-EGFP*, an *EGFP* expression vector driven by *UTF1* genetic elements.

pBS-UTF1-EGFP-noEnh

To gauge the necessity of the 3' *UTF1* enhancer for the expression of the EGFP gene we used the vector *pBS-UTF1-EGFP-noEnh*, constructed by removing the enhancer from *pBS-UTF1-EGFP*. To achieve this, *pBS-UTF1-EGFP* was digested with PstI and XbaI to remove the *UTF1* 3' enhancer sequence (Fig. III.5A). This 4.7kb backbone was treated with DNA Polymerase I, Large (Klenow) Fragment (NEB) to blunt both ends, self ligated and transformed (Fig. III.5B). *pBS-UTF1-EGFP-noEnh* was checked via restriction digests with KpnI and BamHI yielding three fragments (3.3kb, 1.2kb and 245bp), and the single-cutter NcoI (Fig. III.5C).

Figure III.5. *pBS-UTF1-EGFP-noEnh* generation. (A) The 1.3kb 3' *UTF1* enhancer was removed from *pBS-UTF1-EGFP* via PstI and XbaI digestion. (B) *pBS-UTF1-EGFP-noEnh* was produced by self-ligating the blunt-ended 4.7kb fragment. (C) Separate KpnI with BamHI, and NcoI digests were conducted to authenticate *pBS-UTF1-EGFP-noEnh*.

pTZ-UTF1-EGFP

To investigate if a different plasmid backbone affects *UTF1*-driven EGFP expression, we utilized the plasmid *pTZ-UTF1-EGFP*, generated via cloning the entire *UTF1-EGFP* insert into a *pTZ-18R* backbone. The plasmid *pBS-UTF1-EGFP* was digested with EcoRI

and XbaI REs to extract the *UTF1-EGFP* expression cassette, and this was ligated to a similarly-digested *pTZ-18R* (lab stock) phagemid backbone. This *pTZ-UTF1-EGFP* plasmid was transformed, expanded and digested with PstI to confirm its identity (**Fig. III.6A**). The expected two fragments (3.9kb and 1.3kb) were compared with the similarly-digested *pBS-UTF1-EGFP*, which resulted in a linearized 6kb DNA fragment.

pTZ-UTF1-Neo

Using a derivative of *pTZ-UTF1-EGFP*, we sought to generate a plasmid that could be used for the creation of stable cell lines that would be selectable for high UTF1 expression. With neomycin antibiotic resistance our marker of choice, we proceeded to replace the *EGFP* gene. The neomycin resistance gene (*Neo*) was amplified with flanking KpnI and BsrGI RE sites via PCR with *pPGKssNeobpa* (lab stock) as the template and Neo-F (CTGGGTACCCCATGATTGAACAAGATGG) and Neo-R (ACTGTGTACATCAGAAGAACTCGTCAAGAA) primers. The PCR reaction was carried out with *Pfu* polymerase, 50ng of template plasmid, at 60°C annealing temperature and a 2min extension step. The *pTZ-UTF1-EGFP* backbone was digested with KpnI and BsrGI REs to remove the *EGFP* gene, and ligated with the similarly-digested 800bp *Neo* fragment. The resultant clones were digested with NcoI, with two bands expected at around 4.4kb and 0.9kb. This *pTZ-UTF1-Neo* plasmid (**Fig. III.6B**) was transformed and expanded. The region from the 5' *UTF1* promoter to the *UTF1* polyA signal (22-1129bp) was sequenced to confirm its identity.

Figure III.6. *pTZ-UTF1-EGFP* and *-Neo* construction strategy. (A) *pTZ-UTF1-EGFP* was digested with *Pst*I (3.9kb and 1.3kb) and compared to *Pst*I-cleaved *pBS-UTF1-EGFP* (6kb). (B) *Neo* gene PCR product digested with *Kpn*I and *Bsr*GI (799bp) was ligated to similarly-digested *pTZ-UTF1-EGFP* backbone (4.5kb) to generate *pTZ-UTF1-Neo*.

pTZ-UTF1-EGFP-noEnh

The plasmids *pTZ-UTF1-EGFP-noEnh* and *pBS-UTF1-EGFP-noEnh* were used to compare the EGFP expression levels resulting from the removal of the 3' *UTF1* enhancer from two different backbones. The *pTZ-18R* version of *pBS-UTF1-EGFP-noEnh* was generated as follows: *pTZ-UTF1-EGFP* was digested with *Pst*I to remove the *UTF1* enhancer, self-ligated and transformed. The resulting *pTZ-UTF1-EGFP-noEnh* plasmid was then checked via an *Nco*I restriction digest, which yielded a linearized 3.9kb fragment (Fig. III.7A).

Figure III.7. *pTZ-UTF1-EGFP-noEnh* and *-noPro* construction strategy. (A) *pTZ-UTF1-EGFP* was digested with *PstI* and re-ligated to generate *pTZ-UTF1-EGFP-noEnh*, which can be linearized with *NcoI*. (B) *pTZ-UTF1-EGFP* was digested with *EcoRI* and *KpnI*, blunted with Klenow fragment, and re-ligated to generate *pTZ-UTF1-EGFP-noPro*. *NcoI* digestion of this results in two expected fragments (3.7kb and 1.3kb).

pTZ-UTF1-EGFP-noPro

To gauge the necessity of the 5' *UTF1* promoter for the expression of the EGFP gene we used the vector *pTZ-UTF1-EGFP-noPro*, constructed by removing the promoter from *pTZ-UTF1-EGFP*. *pTZ-UTF1-EGFP* was digested with both *EcoRI* and *KpnI* to remove the *UTF1* promoter. This is then incubated with DNA Polymerase I, Large (Klenow) Fragment (NEB) to blunt and fill in the ends, self-ligated and transformed. The resulting

pTZ-UTF1-EGFP-noPro plasmid was then checked using *NcoI*, which yielded two expected fragments of 3.7kb and 1.3kb (**Fig. III.7B**).

pTZ-UTF1-EGFP-M1mut

We proceeded to mutagenize the M1 sequence on the 3' *UTF1* enhancer to gauge its effects on *UTF1*-driven EGFP expression, with the same sequence as previously reported for the *Nanog* gene (Rodda *et al.*, 2005). 50ng of plasmid, *pTZ-UTF1-EGFP* was used as the template for cloning the mutant M1 sequence (GTAGTGGT; original sequence is GTCTGGGT) via PCR. Flanking arms 200bp from M1 were amplified using two primer pairs, Left arm (near the 5' *XcmI* RE site): UTF1-M1F (GACCCAGGAGTAGCATG) and UTF1-M1mR (CTAGGCCCACCACTACCAGAGCCAC), and Right arm (near the 3' *NheI* RE site): UTF1-M1mF (GTGGCTCTGGTAGTGGTGGGCCTAG) and UTF1-M1R (CATTGTTATGCTAGCGG) (**Fig. III.8A**). These were then assembled via PCR with the two arms as template (2 μ l from each PCR reaction) and UTF1-M1F and UTF1-M1R for primers. The resulting 370bp fragment was digested with *XcmI* and *NheI*, and ligated to a similarly-cleaved *pTZ-UTF1-EGFP* to replace the original M1 sequence (**Fig. III.8B**). These PCR reactions were carried out with *Pfu* polymerase at 55°C annealing temperature and a 1min extension step. This *pTZ-UTF1-EGFP-M1mut* plasmid (**Fig. III.8C**) was transformed, expanded and the region from the end of the EGFP gene to after the mutated M1 site (954-1688bp) sequenced to confirm its identity (cloning work performed by exchange student C.-T. Lee). More information can be found in Appendix II (**Fig. VIII.1**).

Figure III.8. *pTZ-UTF1-EGFP-M1mut* construction strategy. (A) The *M1* sequence, flanked by *XcmI* and *NheI* restriction sites, was mutated via assembly PCR. (B) The *M1mut* insert was digested with *XcmI* and *NheI*, and ligated to a similarly-digested *pTZ-UTF1-EGFP* backbone, (C) creating *pTZ-UTF1-EGFP-M1mut*, which was sequenced.

III.2 Cell culture

III.2.1 Routine cell line maintenance

Cell lines:

Name	Details	Source
HeLa	Human cervical carcinoma cells	Lab stock
A549	Human alveolar basal epithelial carcinoma cells	Lab stock
F3	Primary human foreskin fibroblast cells	Technion-Israel Institute of Technology
C57Bl	MEF cells	Derived
Balb/C	MEF cells	Derived
K4	MEF cells	ES Cell International
PMEF	MEF cells	Speciality Media
PMEF-N	Neomycin-resistant MEF cells	Speciality Media
PMEF-H	Hygromycin-resistant MEF cells	Speciality Media

Materials:

Type	Composition	Source
Culture media	Dulbecco's modified eagle medium (DMEM) 10% (v/v) fetal bovine serum (FBS) 2mM L-glutamine 100unit/ml penicillin/streptomycin	All from Gibco
Freezing media	Fetal bovine serum (FBS) 10% (v/v) Dimethyl sulfoxide (DMSO)	Gibco Sigma

Cell lines were cultured on tissue-culture grade dishes and flasks (Nunc) at 37°C and 5% CO₂, and passaged when confluency was around 90% in a 1 to 5 split ratio. Phosphate buffered saline, PBS (Gibco), was first used to wash the cells before the addition of 0.05% trypsin-ethylenediaminetetraacetic acid (EDTA) (Gibco). Cells were then incubated at 37°C for 5min before dissociation by pipetting, resuspended in culture media to neutralize trypsin and collected by centrifugation at 600g for 3min, and plated out with new culture media on new dishes/flasks.

These cell lines were frozen via the standard slow cooling protocol. Briefly, cultures were centrifuged at 600g for 3min after trypsinization and neutralization. This pellet was then resuspended in freezing media and split into cryovials (Nalgene; 500 μ l per vial). The vials were placed into a Mr. Frosty freezing aid (Nalgene) and incubated at -80°C overnight. The next day, they were transferred to either a -150°C freezer or liquid nitrogen tanks. For thawing, frozen cryovials were incubated in a 37°C water bath until the frozen contents were almost fully thawed and then added drop-wise to 10ml of culture media. Cells were collected by centrifugation at 600g for 3min, re-suspended in culture media, and plated out.

III.2.2 Derivation of mouse embryonic fibroblasts

Primary MEFs were derived from the C57Bl and Balb/C mouse strains. 13.5 day post-coital mice were purchased from the National University of Singapore's Laboratory Animal's Centre. Pregnant mice were sacrificed and washed with 70% ethanol. Sterilized surgical scissors were used to cut skin and peritoneal walls, allowing access to and removal of the embryos, which were washed in PBS. The placenta, head and internal viscera were removed from each embryo, and washed twice with PBS in a petri dish. Embryos were collected and minced finely in 2ml PBS and transferred into a 50ml falcon tube. 2ml of 0.25% trypsin- EDTA per embryo was added to the above mixture and incubated for 30min at 37°C. 35ml of culture media was added to neutralize the trypsin. This was then mixed via pipetting and passed through a 40 μ m cell strainer (BD Falcon). The cells were collected, topped up to 60ml with culture media and plated onto a T175 tissue culture flask (Nunc) (about 7-8 embryos per T175 flask). The primary MEFs were

allowed to expand and were passaged when they reached 90% confluency at a 1:3 split ratio. They were frozen as stock at passage 4, and used to plate MEF-monolayered dishes for hESC culture at passage 7.

III.2.3 Human embryonic stem cell maintenance

Cell lines:

Name	Karyotype	Race	Year Derived	Origin
hES2	46, XX	Chinese	1997	ES Cell International Pte Ltd
HUES7	46, XY	Unknown	2003	Melton hES Cell Facility, Harvard University
HUES8	46, XY	Unknown	2003	
HUES9	46, XX, inv 9	Unknown	2003	

Culture media:

Type	Composition	Source
hES culture media (ES Cell International)	Dulbecco's modified eagle medium (DMEM) 10% (v/v) defined fetal bovine serum (dFBS) 2mM L-glutamine 1X Insulin-Transferrin-Selenium (ITS) Non-Essential Amino Acids (NEAA) 0.1% (v/v) β -Mercaptoethanol 100unit/ml penicillin/streptomycin	Gibco Hyclone Gibco Gibco Gibco Sigma Gibco
HUES culture media (Harvard University)	Knockout DMEM 10% (v/v) Knockout Serum Replacement 10% (v/v) Plasmanate 2mM Gluta-MAX Non-Essential Amino Acids (NEAA) 10ng/mL basic fibroblast growth factor (bFGF) 0.1% (v/v) β -Mercaptoethanol 100unit/ml penicillin/streptomycin	Gibco Gibco Bayer Gibco Gibco Invitrogen Sigma Gibco

Freezing media:

Type	Composition	Source
hES ES- HEPES media	Dulbecco's modified eagle medium (DMEM) 20% (v/v) defined fetal bovine serum (dFBS) 20mM HEPES	Gibco Hyclone Gibco
hES 10% Vitrification solution	ES-HEPES media 10% (v/v) Ethylene Glycol 10% (v/v) Dimethyl sulfoxide (DMSO)	See above Sigma Sigma
hES 20% Vitrification solution	ES-HEPES media 0.3M Sucrose 20% (v/v) Ethylene Glycol 20% (v/v) Dimethyl sulfoxide (DMSO)	See above Sigma Sigma Sigma
HUES Freezing media	Defined fetal bovine serum (dFBS) 10% (v/v) Dimethyl sulfoxide (DMSO)	Hyclone Sigma

Gelatin- and matrigel-coated plate preparation

1% gelatin (w/v, diluted with distilled water) (Sigma) stock solution was autoclaved and stored at -20°C. Plates were then coated by incubating 0.1% gelatin solution on them for at least one hour in a 5% CO₂ incubator. Growth factor-reduced matrigel matrix (BD Biosciences) was diluted 1:20 in cold DMEM (Gibco), dispensed onto dishes, and plates prepared by incubating at 4°C overnight to form a thin matrigel layer on the culture dish.

Mouse embryonic fibroblast feeder plate preparation

MEFs were cultured until passage 7, when they were mitotically inactivated. A confluent T175 flask was incubated with 10µg/ml mitomycin C (Sigma) in culture media for three hours in a 5% CO₂ incubator. These inactivated MEFs were washed with culture media and collected via trypsinization, and quantified using a haemocytometer. 1.2 x10⁶ cells were frozen down per vial (enough to plate 6 Falcon centre well organ culture dishes or three 3cm dishes). MEFs were either thawed or seeded directly from fresh, collected cultures, and 1.5 x10⁵ cells were plated onto each gelatin-coated organ culture dish (twice

this number for a 3cm dish). These were incubated overnight to allow for cell attachment before they were ready for hESC plating.

hES2 cell culture

hES2 cells were cultured on K4 MEF-plated organ culture dishes (Falcon). hES2 colonies were mechanically cut and dissociated using a 27G½ needle as a scalpel (6 pieces from each large colony), and plated onto MEF dishes (colonies or colony pieces that looked differentiated were not passaged; 6-8 colonies were plated per organ culture dish). Media was changed everyday; hES2 colonies were passaged every 7 days. Colony manipulation was performed on an Olympus microscope equipped with a heated stage at 37°C at 12.5X magnification.

hES2 vitrification

hES2 cells were frozen via the vitrification process. Briefly, hES2 colonies were dissociated into large pieces (4 pieces from each large colony), and placed in ES-HEPES media for 5min. The pieces were transferred to the 10% vitrification solution for 1min, and then to the 20% vitrification solution for 25sec. A maximum of 8 colony pieces were transferred together via micropipette into a small 20µl droplet. All the colony pieces were then transferred to a smaller 3µl droplet. Subsequently, the colony pieces were drawn into a vitrification straw via capillary action, flash frozen in liquid nitrogen, and placed in a 5ml perforated cryovial (Nalgene) for storage in liquid nitrogen.

hES2 thawing

hES2 colony pieces were thawed by placing the vitrification straw directly from liquid nitrogen into 0.2M sucrose/ES-HEPES solution at 37°C. All the colony pieces, which were expelled from the straw by the thawing vitrification solution, were then transferred to another well containing 0.2M sucrose/ES-HEPES solution. Thereafter, they were transferred into ES-HEPES media for 5min, and transferred to another well with ED-HEPES media for an additional 5min, before finally being plated onto a MEF-layered dish. Media was changed everyday thereafter, and hES2 colonies were ready to be passaged after 10 to 14 days.

HUES cell culture

HUES hESCs were cultured on PMEF or PMEF-N MEF-layered 3cm dishes and passaged via trypsinization. Confluent cultures were washed with PBS and dissociated via incubation with 600µl of 0.05% trypsin-EDTA per 3cm dish for 5min and resuspended with a 1ml micropipette. 600µl of HUES culture media was then added to the cell suspension to neutralize the trypsin. 240µl of this was seeded into each MEF-layered 3cm dish (equivalent to a 1:5 split), and topped up to 2ml with HUES culture media. Media was changed everyday and cultures were usually confluent and ready for the next passage in 5-7 days. Neomycin-resistant stable HUES8 lines were cultured with 200µg/ml (pλIR and pλER cell lines) or 500µg/ml (ZUN cell lines) of G418 antibiotic (Gibco) supplemented in HUES culture media.

HUES cells were frozen via the slow cooling method. Briefly, HUES cultures were washed with PBS, trypsinized and neutralized with HUES culture media. The cells were then resuspended in HUES freezing media distributed into cryovials (500 μ l per vial). The vials were placed into a Mr. Frosty freezing aid and incubated at -80°C overnight. The next day, they were transferred to a -150°C freezer. For thawing, frozen cryovials were incubated in a 37°C water bath until the frozen contents were almost fully thawed and then added drop-wise to 10ml of HUES culture media. Cells were collected by centrifugation at 600g for 3min, re-suspended in 2ml of HUES culture media, and plated out onto fresh MEF-layered 3cm dishes.

III.2.4 Differentiation protocols

Type	Composition	Source
Differentiation media	Knockout DMEM	Gibco
	10% (v/v) Fetal bovine serum	Gibco
	10% (v/v) Plasmanate	Bayer
	2mM Gluta-MAX	Gibco
	Non-Essential Amino Acids (NEAA)	Gibco
	0.1% (v/v) β -Mercaptoethanol	Sigma
	100unit/ml penicillin/streptomycin	Gibco

Spontaneous differentiation

hESCs were allowed to differentiate spontaneously without MEFs on two different substrates, gelatin (Sigma) or matrigel (BD Biosciences), whilst keeping them in HUES culture media without G418 for 12 days. In ablation experiments, where 500 μ g/ml of G418 was added from the 12th to the 14th day of differentiation, alkaline phosphatase staining (see below) was performed on the 14th day to gauge pluripotency of these cultures. Long term matrigel differentiation was achieved by keeping these cells in

continuous culture for 25 days, passaging whenever confluency reached 90%. 500µg/ml of G418 was then added till the 30th day, when comparative photomicrographs were taken.

Embryoid body formation

Embryoid body formation was induced by plating 2×10^6 hESCs onto a 60mm low-cell binding plate (Nunc) in differentiation media for 21 days without changing media, before photomicrographs were taken and cells harvested for analysis.

Teratoma formation

Serially passaged hESCs were manually harvested and injected as clumps with an approximate cell dose of $2-4 \times 10^6$ cells in a volume of 50 µl into the quadriceps of the right hind limb of a male SCID mouse (3 mice per hESC line). Mice were maintained under controlled conditions in accordance with the National Institutes of Health (NIH) and National Advisory Committee for Laboratory Animal Research (NACLAR) guidelines, and with approval of the Biopolis Institutional Animal Care and Use Committee (Biopolis IACUC approval 050008, National University of Singapore Institutional Review Board 05-020). Teratoma formation was monitored visually using a simple grading system that was confirmed by caliper measurements (grade 0 - no detectible enlargement, grade 1 - enlargement just detectible, grade 2 - obvious enlargement, grade 3 - enlargement impedes locomotion). When teratomas reached grade 3 (after 6-8 weeks), teratomas weighing 1-2 g were excised and analysed via hematoxylin and eosin staining (work was performed by S.T. Wang and H. Hentze at ESI, Singapore).

Directed differentiation

Human ESCs cultured on MEFs were directed to differentiate via addition of 10 μ M RA (for 12 days) or 1% DMSO (for 7 days) to differentiation media (without G418). These were then subjected to qRT-PCR analysis, immunostaining, or ablation experiments where 500 μ g/ml of G418 was added for two days, and photomicrographs compared.

Long term directed differentiation was achieved by keeping these cells in continuous culture for 30 days, passaging whenever confluency reached 90%. Alkaline phosphatase staining was conducted on the 30th day to detect remaining pluripotent hESCs. In parallel, 500 μ g/ml of G418 was added on the 25th day, and by the 30th day there were no surviving cells for both treatments.

III.3 Transfection

III.3.1 Electroporation

HeLa cultures

HeLa cells were passaged the day before electroporation. 1×10^7 cells per transfection experiment were collected, neutralized, and re-suspended in 400 μ l of RPMI 1640 (Gibco) containing 10 μ g of each plasmid. This DNA/cell suspension was subjected to an exponential pulse (300V, 950 μ F, 300 Ω) in a 2mm cuvette using a GenePulser (Biorad). Floating cell debris was then removed with a 1ml pipette, and the remaining cell solution was transferred into a 10cm dish with culture media.

HUES8 cultures

HUES8 cells (about 8×10^5 per transfection well of a 12-well plate) were trypsinized, neutralized and re-suspended in 800 μ l of HUES culture media containing 10 μ g of DNA. This DNA/cell suspension was subjected to a single 320V, 200 μ F pulse in a 4mm cuvette using a GenePulser (Biorad), and subsequently plated onto MEF-layered 12-well plates. HUES culture media was changed for the following two days before cultures were harvested via trypsinization for analysis, or maintained for generation of stable cell lines (see below).

III.3.2 Reagent-based protocols

Name	Nature	Source
Exgen 500	Polymer (polyethylenimine)	Fermentas
GenePorter2	Cationic lipid	Gene Therapy Systems
Lipofectamine 2000	Cationic liposome	Invitrogen
Fugene 6	Propriety lipid blend	Roche
Effectene	Non-liposomal lipid	Qiagen
SAINT-MIX	Synthetic amphiphilic pyridinium surfactant	Synvolux Therapeutics
DharmaFECT 2	Cationic lipid	Dharmacon

hES2: General transfection protocol

The hES2 colonies were passaged onto matrigel-coated 4-well dishes (4 colonies or about 2×10^5 cells each well at the time of transfection) 4 days before transfection to remove as much MEF contamination as possible. Transfection media was removed after 24hr and hES culture media added for another 24hr before cultures were harvested via trypsinization for analysis.

hES2: Exgen 500

Exgen 500 was used at both 7:1 and 9:1 N/P molecular ratios (Nitrogen in Exgen 500 to Phosphate in DNA); amount of DNA transfected was kept constant at $2 \mu\text{g}$. The DNA was first diluted in $100 \mu\text{l}$ of 150mM NaCl (Sigma), mixed with appropriate amounts of Exgen 500, and incubated for 10min. This mixture was then added to the hES2 colonies on matrigel-coated 4-well dishes (to $500 \mu\text{l}$) for 24hr.

hES2: GenePorter2

GenePorter2 was used with both diluents A and B. 10 μ l of GenePorter2 was added to 40 μ l of hES culture media and 2 μ g of DNA was added to 50 μ l of diluent separately and incubated for 5min. Geneporter/DNA lipoplexes were formed after mixing the above reagents and incubating for 10min. This mixture was then added to the hES2 colonies on matrigel-coated 4-well dishes (to 500 μ l) for 24hr.

hES2: Lipofectamine 2000

Lipofectamine 2000 was used at 4:2 or 6:2 ratios (μ l of Lipofectamine to μ g of DNA). Lipofectamine and DNA were incubated separately in 50 μ l of hES culture media each for 15min. Lipofectamine/DNA complexes were formed by mixing the above reagents and incubating for 15min. This mixture was diluted into 500 μ l of hES culture media and added to the hES2 colonies on matrigel-coated 4-well dishes for 24hr.

hES2: Fugene 6

Fugene 6 (Roche) was used at 15:1, 20:1 and 30:1 ratios (μ l of Fugene to μ g of DNA); amount of DNA transfected was either 2 μ g or 4 μ g for transfection efficiency or recombination efficiency experiments respectively (2 μ g of substrate and expression plasmids each). Fugene 6 was first diluted in 500 μ l of hES culture media and incubated for 5min. DNA was then added to the above, and incubated for a further 15min. Thereafter, this mixture was added to the hES2 colonies on matrigel-coated 4-well dishes for 24hr.

hES2: Effectene

Effectene was used at 10:1, 25:1 and 50:1 ratios (μl of Effectene to μg of DNA); amount of DNA transfected was either $2\mu\text{g}$ or $4\mu\text{g}$ for transfection efficiency or recombination efficiency experiments respectively. DNA was diluted in $150\mu\text{l}$ of Buffer EC and $8\mu\text{l}$ of Enhancer per μg of DNA was added to this mixture and incubated for 5min. Effectene was then added to the above, and incubated for a further 10min for complex formation. This mixture was diluted into $500\mu\text{l}$ of hES culture media and added to the hES2 colonies on matrigel-coated 4-well dishes for 24hr.

hES2: Effectene with silica microspheres

The hES2 effectene transfection protocol above was modified to include inert silica microspheres (Cornell University) (Luo and Saltzman, 2000). 1×10^9 silica microspheres were added 5min after the start of the final incubation period, with the rest of the protocol unchanged.

HUES8: SAINT-MIX and SAINT-18

HUES8 cells were passaged onto MEF-layered 24-well dishes (about 3×10^5 cells per well at the time of transfection) 3 days before transfection. Both SAINT-MIX and SAINT-18 were tested with $2\mu\text{g}$ of DNA. DNA was diluted in $100\mu\text{l}$ of HEPES buffered salt, and $40\mu\text{l}$ of either SAINT reagent was added to $60\mu\text{l}$ of HEPES buffered salt separately and incubated for 5min. SAINT/DNA complexes were formed by mixing the above reagents, incubating for 15min, and then adding this, plus $300\mu\text{l}$ of HUES culture media, to the

cells. Transfection media was removed after 24hr and HUES culture media added for another 24hr before cultures were harvested via trypsinization for analysis.

HUES8: Effectene

HUES8 cells were passaged onto MEF-layered 24-well dishes (about 3×10^5 cells per well at the time of transfection) 3 days before transfection. Effectene transfection was then performed as in the hES2 protocol. Transfection media was removed after 24hr and HUES culture media added for another 24hr before cultures were harvested via trypsinization for analysis.

HUES8: DharmaFECT 2

HUES8 cells were transfected with DharmaFECT 2 in suspension. $2\mu\text{g}$ of DNA was diluted in $50\mu\text{l}$ of HUES culture media, and $10\mu\text{l}$ of DharmaFECT 2 was added to another $40\mu\text{l}$ of hES culture media separately and incubated for 5min. DharmaFECT/DNA complexes were formed by mixing the above reagents and incubating for 20min. HUES8 cultures were trypsinized and neutralized, mixed with the above complexed reagents, and plated onto MEF-layered 24-well dishes (about 3×10^5 cells in a total of $500\mu\text{l}$ of media in each well). Transfection media was removed after 24hr and HUES culture media added for another 24hr before cultures were harvested via trypsinization for analysis.

III.3.3 Flow cytometry

Cells were washed with PBS 48 hours after transfection and harvested via incubation with 0.05% Trypsin/EDTA for 5 minutes (both Gibco). Cells were collected, washed once with 5ml of PBS, and re-suspended in 100 μ l or 200 μ l of PBS (depending on cell density) for analysis in a FACSCalibur Flow Cytometer (Becton Dickson). A minimum of 20,000 total events were recorded per sample and analyzed using CELLQuest software (Becton Dickson). Briefly, dot plots of side scatter (SSC, indicative of cell granularity) versus forward scatter (FSC, indicative of cell size) were constructed. This was used to gate live, intact cells (R1) from cell debris at the bottom left, and cell aggregates at the top right of the dot plot (**Fig. III.9**). The percentage of EGFP or FITC-positive cells was then elucidated by looking at FL1 levels within all R1-gated cells, plotted against cell size (FSC). All transfection efficiency values were obtained after gating mock transfected cells to zero from the dot plot. Transfection or recombination efficiencies were then gleaned from the percentage of cells in top right quadrant. Positive control values were taken as 100% for efficiency quantification experiments.

Figure III.9. Schematic of flow cytometry analysis. Samples were first gated to select for live, intact cells (R1). FL1 levels indicate the intensity of fluorescence in a particular cell, and X% positive values were obtained by gating the mock transfected cells to 0%.

III.3.4 Stable cell line generation

HeLa stable cell line generation

One passage after electroporation, *neo*-expressing cells were selected via the addition of 200µg/ml of G418 to culture media, which is then changed frequently to discard dead cells. Colonies that grow out after two weeks were mechanically picked, transferred individually into 96-well plates, and expanded.

hES2: Cut/Paste protocol

Construction of EGFP-positive transgenic hES2 cell lines was attempted via a Cut/Paste strategy (**Fig. III.10A**). After Effectene transfection, the locations of EGFP-positive cells were marked, and small colony pieces were cut around these cells. The pieces were then transferred to another MEF-layered dish and allowed to expand.

HUES8: Select/Expand protocol

p λ IR, p λ ER and ZUN neomycin-resistant transgenic HUES8 cell lines were constructed via a Select/Expand strategy (**Fig. III.10B**). Transfection reagent Effectene was used to introduce 2 μ g of circular plasmid DNA (p λ IR, p λ ER and pTZ-UTF1-Neo respectively) into each well of a 24-well plate containing about 3×10^5 HUES8 cells as described above. A total of 24 wells were transfected for each plasmid. After three days, these transfected HUES8 cultures were trypsinized and passaged onto G418-resistant PMEF-N-layered 3cm dishes (cells from every two wells of a 24-well plate were combined onto one 3cm dish). 200 μ g/ml of G418 antibiotic selection was initiated on day 5, and G418-resistant, expanding colonies become obvious on day 15. Large colonies were mechanically dissociated with a pipette tip on day 21 and transferred to a PMEF-N-layered 96-well plate. These colonies were expanded via trypsin to progressively larger wells, and maintained on PMEF-N-layered 3cm dishes, with HUES culture media supplemented with 200 μ g/ml G418. p λ IR and p λ ER cell lines were then characterized and frozen. Before the same was done for ZUN cell lines, an additional elevated selection step was performed. G418 concentration was increased to 500 μ g/ml after 5 passages to select for cells expressing a higher level of the *Neo* transgene.

Figure III.10. Protocols for the generation of stable hESC lines. (A) Cut/Paste protocol for hES2 cells. EGFP-positive cells were isolated via mechanical dissociation and re-plated. However, no EGFP-positive cells were detected in these transferred colonies. (B) Select/Expand protocol for HUES8 cells. Neomycin-resistant cells were selected for and expanded via the addition of 200 μ g/ml of G418 antibiotic.

III.4 Characterization of cell lines

III.4.1 Pluripotency assays

Alkaline phosphatase staining

Alkaline phosphatase (AP) is a hydrolase enzyme responsible for catalyzing dephosphorylation and is highly expressed in embryonic stem cells. AP staining was performed on hESCs using the ES cell characterization kit (Chemicon). Briefly, the cells were fixed with 90% Methanol/10% Formalin (Merck) for 1min, washed with rinse buffer (see below), and incubated with Napthol/Fast Red Violet/water at a 2:1:1 ratio for 15min in the dark. The cells were then washed with rinse buffer and photomicrographs were taken using the IX71 and DP70 digital imaging system (Olympus).

Immunostaining

Materials:

Type	Composition	Source
Rinse buffer (TBST)	20mM Tris-HCl, pH 7.4 0.15M NaCl 0.05% Tween-20	All from Sigma
Primary antibody	Rat anti-mouse Thy-1, IgG-Texas Red conjugated	Santa Cruz
Primary antibody	Mouse SSEA-1, IgM, monoclonal	Chemicon
Primary antibody	Mouse SSEA-4, IgG, monoclonal	Chemicon
Primary antibody	Mouse TRA-1-60, IgM, monoclonal	Chemicon
Primary antibody	Mouse TRA-1-81, IgM, monoclonal	Chemicon
Primary antibody	Mouse Oct4, IgG, monoclonal	Santa Cruz
Primary antibody	Mouse Sox2, IgG, monoclonal	R&D Systems
Secondary antibody	Anti-Mouse F(ab') ₂ IgG-FITC conjugated	Chemicon
Secondary antibody	Anti-Mouse IgM-FITC conjugated	Santa Cruz

Human ESCs to be stained were cultured on MEF-layered 96-well black/clear bottom plates (Greiner). At 70% confluency, they were first fixed in 4% Paraformaldehyde

(Sigma)/PBS for 15min, washed twice with rinse buffer, and then permeabilized with 0.1% Triton X-100 (Sigma)/PBS for 10min. Subsequently, the cells were washed twice with rinse buffer and blocked with 4% goat serum (Santa Cruz)/PBS for 30min. The hESCs were incubated with the primary antibodies (1:20 dilution with blocking solution) for 60min and washed thrice with rinse buffer. Secondary antibodies were then added (1:50 dilution with PBS) to the hESCs for 60min and washed thrice with rinse buffer. Photomicrographs were taken using the IX71 and DP70 digital imaging system (Olympus).

Immunostaining for flow cytometry

For flow cytometric analysis of Oct4 and Sox2 immunostaining, 2µl each of primary and secondary monoclonal antibodies were pre-incubated for 30min in 3% BSA/PBS to a total of 50µl per experiment. The hESCs were collected via trypsinization and washed once with PBS. These cells were then fixed with 500µl of 2% Paraformaldehyde/PBS for 20min, washed with 3ml of PBS, and permeabilized with 0.1% Triton X-100/PBS for 5min. After another 3ml of PBS was added to wash the cells, 5×10^5 cells were resuspended in 150µl of 3% BSA/PBS for each experiment. The 50µl antibody mix was incubated with this for 30min before analysis via flow cytometry. For TRA antibody flow cytometry immunostaining, fixation and permeabilization steps were removed from the above protocol. These experiments were conducted with at least two biological and three assay repeats, with corresponding samples incubated only with relevant secondary antibodies as negative controls. The Student's T-test was used to gauge statistical significance between samples.

SCID mouse teratoma assay

To test the differentiation potential of hESC lines *in vivo*, teratomas were generated using SCID mice as detailed above. When teratomas reached grade 3 (after 6-8 weeks), teratomas weighing 1-2 g were excised, fixed (Bouin's solution), paraffin-embedded and sectioned. These sectioned teratomas were then stained with hematoxylin and eosin (H&E), and histologically analysed via comparison to standard H&E staining patterns for differentiated tissue. Additionally, teratomas were assessed by a pathologist at the Department of Pathology, National University of Singapore (work was performed by S.T. Wang and H. Hentze at ESI, Singapore).

III.4.2 Transgene validation

p λ IR and p λ ER Cell lines

Stable cell lines randomly integrated with p λ IR or p λ ER were generated with both HeLa and HUES8 cell lines via G418 selection as described. Nested PCR was used to detect the presence of the transgenic cassettes in the genome. Briefly, genomic DNA from individual cell lines was isolated using the DNeasy Blood and Tissue Kit (Qiagen). 1 μ g of this was used as a template for the primary PCR with 50pMol each of pLI/ER-F (TACGGTGGGAGGTCTATA) and pLI/ER-R (CTTGAAGAAGTCGTGCTG) primers in a 50 μ l PCR reaction, conducted with *Taq* DNA polymerase and Q solution (Qiagen) at 55°C annealing temperature and a 6min extension step. Nested PCR was conducted using 2 μ l of the primary PCR as template and 50pMol each of pLI/ER-NF (GCAGAGCTGGTTTAGTGAA) and pLI/ER-NR (TTCATGTGGTCGGGGTAG) primers, with the same PCR conditions. Products were viewed on an agarose gel; a size

of 4.1kb was expected if the whole cassette was successfully integrated into the target genome. These products were sent for sequencing to verify the genomic integrity of the transgenic insert.

ZUN Cell lines

Stable cell lines randomly integrated with *pTZ-UTF1-Neo* were generated with HUES8 via G418 selection as described. Genomic PCR was used to detect the presence of the transgenic cassette in the genome. Briefly, genomic DNA from individual cell lines was isolated using the DNeasy Blood and Tissue Kit (Qiagen). 500ng of this was used as a template for each PCR reaction. The UTF1P-S3F (AGGACCGACCCCTTAGGAC) and Neo-R (ACTGTGTACATCAGAAGAACTCGTCAAGAA) primer pair (50pMol each) was used to detect the *UTF1* promoter and *Neo* transgene, while the UTF1P-S3F and UTF1-S10R (GGGCAAACCACTTGTCTTTC) primer pair was employed to detect the full transgenic cassette in separate, standard 20 μ l PCR reactions. These were conducted using *Taq* DNA polymerase and Q solution (Qiagen), at 55°C annealing temperature, a 4min extension step and 30 cycles. Products were then viewed on an agarose gel; a size of 1.9kb was expected if the whole cassette was successfully integrated into the target genome. These products were sent for sequencing to verify the genomic integrity of the transgenic insert.

Southern blot

Southern blotting was performed using a well-established protocol in our lab, as previously described (Christ and Dröge, 2002; Corona *et al.*, 2003). 10 μ g of isolated genomic DNA

from each sample (HUES8, ZUN1 and ZUN2) was isolated using the DNeasy Blood and Tissue Kit (Qiagen) and digested with 100U of EcoRI (NEB) overnight in a 20 μ l reaction. These three samples were subsequently loaded and separated on a 0.8%, 0.5X TBE agarose gel overnight. The gel was depurinated with 0.25M HCl for 10min and denatured with 0.4M NaOH, 1M NaCl for 40min. In the following alkaline transfer step, DNA fragments were transferred onto a positively-charged nylon membrane overnight with 0.4M NaOH, 1M NaCl. The membrane was then neutralized with two 1x SSC washing steps, and treated with pre-hybridization buffer (1M NaCl, 50mM Tris-Cl, pH=7.5, 1% SDS, 0.5mg/ml salmon testes DNA, 10% dextran sulfate) for 3hr at 65°C.

In the meantime, the *Neo*-specific radioactive probe was amplified from primers pPR-F (CAAGATGGATTGCACGCAGG) and pPR-R (AGCCGGCCACAGTCGATG) using the standard protocol from the MegaPrime random labeling kit (Amersham). Hybridization was performed by adding 10⁷ cpm of P32-labeled probe to the pre-hybridization buffer (0.1ml for each 1cm² of membrane area) and incubated overnight together with the membrane at 65°C. The following day, the membrane was washed from low to high stringency: 2x SSC, 0.1% SDS (15min), 1xSSC, 0.1% SDS (15min) and 0.5x SSC, 0.1% SDS (15min), and visualized using Biorad's Personal Molecular Imager System (1xSSC: 0.15M NaCl, 0.015M sodium citrate, pH 7.0; All chemicals from Sigma; Radioactive work was performed by R.J. Zhou).

III.4.3 Karyotyping

Karyotyping of HUES8 and ZUN cell lines via Giemsa staining was performed by the Cytogenetics Lab in KK Women's and Children's Hospital, Singapore. 50 μ l of BrdU/Colcemid (50:50) mixture was added to an 80% confluent 35mm dish containing hESCs with 2ml of media the night before staining and analysis. More than 40 distinct, separate, normal chromosomal spreads were analyzed per sample before a normal karyotype was declared.

III.4.4 Proliferation assay

To analyze the growth rate of HUES8 and ZUN cell lines, the cell proliferation reagent WST-1 (Roche) was used. This colorimetric assay is based on the cleavage of the tetrazolium salt WST-1 to formazan by mitochondrial dehydrogenases found in viable cells. Expansion in the number of viable cells results in an increase in the overall activity of the mitochondrial dehydrogenases in the sample, thus increasing the amount of formazan dye produced by these cells which can then be quantified by measuring the absorbance of the dye solution at 440nm using a standard photoplate reader. For our purposes, 3x10⁴ hESCs were seeded onto each MEF-layered 96-well at day 0. WST-1 reagent was added to the wells for 30min the next day, and individual absorbances were recorded. This was repeated for a total of 4 days with different plates; each cell line had five replicates per time-point.

III.4.5 Modulation of selection pressure

The selection pressure for ZUN hESCs was removed and subsequently reinstated to analyze the effects of antibiotic withdrawal on the pluripotency of these cells. ZUN cells were cultured with HUES culture media (without G418) for 30 days or 60 days. A third ZUN sample set was cultured with HUES culture media (without G418) for 30 days, with the re-introduction of 500 μ g/ml of G418 selection pressure throughout the following 30 days. ZUN cultures were passaged as per normal during this treatment.

III.4.6 Transcript knockdown

The amount of functional *UTF1* transcripts in HUES8 hESCs was reduced via RNA interference (RNAi) technology. ON-TARGETplus SMARTpool UTF1 short interfering RNA (siRNA) (accession number: NM_003577) was transfected via DharmaFECT 2 (both Dharmacon) according to the manufacturer's protocol. 3×10^5 HUES8 cells in a total of 500 μ l in a 24-well were transfected for three consecutive days. The cells were transfected in suspension on the first day and as adherent cells for the subsequent two days. Two days thereafter, cells are photographed using the IX71 and DP70 digital imaging system (Olympus), and harvested for transcript analysis via qRT-PCR.

III.4.7 Transcript analysis

cDNA amplification

cDNA amplified from total RNA was used as the template for RT-PCR reactions to analyze transcript expression levels. After the collection of cells via trypsinization, total RNA was recovered from cell pellets using the RNA extraction kit (Qiagen) following

the manufacturer's protocol, and quantified via spectrometric analysis. All RNA samples had an ideal purity, with an absorbance ratio (A_{260}/A_{280}) of between 1.8 and 2.2. cDNA was synthesized from 500ng of total RNA per sample via the SSII reverse transcriptase kit (Invitrogen), primed with oligo dTs in a reaction volume of 20 μ l according to the manufacturer's protocol. This effectively generates cDNA for all the mRNA transcripts in each sample. The resulting cDNA mixture was diluted 4x to 80 μ l.

Reverse transcription PCR

We first used reverse transcription PCR (RT-PCR) to analyze and compare transcript expression levels. RT-PCR was performed using 4 μ l of diluted cDNA from the reverse transcription reaction above as template, *Taq* DNA polymerase, Q Solution (Qiagen) and a standard thermocycler. PCR conditions were: 30sec for denaturation at 95°C, followed by 30sec annealing at 60°C, 35sec extension time at 72°C and 30 cycles in a 20 μ l reaction volume with 20pMol of each gene-specific primer pair. No final extension step was included. 5 μ l of each completed reaction was run on a 1.5% agarose gel (Biorad) and stained with ethidium bromide (Biorad) for 10min before visualization with Biorad's gel documentation system.

Quantitative real-time PCR

Quantitative real-time PCR (qRT-PCR) was used to provide a higher quantitative resolution compared to RT-PCR for transcript expression levels. The SYBR Green qRT-PCR system (Qiagen) was used, with 20 μ l reaction samples prepared for each experiment as per manufacturer's protocol, with 4 μ l of diluted cDNA from the reverse transcription

reaction above as template. The qRT-PCR was conducted using the 7500 Real-Time PCR System. PCR conditions were: 40 cycles, 30sec for denaturation at 95°C and annealing at 60°C, 35sec for extension at 72°C, plus the programmed dissociation step (95°C/15sec followed by 60°C/60sec and 95°C/15sec) to determine the specificity of the PCR primers. Results were analyzed using sequence detection software version 1.2.3 (SDS, both Applied Biosystems, USA), with the threshold set at 0.08. All experiments had at least biological duplicates and qRT-PCR assay triplicates (with only one dissociation peak). Threshold cycle (C_T) values were normalized to corresponding GAPDH results to achieve ΔC_T values, and transcript expression fold change quantitated using the $\Delta\Delta C_T$ method. Validation agarose gels were also run separately for each of the gene-specific primer pairs with their respective templates to ensure that only one PCR product was amplified and thus detected by the software.

Primers used for transcriptional profiling:

Gene	Primers	Product Size (bp)	Source
GAPDH	CCCACTCCTCCACCTTTGAC CTCCCCTCTTCAAGGGGTCT	283	This study
Nanog	AGTCCCAAAGGCAAACAACCCACTTC ATCTGCTGGAGGCTGAGGTATTTCTGTCTC	164	Bhattacharya <i>et al.</i> , 2004
Oct4	CTTGCTGCAGAAGTGGGTGGAGGAA CTGCAGTGTGGGTTTCGGGCA	171	Ginis <i>et al.</i> , 2003
Sox2	ATGCACCGCTACGACGTGA CTTTTGCACCCCTCCCATT	436	Ginis <i>et al.</i> , 2003
UTF1	AGCAGATCCGGAAGCTCATGGG TCCTCGGGGATGCAGGTG	295	This study
UTF1-Neo	GGGTACCCCATGATTGAACA CAGGTCGGTCTTGACAAAAAG	168	This study
Sox9	GAAGCTCGCGGACCAGTA CGTTCTTCACCGACTTCCTC	190	This study
Nestin	CGTTGGAACAGAGGTTGGAG GAGCGATCTGGCTCTGTAGG	116	This study
ID2	CTGGACTCGCATCCCACTAT CACACAGTGCTTTGCTGTCA	166	This study
HAND1	TGCCTGAGAAAGAGAACCAG ATGGCAGGATGAACAAACAC	274	Cai <i>et al.</i> , 2006
IGF2	TCCTCCCTGGACAATCAGAC AGAAGCACCAGCATCGACTT	247	Cai <i>et al.</i> , 2006
α 1AT	AGACCCTTTGAAGTCAAGGACACCG CCATTGCTGAAGACCTTAGTGATGC	360	Schuldiner <i>et al.</i> , 2000

IV. RESULTS

IV.1 Characterization of human embryonic stem cells

IV.1.1 Morphology

The two hESC lines used mainly in this study were hES2, passaged via mechanical dissociation (Reubinoﬀ *et al.*, 2000), and HUES8, passaged by trypsin digestion (Cowan *et al.*, 2004). The two lines exhibited standard hESC morphological characteristics such as a high nuclear to cytoplasm ratio, prominent nucleoli and a highly compact cell size (Amit *et al.*, 2000). However, individual hESCs in hES2 colonies were more compact than their HUES8 counterparts due to the denser nature of the colonies. By virtue of differences in passaging protocol, hES2 colonies (**Fig. IV.1A**) were much larger than that of HUES8 (**Fig. IV.1B**). In addition, a central “button” consisting of packed, malnourished and differentiated hESCs was usually observed in hES2 colonies. These cells were not propagated and were discarded at every passage. In general, the culture of HUES8 is less labor-intensive and relatively straightforward compared to hES2.

Figure IV.1. hES2 and HUES8 colony morphologies. (A) Single hES2 colony on MEFs with differentiated “button”. Scale bar = 320 μ m. (B) HUES8 colonies on MEFs. Scale bar = 125 μ m.

Figure IV.2. HUES8 on MEFs stained with anti-mouse Thy-1. (A) Representative flow cytometry dot plot showing MEF-positive staining of about 10% in standard hESC co-culture. FL2 = Texas Red detection, FSC = Cell size. (B) Low magnification fluorescence photomicrographs showing MEF-specific Thy-1 expression (Texas Red) and unstained hESC colonies. Scale bar = 250 μ m. (C) Higher magnification fluorescence photomicrographs showing MEF-specific Thy-1 expression (Texas Red). Scale bar = 100 μ m.

IV.1.2 Mouse embryonic fibroblast quantification

As our laboratory works with hESCs co-cultured on MEF layers, we sought to quantify the percentage of MEF contribution to these cultures via immunostaining and flow cytometry analysis. Thy-1, also known as CD90, is a glycosylphosphatidylinositol (GPI)

-anchored glycoprotein and an important regulator of cell-cell and cell-matrix interactions (Rege and Hagood, 2006). It is also a well characterized fibroblast marker (Esterre *et al.*, 1992). An anti-mouse Thy-1, Texas Red-conjugated antibody was used to detect MEF cells via the standard immunostaining protocol. The fraction of MEF cells detected in standard HUES8 co-cultures was about 5-10% (**Fig. IV.2A**), and thus should not have a profound effect on the quantification of subsequent experiments. The specificity of MEF staining was validated via microscopy, and hESCs were shown to be unstained (**Fig. IV.2B** and **Fig. IV.2C**).

IV.1.3 Pluripotency and karyotyping

Alkaline phosphatase staining

The pluripotency of hESCs in culture was routinely analyzed by a number of assays. A convenient procedure to obtain a snapshot of this status is to analyze expression level of membrane alkaline phosphatase (AP) via staining. AP, a hydrolase enzyme responsible for catalyzing dephosphorylation of natural substrates such as pyrophosphate, pyridoxal 5'-phosphate and bacterial endotoxins (Coburn *et al.*, 1998; Poelstra *et al.*, 1997), has an exceptionally high expression level in embryonic stem cells, and is downregulated upon differentiation (Amit and Itskovitz-Eldor, 2006; Trounson, 2006). This phenomenon has been widely exploited to gauge pluripotency of hESC cultures, and is a relatively simple procedure of incubating the cells with Naphthol/Fast Red Violet mixture for 15min in the dark (Chemicon). Both our hESC lines routinely stained positive for AP (**Fig. IV.3**).

Figure IV.3. Alkaline phosphatase staining for hES2 and HUES8. (A) hES2 colony on MEFs with positive AP staining, except for differentiated centre “button”. Scale bar = 250µm. (B) HUES8 colonies on MEFs showing positive AP staining. Scale bar = 100µm.

Immunostaining

The expression of pluripotency markers is commonly used to characterize the “stemness” of hESC cultures. Cell surface marker proteins specific to hESCs, like the globoseries glycolipid stage specific embryonic antigens (SSEA) and the keratin sulphate-associated tumor recognition antigens (TRA) have been well characterized and documented (Amit and Itskovitz-Eldor, 2006; Trounson, 2006). Transcription factor POU5F1, also known as Oct4, is important in the maintenance of pluripotency and involved in the regulation of Sox2, Nanog, Activin and other pathways regulating hESC differentiation (Babaie *et al.*, 2007). Pluripotent hESCs express Oct4, SSEA-3, TRA-1-60 and TRA-1-81, but not SSEA-1 (Henderson *et al.*, 2002; Carpenter *et al.*, 2004). The presence of these proteins in our hESC cultures was detected routinely via immunostaining. Both our hESC lines continue to exhibit a pluripotent immunostaining profile for this panel of markers, after more than 30 continuous passages and 6 months in culture on MEFs (**Fig. IV.4**). In addition, hES2 cells also exhibited positive immunostaining four days after transfer onto matrigel from normal MEF culture conditions in preparation for transfection experiments.

Figure IV.4. Immunostaining for hES2 and HUES8. (A-E) hES2 cultures on MEFs; (F-J) HUES8 cultures on MEFs. (A,F) SSEA-1; (B,G) SSEA-4; (C,H) TRA-1-60; (D,I) TRA-1-81; (E,J) Oct4. Fluorescent photomicrographs showing negative SSEA-1 expression, but positive expression for all four other markers in both hESC lines. Scale bar = 140 μ m.

Karyotyping

hESCs in long-term culture have the propensity to suffer form karyotypic abnormalities, in particular recurrent gain of chromosomes 17q and 12 (Draper *et al.*, 2004). Indeed, some HUES lines (namely 3 and 4) were reported to feature trisomy of chromosome 12 after prolonged culture, which was accompanied by a more proliferative phenotype (Cowan *et al.*, 2004). Thus, we were concerned that HUES8, our primary hESC line, might be susceptible to karyotypic alterations. We performed a karyotypic analysis of cultures that were in continuous culture for more than three months, or 12 passages (Fig. IV.5). Giemsa staining and analysis, performed by a professional cytogenetics lab, was used to ascertain that there were no karyotypic abnormalities our HUES8 cultures.

Figure IV.5. Karyotypic analysis of HUES8 cultures. Giemsa staining of normal HUES8 hESC karyotypes cultured continuously for three months (representative photo).

IV.2 Transfection of human embryonic stem cells

IV.2.1 Transfection efficiency

Reporter plasmids

After successful culture of undifferentiated hESCs, we proceeded to the task of testing, improving and optimizing transfection procedures. EGFP was chosen as the reporter marker of choice for the readout of both transfection and recombination efficiencies via flow cytometry quantification. The plasmid *pCMVssEGFPbpa*, with a strong, ubiquitously expressed Cyto Megalo Virus (CMV) promoter driving the *EGFP* expression cassette through a hybrid intron consisting of an adenovirus splice donor and an immunoglobulin G splice acceptor (SS) (Choi *et al.*, 1991), was used for this purpose. The empty vector *pCMV* without the *EGFP* gene was employed as a negative control in all experiments involving transfection (Lorbach *et al.*, 2000) (Fig. IV.6).

Figure IV.6. Maps of *pCMV* and *pCMVssEGFPbpa*. *pCMVssEGFPbpa* was used as a reporter plasmid for the analysis of transfection efficiency, and as a positive control for transfection experiments; *pCMV* was used as the mock vector for all transfection experiments (Lorbach *et al.*, 2000).

hES2 transfection efficiency

Transfection efficiencies were determined starting with the hES2 cell line. In a preliminary assay, various concentrations of both polymer-based and lipid-based transfection reagents were used to transfect the *pCMVssEGFPbpa* reporter plasmid. The percentage of EGFP-positive cells was determined by flow cytometry 48 hours following transfection, when EGFP expression would be at an optimal level (Christ *et al.*, 2002). Efficiencies ranging from below 1% to about 6% were obtained for the various protocols (**Fig. IV.7A**). Thereafter, four optimized protocols were selected for a comprehensive comparison: the polymer-based Exgen 500 as a negative control, and lipid-based Fugene 6, Lipofectamine and Effectene. In addition, the Effectene protocol was further enhanced via the addition of silica microspheres, which serves as an uptake-enhancing component that increases the physical concentration of DNA at the cell surface (Luo and Saltzman, 2000). These ternary-complexed nanoparticles are internalized by an endosome-lysosomal route followed by perinuclear accumulation (Gemeinhart *et al.*, 2005), and have been reported to increase Effectene transfection efficiencies in COS-7 African green monkey kidney and Chinese hamster ovary (CHO) cells (Luo and Saltzman, 2000; Luo *et al.*, 2004).

The three lipid-based reagents scored an efficiency of between 4% and 6%, whilst hES2 cultures transfected with the three-component Effectene+Silica protocol consisted of more than 10% EGFP-positive cells (**Fig. IV.7B**). EGFP expression in hESCs after 48 hours was also strong and easily detectable via fluorescence microscopy (**Fig. IV.7C**). Therefore, our plasmid transfection efficiencies in the range of 10% using the commonly

available and easy-to-use Effectene in combination with silica microspheres were comparable to those achieved with more sophisticated protocols, such as nucleofection (Lakshimipathy et al., 2004).

Figure IV.7. Transfection efficiencies of various reagents in hES2. (A) Preliminary testing of various transfection reagents and protocols. (B) Selected reagents were subjected to additional tests with at least three biological repeats each. (C) Representative photomicrograph of a hES2 colony successfully transfected with pCMVssEGFPbpa using the Effectene+Silica protocol, showing EGFP-positive cells. Scale bar = 400μm.

HUES8 transfection efficiency

We proceeded to transfer our transfection expertise to HUES8 cells, and compared the Effectene+Silica protocol to four newer methods. Two polymer-based reagents continued the inefficient trend of such reagents: the SAINT-MIX transfection reagent did not generate many EGFP-positive hESCs, and SAINT-18 did not function at all in hESCs. Transfection efficiencies of both DharmaFECT 2 (lipid-based) and electroporation protocols (around 8% and 14% respectively) were higher than that of Effectene+Silica, which was only about 5% (**Fig. IV.8A**). With the electroporation protocol, efficiencies of up to 15% were routinely achieved, at the expense of cell toxicity indicated by the low number of surviving adherent cells. EGFP expression in successfully transfected HUES8 cells was also visible after 48 hours (**Fig. IV.8B**).

Figure IV.8. Transfection efficiencies of various reagents in HUES8. (A) Four transfection methodologies were compared, with at least three biological repeats each. (B) Representative photomicrograph of HUES8 cells successfully transfected with pCMVssEGFPbpa using the electroporation protocol, showing EGFP-positive cells. Scale bar = 130 μ m.

IV.2.2 Generation of stable cell lines

Following successful optimization of transfection protocols, we attempted to generate stable cell lines via initial transfection of transgenic plasmid DNA with Effectene. Unfortunately, we were unable to isolate hES2 colonies with sustained and stable expression of EGFP using the Cut/Paste technique, where EGFP-positive hES2 cells were selected via mechanical dissociation and manually passaged (**Fig. III.10A**). On the other hand, we were able to generate clonal HUES8 colonies expressing the neomycin resistance marker with the Select/Expand protocol, using the G418 antibiotic to select for cells which incorporated a *Neo* transgene and were consequentially resistant to the antibiotic (**Fig. III.10B**). These HUES8 cells were able to proliferate in the presence of G418, and eventually formed colonies which could be mechanically picked and expanded.

IV.3 Site-specific recombination

IV.3.1 Episomal site specific recombination in hES2

Experimental design

Safe and efficient transgene insertion is a potentially important technique for hESC research. SSR systems possessing high fidelity and stringency could be used to minimize the risk of insertional mutagenesis and increase the efficiency of genomic integration. To achieve this, we started by testing the episomal recombination efficiency of a panel of three site-specific recombinases and progressed to proof-of-concept genomic integration experiments with the λ integrase system in both HeLa and hESCs.

Our laboratory has recently developed mutant phage λ integrase and mutant $\gamma\delta$ resolvase SSR systems for applications in eukaryotic cells, and showed that both of these are recombinant on episomal and genomic DNA substrates (Lorbach *et al.*, 2000; Christ *et al.*, 2002; Christ and Dröge, 2002; Schwikardi and Dröge, 2000, 2001). However, because hESCs differ from other mammalian cells, including mouse ESCs, in features such as cell size, doubling time, and gene expression patterns (Sato *et al.*, 2003; Rao, 2004), we were interested in comparing the λ and $\gamma\delta$ SSR system with the widely used Cre system inside hESCs.

For this, we used substrate vectors that, when co-transfected with the respective recombinase expression vector, lead to EGFP expression as an easy read-out for recombination (**Fig. IV.9**). Specifically, three substrate vectors: *p λ IR* and *p λ ER* (Corona and Dröge, 2002), and *pCH-RLNRLE* were used, along with positive control recombined

vector *pCH-RLE*. In addition, four expression vectors were used: *pPGKssCrebpa*, *pPGK $\gamma\delta$ 102NLS*, *pCMVssInth-218bpa* and *pPGKssIHF2bpa*, for phage P1 Cre recombinase, mutant $\gamma\delta$ resolvase, mutant phage λ integrase and single-chain IHF2 co-factor respectively (**Fig. IV.10**) (Schwikardi and Dröge, 2000; Lorbach *et al.*, 2000; Corona *et al.*, 2003).

Figure IV.9. Recombination substrate vectors and experimental design. (A) *pCH-RLNRLE* was used as substrate for Cre and $\gamma\delta$ resolvase. Recombination between the directly repeated 114-bp Res sites as target sequence for $\gamma\delta$ resolvase, or between the directly repeated 34-bp Lox sites as target sites for Cre, leads to the excision of the neomycin resistance gene (*Neo*). This deletion results in EGFP expression from the *N-terminus-LacZ-EGFP-fused* coding region. (B) *pλIR* contains directly repeated attachment sites *attB* and *attP* as target sequences for the λ integrative pathway. *pλER* instead contains *attL* and *attR* as direct repeats that serve as target sequences for the excisive pathway. Recombination between *attB* and *attP* or between *attL* and *attR* deletes the promoter-*Neo*-transcriptional stop (PGKNeo-TSS) cassette and results in EGFP expression.

Figure IV.10. Maps of expression vectors. (A) *pPGKssCrebpa* encodes for wild type Cre recombinase, (B) *pPGKγδ102NLS* for modified γδ resolvase with a nuclear localization signal (NLS), (C) *pCMVssInth-218bpa* encodes for mutant cofactor-independent λ integrase Int-h/218, and (D) *pPGKssIHF2bpa* for engineered single chain IHF protein scIHF2.

Cre and γδ-mediated recombination in hES2

In the first series of recombination assays, we analyzed the recombination efficiencies catalyzed by Cre and γδ resolvase via the co-transfection of *pCH-RLNRLE* and respective expression vector. As a positive control, the co-transfection of recombined product vector *pCH-RLE* and mock vector *pPGK* was conducted and the percentage of EGFP-expressing

cells was taken as 100% recombination efficiency (**Fig. IV.9**). We found that Cre recombined *pCH-RLNRLE* in more than half of the transfected cells (**Fig. IV.11**). In contrast, only 5% of transfected cells displayed detectable EGFP expression due to recombination of *pCH-RLNRLE* by $\gamma\delta$ resolvase. The latter finding is in agreement with previous results which showed that $\gamma\delta$ resolvase was almost as efficient as Cre in murine cells, but significantly less so in human cells (Schwikardi, 2001).

λ integrase-mediated recombination in hES2

In a second series of assays, we co-transfected *pCMVssEGFPbpa* together with *pCMV* mock vector as positive control and demonstrate that λ Int-h/218 was able to recombine the substrate for integrative recombination, *p λ IR*, in 20% of transfected cells (**Fig. IV.11**), which was similar to that described for HeLa cells (Christ *et al.*, 2002). Surprisingly, this number dropped to approximately 4% with *p λ ER*, the substrate for excisive recombination. This difference in the number of hESCs expressing EGFP was also observed when the transfected colonies were viewed under the microscope (**Fig. IV.12**), and deviated from previous findings with HeLa cells in which this difference was less than twofold (Christ *et al.*, 2002). This interesting observation suggests that λ integrase-mediated recombination in hESCs might be biased towards integrative (*attB* x *attP*) recombination, and deserves further investigation.

Figure IV.11. Comparative analysis of three SSR systems in hESCs. Cells were co-transfected with substrate and respective recombinase expression vector. In each case, the recombination efficiency was determined 48 hours after transfection. Experiments were repeated at least three times. Based on the two-tailed paired Student’s t-test, all values are significant compared with respective negative controls, and p values are always < 0.05.

Figure IV.12. Visual comparison of Int-h/218-mediated integrative and excisive episomal recombination in hESCs. Merged fluorescence photomicrographs showing the higher efficiency of integrative recombination (*pλIR+pCMVssInt-h-218bpa*) compared to excisive recombination (*pλER+pCMVssInt-h-218bpa*) as evidenced by the low number of EGFP-positive cells in the right panel. Scale bar = 75 μ m.

IV.3.2 λ integrase-mediated genomic site-specific recombination

Our laboratory has been successful in using mutant phage λ integrase-mediated genomic SSR to excise an *att*-flanked cassette targeted to the *ROSA26* locus in mouse embryonic stem cells (Christ and Dröge, 2002). Our results above have shown that episomal SSR mediated by mutant λ integrase have similar efficiencies in both human HeLa and embryonic stem cells, with the exception of a more pronounced bias towards *attB* x *attP* over *attL* x *attR* recombination. Therefore, we decided to evaluate the potential of λ integrase SSR-mediated resistance marker gene excisions in the genome of both HeLa cells and HUES8 hESCs. To achieve this, we randomly inserted both λ integrase substrate plasmids *p λ IR* and *p λ ER* (with *attB* and *attP* or *attL* and *attR* flanking the *Neo* gene respectively) into HeLa cells via electroporation, or HUES8 cells via Effectene transfection and subsequent G418 antibiotic selection. These stable cell lines were picked, expanded and presence of genomic inserts characterized via nested PCR (Fig. IV.13).

Figure IV.13. Nested PCR strategy for cell line characterization. The presence of the entire plasmid insert was assessed by first amplifying genomic DNA from individual cell lines using the I/ER primer pair (TACGGTGGGAGGTCTATA and CTTGAAGAAGTCGTGCTG). A secondary nested PCR was then performed with the first as template, and the I/ER-N primer pair (GCAGAGCTGGTTTAGTGAA and TTCATGTGGTCGGGGTAG). A product of 4.1kb would be expected for both HeLa and HUES8 stable cell lines if the entire cassette was inserted into the genome.

Characterization of stable cell lines

The nested genomic PCR assay was conducted as described above on all our HeLa and HUES8 stable cell lines. The presence of the entire cassette, as indicated by the nested PCR product sized at 4.1kb, was detected in a total of three out of 9 stable λ IR lines (1, 4 and 8) and four out of 17 stable λ ER lines (2, 24, 25 and 27) for HeLa cells (**Fig. IV.14**), and three out of four stable λ IR lines (1, 2 and 4) and one out of two stable λ ER lines (2) for hESCs (**Fig. IV.15**). These PCR products were then extracted from the agarose gel, and authenticated via sequencing from I/ER-FN primer site to the end of the *attB/attL* site and from the second *attP/attR* site to I/ER-RN primer site (data not shown); all cell lines contained the functional insert, without any mutations in the essential elements of the *p λ IR* and *p λ ER* cassettes. Therefore, we have successfully obtained a total of 11 stable cell lines on which we could carry out efficiency analyses of λ integrase-mediated genomic SSR.

Figure IV.14. Nested PCR gel for stable λ IR and λ ER HeLa lines. Agarose gel showing 4.1kb positive nested PCR products for all cell lines. Shown also are HeLa negative and positive plasmid controls *p λ IR* and *p λ ER*.

Figure IV.15. Nested PCR gel for stable λ IR and λ ER HUES8 lines. Agarose gel showing 4.1kb positive nested PCR products for some cell lines. Shown also are HUES8 negative and plasmid p λ IR positive controls.

λ integrase-mediated genomic recombination efficiency

Successful λ integrase-mediated genomic recombination will result in the excision of the PGK promoter, neomycin resistance gene (*Neo*) and transcriptional stop signal, initiating the transcription of the EGFP gene via the CMV promoter (**Fig. IV.9B**). EGFP-positive cells can then be quantified via flow cytometry and compared against transfection efficiency of *pCMVssEGFPbpa* positive control plasmid, to arrive at the genomic recombination efficiency. To introduce Int/h-218 protein, mutant λ integrase expression plasmid *pCMVssInt-h-218bpa* was electroporated into each of the 3 different λ IR (integrative recombination substrate vector) and λ ER (excisive recombination substrate vector) stable HeLa cell lines. The plasmid *pPGKssIHF2bpa*, encoding the engineered Int co-factor scIHF2, was also co-transfected with Int-h/218 in a separate set of experiments. Although this Int mutant does not require IHF as a co-factor to catalyze SSR reactions, we anticipated that scIHF2 might be able increase the efficiency of the genomic recombination reaction.

HeLa stable cell lines

To minimize the possibility of false positives, experimental FACS results were normalized to that of *pCMV*-transfected negative control samples (usually no visible EGFP-positive cells in culture). Flow cytometry analysis 48 hours after transfection revealed that λ integrase-mediated genomic integrative recombination was much more efficient compared to excisive recombination (**Fig. IV.16**). The three HeLa- λ IR lines exhibited up to 15% recombination efficiency, compared to a maximum of only 2% for the three HeLa- λ ER lines. Without scIHF2, genomic integrative recombination efficiencies ranged from 5-8%, roughly three-fold less than episomal efficiencies in HeLa cells (Christ *et al.*, 2002). The reduced efficiency of excisive recombination could indicate that genomic λ integrase-mediated recombination prefers *attB* and *attP* sites, as compared to *attL* and *attR*. A possible reason for this is the need for Int proteins to assemble on both *attL* and *attR* for excisive recombination, as compared to only *attP* for integrative recombination. Interestingly, the addition of engineered Int cofactor scIHF2 improved genomic recombination by as much as three-fold. This could be achieved by the binding of scIHF2 to inserted genomic *att* sites, which would result in the binding and bending of DNA at those loci.

Figure IV.16. λ integrase-mediated genomic recombination in stable HeLa lines. Cells were transfected with Int-h/218 expression vector, with and without co-transfection of scIHF2 expression vectors; individual experiments were repeated three times, normalized to *EGFP*-transfected positive control. Genomic integrative recombination was much more efficient than excisive recombination. Int = Int-h/218.

HUES8 stable cell lines

Among the four transgenic hESC lines tested, only HUES8- λ IR4 displayed positive EGFP expression 48 hours after transfection, representing successful genomic recombination after Int-h/218 and scIHF2 expression plasmid introduction by Effectene. However, these events were rare (**Fig. IV.17**), and we were unable to quantify them via flow cytometry with sufficient significance. Therefore, it is possible that genomic recombination in hESCs is much less efficient when compared to HeLa cells. This might be due to the difference in chromatin structure or the disruption to SSR intermediates caused by the high frequency of replication fork passages in hESCs. In addition, the

limited transfection efficiency in hESCs (less than 10% compared to in excess of 80% with the HeLa electroporation protocol) resulted in few successfully-transfected cells. This could impact the expression level of Int-h/218 by virtue of the low number of plasmids transfected per cell, thus potentially contributing to the low efficiency of genomic λ integrase-mediated recombination in hESCs. As the amount of episomal DNA recovered from these transient transfections is usually too small, we were unable to proceed with further analysis at the DNA level. Improvements to transfection and recombination efficiencies should allow us to perform Southern and PCR analyses in the future to bolster this result.

Figure IV.17. Representative photomicrographs of HUES8- λ IR4 hESCs after recombination. Rare recombination events were detected via EGFP expression 48 hours after transfection of Int-h/218 and scIHF2 expression vectors.

IV.4 Characterization of *UTF1*

IV.4.1 Analysis of transcript expression in various cell lines

UTF1-EGFP vectors

Together with Oct4, Sox2 and Nanog, the transcription factor UTF1 is regarded as one of the most important markers defining the pluripotent stem cell niche (Bhattacharya *et al.*, 2004; Trounson, 2006). To further characterize UTF1 expression in human cells, we cloned and sequenced the entire gene, inclusive of the 5' promoter and 3' enhancer regions from HeLa cells, and replaced the *UTF1*-coding region with that of *EGFP*. This would enable us to detect transient *UTF1*-driven EGFP expression in various cell lines after transfection. Initially, *pBS-UTF1-EGFP* and its enhancer-less variant were used to analyze transient expression of UTF1 in HUES8 cells. However, the removal of the *UTF1* enhancer did not cause a significant downregulation of EGFP expression. Upon transfer to a *pTZ-18R*-based backbone, *UTF1*-driven EGFP expression could be abolished by the removal of the *UTF1* enhancer region (**Fig. IV.18A**).

Interestingly, the only difference between the *pBS*- and *pTZ*-based *UTF1-EGFP* expression vector sequences was a short 90bp segment originating from the multiple cloning site and β -galactosidase α -fragment of the *pBluescript* plasmid. However, further bioinformatics analysis using AliBaba2.1 and Match (both from <http://www.gene-regulation.com>) revealed that this sequence contains two potential Sp1-binding sites. Considering that the endogenous *UTF1* promoter itself contains four Sp1-binding GC boxes, it is possible that the removal of the *UTF1* enhancer region from *pBS-UTF1-EGFP* did not affect expression of the EGFP gene, which was now constitutively driven

by a total of 6 closely-spaced Sp1-specific binding sites or GC boxes. These are spaced three abreast in a paired fashion to constitute a possible hybrid promoter (**Fig. IV.18B**). It is possible that this hybrid promoter is also sensitive to differentiation as Sp1-like transcription factors that bind to GC boxes are currently thought to be regulators of embryonic development in vertebrates (Zhao and Ming, 2005). Therefore, we discarded the *pBS*-based vectors, and switched to *pTZ*-based UTF1 vectors instead.

Figure IV.18. *pBS* and *pTZ*-backboned UTF1-EGFP expression differences in HUES8 cells. (A) The removal of the UTF1 3' enhancer region only has an effect on *pTZ*-based, but not *pBS*-based UTF1-EGFP vectors. (B) Schematic showing the additional difference of 90bp (red) between *pBluescript* and *pTZ-18R* backbones. Note the addition of A&B paired SP1 binding sites (red) to the original two pairs 1&2 and 3&4, creating a possible hybrid promoter. (C) Detailed sequence.

UTF1-EGFP and UTF1-EGFP mutant expression

Transient transfection of *pTZ-UTF1-EGFP*, generated from *pTZ-UTF1*, was used to determine endogenous UTF1 expression in a number of cell lines. In addition, three variants of *pTZ-UTF1-EGFP* were designed: one without the minimal 5' promoter, one lacking the 3' enhancer, and one containing an altered *M1* sequence (**Fig. IV.19A**). The *M1* octamer sequence was found to be conserved in the *Nanog* promoter of many species (Rodda *et al.*, 2005), and we identified it recently upstream of the Oct4/Sox2 binding site within the 3' *UTF1* enhancer. These mutants were transiently transfected into HUES8 hESCs via Effectene to test the importance of each mutated component for endogenous UTF1 expression.

The results demonstrated the hESC-specific expression of full-length *pTZ-UTF1-EGFP* after transfection into various human cell lines and hESC line HUES8 (**Fig. IV.19B**). We next introduced three mutant *pTZ-UTF1-EGFP* versions into HUES8 cells and showed that the 5' promoter and 3' enhancer regions were essential for EGFP expression (**Fig. IV.19B**). In addition, the altered *M1* element reproducibly led to a significantly smaller number of EGFP expressing hESCs, which was similar in magnitude to its reported effect on *Nanog* transcription (Rodda *et al.*, 2005). Together, these data confirmed that *UTF1* transcription is a very sensitive pluripotency marker specific for hESCs and that its regulation in hESCs involves *M1*-binding factors in addition to Oct4 and Sox2.

Figure IV.19. UTF1 expression analysis in various cell lines. (A) Schematic of parental vector *pTZ-UTF1* and red labeling showing the nature of the various mutants of *pTZ-UTF1-EGFP*. (B) Flow cytometry of *UTF1*-driven EGFP expression in various cell lines following transient transfection of *pTZ-UTF1-EGFP*. Transfection efficiencies were normalized to those determined in parallel using *pCMVssEGFPbpa* and values were an average of biological triplicates.

IV.4.2 Transcript knockdown

RNA interference (RNAi), a mechanism used by most eukaryotic cells to eradicate foreign double-stranded molecules, has been exploited in mammalian cells to alter gene expression. More recently, this technique has been applied to study stem cell function via the suppression of gene expression (Zou and Yoder, 2005). As the downregulation of Nanog via RNAi has been shown to induce differentiation of hESCs (Hyslop *et al.*, 2005), we were curious to see if a similar result could be obtained for UTF1.

We transfected a commercially-designed *UTF1* transcript-targeting siRNA from Dharmacon (Boese *et al.*, 2005) into HUES8 hESCs for three consecutive days and total RNA was collected to analyze pluripotency transcript levels on the fourth day. Using our reverse transcription protocol, cDNA was synthesized from these RNA samples and used as a template in SYBR Green quantitative real-time PCR (qRT-PCR) reactions with specific primers for Oct4, Nanog, Sox2 and UTF1 transcripts. Subsequent analysis using 7500 System SDS Software (ABI) at a threshold level of 0.08 revealed the threshold cycle (C_T) reading for each reaction (see Appendix IV, **Fig. VIII.4** for representative amplification curves and Appendix V for raw data calculations). We compared these C_T values for siUTF1 treatment versus siNT (non-targeting) and converted results to relative expression levels.

Successful short-term knockdown of UTF1 transcripts with this siRNA protocol was achieved by up to five-fold when compared to mock non-targeting siRNA controls (siNT) (**Fig. IV.20**). This resulted in a slight but reproducible downregulation of Nanog transcripts (30%). However, this extent of UTF1 downregulation was not sufficient to downregulate Oct4 or Sox2, and did not appear to trigger hESC differentiation. This was supported by the observation that hESC cultures in both siUTF1 or siNT experiments displayed similar cellular and colony morphologies (**Fig. IV.20**). These results confirmed that UTF1 is downstream of both these pluripotent markers, and indicated that *UTF1* and *Nanog*, along with *MI*-binding factor(s), might be involved in the same pluripotency or proliferation expression network.

Figure IV.20. Short-term siRNA-mediated knockdown of *UTF1* transcripts. Morphology comparison of non-targeting (siNT) with siUTF1 treatment (top); qRT-PCR transcript analysis of important pluripotency markers, normalized to siNT-treated HUES8 (bottom) and were an average of three biological and assay repeats. Scale bar = 130 μ m.

IV.4.3 Transcript expression after RA/DMSO treatment

We evaluated next the sensitivity of *UTF1* expression with respect to induced differentiation using qRT-PCR to profile transcriptional expression levels. Various hESC lines were triggered to differentiate via the addition of either retinoic acid (RA) for 12 days, which has been used to induce the derivation of neurons from ESCs (Glaser and Brüstle, 2005) or dimethyl sulfoxide (DMSO) for 7 days, a potent morphogen (Adler *et*

al., 2006). The C_T value of each reaction was recorded (see Appendix IV, **Fig. VIII.5** for representative amplification curves and Appendix V for raw data calculations), and converted to expression percentages relative to the respective undifferentiated hESC lines. The results show that these differentiation protocols almost completely abolished *UTF1* expression; a response that was not observed for *Nanog* and *Oct4* (**Fig. IV.21**). Transcript analysis after these two treatments confirm previous reports that endogenous *UTF1* is downregulated faster than *Oct4* or *Nanog* at the onset of differentiation to nearly undetectable levels. Although our quantitative real-time PCR protocol included a dissociation step in which the number of end product species could be discerned, we confirmed these results by validating the specificity of our primers by running the products on an agarose gel and showed only one product for each pair of specific primers (see Appendix III, **Fig. VIII.2**).

Together, these data confirmed that *UTF1* transcription is a very sensitive and reliable marker specific for hESC pluripotency and that its regulation in hESCs involves *MI*-binding factors in addition to *Oct4* and *Sox2*. Given these properties, a *UTF1*-based selectivity tool should be generally applicable to select for homogeneously pluripotent human stem cell cultures. In addition, the < 2kb length of the entire human *UTF1* promoter/enhancer element represents a significant advantage for genetic manipulation of hESCs, or other human cell types, if one compares this to an unwieldy 8kb of the oft-used human *Oct4* control element.

Figure IV.21. Effect of RA and DMSO differentiation in untransformed hESCs. Pluripotency marker qRT-PCR transcript expression analysis of hESCs on MEFs induced to differentiate via RA or DMSO differentiation protocols, normalized to that of respective undifferentiated hESC lines, with two biological replicates for each sample. *UTF1* transcripts are downregulated the most significantly.

IV.5 Characterization of ZUN cell lines

IV.5.1 Genomic characterization

Sequencing of insert

We established stable *pTZ-UTF1-Neo* cell lines from HUES8 cells via random genomic transgene insertion after Effectene transfection, and subsequent G418 antibiotic selection. Cloning efficiency was around 10^{-5} , with a total of three so-called “ZUN” cell lines successfully isolated and subsequently analyzed for the presence of the insert. All of these contained the *UTF1* 5' promoter and *Neo* transgene, but only ZUN1 and ZUN2 cell lines contained genomic inserts that encompassed the full-length transgene inclusive of the *UTF1* 3' enhancer (Fig. IV.22). The full-length sequence of Genomic PCR B products in ZUN1 and ZUN2 were then confirmed to be correct via sequencing (data not shown). These ZUN hESC lines were maintained under constant G418 antibiotic selection pressure.

Copy number

We then evaluated the copy number of *UTF1-Neo* transgenes in ZUN1 and ZUN2 stable hESC lines. Copy number analysis was achieved via Southern blot of EcoRI-digested ZUN genomic DNA, using the *Neo* gene as a probe. Both ZUN cell lines contained only one copy of the insert each as represented by a single strong band each in the gel, whilst the weaker background band in all three lanes was indicative of the *Neo* gene in the neomycin-resistant PMEF-N MEF genome, which constitute about 10% of the cells in hESC-MEF co-cultures (Fig. IV.23).

Figure IV.22. Detection of transgene in ZUN cell lines. (A) Schematic of *pTZ-UTF1-Neo* and PCR strategy. (B) Agarose gel of PCR A showing a positive 1kb band for all three ZUN lines. (C) Agarose gel of PCR B showing an endogenous UTF1 band for all three lines. However the full insert was only present in ZUN1 and ZUN2. Ladder = 1kb.

Figure IV.23. Southern blot of ZUN cell lines. The Southern blot showed that both ZUN1 and ZUN2 have one copy each of the *pTZ-UTF1-Neo* insert (arrows), relative to the linearized 5.5kb plasmid control, loaded in two different amounts (+ve and +ve(10x)). The faint band observed in all three samples (labeled with *) was most likely the signal from neomycin-resistant MEFs (~10% in hESC cultures).

Figure IV.24. Karyotypic analysis of ZUN hESC lines. Representative chromosomal spreads showing a normal 46XY karyotype for both ZUN lines after culture of at least 10 passages. At least 40 individual cells were analyzed per sample.

IV.5.2 Pluripotency and proliferation

Reverse transcription PCR

Endpoint reverse transcription PCR (RT-PCR) was first used to qualitate differential transcript levels between HUES8 and ZUN2 hESC cultures. Total RNA was collected from the various cell samples and cDNA was generated via reverse transcription. This was used as a template for a 30-cycle PCR reaction. Equal volumes of PCR product were then run on an agarose gel to compare band intensities which were a rough indication of the amount of transcripts present for each of the various markers. Transcript levels of undifferentiated cultures were compared to spontaneously differentiating cultures plated on matrigel for five or 10 days without MEFs or G418 antibiotic. Both *GAPDH* (housekeeping gene) and *Oct4* expression remained relatively constant under spontaneous differentiation, whilst *UTF1* expression decreased progressively with increased duration of culture on matrigel. Nanog was downregulated in HUES8 cultures, but relatively constant in ZUN2 cultures. The mesodermal marker *IGF2* (Morali *et al.*, 2000) was expressed after spontaneous differentiation in both HUES8 and ZUN hESC cultures (**Fig. IV.25**). These preliminary results demonstrated that UTF1 is indeed a sensitive pluripotency marker, and that ZUN hESC populations potentially differentiate at a globally slower pace compared to cultures from the parental HUES8 line. Although the RT-PCR technique could be used to detect large transcript level changes, a more quantitative assay that would be more sensitive to subtle changes was required to verify these results. Therefore, all following transcript expression analyses were conducted using quantitative real-time PCR (qRT-PCR).

Figure IV.25. Preliminary reverse transcription PCR. Agarose gel comparing RT-PCR products for pluripotency markers *Oct4*, *Nanog* and *UTF1* and mesodermal marker *IGF2*. MG = culture on matrigel for induction of spontaneous differentiation.

Expression of pluripotency markers

To assess the pluripotency status of our transgenic hESC lines, pluripotency marker transcripts were compared across HUES8 and both ZUN cell lines via qRT-PCR. The C_T of each reaction was recorded (see Appendix IV, **Fig. VIII.6** for representative amplification curves) and subsequently converted to reflect ZUN transcript expression relative to HUES8 values. ZUN cell lines express a similar level of *Nanog* transcripts as the parental HUES8. However, a higher expression level of *Oct4*, *Sox2* and *UTF1* was reproducibly detected by quantitative real-time PCR (**Fig. IV.26**). This correlation of expression supports the notion that the control of *UTF1* involves factors Oct4 and Sox2, but most likely not Nanog (Boyer *et al.*, 2005). In addition, *UTF1*-driven *Neo* was expressed exclusively in ZUN cells (**Fig. IV.26**).

Figure IV.26. Transcript analysis of ZUN hESCs. qRT-PCR comparison of pluripotency markers across the three hESC cultures, normalized to HUES8 expression and performed with biological triplicates.. Nanog expression was similar across all cell lines, whilst *Oct4*, *Sox2* and *UTF1* were expressed at a higher level in ZUN hESC lines. *UTF1-Neo* was expressed specifically in both ZUN cell lines.

Expression of surface markers of pluripotency

We next wanted to find out if this increase in expression was also detectible at the protein level. To achieve this, the immunostaining of various pluripotency markers using pre-incubated primary and secondary antibodies was conducted and these samples analyzed by flow cytometry. Negative control (secondary antibody only) samples were used to set the gating at zero and the percentage of positively-stained cells was then quantitated. This experiment revealed that the observed higher level of global *Oct4*, *Sox2*, and *UTF1* transcript expression was due to the presence of a larger fraction of cells that express these markers. A statistically significant higher percentage of ZUN cells stained positive for Oct4 and Sox2 compared to HUES8 cultures (**Fig. IV.27**). We could also demonstrate that ZUN cultures contained more cells which stain positive for surface markers TRA-1-

60 and TRA-1-81 (**Fig. IV.28**). These results indicate that HUES8 cultures contained more hESCs that were no longer pluripotent thus did not express these pluripotency markers. In contrast, such cells have been ablated in ZUN cultures via G418 selection. In this way, ZUN cultures consist of a more homogeneous population of pluripotent hESCs.

Figure IV.27. Oct4 and Sox2 expression analysis of ZUN hESCs. Percentage of hESCs expressing Oct4 and Sox2. ZUN cultures consisted of significantly more Oct4- and Sox2-expressing cells compared to HUES8, based on the two-tailed paired Student’s *t*-test (*p* values <0.05). Average values displayed were obtained from three biological replicates.

Figure IV.28. Surface marker flow cytometry analysis of ZUN hESCs. Percentage of positively stained hESCs for TRA-1-60 and TRA-1-81; both ZUN lines consisted of significantly more positively stained cells compared to parental HUES8 based on the two-tailed paired Student’s *t*-test (*p* values <0.05). Average values displayed were obtained from three biological replicates.

Alkaline phosphatase staining

AP staining, a common standard of pluripotency (Amit and Itskovitz-Eldor, 2006; Trounson, 2006), was used to gauge the general pluripotency of ZUN cell lines. HUES8, ZUN1 and ZUN2 cell lines demonstrated a similar degree of positive staining, indicating a high level of pluripotency for all three hESC lines (**Fig. IV.29**).

Figure IV.29. Alkaline phosphatase staining. Positive AP staining of HUES8, ZUN1 and ZUN2 hESCs. Scale bar: 150 μ m.

Proliferation assay

As UTF1 is thought to be involved in cell proliferation control (Nishimoto *et al.*, 2005), we sought to examine if ZUN cell lines had a proliferative advantage over their parental HUES8 counterparts. Cell proliferation rates of the ZUN hESC cultures were compared with that of parental HUES8 via the WST-1 reagent (Roche). This colorimetric assay is based on the cleavage of the tetrazolium salt WST-1 to formazan by mitochondrial dehydrogenases found in viable cells. Essentially there was no difference in proliferation rates between HUES8 and ZUN cells, as indicated by a similar increase in formazan dye quantitated by a photoplate reader for all three lines across four days (**Fig. IV.30**). This result suggested that a slightly higher level of UTF1 expression in ZUN cultures does not confer a proliferative advantage over non-transformed hESCs.

Figure IV.30. Cell proliferation assay. HUES8, ZUN1 and ZUN2 hESCs were plated on 96-well plates and analyzed daily at 440nm after 1 hour of incubation with WST-1 reagent (average of 2 data sets, 4 wells per reading). Relative absorbance (normalized to control wells containing only MEFs) showed a similar gradient for all three hESC lines over four days, indicating a similar rate of proliferation.

IV.5.3 Modulation of selection pressure

Expression of pluripotency marker transcripts

ZUN cultures have been established to consist of homogeneous populations of pluripotent cells. Subsequently we investigated if G418 removal would cause the downregulation of pluripotency markers in these hESC lines. This was analyzed via the effect of sequential removal and restoration of G418 selection on both ZUN lines and employed qRT-PCR of pluripotency markers *Nanog*, *Oct4*, *Sox2*, and *UTF1* for the detection of the onset of differentiation (Cai *et al.*, 2006). We also included the transgene *UTF1-Neo* in this analysis because its regulation should be linked to that of the endogenous *UTF1*. The C_T

value of each reaction was recorded (see Appendix IV, **Fig. VIII.7** for representative amplification curves) and converted to expression percentages relative to respective cultures with constant G418 selection pressure.

The results showed that spontaneous differentiation could be detected in cultures during 60 days of G418 withdrawal. This was indicated by the progressive reduction in signals obtained from all markers, with the exception of *Sox2* (**Fig. IV.31**). However, upon re-application of antibiotic selection, original expression levels could be restored, and pluripotency in these cultures rescued by the ablation of differentiated hESCs which had downregulated *UTF1* and, consequently, *UTF1-Neo*. In fact, we were able to reinstate previous levels of all four pluripotency marker transcripts in ZUN cells (**Fig. IV.31**).

Using a transgenic hESC line in which *Neo* is expressed from the elongation factor 1 alpha (*EF-1 α*) promoter as a control (EBN14, generated by K. Neef), we confirmed that G418 withdrawal has no effect on the expression of this set of markers which clearly links this effect to the presence of the *UTF1-Neo* transgene in hESCs (**Fig. IV.32**). Together, these data showed that ZUN cultures were more homogeneously pluripotent than the parental cultures and that G418 addition and subsequent ablation of differentiated ZUN cells was required to sustain homogeneous cultures over longer periods of time.

Figure IV.31. qRT-PCR analysis of removal and re-introduction of G418 on ZUN hESCs. ZUN cell lines were cultured with G418 antibiotic for 90 days (+G418), G418 was removed for 30 (-G418) or 60 (--G418) days, G418 was removed for 30 days and re-introduced for the subsequent 30 days (-+G418), normalized to +G418 values respectively. All markers except for *Sox2* showed sensitivity to change in G418 removal and addition and presented as an average of three experimental repeats.

Figure IV.32. qRT-PCR analysis of G418 removal on EBN14 hESCs constitutively expressing Neo through the human *EF-1 α* promoter. EBN14 cells were cultured with (+G418) or without G418 antibiotic (-G418) for 30 days. Results were normalized to +G418 values. All markers showed no significant change in response to G418 antibiotic removal. Experiments were performed with biological and assay triplicates, and presented as expression level (%) relative to EBN14 -G418 values.

IV.6 Comparison of ZUN differentiation

IV.6.1 Spontaneous differentiation

Expression of pluripotency marker transcripts

We have demonstrated so far that ZUN cultures are more homogeneously pluripotent than the parental cultures. We were interested, therefore, in investigating the differentiation behavior of ZUN cells, as compared to their parental HUES8 counterparts. The spontaneous differentiation of ZUN cultures was triggered by plating them on matrigel or gelatin without MEFs and in the absence of G418 antibiotic. This was achieved via qRT-PCR analysis as detailed before. In addition, we confirmed the specificity of our primers by running the endpoint products on an agarose gel and showed only one product for each pair of specific primers (see Appendix III, **Fig. VIII.3**).

After 12 days, expression levels of most pluripotency markers remained elevated in ZUN lines compared to parental HUES8 cells. In the matrigel protocol, HUES8 cells quickly initiated differentiation, as indicated by the 5-fold downregulation of our panel of markers. In the matrigel protocol, HUES8 cells quickly initiated differentiation as indicated by the 5-fold downregulation of our panel of markers (**Fig. IV.33**, see and Appendix V for raw data calculations). However, both ZUN lines maintained significantly higher overall expression levels, even after 12 days of culture. Similarly, the gelatin protocol induced a two-fold downregulation of *Oct4* and *UTF1* in HUES8, but not in ZUN cultures.

Figure IV.33. Differentiation of ZUN hESC lines on matrigel or gelatin. Pluripotency marker transcript analysis of spontaneous differentiation: qRT-PCR of hESCs without MEFs plated on gelatin or matrigel (with hESC media and without G418 antibiotic) for 12 days. Experiments were performed with biological duplicates, and presented as expression levels relative to that of respective undifferentiated cell lines.

Cell and colony morphology

Visual inspection of cell and colony morphology showed that spontaneous differentiation of HUES8 cells with both protocols displayed visible indication of hESC differentiation.

An increase of cytoplasm to nuclear ratio and the physical spreading out of individual cells and colonies was observed in HUES8 but not ZUN hESCs (**Fig. IV.34**).

Figure IV.34. Cell and colony morphology under spontaneous differentiation. Unlike the differentiated HUES8, ZUN hESCs display compact pluripotent morphology when passaged without MEFs and plated onto gelatin or matrigel (with hESC media and without G418 antibiotic) for 12 days. Scale bar: 60 μ m.

Alkaline phosphatase staining after antibiotic selection

AP staining was then used as a means to identify pluripotent cells within these spontaneously differentiating cultures, and to visualize the effects of G418 ablation of differentiating ZUN cells. Under these spontaneous differentiation conditions, ZUN cultures have a larger number of colonies consisting of cells which displayed a more compact morphology and exhibited stronger AP staining, and a reduced occurrence of differentiating cells at colony edges. Upon subsequent addition of G418 antibiotic for 2 days, only pluripotent cells that stain for AP survived in ZUN cultures, demonstrating

that differentiating ZUN cells are rapidly and efficiently ablated by G418 selection (**Fig. IV.35**).

Figure IV.35. AP staining of spontaneous differentiation and subsequent G418 treatment. ZUN hESCs consists of more AP positive cells compared to HUES8 after 12 days of spontaneous differentiation; AP negative ZUN cells are efficiently ablated upon subsequent G418 treatment for 2 days. Scale bar: 70 μ m.

IV.6.2 Directed differentiation

Expression of pluripotency marker transcripts

Small bioorganic molecules like RA have been shown to be able to direct the differentiation of ESCs into neuronal and insulin-producing lineages (Glaser and Brüstle, 2005; Roche *et al.*, 2006), and solvents like DMSO are known to induce ESC differentiation into immature hepatocytes and various other cell types (Brandenberger *et al.*, 2004; Adler *et al.*, 2005). Thus, the effect of directed differentiation induced via culture with RA or DMSO differentiation media on ZUN hESC lines, in the absence of

G418 was analyzed subsequently. After 12 days (for RA treatment) and 7 days (for DMSO treatment) of culture, qRT-PCR analysis showed that ZUN cultures exhibited a lower rate of differentiation in comparison to HUES8 cultures. Notably, *UTF1* and *UTF1-Neo* expression diminished by the largest magnitude among the pluripotency markers (**Fig. IV.36**, see and Appendix V for raw data calculations).

Figure IV.36. Differentiation of ZUN hESC lines with RA or DMSO. qRT-PCR Pluripotency marker transcript analysis of directed differentiation. The differentiation of hESCs on MEFs was induced via addition of RA (for 12 days) or DMSO (for 7 days) and differentiation media. Experiments were performed with biological duplicates, and presented as expression levels relative to that of respective undifferentiated cell lines.

Expression of differentiation marker transcripts

The expression levels of certain differentiation markers also indicated a slower differentiation rate of ZUN lines, in agreement with the expression differences as shown with pluripotency markers as shown above. The C_T value of each qRT-PCR reaction was recorded and expressed as fold change in comparison with the same transcript of the respective undifferentiated hESC cultures. We used the following markers to judge the differentiation status of our hESCs under RA or DMSO treatments: Nestin and Sox9 have been shown to be important in ectodermal development and neural crest formation (Lenka *et al.*, 2002; Cheung and Briscoe, 2003), ID2 was recently identified as important for gut organogenesis and is thus an endodermal marker (Torihashi *et al.*, 2006), and HAND1 and IGF2 have been identified as early mesodermal markers (Risebro *et al.*, 2006; Morali *et al.*, 2000).

Quantitative real-time PCR results demonstrated that HUES8 cultures displayed a greater (2 to 10 fold) upregulation of *Sox9*, *Nestin* and *ID2* due to RA treatment and of *HAND1*, *IGF2* and *Nestin* after exposure to DMSO (**Fig. IV.37**, see and Appendix V for raw data calculations). The specificity of the primers used in this qRT-PCR experiment were confirmed by running the endpoint products on an agarose gel, with each pair of specific primers resulting in only one product (see Appendix III, **Fig. VIII.3**).

Figure IV.37. Differentiation of ZUN hESC lines with RA or DMSO. qRT-PCR differentiation marker transcript analysis of directed differentiation. The differentiation of hESCs on MEFs was induced via addition of RA (for 12 days) or DMSO (for 7 days) and differentiation media. Experiments were performed with biological duplicates, and presented as expression levels relative to that of respective undifferentiated cell lines.

Expression of surface markers of pluripotency

The expression of surface markers of pluripotency, TRA-1-60 and TRA-1-81, complements the transcript analyses above, showing that HUES8 cells exhibited a more muted staining pattern compared to ZUN colonies (**Fig. IV.38**). These results confirmed our previous conclusion that ZUN hESCs exhibit a more robust phenotype against the onset of differentiation than parental cultures.

Figure IV.38. Surface marker immunostaining in directed differentiation conditions. TRA-1-60 and TRA-1-81 surface marker expression, commonly used as determinants of pluripotency, complement qRT-PCR data for both treatments; Both ZUN cultures demonstrated a larger number of hESCs positively stained with a higher intensity than parental HUES8 cultures. Scale bar: 70 μ m.

Colony morphology and G418 treatment

We next analyzed the colony morphology after both RA and DMSO differentiation protocols. After 12 days of differentiation, almost all RA-treated HUES8 colonies were diffuse and sprawling, and exhibited virtually zero small, tightly-packed colonies characteristic of pluripotent hESCs. However, both ZUN cell lines still retained some of these pluripotent colonies. In contrast, DMSO treatment for 7 days showed less significant differences between HUES8 and ZUN colony morphology (Fig. IV.39). ZUN cells treated with both induction protocols could be ablated with the addition of G418 antibiotic, leaving behind mainly neomycin-resistant MEFs (Fig. IV.40). This confirmed our transcript expression level results above, and showed again that differentiating ZUN hESCs that downregulate *UTF1*, and consequently *Neo*, were ablated upon G418 treatment.

Figure IV.39. Colony morphology of hESCs after directed differentiation treatments. Small, compact colonies were observed as tight white dots are pluripotent colonies; large diffuse colonies consist of differentiated hESCs. Scale bar: 150 μ m.

Figure IV40. Addition of G418 after RA or DMSO differentiation. Ablation of differentiated ZUN hESCs as demonstrated before (top) and after (bottom) addition of G418 antibiotic. Scale bar: 80 μ m.

IV.6.3 Long-term differentiation

Matrigel differentiation

Various reports indicate that pluripotent hESCs were able to be cultured on matrigel (Carpenter *et al.*, 2004; Xu *et al.*, 2005; Yao *et al.*, 2006). However, all of these protocols require either MEF-conditioned media or other additives like bFGF to maintain hESC pluripotency. We thus sought to select for ZUN hESCs that were pluripotent on matrigel via G418 selection without the need for conditioned medium. Both ZUN and parental HUES8 hESC lines were able to generate pluripotent-like colonies after 25 days on matrigel without MEFs and without G418. However, upon addition of G418 for 5 days, these cells in ZUN cultures were ablated, indicating that these colonies had downregulated UTF1 and were thus no longer pluripotent hESCs (**Fig. IV.41.**). This demonstrated that our standard HUES8 media was not sufficient to enable feeder-free pluripotent hESC culture on matrigel. Also, it illustrates the need for careful analysis of hESC colonies in new culture conditions, as the appearance of hESC-like cell and colony morphologies do not necessarily mean that those colonies are actually pluripotent.

Figure IV.41. Long term matrigel treatment. Viable hESC-like colonies survived for 30 days on matrigel in all three cell lines, but those in ZUN cultures were ablated after 5 days of subsequent G418 treatment. White scale bar: 200 μ m; Black scale bar: 50 μ m.

RA and DMSO differentiation

The observed phenotype of increased robustness against the onset of differentiation led us to investigate ZUN cultures under long term differentiation conditions. Our results showed that ZUN and HUES8 cells cultured with either RA or DMSO differentiation protocols for 30 days displayed similar morphologies and lack of AP staining (**Fig. IV.42.**). These cells were completely ablated upon addition of G418, as were those in equally long-term gelatin and matrigel cultures (data not shown).

Figure IV.42. AP staining of long term RA and DMSO treated hESCs. Differentiated cells derived from hESCs survived for 30 days with constant passaging on MEFs; all three cell lines displayed a similar lack of AP staining. Scale bar: 70 μ m.

Embryoid body formation

In addition, 21 days of embryoid body (EB) formation in suspension cultures caused a significant upregulation of differentiation markers from all three germ layers (**Fig. IV.43,** see and Appendix V for raw data calculations). Compared to undifferentiated hESCs,

HUES8 and both ZUN lines significantly increased the transcript expression levels of *IGF2* and $\alpha 1$ anti-trypsin (*$\alpha 1AT$*), a sensitive early endodermal marker (Schuldiner *et al.*, 2000; Hamazaki *et al.*, 2001). EBs across all three cell lines also showed a considerable upregulation of *HAND1* and *ID2* mRNA, and a consistent increase in the expression of ectodermal marker *Sox9*. Morphologically, embryoid bodies generated via suspension culture were similar across all three cell lines, with dense cell clumps and cavitated EBs evident in all cultures (Fig. IV.44).

Figure IV.43. Transcript analysis of embryoid body (EB) formation. qRT-PCR of EBs formed from hESCs induced to differentiate in suspension for 21 days. All three cell lines showed relatively even upregulation of selected differentiation markers for all three germ layers. Experiments were performed with biological duplicates, and presented as expression levels relative to that of respective undifferentiated cell lines.

Figure IV.44. Morphological analysis of EB formation. Representative photomicrographs of EBs formed from hESCs induced to differentiate in suspension for 21 days. All three cell lines displayed similar morphological characteristics and growth rates after 21 days of EB-induced differentiation. Scale bar: 500 μ m.

Teratoma formation

Furthermore, we showed by a teratoma assay that the parental HUES8 line as well as ZUN1 and ZUN2 formed tissue of all three germ layers (i.e., ectodermal structures, mesoderm such as cartilage and bone, as well as mesodermal glandular structures) when transplanted into immune-deficient SCID mice (**Fig. IV.45**). As in EB formation, the morphological characteristics and growth rates of HUES8 and ZUN cell-derived teratomas were indistinguishable. We conclude, therefore, that the robustness which our ZUN cultures displayed against the onset of differentiation induced by various protocols does not compromise their potential to differentiate, like parental cultures, into all three germ layers *in vitro* and *in vivo*.

Figure IV.45. Analysis of teratoma formation in SCID mice. The hESC lines ZUN1 (A), ZUN2 (B) and the parental line HUES8 (C) were injected into the hind leg of male SCID mice and, when fully developed (1–2 g tissue weight, 6–8 weeks), they were harvested and processed by routine histological procedures, followed by heamatoxilin and eosin staining. (D) Detailed histology of ectodermal (epithelium, neuroectoderm), mesodermal (bone in red, cartilage in blue) and endodermal (glandular structures) structures are shown for each hESC line.

V. DISCUSSION

V.1 Human embryonic stem cell culture

Status of hESC lines

We have established culture conditions for the maintenance of undifferentiated hESCs on MEFs and have extensively characterized both hESC lines used for further experiments. Under these optimal conditions and for over 30 passages, both hES2 and HUES8 lines display pluripotent morphology, stain positively for alkaline phosphatase, and express stem cell-related surface markers SSEA-4, TRA-1-60, TRA-1-81 and essential pluripotency marker Oct4. HUES8 was also found to have no karyotypic abnormalities after prolonged culture. In addition, we show that MEF cells account for only 5-10% of the total cells in our co-culture system, and thus should not adversely affect the quantification of downstream EGFP flow cytometry or transcriptional profiling.

Transfection efficiencies

Unlike mouse ESCs, hESCs are rather difficult to transfect efficiently. Protocols using lentiviral vectors (Gropp *et al.*, 2003; Ma *et al.*, 2003), electroporation (Zwaka and Thomson, 2003; Mohr, de Pablo and Palecek, 2006), and various transfection reagents (Eiges *et al.*, 2001; Vallier *et al.*, 2004; Park *et al.*, 2003) have been developed, but comparative analyses are scarce. Currently, lentiviral-based approaches appear to be the preferred choice to achieve high transfection efficiencies. However, these vectors are quite difficult to generate and have size limitations for transgenes (Vallier *et al.*, 2004). Here we compare plasmid transfection efficiencies obtained with more popular reagents

and protocols, and demonstrate that both hES2 and HUES8 cells exhibited a clear preference for lipid-based transfection reagents.

Optimized Fugene 6, Effectene, Lipofectamine 2000 and DharmaFECT 2 showed similar transfection efficiencies of 5-7%. This number decreased to less than 1% with polymer-based reagents Exgen 500 or SAINT-MIX. We also tested the addition of silica microspheres, based on the theory that DNA/reagent/silica complexes increase the DNA concentration at the cell surface, to lipid-based transfection reagents and found that it resulted in a close to threefold increase in plasmid uptake in combination with Effectene in hES2 cells. A similar finding has been reported with COS-7 and CHO cells (Luo and Saltzman, 2000; Luo *et al.*, 2004). On the other hand, addition of microspheres showed only a limited effect with Lipofectamine 2000 and none with Fugene 6 (data not shown).

In addition to being rather efficient transfection reagents, Effectene with silica microspheres and DharmaFECT 2 also showed the least cytotoxicity based on flow cytometry flow rates (data not shown). Our plasmid transfection efficiencies of up to 10% using the transfection reagents DharmaFECT 2 or Effectene in combination with silica microspheres are comparable with previously reported protocols such as nucleofection or electroporation (Lakshimpathy *et al.*, 2004; Mohr, de Pablo and Palecek, 2006).

In our hands, electroporation proved to be the most efficient method of transfection at a 15% rate of efficiency, similar to that achieved previously (Mohr, de Pablo and Palecek, 2006). Due to the need to dissociate the cells rather thoroughly, only the HUES8 line

could be subjected to this protocol. However, the large percentage of cell death and the uncertainty of its effect on pluripotency led us to disregard the use of electroporation for the generation of stable cell lines.

Stable cell line generation

It was not possible to generate stable EGFP-expressing hES2 lines via the Cut/Paste protocol. The initial idea was to manually isolate EGFP-positive hESCs in the hope that these cells would expand, and that successive manual dissections and passages of these would eventually lead to a heterogeneous, non-clonal, stable cell line. Unfortunately, groups of EGFP-positive cells did not respond well to mechanical dissociation and failed to proliferate after being transferred. Therefore, we focused our efforts on the HUES8 line as these cells are relatively easier to culture, can be dissociated into smaller clumps and will survive in much smaller colonies. Upon successful transfection of neomycin resistance vectors, HUES8 cultures were dissociated by trypsinization again into small cell clumps, and allowed to expand. The addition of antibiotic selection then allowed *Neo*-expressing cells to proliferate whilst ablating hESCs without the integrated vector, generating clonal transgenic colonies. Thus, Effectene transfection coupled with the Select/Expand protocol in HUES8 enabled us to select for stable, clonal hESC lines that were used in subsequent experiments. This method routinely resulted in about 5 colonies per 3cm well, with 6×10^5 cells initially transfected. This indicated that the cloning efficiency using our Effectene-based protocol was about 10^{-5} , similar to that reported by other studies which were based on either Exgen transfection or electroporation methodologies (Eiges *et al.*, 2001; Zwaka and Thomson, 2003).

V.2 Site-specific recombination in human embryonic stem cells

Episomal SSR efficiency

We compared the ability of two SSR systems developed by our laboratory to catalyze recombination reactions on episomal substrates in hES2 hESCs, namely mutant phage λ integrase and mutant $\gamma\delta$ resolvase $\gamma\delta 102\text{NLS}$, with that of the Cre SSR system. It is interesting to note that Cre recombinase is exceptionally efficient in hESCs. This is similar to results reported with cell-permeant Cre, which attains almost 100% recombination efficiency on genomic substrates (Nolden *et al.*, 2006). Therefore, this indicates that Cre recombinase is a good candidate for future hESC genome manipulations such as removal of marker genes.

λ integrase-mediated episomal integrative SSR reactions in hESCs were similarly efficient to that in HeLa cells (Christ *et al.*, 2002). On the other hand, excisive recombination was particularly inefficient in hESCs. Because integrative recombination between attachment sites *attB* and *attP* generates sites *attL* and *attR*, which are then substrates for excisive recombination, our results imply that cofactor-independent λ integrases display a significant degree of directionality inside hESCs. The molecular basis for this pronounced directionality in hESCs is not understood at the moment and deserves further investigation.

In direct comparison with Cre, the resolvase mutant $\gamma\delta 102\text{NLS}$ is significantly less active in hESCs than in Chinese hamster ovary (CHO) cells and, therefore, may be useful only for specific applications that do not require high recombination efficiencies. It is not clear

at present what causes the reduced activity of $\gamma\delta 102\text{NLS}$ in human cells. However, differences in nuclear localization can be excluded because both Cre and $\gamma\delta 102\text{NLS}$ contain functional nuclear localization signals (Schwikardi and Dröge, 2000; Le, Gagnetten and Tombaccini, 1999). One possible factor to consider may be protein modification of $\gamma\delta$ resolvase in human cells.

Practically, we are unable to determine recombination efficiencies at plasmid level because the copy number of internalized substrate DNA molecules accessible for the recombinase is unknown. This number varies per cell and not all the internalized plasmids in each cell will be recombined. Thus it is reasonable to assume that a single recombined copy will suffice to produce an EGFP-positive cell. This means that the actual percentage of recombined substrate molecules could be lower than the percentage of EGFP expressing cells normalized to the positive control.

Genomic λ integrase-mediated recombination efficiency

We have shown that mutant λ integrase is able to catalyze episomal recombination reactions in hESCs. We take this system a step further by analyzing the recombination efficiencies at the chromosomal level. Stable cell lines were generated via the random integration of reporter plasmids $p\lambda IR$ or $p\lambda ER$. In HeLa cells, genomic integrative recombination between $attB/attP$ sites was more efficient than excisive recombination (7% vs 1%) (**Fig. IV.16**). This trend of Int-h/218 favoring the integrative recombination pathway in genomic substrates has also been observed in BL60 cells (Lorbach *et al.*, 2000). In our hands, the efficiencies of both pathways could be enhanced by engineered

scIHF2, an accessory factor required for wild-type λ integrase activity but not for Int-h/218, by up to three-fold. Without additional *in vitro* experiments, we can only speculate that the engineered scIHF2 improves accessibility of genomic *att* sites to Int-h/218 via binding to and bending of genomic DNA.

When combined with these results, the hypothesis of genomic loci accessibility suggests that there might be an upper limit to recombination efficiency. In cases where accessibility to Int-h/218 might be optimal, like HeLa- λ IR4 and HeLa- λ IR8, the effect of scIHF2 on recombination efficiency was not so pronounced (**Fig. IV.16**). Conversely, recombination reactions at less accessible loci like that of HeLa- λ IR1 might be greatly enhanced by the binding and bending action of scIHF2, modifying chromatin/DNA structure such that it becomes more accessible to Int-h/218. HeLa- λ ER cell lines also seem to have an upper limit, in this case 2% with Int-h/218 and scIHF2 co-expression. This merits further investigation as scIHF2 could be a useful tool to increase the efficiency of λ integrase-mediated genomic site-specific recombination.

Genomic recombination in hESCs mediated by mutant λ integrase was relatively inefficient when compared to HeLa cells. Rare recombination events in HUES8- λ IR4 could be observed consistently after Int-h/218 transfection; however the percentage of EGFP expression was difficult to quantify statistically. Taking into account the low number of stable cell lines with the integrated cassette analyzed, we can conclude that genomic integration via the λ SSR system is possible in hESCs, but at a very low

efficiency. More cell lines with the integrated *att* sites at different genomic loci are needed to confirm this result.

It is interesting to note that genomic recombination for accessible sites are much less efficient when compared to episomal substrates, by as much as five fold, in both HeLa cells and hESCs. This might be due to a variety reasons, for example integrated recombination cassette malfunction or inaccessibility of various genomic loci. Another possible explanation for this lack of efficiency is that there might be many more accessible DNA substrate molecules available in episomal experiments, compared to the one or few integrated copies per cell in genomic experiments. These problems could be addressed by integration of the recombination cassette or *att* site at a defined, accessible genomic locus via homologous recombination. Also, improvements to the reporter system could lead to higher genomic recombination efficiencies, for example introducing chromatin insulators flanking the integrated cassette. In this way, the true genomic recombination efficiency can be elucidated.

Alternative SSR systems in hESCs

Aside from λ integrase-mediated SSR, other systems such as Cre and Φ C31 have been described to be functional in hESCs, with successful recombination events detected at the chromosomal level. (Nolden *et al.*, 2006; Thyagarajan *et al.*, 2007). These recombinases offer interesting alternatives, with Cre boasting a very high recombination efficiency (almost 100%), and Φ C31 having the benefit of being able to integrate into endogenous pseudo-*attP* sites. However, the shortcomings of either are not trivial: Cre-mediated

recombination utilizes a single *att* site and is therefore non-directional and more suited to transgene excision as opposed to insertion (Kolb, 2002). On the other hand Φ C31, whilst being a directional system, can insert targeting constructs into endogenous pseudo-*att* sites that occur in the human genome. In fact, the above study identified more than 20 sites, with some of these classified as “recombination hotspots” (Thyagarajan *et al.*, 2007). For both these systems, there would be an increased risk of unwanted genome alterations due to random DNA insertions or rearrangements (Zhou and Dröge, 2006), and consequently toxicity. Therefore with further increases in efficiency, λ -mediated SSR, with its intrinsic benefits of specificity and recombination directionality, has the potential to become the superior tool for genetic manipulation in hESCs.

The future for λ integrase-mediated genomic SSR in hESCs

The λ integrative pathway catalyzed by Int-h/218 is quite efficient in hESCs. Both episomal and genomic integrative intramolecular recombination seem to be particularly favored. Although the efficiency of genomic recombination is lacking in hESCs, experiments in HeLa have demonstrated that higher efficiencies are possible. When optimized, this SSR system could prove very valuable for targeted gene insertions into genomic attachment sites, due to the directionality that λ Int-h/218 exhibits. These *att* sites could be artificially introduced or occur naturally in the human genome. Examples for λ integrase-mediated recombination genome insertions have been reported with human Burkitt’s lymphoma cells and mammalian artificial chromosomes (Dröge, Christ and Lorbach, 2003; Lindenbaum *et al.*, 2004). An innovative strategy for λ integrase-mediated SSR in hESCs would be to utilize possible human genome sequence tracts as

potential target sites (“*attH*”) that are homologues to the 21-bp *attB* site, which can be targeted via a modified *attP*’ targeting vector (Fig. V.1). Such tracts should contain inverted consensus λ integrase recognition sites separated by a 7-bp–long overlap region, or spacer. A preliminary bioinformatics analysis of the human genome identified more than a thousand such sites, nearly all of which are single copy because of a unique spacer sequence (Dröge and Li, unpublished results).

Based on the high fidelity of the λ system in mammalian cells (Dröge, unpublished results), it should be possible to specifically direct the insertion of vectors into a few selected sites that may be more accessible in hESC chromatin. Because a strategy of gene insertion employing the pair *attB/attP*, or derivatives thereof, will generate *attL/attR* sites in the genome, integration could be stable even in the continued presence of the recombinase. This should allow for sequential targeting events into different genomic loci.

Figure V.1. Endogenous *attH* λ integrative targeting strategy. Endogenous “*attH*” sequences in the human genome can be targeted via specific modified *attP*⁺ targeting vectors. Reporter gene or genetic payloads can be inserted into these loci by λ integrase-mediated SSR.

V.3 Characterization of *UTF1*

Genomic elements

Expression of *UTF1* is controlled by a 5' TATA-less promoter consisting of four GC boxes. The 3' enhancer element harbors a twin octamer sequence where the synergistic binding of Oct4 and Sox2 is essential for *UTF1* expression in both mouse and human ESCs (Nishimoto *et al.*, 1999; Boyer *et al.*, 2005). Sp1-like transcription factors that bind to GC boxes present in the *UTF1* promoter are most likely involved in the regulation of *UTF1*, since they are regarded as regulators of embryonic development in vertebrates (Zhao and Meng, 2005) and shown to be involved in the transcriptional control of both *Oct4* and *Nanog* (Yang *et al.*, 2005; Wu and Yao, 2006). Interestingly, the addition of a pair of SP1-binding sites 5' of the endogenous *UTF1* promoter enables our *UTF1-EGFP* construct to transcribe *EGFP* even with the removal of the 3' *UTF1* enhancer. More studies could be done on the characterization of these SP1-binding sites and the way they influence transcription of pluripotency genes in hESCs.

Expression and transcript knockdown

We show here that *UTF1* is indeed expressed in hESCs, but not in human primary fibroblasts and carcinoma cell lines. Further, both 5' promoter and 3' enhancer are needed for expression. A novel finding is that an octamer sequence (*M1*) which is conserved in the *Nanog* promoter of many species (Rodda *et al.*, 2005) is also present upstream of the analogous human *Oct4/Sox2* cognate sequence in the 3' *UTF1* enhancer. The fact that *M1* mutations resulted in equally muted expression levels of *UTF1* and *Nanog*, strongly suggests that similar *M1*-binding factor(s) are involved in the control of

both pluripotency marker genes. Our results also confirm previous reports showing that endogenous *UTF1* is downregulated faster than *Oct4* or *Nanog* at the onset of differentiation to nearly undetectable levels. In addition, detailed analysis of SymAtlas data from the Genomics Institute of the Novartis Research Foundation showed that both *UTF1* and *Nanog* expression are greatly diminished at day 8.5 of mouse embryo formation, unlike *Oct4* or *Sox2* (<http://symatlas.gnf.org>). Together, these results firmly establish *UTF1* as a sensitive and reliable pluripotency marker for hESCs.

Recent reports revealed that *UTF1* is also an important marker expressed in other stem cell types. For example, spermatogonial stem cells isolated from adult mouse testis were shown to exhibit features reminiscent of pluripotent ESCs. These cells express *Oct4*, *Nanog*, *Rex1*, and *UTF1* and, interestingly, also downregulate *UTF1* significantly faster than these other markers upon embryoid body formation (Guan *et al.*, 2006). Further, pluripotent stem cells derived from re-programmed primary mouse fibroblasts through co-expression of transgenes encoding Oct4, Sox2, c-Myc, and Klf4 also express *UTF1*. It is noteworthy that the expression level of *UTF1* in pluripotent stem cells derived from different sources is always more tightly linked with that of the key marker *Oct4*, and not with other markers like *Cripto* or *Nanog* (Takahashi and Yamanaka, 2006). These data, in conjunction with results presented here, indicate that our *UTF1*-based selectivity tool should be widely applicable to select for pluripotent human stem cell lines derived from different sources.

V.4 ZUN cell line characterization and differentiation

ZUN hESC characterization

We demonstrated that the control of both endogenous and exogenous *UTF1* in ZUN cells is linked, as expected, and that *UTF1*-driven *Neo* expression in hESCs in conjunction with G418 selection can be used to efficiently ablate differentiating cells in standard culture conditions. This resulted in ZUN cultures which display a higher global expression level of key pluripotency markers. We showed that this elevated level is most likely due to an increase in the fraction of Oct4 and Sox2 positive hESCs. Furthermore, the number of cells staining positive for surface markers TRA-1-60 and TRA-1-81 was also elevated which, if taken together, indicates that ZUN cultures are more homogeneously pluripotent than the parental cultures (**Figs. IV.26-28**). Indeed, parental and ZUN cultures formed EBs and teratomas that were indistinguishable (**Figs. IV.44-45**), thus demonstrating that *UFTI* transgene insertion did not compromise the pluripotency of these ZUN lines.

Upon removal of selection pressure from ZUN cultures, the overall expression level of these markers declined, which indicates loss of pluripotency and, consequently, more heterogeneous hESC cultures. Interestingly, the absence of selection pressure for 60 days did not cause widespread silencing of the exogenous *UTF1-Neo*. This enabled us to add G418 only periodically in order to select against differentiating hESCs in ZUN cultures. Hence, the *UTF1* promoter/enhancer combination could be a generally applicable control element for sustained expression of transgenes in pluripotent hESCs, even in the absence of selection.

ZUN hESC utility

We showed that the generation of ZUN lines harnessed the sensitivity of *UTF1* as a pluripotency marker which resulted in rapid and efficient ablation of differentiating cells. An unexpected finding was the refractoriness of ZUN cultures to the onset of differentiation. We attribute this phenotype to the increased fraction of pluripotent hESCs in homogeneous ZUN cultures compared to typical, more heterogeneous hESC cultures. Given the propensity for heterogeneous cultures to spontaneously differentiate (Mountford *et al.*, 1998), the increased stability of human stem cell cultures with respect to the uniformity in pluripotency combined with efficient ablation of differentiating cells offered by the ZUN system could significantly improve large scale growth for many future hESCs applications.

An intrinsic advantage of the *UTF1*-based selection system over other pluripotency marker-based strategies is its enhanced sensitivity towards many differentiation pathways. Although the well-characterized pluripotency marker *Oct4* has been a component in strategies for both mouse and human ESCs (Mountford *et al.*, 1998; Zwaka and Thomson, 2003), the use of *Oct4* in this context can be problematic. Apart from being less sensitive to the onset of differentiation, its expression level is also very tightly controlled. A less than two-fold increase in *Oct4* expression alone leads to differentiation into primitive endodermal and mesodermal lineages in mouse ESCs (Niwa *et al.*, 2000). This implies that the *Oct4-Neo* system is potentially less sensitive to these differentiation pathways, and, thus, will not be able to efficiently ablate cells that are undergoing differentiation into these lineages. In addition, the < 2kb length of the entire human *UTF1*

promoter/enhancer element represents a significant advantage for genetic manipulation of hESCs, or other human cell types, if one compares this to an unwieldy 8kb of the *Oct4* control element.

Potential ZUN hESC applications

Our ZUN system has a large range of potential applications. In its current context as a stable transgene in hESCs, it can be used for the routine maintenance of homogeneously pluripotent cultures. In combination with gene targeting technologies, this application can be upscaled via bioreactor culture automation to meet the larger cell quantities ultimately needed for stem cell-based therapies (Terstegge *et al.*, 2007). The ablative properties of the ZUN system has huge pharmaceutical potential in the field of cell therapy where it can be harnessed in large-scale, sensitive, high-throughput screens for small compound- or drug-induced differentiation (Pouton and Haynes, 2005). In addition, the optimization of combined selection/differentiation protocols with this system could lead to the derivation of tissue-specific multipotent stem cells. These cells could then be used to repopulate stem cell niches in patients or as a pre-clinical test system for new drugs or other potential therapeutics.

ZUN lines can be employed to screen and characterize synthetic factors, media components or surface matrices that permit maintenance of pluripotency, and it can be used to test 3D scaffolds or aid in the derivation of new hESC lines. As indicated above, the ZUN system shows potential in selection protocols in order to isolate adult stem cells from primary cultures. It can also be used in hESCs in conjunction with cell

differentiation strategies as a component of positive-negative selection. The system can also function as a tool in somatic cell reprogramming experiments to identify successfully reprogrammed pluripotent cells. It can also be used to improve future human somatic cell nuclear transfer protocols.

With our first demonstration of its utility here and its many possible applications, we envision that our ZUN system will become a valuable selectivity tool for many facets of stem cell research and an important cell culture tool in the pharmaceutical industry.

VI. CONCLUSION

We have successfully established a fully-functional hESC lab in NTU, with capabilities ranging from cell culture and characterization to cell line generation and analysis. This has enabled us to develop two different technologies for use in hESCs and potentially other stem cell types. The λ integrase SSR genomic manipulation system was demonstrated to function in both HeLa and hESCs, albeit at low efficiencies and our *UTF1*-based ZUN system in hESCs was utilized to select for more homogeneously pluripotent cultures.

The efficiency of mutant λ integrase in catalyzing genomic recombination was minimal. Further development in many areas such as enzyme efficiency, reporter system design, *att* site and accessory factor optimization have to be improved upon to increase the efficiency of this high-precision tool for use in hESC genomic surgery. In addition, more stable hESC lines in future experiments should be generated so as to increase the number of permissive loci that can be analyzed.

As a tool to select for homogeneously pluripotent populations, the ZUN system has performed admirably in the hESC context. The next step would be to attempt to scale up ZUN hESC culture incrementally to reach bioreactor quantities, transfer this system into somatic cells for use as an assay system for successful reprogramming, and to use it as a sensitive marker to select for, or derive adult stem cells.

VII. REFERENCES

Adler S, Pellizzer C, Paparella M, Hartung T, Bremer S. The effects of solvents on embryonic stem cell differentiation. *Toxicol In Vitro*. 2006; 20(3):265-71.

Alberio R, Campbell KH, Johnson AD. Reprogramming somatic cells into stem cells. *Reproduction*. 2006; 132(5):709-20.

Amit M, Carpenter MK, Inokuma MS, Chiu CP, Harris CP, Waknitz MA, Itskovitz-Eldor J, Thomson JA. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol*. 2000; 227(2):271-8.

Amit M, Itskovitz-Eldor J. Sources, Derivation, and Culture of Human Embryonic Stem Cells. *Semin Reprod Med*. 2006; 24:298-303.

Amit M, Margulets V, Segev H, Shariki K, Laevsky I, Coleman R, Itskovitz-Eldor J. Human feeder layers for human embryonic stem cells. *Biol Reprod*. 2003; 68(6):2150-6.

Aranda M, Kanellopoulou C, Christ N, Peitz M, Rajewsky K, Dröge P. Altered directionality in the cre-loxP site-specific recombination pathway. *J Mol Biol*. 2001; 311:453-9.

Atlasi Y, Mowla SJ, Ziaee SA, Bahrami AR. OCT-4, an embryonic stem cell marker, is highly expressed in bladder cancer. *Int J Cancer*. 2007; 120(7):1598-602.

Babaie Y, Herwig R, Greber B, Brink TC, Wruck W, Groth D, Lehrach H, Burdon T, Adjaye J. Analysis of oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem Cells*. 2007; 25(2):500-10.

Ben-Dor I, Itsykson P, Goldenberg D, Galun E, Reubinooff BE. Lentiviral vectors harboring a dual-gene system allow high and homogeneous transgene expression in selected polyclonal human embryonic stem cells. *Mol Ther*. 2006; 14(2):255-67.

Benzing C, Segschneider M, Leinhaas A, Itskovitz-Eldor J, Brüstle O. Neural conversion of human embryonic stem cell colonies in the presence of fibroblast growth factor-2. *Neuroreport*. 2006; 17(16):1675-81.

Bhattacharya B, Miura T, Brandenberger R, Mejido J, Luo Y, Yang AX, Joshi BH, Ginis I, Thies RS, Amit M, Lyons I, Condie BG, Itskovitz-Eldor J, Rao MS, Puri RK. Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood*. 2004; 103(8):2956-64.

Boese Q, Leake D, Reynolds A, Read S, Scaringe SA, Marshall WS, Khvorova A. Mechanistic insights aid computational short interfering RNA design. *Methods Enzymol*. 2005; 392:73-96.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 2005; 122(6):947-56.

Boyer LA, Mathur D, Jaenisch R. Molecular control of pluripotency. *Curr Opin Genet Dev*. 2006; 16(5):455-62.

Branda CS, Dymecki SM. Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. *Dev Cell*. 2004; 6:7-28.

Brandenberger R, Wei H, Zhang S, Lei S, Murage J, Fisk GJ, Li Y, Xu C, Fang R, Guegler K, Rao MS, Mandalam R, Lebkowski J, Stanton LW. Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat Biotechnol*. 2004; 22(6):707-16.

Brimble SN, Zeng X, Weiler DA, Luo Y, Liu Y, Lyons IG, Freed WJ, Robins AJ, Rao MS, Schulz TC. Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001. *Stem Cells Dev*. 2004; 13(6):585-97.

Brolen GK, Heins N, Edsbagge J, Semb H. Signals from the embryonic mouse pancreas induce differentiation of human embryonic stem cells into insulin-producing beta-cell-like cells. *Diabetes*. 2005; 54(10):2867-74.

Cabrera CM, Cobo F, Nieto A, Concha A. Strategies for preventing immunologic rejection of transplanted human embryonic stem cells. *Cytotherapy*. 2006; 8(5):517-8.

Cai J, Chen J, Liu Y, Miura T, Luo Y, Loring JF, Freed WJ, Rao MS, Zeng X. Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells*. 2006; 24(3):516-30.

Carpenter MK, Rosler ES, Fisk GJ, Brandenberger R, Ares X, Miura T, Lucero M, Rao MS. Properties of four human embryonic stem cell lines maintained in a feeder-free culture system. *Dev Dyn*. 2004; 229(2):243-58.

Cheng L, Hammond H, Ye Z, Zhan X, Dravid G. Human adult marrow cells support prolonged expansion of human embryonic stem cells in culture. *Stem Cells*. 2003; 21(2):131-42.

Cheung M, Briscoe J. Neural crest development is regulated by the transcription factor Sox9. *Development*. 2003; 130(23):5681-93.

Choi T, Huang M, Gorman C, Jaenisch R. A generic intron increases gene expression in transgenic mice. *Mol Cell Biol.* 1991; 11(6):3070-4.

Christ N, Corona T, Dröge P. Site-specific recombination in eukaryotic cells mediated by mutant λ integrases: implications for synaptic complex formation and the reactivity of episomal DNA segments. *J Mol Biol.* 2002; 319:305-14.

Christ N, Dröge P. Genetic manipulation of mouse embryonic stem cells by mutant λ integrase. *Genesis.* 2002; 32:203-8.

Coburn SP, Mahuren JD, Jain M, Zubovic Y, Wortsman J. Alkaline phosphatase (EC 3.1.3.1) in serum is inhibited by physiological concentrations of inorganic phosphate. *J Clin Endocrinol Metab.* 1998; 83(11):3951-7.

Corona T, Bao Q, Christ N, Schwartz T, Li J, Dröge P. Activation of site-specific DNA integration in human cells by a single chain integration host factor. *Nucleic Acids Res.* 2003; 31(17):5140-8.

Costa M, Dottori M, Ng E, Hawes SM, Sourris K, Jamshidi P, Pera MF, Elefanty AG, Stanley EG. The hESC line Envy expresses high levels of GFP in all differentiated progeny. *Nat Methods.* 2005; 2(4):259-60.

Cowan CA, Klimanskaya I, McMahon J, Atienza J, Witmyer J, Zucker JP, Wang S, Morton CC, McMahon AP, Powers D, Melton DA. Derivation of embryonic stem-cell lines from human blastocysts. *N Engl J Med.* 2004; 350(13):1353-6.

D'Amour KA, Bang AG, Eliazer S, Kelly OG, Agulnick AD, Smart NG, Moorman MA, Kroon E, Carpenter MK, Baetge EE. Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat Biotechnol.* 2006; 24(11):1392-401.

Darr H, Mayshar Y, Benvenisty N. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development.* 2006; 133(6):1193-201.

Draper JS, Smith K, Gokhale P, Moore HD, Maltby E, Johnson K, Meisner L, Zwaka TP, Thomson JA, Andrews PW. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat Biotechnol.* 2004; 22(1):53-4.

Dröge P, Christ N, Lorbach E. (2003) Sequence-specific DNA recombination in eukaryotic cells. U.S. Patent Application Publication US20030027337A1.

Drukker M, Benvenisty N. The immunogenicity of human embryonic stem-derived cells. *Trends Biotechnol.* 2004; 22(3):136-41.

Eifert C, Sangster-Guity N, Yu LM, Chittur SV, Perez AV, Tine JA, McCormick PJ. Global gene expression profiles associated with retinoic acid-induced differentiation of embryonal carcinoma cells. *Mol Reprod Dev.* 2006; 73(7):796-824.

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. *Curr Biol.* 2001; 11(7):514-8.

Esterre P, Melin M, Serrar M, Grimaud JA. New specific markers of human and mouse fibroblasts. *Cell Mol Biol.* 1992; 38(3):297-301.

Ezeh UI, Turek PJ, Reijo RA, Clark AT. Human embryonic stem cell genes OCT4, NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma. *Cancer.* 2005; 104(10):2255-65.

Forsyth NR, Musio A, Vezzoni P, Simpson AH, Noble BS, McWhir J. Physiologic oxygen enhances human embryonic stem cell clonal recovery and reduces chromosomal abnormalities. *Cloning Stem Cells.* 2006; 8(1):16-23.

Fukushima A, Okuda A, Nishimoto M, Seki N, Hori TA, Muramatsu M. Characterization of functional domains of an embryonic stem cell coactivator UTF1 which are conserved and essential for potentiation of ATF-2 activity. *J Biol Chem.* 1998; 273(40):25840-9.

Gallagher EJ, Lodge P, Ansell R, McWhir J. Isolation of murine embryonic stem and embryonic germ cells by selective ablation. *Transgenic Res.* 2003; 12(4):451-60.

Gerrard L, Zhao D, Clark AJ, Cui W. Stably transfected human embryonic stem cell clones express OCT4-specific green fluorescent protein and maintain self-renewal and pluripotency. *Stem Cells.* 2005; 23(1):124-33.

Ginis I, Luo Y, Miura T, Thies S, Brandenberger R, Gerecht-Nir S, Amit M, Hoke A, Carpenter MK, Itskovitz-Eldor J, Rao MS. Differences between human and mouse embryonic stem cells. *Dev Biol.* 2004; 269(2):360-80.

Glaser T, Brüstle O. Retinoic acid induction of ES-cell-derived neurons: the radial glia connection. *Trends Neurosci.* 2005; 28(8):397-400.

Glover DJ, Lipps HJ, Jans DA. Towards safe, non-viral therapeutic gene expression in humans. *Nat Rev Genet.* 2005; 6(4):299-310.

Golos TG, Pollastrini LM, Gerami-Naini B. Human embryonic stem cells as a model for trophoblast differentiation. *Semin Reprod Med.* 2006; 24(5):314-21.

Grindley ND, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. *Annu Rev Biochem.* 2006; 75:567-605.

Grinnell KL, Yang B, Eckert RL, Bickenbach JR. De-differentiation of mouse interfollicular keratinocytes by the embryonic transcriptional factor Oct-4. *J Invest Dermatol*. 2007; 127(2):372-80.

Gropp M, Itsykson P, Singer O, Ben-Hur T, Reinhartz E, Galun E, Reubinoff BE. Stable genetic modification of human embryonic stem cells by lentiviral vectors. *Mol Ther*. 2003; 7(2):281-7.

Groth AC, Calos MP. Phage integrases: biology and applications. *J Mol Biol*. 2004; 335(3):667-78.

Guan K, Nayernia K, Maier LS, Wagner S, Dressel R, Lee JH, Nolte J, Wolf F, Li M, Engel W, Hasenfuss G. Pluripotency of spermatogonial stem cells from adult mouse testis. *Nature*. 2006; 440(7088):1199-203.

Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*. 2003; 302:415-8.

Hamazaki T, Iiboshi Y, Oka M, Papst PJ, Meacham AM, Zon LI, Treadwell N. Hepatic maturation in differentiating embryonic stem cells in vitro. *FEBS Lett*. 2001; 497(1):15-9.

Hanahan D. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol.* 1983; 166(4):557-80.

Henderson JK, Draper JS, Baillie HS, Fishel S, Thomson JA, Moore H, Andrews PW. Preimplantation human embryos and embryonic stem cells show comparable expression of stage-specific embryonic antigens. *Stem Cells.* 2002; 20(4):329-37.

Heng BC, Cao T, Haider HK, Wang DZ, Sim EK, Ng SC. An overview and synopsis of techniques for directing stem cell differentiation in vitro. *Cell Tissue Res.* 2004; 315(3):291-303.

Heng BC, Cao T, Tong GQ, Ng SC. Potential utility of cell-permeable transcription factors to direct stem cell differentiation. *Stem Cells Dev.* 2004; 13(5):460-2.

Heng BC, Haider HK, Sim EK, Cao T, Ng SC. Strategies for directing the differentiation of stem cells into the cardiomyogenic lineage in vitro. *Cardiovasc Res.* 2004; 62(1):34-42.

Heng BC, Haider HK, Sim EK, Cao T, Tong GQ, Ng SC. Comments about possible use of human embryonic stem cell-derived cardiomyocytes to direct autologous adult stem cells into the cardiomyogenic lineage. *Acta Cardiol.* 2005; 60(1):7-12.

Heng BC, Liu H, Cao T. Utilising human embryonic stem cells as “catalysts” for biological repair and regeneration. Challenges and some possible strategies. Clin Exp Med. 2005; 5(1):37-9.

Hentze H, Graichen R, Colman A. Cell therapy and the safety of embryonic stem cell-derived grafts. Trends Biotechnol. 2007; 25(1):24-32.

Hewitt Z, Forsyth NR, Waterfall M, Wojtacha D, Thomson AJ, McWhir J. Fluorescence-activated single cell sorting of human embryonic stem cells. Cloning Stem Cells. 2006; 8(3):225-34.

Hirst CE, Ng ES, Azzola L, Voss AK, Thomas T, Stanley EG, Elefanty AG. Transcriptional profiling of mouse and human ES cells identifies SLAIN1, a novel stem cell gene. Dev Biol. 2006; 293(1):90-103.

Hoffman LM, Carpenter MK. Characterization and culture of human embryonic stem cells. Nat Biotechnol. 2005; 23(6):699-708.

Hong Y, Winkler C, Liu T, Chai G, Scharf M. Activation of the mouse *Oct4* promoter in medaka embryonic stem cells and its use for ablation of spontaneous differentiation. Mech Dev. 2004; 121(7-8):933-43.

Hong Y, Winkler C, Scharl M. Pluripotency and differentiation of embryonic stem cell lines from the medakafish (*Oryzias latipes*). *Mech Dev*. 1996; 60(1):33-44.

Hoof DV, Passier R, Oostwaard DW-V, Pinkse MWH, Heck AJR, Mummery CL, Krijgsveld J. A quest for human and mouse embryonic stem cell-specific proteins. *Mol Cell Proteomics*. 2006; 5:1261-73.

Hyslop L, Stojkovic M, Armstrong L, Walter T, Stojkovic P, Przyborski S, Herbert M, Murdoch A, Strachan T, Lako M. Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. *Stem Cells*. 2005; 23(8):1035-43.

Iacovitti L, Donaldson AE, Marshall CE, Suon S, Yang M. A protocol for the differentiation of human embryonic stem cells into dopaminergic neurons using only chemically defined human additives: Studies in vitro and in vivo. *Brain Res*. 2007; 1127(1):19-25.

Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR. Dissecting self-renewal in stem cells with RNA interference. *Nature*. 2006; 442(7102):533-8.

James D, Noggle SA, Swigut T, Brivanlou AH. Contribution of human embryonic stem cells to mouse blastocysts. *Dev Biol*. 2006; 295(1):90-102.

Johnson RD, Jasin M. Double-strand-break-induced homologous recombination in mammalian cells. *Biochem Soc Trans.* 2001; 29(Pt2):196-201.

Jung SM, Chu PH, Shiu TF, Wu HH, Kuo TT, Chu JJ, Lin PJ. Expression of OCT4 in the primary germ cell tumors and thymoma in the mediastinum. *Appl Immunohistochem Mol Morphol.* 2006. 14(3):273-5.

Klimanskaya I, Chung Y, Becker S, Lu SJ, Lanza R. Human embryonic stem cell lines derived from single blastomeres. *Nature.* 2006; 444(7118):481-5.

Klimanskaya I, Chung Y, Meisner L, Johnson J, West MD, Lanza R. Human embryonic stem cells derived without feeder cells. *Lancet.* 2005; 365(9471):1636-41.

Koch P, Siemen H, Biegler A, Itskovitz-Eldor J, Brüstle O. Transduction of human embryonic stem cells by ecotropic retroviral vectors. *Nucleic Acids Res.* 2006; 34(18):e120.

Kolb AF. Genome engineering using site-specific recombinases. *Cloning Stem Cells.* 2002; 4:65–80.

Kuroda T, Tada M, Kubota H, Kimura H, Hatano SY, Suemori H, Nakatsuji N, Tada T. Octamer and Sox elements are required for transcriptional cis regulation of *Nanog* gene expression. *Mol Cell Biol.* 2005; 25(6):2475-85.

Lakshmipathy U, Pelacho B, Sudo K, Linehan JL, Coucouvanis E, Kaufman DS, Verfaillie CM. Efficient transfection of embryonic and adult stem cells. *Stem Cells*. 2004; 22(4):531-43.

Landy A. Dynamic, structural and regulatory aspects of lambda site-specific recombination. *Ann Rev Biochem*. 1989; 58:913-49.

Landy A. Mechanistic and structural complexity in the site-specific recombination pathways of Int and FLP. *Curr Opin Genet Dev*. 1993; 3:699-707.

Laslett AL, Filipczyk AA, Pera MF. Characterization and culture of human embryonic stem cells. *Trends Cardiovasc Med*. 2003; 13(7):295-301.

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(8):1923-30.

Le Y, Gagneten S, Tombaccini D. Nuclear determinants of the phage P1 Cre DNA recombinase. *Nucleic Acids Res*. 1999; 27:4703-9.

Lenka N, Lu ZJ, Sasse P, Hescheler J, Fleischmann BK. Quantitation and functional characterization of neural cells derived from ES cells using nestin enhancer-mediated targeting in vitro. *J Cell Sci*. 2002; 115(Pt 7):1471-85.

Lindenbaum M, Perkins E, Csonka E et al. A mammalian artificial chromosome engineering system (ACE system) applicable to biopharmaceutical protein production, transgenesis and gene-based cell therapy. *Nucleic Acids Res.* 2004; 32:e172.

Liu YP, Dovzhenko OV, Garthwaite MA, Dambaeva SV, Durning M, Pollastrini LM, Golos TG. Maintenance of pluripotency in human embryonic stem cells stably over-expressing enhanced green fluorescent protein. *Stem Cells Dev.* 2004; 13(6):636-45.

Lorbach E, Christ N, Schwikardi M, Dröge P. Site-specific recombination in human cells catalyzed by phage λ integrase mutants. *J Mol Biol.* 2000; 296:1175-81.

Lu J, Hou R, Booth CJ, Yang SH, Snyder M. Defined culture conditions of human embryonic stem cells. *Proc Natl Acad Sci USA.* 2006; 103(15):5688-93.

Ludwig TE, Levenstein ME, Jones JM, Berggren WT, Mitchen ER, Frane JL, Crandall LJ, Daigh CA, Conard KR, Piekarczyk MS, Llanas RA, Thomson JA. Derivation of human embryonic stem cells in defined conditions. *Nat Biotechnol.* 2006; 24(2):185-7.

Luo D, Han E, Belcheva N, Saltzman WM. A self-assembled, modular DNA delivery system mediated by silica nanoparticles. *J Control Release.* 2004; 95(2):333-41.

Luo D, Saltzman WM. Enhancement of transfection by physical concentration of DNA at the cell surface. *Nat Biotechnol.* 2000; 18(8):893-5.

Ma Y, Ramezani A, Lewis R, Hawley RG, Thomson JA. High-level sustained transgene expression in human embryonic stem cells using lentiviral vectors. *Stem Cells*. 2003; 21(1):111-7.

McNeish J. Embryonic stem cells in drug discovery. *Nat Rev Drug Discov*. 2004; 3(1):70-80.

McWhir J, Schnieke AE, Ansell R, Wallace H, Colman A, Scott AR, Kind AJ. Selective ablation of differentiated cells permits isolation of embryonic stem cell lines from murine embryos with a non-permissive genetic background. *Nat Genet*. 1996; 14(2):223-6.

Menendez P, Wang L, Bhatia M. Genetic manipulation of human embryonic stem cells: a system to study early human development and potential therapeutic applications. *Curr Gene Ther*. 2005; 5(4):375-85.

Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*. 2003; 113:631-42.

Mohr JC, de Pablo JJ, Palecek SP. Electroporation of human embryonic stem cells: Small and macromolecule loading and DNA transfection. *Biotechnol Prog*. 2006; 22(3):825-34.

Morali OG, Jouneau A, McLaughlin KJ, Thiery JP, Larue L. IGF-II promotes mesoderm formation. *Dev Biol.* 2000; 227(1):133-45.

Mountford P, Nichols J, Zevnik B, O'Brien C, Smith A. Maintenance of pluripotential embryonic stem cells by stem cell selection. *Reprod Fertil Dev.* 1998; 10(7-8):527-33.

Murnane JP, Yezzi MJ, Young BR. Recombination events during integration of transfected DNA into normal human cells. *Nucleic Acids Res.* 1990; 18(9):2733-8.

Nishimoto M, Fukushima A, Miyagi S, Suzuki Y, Sugano S, Matsuda Y, Hori T, Muramatsu M, Okuda A. Structural analyses of the UTF1 gene encoding a transcriptional coactivator expressed in pluripotent embryonic stem cells. *Biochem Biophys Res Commun.* 2001; 285(4):945-53.

Nishimoto M, Fukushima A, Okuda A, Muramatsu M. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol.* 1999; 19(8):5453-65.

Nishimoto M, Miyagi S, Yamagishi T, Sakaguchi T, Niwa H, Muramatsu M, Okuda A. Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the *UTF1* locus. *Mol Cell Biol.* 2005; 25(12):5084-94.

Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet.* 2000; 24(4):372-6.

Nolden L, Edenhofer F, Haupt S, Koch P, Wunderlich FT, Siemen H, Brüstle O. Site-specific recombination in human embryonic stem cells induced by cell-permeant Cre recombinase. *Nat Methods.* 2006; 3(6):461-7.

Odorico JS, Kaufman DS, Thomson JA. Multilineage differentiation from human embryonic stem cell lines. *Stem Cells.* 2001; 19(3):193-204.

Okuda A, Fukushima A, Nishimoto M, Orimo A, Yamagishi T, Nabeshima Y, Kuro-o M, Nabeshima Y, Boon K, Keaveney M, Stunnenberg HG, Muramatsu M. UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *EMBO J.* 1998; 17(7):2019-32.

Okumura-Nakanishi S, Saito M, Niwa H, Ishikawa F. Oct-3/4 and Sox2 regulate *Oct-3/4* gene in embryonic stem cells. *J Biol Chem.* 2005; 280(7):5307-17.

Olivier EN, Qiu C, Velho M, Hirsch RE, Bouhassira EE. Large-scale production of embryonic red blood cells from human embryonic stem cells. *Exp Hematol.* 2006; 34(12):1635-42.

Park S, Kim EY, Ghil GS, Joo WS, Wang KC, Kim YS, Lee YJ, Lim J. Genetically modified human embryonic stem cells relieve symptomatic motor behavior in a rat model of Parkinson's disease. *Neurosci Lett*. 2003; 353(2):91-4.

Pera MF, Reubinooff B, Trounson A. Human embryonic stem cells. *J Cell Sci*. 2000; 113(Pt 1):5-10.

Perrier AL, Tabar V, Barberi T, Rubio ME, Bruses J, Topf N, Harrison NL, Studer L. Derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc Natl Acad Sci USA*. 2004; 101(34):12543-8.

Perry P, Sauer S, Billon N, Richardson WD, Spivakov M, Warnes G, Livesey FJ, Merkenschlager M, Fisher AG, Azuara V. A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction. *Cell Cycle*. 2004; 3(12):1645-50.

Poelstra K, Bakker WW, Klok PA, Hardonk MJ, Meijer DK. A physiologic function for alkaline phosphatase: endotoxin detoxification. *Lab Invest*. 1997; 76(3):319-27.

Pouton CW, Haynes JM. Pharmaceutical applications of embryonic stem cells. *Adv Drug Deliv Rev*. 2005; 57(13):1918-34.

Rao M. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev Biol.* 2004; 275:269-86.

Raymond CS, Soriano P. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS ONE.* 2007; 2:e162.

Rege TA, Hagood JS. Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis. *FASEB J.* 2006; 20(8):1045-54.

Reubinoff BE, Pera MF, Fong CY, Trounson A, Bongso A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotechnol.* 2000; 18(4):399-404.

Richards M, Tan SP, Chan WC, Bongso A. Reverse serial analysis of gene expression (SAGE) characterization of orphan SAGE tags from human embryonic stem cells identifies the presence of novel transcripts and antisense transcription of key pluripotency genes. *Stem Cells.* 2006; 24(5):1162-73.

Risebro CA, Smart N, Dupays L, Breckenridge R, Mohun TJ, Riley PR. Hand1 regulates cardiomyocyte proliferation versus differentiation in the developing heart. *Development.* 2006; 133(22):4595-606.

Roche E, Jones J, Arribas MI, Leon-Quinto T, Soria B. Role of small bioorganic molecules in stem cell differentiation to insulin-producing cells. *Bioorg Med Chem*. 2006; 14(19):6466-74.

Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P. Transcriptional regulation of *nanog* by OCT4 and SOX2. *J Biol Chem*. 2005; 280(26):24731-7.

Rosler ES, Fisk GJ, Ares X, Irving J, Miura T, Rao MS, Carpenter MK. Long-term culture of human embryonic stem cells in feeder-free conditions. *Dev Dyn*. 2004; 229(2):259-74.

Sato N, Sanjuan IM, Heke M et al. Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol*. 2003; 260:404-13.

Sauer B, Henderson N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci USA*. 1988; 85:5166-70.

Schuldiner M, Itskovitz-Eldor J, Benvenisty N. Selective ablation of human embryonic stem cells expressing a "suicide" gene. *Stem Cells*. 2003; 21(3):257-65.

Schuldiner M, Yanuka O, Itskovitz-Eldor J, Melton DA, Benvenisty N. Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. *Proc Natl Acad Sci USA*. 2000; 97(21):11307-12.

Schwikardi M. Sequenz-spezifische DNA Rekombination durch $\gamma\delta$ resolvase in eukaryotischen Zellen. Dissertation, University of Cologne, Germany, 2001.

Schwikardi M, Dröge P. Site-specific recombination in mammalian cells catalyzed by $\gamma\delta$ resolvase mutants: implications for the topology of episomal DNA. FEBS Lett. 2000; 471:147–50.

Schwikardi M, Dröge P. Use of site-specific recombination as a probe of nucleoprotein complex formation in chromatin. Eur J Biochem. 2001; 268:6256–62.

Segev H, Kenyagin-Karsenti D, Fishman B, Gerecht-Nir S, Ziskind A, Amit M, Coleman R, Itskovitz-Eldor J. Molecular analysis of cardiomyocytes derived from human embryonic stem cells. Dev Growth Differ. 2005; 47(5):295-306.

Sen G, Wehrman TS, Myers JW, Blau HM. Restriction enzyme-generated siRNA (REGS) vectors and libraries. Nat Genet. 2004; 36(2):183-9.

Siemen H, Nix M, Endl E, Koch P, Itskovitz-Eldor J, Brüstle O. Nucleofection of human embryonic stem cells. Stem Cells Dev. 2005; 14(4):378-83.

Skottman H, Stromberg A-M, Matilainen E, Inzunza J, Hovatta O, Lahesmaa R. Unique gene expression signature by human embryonic stem cells cultured under serum-free conditions correlates with their enhanced and prolonged growth in an undifferentiated stage. *Stem Cells*. 2005; 24(1):151-67.

Solter D, Knowles BB. Immunosurgery of mouse blastocyst. *Proc Natl Acad Sci USA*. 1975;72(12):5099-102.

Spagnoli FM, Hemmati-Brivanlou A. Guiding embryonic stem cells towards differentiation: lessons from molecular embryology. *Curr Opin Genet Dev*. 2006; 16(5):469-75.

Stewart MH, Bosse M, Chadwick K, Menendez P, Bendall SC, Bhatia M. Clonal isolation of hESCs reveals heterogeneity within the pluripotent stem cell compartment. *Nat Methods*. 2006; 3(10):807-15.

Stojkovic P, Lako M, Stewart R, Przyborski S, Armstrong L, Evans J, Murdoch A, Strachan T, Stojkovic M. An autogeneic feeder cell system that efficiently supports growth of undifferentiated human embryonic stem cells. *Stem Cells*. 2005; 23(3):306-14.

Stojkovic M, Lako M, Strachan T, Murdoch A. Derivation, growth and applications of human embryonic stem cells. *Reproduction*. 2004; 128(3):259-67.

Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A, et al. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci USA*. 2002; 99:4465-70.

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA*. 2004; 101:6062-7.

Sun Y, Li H, Liu Y, Shin S, Mattson MP, Rao MS, Zhan M. Cross-species transcriptional profiles establish a functional portrait of embryonic stem cells. *Genomics*. 2007; 89(1):22-35.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006; 126(4):663-76.

Tanaka N, Takeuchi T, Neri QV, Sills ES, Palermo GD. Laser-assisted blastocyst dissection and subsequent cultivation of embryonic stem cells in a serum/cell free culture system: applications and preliminary results in a murine model. *J Transl Med*. 2006; 4:20.

Terstegge S, Laufenberg I, Pochert J, Schenk S, Itskovitz-Eldor J, Endl E, Brüstle O. Automated maintenance of embryonic stem cell cultures. *Biotechnol Bioeng*. 2007; 96(1):195-201.

Tirumalai RS, Kwon HJ, Cardente EH, Ellenberger T, Landy A. Recognition of core-type DNA sites by lambda integrase. *J Mol Biol.* 1998; 279(3):513-27.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. *Science.* 1998; 282(5391):1145-7.

Thyagarajan B, Liu Y, Shin S, Lakshmipathy U, Scheyhing K, Xue H, Ellerstrom C, Strehl R, Hyllner J, Rao MS, Chesnut JD. Creation of engineered human embryonic stem cell lines using phiC31 Integrase. *Stem Cells.* 2007; 0283.

Torihashi S, Kuwahara M, Ogaeri T, Zhu P, Kurahashi M, Fujimoto T. Gut-like structures from mouse embryonic stem cells as an in vitro model for gut organogenesis preserving developmental potential after transplantation. *Stem Cells.* 2006; 24(12):2618-26.

Trono D. Lentiviral vectors: turning a deadly foe into a therapeutic agent. *Gene Ther.* 2000; 7(1):20-3.

Trounson A. The production and directed differentiation of human embryonic stem cells. *Endocr Rev.* 2006; 27(2):208-19.

Urbach A, Schuldiner M, Benvenisty N. Modeling for Lesch-Nyhan disease by gene targeting in human embryonic stem cells. *Stem Cells*. 2004; 22(4):635-41.

Vallier L, Rugg-Gunn PJ, Bouhon IA, Andersson FK, Sadler AJ, Pedersen RA. Enhancing and diminishing gene function in human embryonic stem cells. *Stem Cells*. 2004; 22:2–11.

Vasileva A, Jessberger R. Precise hit: Adeno-associated virus in gene targeting. *Nat Rev Microbiol*. 2005; 3(11): 837-47.

Wei CL, Miura T, Robson P, Lim SK, Xu XQ, Lee MY, Gupta S, Stanton L, Luo Y, Schmitt J, et al. Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. *Stem Cells*. 2005; 23(2):166-85.

Wobus AM, Boheler KR. Embryonic stem cells: Prospects for developmental biology and cell therapy. *Physiol. Rev*. 2005; 85:635–78.

Wu DY, Yao Z. Functional analysis of two Sp1/Sp3 binding sites in murine Nanog gene promoter. *Cell Res*. 2006; 16(3):319-22.

Xu C, Police S, Rao N, Carpenter MK. Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells. *Circ Res*. 2002; 91(6):501-8.

Xu RH, Peck RM, Li DS, Feng X, Ludwig T, Thomson JA. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Methods*. 2005; 2(3):185-90.

Xu C, Rosler E, Jiang J, Lebkowski JS, Gold JD, O'Sullivan C, Delavan-Boorsma K, Mok M, Bronstein A, Carpenter MK. Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium. *Stem Cells*. 2005; 23(3):315-23.

Yang HM, Do HJ, Oh JH, Kim JH, Choi SY, Cha KY, Chung HM, Kim JH. Characterization of putative cis-regulatory elements that control the transcriptional activity of the human Oct4 promoter. *J Cell Biochem*. 2005; 96(4):821-30.

Yang X, Smith SL, Tian XC, Lewin HA, Renard JP, Wakayama T. Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. *Nat Genet*. 2007; 39(3):295-302.

Yao S, Chen S, Clark J, Hao E, Beattie GM, Hayek A, Ding S. Long-term self-renewal and directed differentiation of human embryonic stem cells in chemically defined conditions. *Proc Natl Acad Sci USA*. 2006; 103(18):6907-12.

Yates F, Daley GQ. Progress and prospects: gene transfer into embryonic stem cells. *Gene Ther*. 2006; 13(20):1431-9.

Zhang J, Tam WL, Tong GQ, Wu Q, Chan HY, Soh BS, Lou Y, Yang J, Ma Y, Chai L, et al. *Sall4* modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of *Pou5f1*. *Nat Cell Biol.* 2006; 8(10):1114-23.

Zhao C, Meng A. Sp1-like transcription factors are regulators of embryonic development in vertebrates. *Dev Growth Differ.* 2005; 47(4):201-11.

Zhou R, Dröge P. High-precision genome surgery in human stem cells. *Current Genomics.* 2006; 7(7):427-33.

Zou GM, Yoder MC. Application of RNA interference to study stem cell function: current status and future perspectives. *Biol Cell.* 2005; 97(3):211-9.

Zwaka TP, Thomson JA. Homologous recombination in human embryonic stem cells. *Nat Biotechnol.* 2003; 21(3):319-21.

Appendix I DNA sequences for constructed plasmids***pCH-RLNRLE***

```

1      GGGCCCAAAA AGTCGCATAA AAGTGTATCC TAAATCAAAT ATCGGACAAG
      CAGTGTCTGT TATAACAAAA AATCGATTTA ATAGACACAT TAACAGCACT
101    GTTTTTTAAGT GTGCGATAAT TTATAATATT TCGGACGGTT GCAAAAATTGT
      TAGATCTATA ACTTCGTATA GCATACATTA TACAAGTTAT AGCATGATTG
201    AACAAGATGG ATTGCACGCA GGTTCTCCGG CCGCTTGGGT GGAGAGGCTA
      TTCGGCTATG ACTGGGCACA ACAGACAATC GGCTGCTCTG ATGCCGCCGT
301    GTTCCGGCTG TCAGCGCAGG GCGCCCCGGT TCTTTTTGTC AAGACCGACC
      TGTCCGGTGC CCTGAATGAA CTGCAGGACG AGGCAGCGCG GCTATCGTGG
401    CTGGCCACGA CGGGCGTTCC TTGCGCAGCT GTGCTCGACG TTGTCACTGA
      AGCGGGAAGG GACTGGCTGC TATTGGGCGA AGTGCCGGGG CAGGATCTCC
501    TGTCATCTCA CCTTGCTCCT GCCGAGAAAG TATCCATCAT GGCTGATGCA
      ATGCGGCGGC TGCATACGCT TGATCCGGCT ACCTGCCCAT TCGACCACCA
601    AGCGAAACAT CGCATCGAGC GAGCACGTAC TCGGATGGAA GCCGGTCTTG
      TCGATCAGGA TGATCTGGAC GAAGAGCATC AGGGGCTCGC GCCAGCCGAA
701    CTGTTGCGCA GGCTCAAGGC GCGCATGCCC GACGGCGAGG ATCTCGTCGT
      GACCCATGGC GATGCCTGCT TGCCGAATAT CATGGTGGAA AATGGCCGCT
801    TTTCTGGATT CATCGACTGT GGCCGGCTGG GTGTGGCGGA CCGCTATCAG
      GACATAGCGT TGGCTACCCG TGATATTGCT GAAGAGCTTG GCGGCGAATG
901    GGCTGACCGC TTCTCGTGC TTTACGGTAT CGCCGCTCCC GATTTCGCAGC
      GCATCGCCTT CTATCGCCTT CTTGACGAGT TCTTCTGAGC GGGACTCTGG
1001   GGTTCGAAAT GACCGACCAA GCGACGCCCA ACCTGCCATC ACGAGATTTT
      GATTCCACCG CCGCCTTCTA TGAAAGGTTG GGCTTCGGAA TCGTTTTCCG
1101   GGACGCCGGC TGGATGATCC TCCAGCGCGG GGATCTCATG CTGGAGTTCT
      TCGCCCACCC CGGATTGATA CCCGTGCGTA ACCAAAAAGT CGCATAAAAA
1201   TGTATCCTAA ATCAAATATC GGACAAGCAG TGTCTGTTAT AACAAAAAAT
      CGATTTAATA GACACATTAA CAGCACTGTT TTTATGTGTG CGATAATTTA
1301   TAATATTTTCG GACGGTTGCA AAATTGTTAT AACTTCGTAT AGCATACATT
      ATACAAGTTA TAAGCTTGCC GTAGATAAAC AGGCTGGGAC ACTTCACATG
1401   AGCGAAAAAT ACATCGTCAC CTGGGACATG TTGCAGATCC ATGCACGTAA
      ACTCGCAAGC CGACTGATGC CTTCTGAACA ATGGAAAGGC ATTATTGCCG
1501   TAAGCCGTGG CGGTCTGGTA CCGGTGGAA GCGGCGGTGT GAGCAAGGGC

```

	GAGGAGCTGT	TCACCGGGGT	GGTGCCCATC	CTGGTCGAGC	TGGACGGCGA
1601	CGTAAACGGC	CACAAGTTCA	GCGTGTCCGG	CGAGGGCGAG	GGCGATGCCA
	CCTACGGCAA	GCTGACCCTG	AAGTTCATCT	GCACCACCGG	CAAGCTGCCC
1701	GTGCCCTGGC	CCACCCTCGT	GACCACCCTG	ACCTACGGCG	TGCAGTGCTT
	CAGCCGCTAC	CCCGACCACA	TGAAGCAGCA	CGACTTCTTC	AAGTCCGCCA
1801	TGCCCCAAGG	CTACGTCCAG	GAGCGCACCA	TCTTCTTCAA	GGACGACGGC
	AACTACAAGA	CCCGCGCCGA	GGTGAAGTTC	GAGGGCGACA	CCCTGGTGAA
1901	CCGCATCGAG	CTGAAGGGCA	TCGACTTCAA	GGAGGACGGC	AACATCCTGG
	GGCACAAGCT	GGAGTACAAC	TACAACAGCC	ACAACGTCTA	TATCATGGCC
2001	GACAAGCAGA	AGAACGGCAT	CAAGGTGAAC	TTCAAGATCC	GCCACAACAT
	CGAGGACGGC	AGCGTGCAGC	TCGCCGACCA	CTACCAGCAG	AACACTCCCA
2101	TCGGCGACGG	CCCCGTGCTG	CTGCCCCGACA	ACCACTACCT	GAGCACCCAG
	TCCGCCCTGA	GCAAAGACCC	CAACGAGAAG	CGCGATCACA	TGGTCTTGCT
2201	GGAGTTCGTG	ACCGCCGCCG	GGGTCACTCT	CGGCATGGAC	GAGCTGTACA
	AGTAATGATC	TAGAGCTCGC	TGATCAGCCT	CGACTGTGCC	TTCTAGTTGC
2301	CGGCCATCTG	TTGTTTGCCC	CTCCCCCGTG	CCTTCCTTGA	CCCTGGAAGG
	TGCCACTCCC	ACTGTCCTTT	CCTAATAAAA	TGAGGAAATT	GCATCGCATT
2401	GTCTGAGTAG	GTGTCATTCT	ATTCTGGGGG	GTGGGGTGGG	GCAGGACAGC
	AAGGGGGAGG	ATTGGGAAGA	CAATAGCAGG	CATGCTGGGG	ATGCGGTGGG
2501	CTCTATGGCT	TCTGAGNNNG	AAAGAACCAG	CTGGGGCTCG	AGATCCACTA
	GTTCTAGCCT	CGGATCCCCA	GGAAGCTCCT	CTGTGTCCTC	ATAAACCCCTA
2601	ACCTCCTCTA	CTTGAGAGGA	CATTCCAATC	ATAGGCTGCC	CATCCACCCT
	CTGTGTCCTC	CTGTTAATTA	GGTCACTTAA	CAAAAAGGAA	ATTGGGTAGG
2701	GGTTTTTTCAC	AGACCGCTTT	CTAAGGGTAA	TTTTTAAATA	TCTGGGAAGT
	CCCTTCCACT	GCTGTGTTCC	AGAAGTGTTG	GTAAACAGCC	CACAAATGTC
2801	AACAGCAGAA	ACATACAAGC	TGTCAGCTTT	GCACAAGGGC	CCAACACCCT
	GCTCATCAAG	AAGCACTGTG	GTTGCTGTGT	TAGTAATGTG	CAAAACAGGA
2901	GGCACATTTT	CCCCACCTGT	GTAGGTTCCA	AAATATCTAG	TGTTTTTCATT
	TTTACTTGGA	TCAGGAACCC	AGCACTCCAC	TGGATAAGCA	TTATCCTTAT
3001	CCAAAACAGC	CTTGTGGTCA	GTGTTTCATCT	GCTGACTGTC	AACTGTAGCA
	TTTTTTTGGGG	TTACAGTTTG	AGCAGGATAT	TTGGTCCTGT	AGTTTGCTAA
3101	CACACCCTGC	AGCTCCAAAG	GTTCCCCACC	AACAGCAAAA	AAATGAAAAT
	TTGACCCTTG	AATGGGTTTT	CCAGCACCAT	TTTCATGAGT	TTTTTGTGTC
3201	CCTGAATGCA	AGTTTAAACAT	AGCAGTTACC	CCAATAACCT	CAGTTTAAAC
	AGTAACAGCT	TCCCACATCA	AAATATTTCC	ACAGGTAAAG	TCCTCATTTA

3301 AATTAGGCAA AGGAATTCTT GAAGACGAAA GGGCCTCGTG ATACGCCTAT
 TTTTATAGGT TAATGTCATG ATAATAATGG TTTCTTAGAC GTCAGGTGGC
 3401 ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT
 ACATTCAAAT ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC
 3501 AATAATATTG AAAAAGGAAG AGTATGAGTA TTCAACATTT CCGTGTCGCC
 CTTATTCCCT TTTTTCGGC ATTTTGCCTT CCTGTTTTTG CTCACCCAGA
 3601 AACGCTGGTG AAAGTAAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG
 GTTACATCGA ACTGGATCTC AACAGCGGTA AGATCCTTGA GAGTTTTCGC
 3701 CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTTAAAGTTC TGCTATGTGG
 CGCGGTATTA TCCCGTGTG ACGCCGGGCA AGAGCAACTC GGTCGCCGCA
 3801 TACACTATTC TCAGAATGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG
 CATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG CTGCCATAAC
 3901 CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG ATCGGAGGAC
 CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACTCGC
 4001 CTTGATCGTT GGGAACCGGA GCTGAATGAA GCCATACCAA ACGACGAGCG
 TGACACCACG ATGCCTGCAG CAATGGCAAC AACGTTGCGC AAACATATTAA
 4101 CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAATTAAT AGACTGGATG
 GAGGCGGATA AAGTTGCAGG ACCACTTCTG CGCTCGGCCC TTCCGGCTGG
 4201 CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG TCTCGCGGTA
 TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC
 4301 TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC
 TGAGATAGGT GCCTCACTGA TTAAGCATTG GTAACGTGCA GACCAAGTTT
 4401 ACTCATATAT ACTTTAGATT GATTTAAAAC TTCATTTTTTA ATTTAAAAGG
 ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG
 4501 TGAGTTTTCG TTCCACTGAG CGTCAGACCC CGTAGAAAAG ATCAAAGGAT
 CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT GCAAACAAAA
 4601 AAACCACCGC TACCAGCGGT GGTTTGTTTG CCGGATCAAG AGCTACCAAC
 TCTTTTTCCG AAGGTAAC TGCTTCAGCAG AGCGCAGATA CCAAATACTG
 4701 TCCTTCTAGT GTAGCCGTAG TTAGGCCACC ACTTCAAGAA CTCTGTAGCA
 CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG CTGCTGCCAG
 4801 TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAAGACGA TAGTTACCGG
 ATAAGGCGCA GCGGTCGGGC TGAACGGGGG GTTCGTGCAC ACAGCCCAGC
 4901 TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCTACAGC GTGAGCATTG
 AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GGCGGACAGG TATCCGGTAA
 5001 GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGGAAAC

```

GCCTGGTATC TTTATAGTCC TGTCGGGTTT CGCCACCTCT GACTTGAGCG
5101 TCGATTTTTG TGATGCTCGT CAGGGGGGCG GAGCCTATGG AAAAACGCCA
GCAACGCGGC CTTTTTACGG TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC
5201 ATGTTCTTTC CTGCGTTATC CCCTGATTCT GTGGATAACC GTATTACCGC
CTTTGAGTGA GCTGATACCG CTCGCCGAG CCGAACGACC GAGCGCAGCG
5301 AGTCAGTGAG CGAGGAAGCG GAAGAGCGCC TGATGCGGTA TTTTCTCCTT
ACGCATCTGT GCGGTATTTT ACACCGCATA TGGTGCACTC TCAGTACAAT
5401 CTGCTCTGAT GCCGCATAGT TAAGCCAGTA TACACTCCGC TATCGCTACG
TGACTGGGTC ATGGCTGCGC CCCGACACCC GCCAACACCC GCTGACGCGC
5501 CCTGACGGGC TTGTCTGCTC CCGGCATCCG CTTACAGACA AGCTGTGACC
GTCTCCGGGA GCTGCATGTG TCAGAGGTTT TCACCGTCAT CACCGAAACG
5601 CGCGAGGCAG CTGTGGAATG TGTGTCAGTT AGGGTGTGGA AAGTCCCCAG
GCTCCCCAGC AGGCAGAAGT ATGCAAAGCA TGCATCTCAA TTAGTCAGCA
5701 ACCAGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG
CATGCATCTC AATTAGTCAG CAACCATAGT CCCGCCCCTA ACTCCGCCCA
5801 TCCCGCCCCC AACTCCGCCC AGTTCCGCCC ATTCTCCGCC CCATGGCTGA
CTAATTTTTT TTATTTATGC AGAGGCCGAG GCCGCCTCGG CCTCTGAGCT
5901 ATTCCAGAAG TAGTGAGGAG GCTTTTTTGG AGGCCTAGGC TTTTGCAAAA
AGCTT

```

pTZ-UTF1

1 TTAATACGAC TCACTATAGG GAATTCAGCG CCAGGACCGA CCCCTTAGGA
 CCCGGCGGGC GGGGCGGGGC GCGCGGGGCG GGGCGGGGAT GGTGGGAGGG
 101 TCGCGTAGAC CCGGCTGGGA CGGAGGGACG GCGGCGCGGG ATGCTCAGAA
 CGCGTGTGGG CGGCCCAGGC GCGCTCTGGC CCTCTCCCCA TCCTCGCGCG
 201 CCGCGCCCCA GCCCCGGGAT GCTGCTCCGG CCCCAGAGGC CGCCCCCGCT
 CGCGCCCCCC GCGCCGCCCT CGCCCGCCAG CCCCAGACCC GAGCCGCGGA
 301 CACCCGGAGA CGCCCCGGGG ACCCCGCCCC GGAGGCCCGC CTCGCCCAGC
 GCGCTGGGGG AACTCGGGTT GCCGGTGTCC CCGGGCTCGG CGCAGCGCAC
 401 GCCCTGGAGC GCCCAGGAGA CGGAGCTGCT GCTGGGGACG CTGCTGCAAC
 CGGCCGTGTG GCGCGCGCTG CTCCTGGACC GCCGCCAGGC CCTGCCACCC
 501 TACCGCCGCG TGTCGGCCGC GCTGGCCCAG CAGCAGGTGC GCCGCACCCC
 CGCGCAGTGC CGCCGCCGCT ACAAGTTCCT TAAAGACAAG TTTCGCGAGG
 601 CGCACGGCCA GCCGCCCAGG CCCTTCGACG AGCAGATCCG GAAGCTCATG
 GGGCTGCTGG GCGACAACGG GCGCAAACGG CCTCGCCGCC GCTCCCCGGG
 701 GTCCGGGCGC CCCCAGCGCG CCCGCCGCC GGTCCCCAAC GCGCACGCGC
 CGGCTCCCAG CGAACCAGGT AGGCGGGGGA CTGGGGGGCC AGGTGGGGCC
 801 GAGGACTGCG GAGGAACCGG GGGCTGCCGA GGGCTGCGGG GGAGGGGCGG
 GCGCGCCCCG CGAAAATCCG CCCGACCGCC TGCTCCGCGT TCCCAGACGC
 901 CACCCCGCTG CCCACCGCCC GCGACCGCGA CGCGGACCCC ACCTGGACGC
 TCCGCTTCAG CCCGTCCCCA CCGAAGTCTG CGGACGCCTC CCCCAGCCCC
 1001 GGCTCCCCGC CAGCTCCCGC CCCGACCGCC CTCGCCACCT GCATCCCCGA
 GGACCGCGCG CCCGTCCGCG GCCCCGGGTC CCCGCCGCCA CCCCAGGCCC
 1101 GCGAAGACCC CGACTCGCCG CCCGGCCGCC CCGAGGACTG CGCGCCCCCT
 CCGGCCGCGC CCCCCTCGCT GAACACCGCC CTGCTGCAGA CCCTGGGGCA
 1201 CCTGGGCGAC ATCGCGAACA TCCTGGGCCC GCTGCGCGAC CAGCTGCTGA
 CCTTGAACCA GCACGTGGAG CAGCTGCGCG GCGCCTTCGA CCAGACAGTG
 1301 TCCCTGGCCG TGGGCTTCAT TCTGGGCAGC GCGGCCGCCG AGCGAGGGGT
 CCTCAGGGAC CCGTGCCAGT GAGTCCCGGC TGCGGCCAGT CTCCCCTCTC
 1401 CAGGCCGAGG CCCCTCCGCC CTGACCCCTC CTGCTGCCTT CCCACCCCCG
 TTCTTGGGTA TGTTCATAA AAGGATTGTT TTCTAGAGT CCGGGCTGTG
 1501 TAGGAGTGCG GGTCTGGAGA CGGGGAGGGA GTGGCGCCCT GCAGTGCTGG
 GGCTGCAAAG GTCAGGGCAG GGTGCGACG CTCTCATCAC AGAGAGGGGC

1601	TCCCACTCGC	TCCTAAGGTG	TAGACAGAGC	TCCGTGGGGC	GCACCCAAAT
	GTAGGCCGAG	GAGGCCAGGC	TCGAGGGTGG	TGGGCGTGGT	GCCTGCTGCC
1701	GGGTCTTGGT	CACACACTCG	GTTCTCAAGG	CCACACCAGA	CTCTCCCTGC
	CCCTCCCGGG	GAGCCCCACT	GAGTAAACG	AAGTGAACGG	GTGGGCTCCG
1801	CTGGGGGTCC	GGAGAGAACC	TGAGGCTGAC	CCCACCTCTT	CACTCAATGC
	AGGGGTAGGG	GTTAGAGTGG	TCGCCTCCAC	ACTGCCCAGC	AGAGAACCCC
1901	CTTTGACCCA	GGAGTAGCAT	GGGGGGTTCC	GAAGGTCCCTC	CCATGGCCTC
	GGGCAGCCTT	GGACCACCCG	GCTGAACTCT	CCCAGGTCCT	GGGAAACAAA
2001	AGGGGAAGAG	AAGCCACAGA	GCTCCAGGGA	GGATGCTGGT	TTTGTCTCCA
	GGGAGGTGGC	TCTGGTCTGG	GTGGGCCTAG	GGAAGCGCTG	AACGTCTGGG
2101	GTGTGATCCT	TGGGCAGAGG	CTGAGGGCAG	GGTCCTTGCA	ATTTTGAGAA
	AACCGAAGCA	AACCTGCCCA	GGAAGTATTT	TCCAAGTAGC	AATAGAGGGC
2201	CCCACCCTGG	ACCCCTGAGT	CAGAAGAGGG	ATTGGCTGGA	GCCGGAGATA
	CAAATGTAGC	CTGGAGGAGT	TCCGCTAGCA	TAACAATGCC	TGCAAAATTAG
2301	TGGGGCCATC	CAGAAAGACA	AGTGGTTTGC	CCAGGGCCAT	GCAGCTAAAT
	GGGGCTGGGG	GGGTCCCCGC	TGGCCCAGCA	CCCAGCAGTC	CTGAGGGGTT
2401	CCCTGTGGGT	GGTACGGTTG	CAGCCCCCTC	TGCACATGCT	TTTCTGGGGA
	TCCAGAAGAA	TAGTTCCCCG	TAGTGCTAAA	GGCTGCTCCT	GGGGAAATGA
2501	ACGGCTTGCC	CAGGTCTGAG	GGGGAGGCCT	GTGAGACCAC	GCCTGCAAGA
	ACTGACTCCA	AGCCTCACCT	CATTTAAGAA	ACTGAGGCGC	CGGACGCTGT
2601	GGCTCACGCC	TGTAATCCTA	ACACTTTGGG	AGGCTGAGGT	GGGTGGATTG
	CCTGAGCTCG	CTCAGGAGTT	CGAGACCAGC	CTGGGCAACA	TGGTGAAACC
2701	CCATCTCTAC	TAAAAATACA	AAAATTAGCC	GGGTGTGGTG	GCGGGTGCCT
	GTAGTCCCAG	TTACTCGGGA	GGCTGAGGCA	GGAGAAGTGC	TGAACCCTCT
2801	AGAGTCGACC	TGCAGGCATG	CAAGCTTGGC	ACTGGCCGTC	GTTTTACAAC
	GTCGTGACTG	GGAAAACCCT	GGCGTTACCC	AACTTAATCG	CCTTGCAGCA
2901	CATCCCCCTT	TCGCCAGCTG	GCGTAATAGC	GAAGAGGCCC	GCACCGATCG
	CCCTTCCCAA	CAGTTGCGCA	GCCTGAATGG	CGAATGGGAA	ATTGTAAACG
3001	TTAATATTTT	GTTAATATTT	TGTTAAAAAT	CGCGTTAAAT	TTTTGTTAAA
	TCAGCTCATT	TTTTAACCBA	TAGGCCGAAA	TCGGCAAAAT	CCCTTATAAA
3101	TCAAAGAAT	AGACCGAGAT	AGGGTTGAGT	GTTGTTCCAG	TTTGGAACAA
	GAGTCCACTA	TTAAAGAACG	TGGACTCCAA	CGTCAAAGGG	CGAAAAACCG
3201	TCTATCAGGG	CGATGGCCCA	CTACGTGAAC	CATCACCTTA	ATCAAGTTTT
	TTGGGGTCGA	GGTGCCGTAA	AGCACTAAAT	CGGAACCCTA	AAGGGATGCC
3301	CCGATTTAGA	GCTTGACGGG	GAAAAGCCGG	CGAACGTGGC	GAGAAAGGAA

3401 GGGAAGAAAG CGAAAGGAGC GGGCGCTAGG GCGCTGGCAA GTGTAGCGGT
 CACGCTGCGC GTAACCACCA CACCCGCCGC GCTTAATGCG CCGCTACAGG
 GCGCGTCAGG TGGCACTTTT CGGGGAAATG TGCGCGGAAC CCTATTGT
 3501 TTATTTTCT AAATACATTC AAATATGTAT CCGCTCATGA GACAATAACC
 CTGATAAATG CTTCAATAAT ATTGAAAAAG GAAGAGTATG AGTATTCAAC
 3601 ATTTCCGTGT CGCCCTTATT CCTTTTTTTG CGGCATTTTG CCTTCCTGTT
 TTTGCTCACC CAGAAACGCT GGTGAAAGTA AAAGATGCTG AAGATCAGTT
 3701 GGGTGCACGA GTGGGTTACA TCGAACTGGA TCTCAACAGC GGTAAGATCC
 TTGAGAGTTT TCGCCCCGAA GAACGTTTTT CAATGATGAG CACTTTTAA
 3801 GTTCTGCTAT GTGGCGCGGT ATTATCCCGT ATTGACGCCG GGCAAGAGCA
 ACTCGGTCGC CGCATACACT ATTCTCAGAA TGACTTGGTT GAGTACTCAC
 3901 CAGTCACAGA AAAGCATCTT ACGGATGGCA TGACAGTAAG AGAATTATGC
 AGTGCTGCCA TAACCATGAG TGATAACACT GCGGCCAACT TACTTCTGAC
 4001 AACGATCGGA GGACCGAAGG AGCTAACCGC TTTT'TTGCAC AACATGGGGG
 ATCATGTAAC TCGCCTTGAT CGTTGGGAAC CGGAGCTGAA TGAAGCCATA
 4101 CCAAACGACG AGCGTGACAC CACGATGCCT GTAGCAATGG CAACAACGTT
 GCGCAAAC TA TAACTGGCG AACTACTTAC TCTAGCTTCC CGGCAACAAT
 4201 TAATAGACTG GATGGAGGCG GATAAAGTTG CAGGACCACT TCTGCGCTCG
 GCCCTTCCGG CTGGCTGGTT TATTGCTGAT AAATCTGGAG CCGGTGAGCG
 4301 TGGGTCTCGC GGTATCATTG CAGCACTGGG GCCAGATGGT AAGCCCTCCC
 GTATCGTAGT TATCTACACG ACGGGGAGTC AGGCAACTAT GGATGAACGA
 4401 AATAGACAGA TCGCTGAGAT AGGTGCCTCA CTGATTAAGC ATTGGTAACT
 GTCAGACCAA GTTTACTCAT ATATACTTTA GATTGATTTA AAAC'TTCATT
 4501 TTTAATTTAA AAGGATCTAG GTGAAGATCC TTTT'TGATAA TCTCATGACC
 AAAATCCCTT AACGTGAGTT TTCGTTCCAC TGAGCGTCAG ACCCCGTAGA
 4601 AAAGATCAAA GGATCTTCTT GAGATCCTTT TTTTCTGCGC GTAATCTGCT
 GCTTGCAAAC AAAAAAACCA CCGCTACCAG CCGTGGTTTTG TTTGCCGGAT
 4701 CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA GCAGAGCGCA
 GATACCAAAT ACTGTCCTTC TAGTGTAGCC GTAGTTAGGC CACCACTTCA
 4801 AGAACTCTGT AGCACC GCCT ACATACCTCG CTCTGCTAAT CCTGTTACCA
 GTGGCTGCTG CCAGTGGCGA TAAGTCGTGT CTTACCGGGT TGGACTCAAG
 4901 ACGATAGTTA CCGGATAAGG CGGAGCGGTC GGGCTGAACG GGGGGTTCGT
 GCACACAGCC CAGCTTGGAG CGAACGACCT ACACCGAACT GAGATACCTA
 5001 CAGCGTGAGC TATGAGAAAG CGCCACGCTT CCCGAAGGGA GAAAGGCGGA
 CAGGTATCCG GTAAGCGGCA GGGTCGGAAC AGGAGAGCGC AGAGGGAGCT

5101 TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCGGG TTTCGCCACC
 TCTGACTTGA GCGTCGATTT TTGTGATGCT CGTCAGGGGG GCGGAGCCTA
5201 TGGAAAAACG CCAGCAACGC GGCCTTTTTTA CGGTTCCCTGG CCTTTTGCTG
 GCCTTTTGCT CACATGTTCT TTCCTGCGTT ATCCCCTGAT TCTGTGGATA
5301 ACCGTATTAC CGCCTTTGAG TGAGCTGATA CCGCTCGCCG CAGCCGAACG
 ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC GCCCAATACG
5401 CAAACCGCCT CTCCCCGCGC GTTGGCCGAT TCATTAATGC AGCTGGCACG
 ACAGGTTTCC CGACTGGAAA GCGGGCAGTG AGCGCAACGC AATTAATGTG
5501 AGTTAGCTCA CTCATTAGGC ACCCCAGGCT TTACACTTTA TGCTTCCGGC
 TCGTATGTTG TGTGGAATTG TGAGCGGATA ACAACAATTT CACACAGGAA
5601 ACAGCTATGA CCATGATTAC GAAT

pTZ-UTF1-EGFP

```

1      TTAATACGAC TCACTATAGG GAATTCAGCG CCAGGACCGA CCCCTTAGGA
      CCCGGCGGGC GGGGCGGGGC GCGCGGGGCG GGGCGGGGAT GGTGGGAGGG
101    TCGCGTAGAC CCGGCTGGGA CGGAGGGACG GCGGCGCGGG ATGCTCAGAA
      CGCGTGTGGG CGGCCCCGGG GCGGTCTGGC CCTCTCCCCA TCCTCGCGCG
201    CCGCGCCCCA GCCCCGGGGT ACCCCATGGT GAGCAAGGGC GAGGAGCTGT
      TCACCGGGGT GGTGCCCATC CTGGTCGAGC TGGACGGCGA CGTAAACGGC
301    CACAAGTTCA GCGTGTCCGG CGAGGGCGAG GGCGATGCCA CCTACGGCAA
      GCTGACCCTG AAGTTCATCT GCACCACCGG CAAGCTGCCC GTGCCCTGGC
401    CCACCCTCGT GACCACCCTG ACCTACGGCG TGCAGTGCTT CAGCCGCTAC
      CCCGACCACA TGAAGCAGCA CGACTTCTTC AAGTCCGCCA TGCCCCAAGG
501    CTACGTCCAG GAGCGCACCA TCTTCTTCAA GGACGACGGC AACTACAAGA
      CCCGCGCCGA GGTGAAGTTC GAGGGCGACA CCCTGGTGAA CCGCATCGAG
601    CTGAAGGGCA TCGACTTCAA GGAGGACGGC AACATCCTGG GGCACAAGCT
      GGAGTACAAC TACAACAGCC ACAACGTCTA TATCATGGCC GACAAGCAGA
701    AGAACGGCAT CAAGGTGAAC TTCAAGATCC GCCACAACAT CGAGGACGGC
      AGCGTGCAGC TCGCCGACCA CTACCAGCAG AACACCCCCA TCGGCGACGG
801    CCCCCTGCTG CTGCCCAGCA ACCACTACCT GAGCACCCAG TCCGCCCTGA
      GCAAAGACCC CAACGAGAAG CGCGATCACA TGGTCCTGCT GGAGTTCGTG
901    ACCGCCGCCG GGATCACTCT CGGCTTGGAC GAGCTGTACA AGTGAGTCCG
      GCTGCGGCCA GTCTCCCCTC TCCAGGCCGA GGCCCCTCCG CCCTGACCCC
1001   TCCTGCTGCC TTCCCACCTT CGTTCTTGGG TACGTTCAAT AAAAGGATTG
      TTTTCCTAGA GTCCGGGCTG TGTAGGAGTG CGGGTCTGGA GACGGGGAGG
1101   GAGTGGCGCC CTGCAGTGCT GGGGCTGCAA AGGTCAGGGC AGGGTTGCGA
      CGCTCTCATC ACAGAGAGGG GCTCCCCTC GCTCCTAAGG TGTAGACAGA
1201   GCTCCGTGGG GCGCACCCAA ATGTAGGCCG AGGAGGCCAG GCTCGAGGGT
      GGTGGGCGTG GTGCCTGCTG CCGGGTCCTG GTCACACACT CGGTTCTCAA
1301   GGCCACACCA GACTCTCCCT GCCCTCCCG GGGAGCCCCA CTGAGTAAAA
      CGAAGTGAAC GGGTGGGCTC CGCTGGGGGT CCGGAGAGAA CCTGAGGCTG
1401   ACCCCACCTC TTTACTCAAT GCAGGGGTAG GGGTTAGAGT GGTCGCCTCC
      AACTGCCCCA GCAGAGAACC CCCTTTGACC CAGGAGTAGC ATGGGGGGTT
1501   CCGAAGGTCC TCCCATGGCC TCGGGCAGCC TTGGACCACC CGGCTGAACT
      CTCCCAGGTC CTGGGAAACA AAAGGGGAAG AGAAGCCACA GAGCTCCAGG

```

```

1601   GAGGATGCTG GTTTTGTCTC CAGGGAGGTG GCTCTGGTCT GGGTGGGCCT
      AGGGAAGCGC TGAACGTCTG GGGTGTGATC CTTGGGCAGA GGCTGAGGGC
1701   AGGGTCCTTG CAATTTTGAG AAAACCGAAG CAAACCTGCC CAGGAAGTAT
      TTTCCAAGTA GCAATAGAGG GCCCCACCCT GGACCCCTGA GTCAGAAGAG
1801   GGATTGGCTG GAGCCGGAGA TACAAATGTA GCCTGGAGGA GTTCCGCTAG
      CATAACAATG CCTGCAAATT AGTGGGCCCA TCCAGAAAGA CAAGTGGTTT
1901   GCCCAGGGCC ATGCAGCTAA ATGGGGCTGG GGGGGTCCCC GCTGGCCCAG
      CACCCAGCAG TCCTGAGGGG TTCCCTGTGG GTGGTACGGT TGCAGCCCCT
2001   CCTGCACATG CTTTTCTGGG GATCCAGAAG AATAGTTCCC GCTAGTGCTA
      AAGGCTGCTC CTGGGGAAAT GAACGGCTTG CCCAGGTCTG AGGGGGAGGC
2101   CTGTGAGACC ACGCCTGCAA GAACTGACTC CAAGCCTCAC CTCATTTAAG
      AAACTGAGGC GCCGGACGCT GTGGCTCACG CCTGTAATCC TAACACTTTG
2201   GGAGGCTGAG GTGGGTGGAT TGCCTGAGCT CGCTCAGGAG TTCGAGACCA
      GCCTGGGCAA CATGGTGAAA CCCCATCTCT ACTAAAAATA CAAAAATTAG
2301   CCGGGTGTGG TGGCGGGTGC CTGTAGTCCC AGTTACTCGG GAGGCTGAGG
      CAGGAGAAGT GCTGAACCCT CTAGAGTCGA CCTGCAGGCA TGCAAGCTTG
2401   GCACTGGCCG TCGTTTTACA ACGTCGTGAC TGGGAAAACC CTGGCGTTAC
      CCAACTTAAT CGCCTTGCA GACATCCCCC TTTCGCCAGC TGGCGTAATA
2501   GCGAAGAGGC CCGCACCGAT CGCCCTTCCC AACAGTTGCG CAGCCTGAAT
      GGCGAATGGG AAATTGTAAA CGTTAATATT TTGTTAATAT TTTGTTAAAA
2601   TTCGCGTTAA ATTTTGTGTA AATCAGCTCA TTTTSTAACC AATAGGCCGA
      AATCGGCAAA ATCCCTTATA AATCAAAAGA ATAGACCGAG ATAGGGTTGA
2701   GTGTTGTTCC AGTTTGGAAC AAGAGTCCAC TATTAAAGAA CGTGGACTCC
      AACGTCAAAG GGCGAAAAAC CGTCTATCAG GGCGATGGCC CACTACGTGA
2801   ACCATCACCC TAATCAAGTT TTTTGGGGTC GAGGTGCCGT AAAGCACTAA
      ATCGGAACCC TAAAGGGATG CCCCATTTA GAGCTTGACG GGGAAAAGCC
2901   GGCGAACGTG GCGAGAAAGG AAGGGAAGAA AGCGAAAGGA GCGGGCGCTA
      GGGCGCTGGC AAGTGTAGCG GTCACGCTGC GCGTAACCAC CACACCCGCC
3001   GCGCTTAATG CGCCGCTACA GGGCGCGTCA GGTGGCACTT TTCGGGGAAA
      TGTGCGCGGA ACCCTATTT GTTTATTTTT CTAAATACAT TCAAATATGT
3101   ATCCGCTCAT GAGACAATAA CCCTGATAAA TGCTTCAATA ATATTGAAAA
      AGGAAGAGTA TGAGTATTCA ACATTTCCGT GTCGCCCTTA TTCCCTTTTT
3201   TGCGGCATTT TGCCTTCCTG TTTTGTCTCA CCCAGAAACG CTGGTGAAAG
      TAAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA CATCGAACTG
3301   GATCTCAACA GCGGTAAGAT CCTTGAGAGT TTTCGCCCCG AAGAACGTTT

```

TCCAATGATG AGCACTTTTA AAGTTC TGCT ATGTGGCGCG GTATTATCCC
 3401 GTATTGACGC CGGGCAAGAG CAACTCGGTC GCCGCATACA CTATTCTCAG
 AATGACTTGG TTGAGTACTC ACCAGTCACA GAAAAGCATC TTACGGATGG
 3501 CATGACAGTA AGAGAATTAT GCAGTGCTGC CATAACCATG AGTGATAACA
 CTGCGGCCAA CTTACTTCTG ACAACGATCG GAGGACCGAA GGAGCTAACC
 3601 GCTTTTTTGC ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGGA
 ACCGGAGCTG AATGAAGCCA TACCAAACGA CGAGCGTGAC ACCACGATGC
 3701 CTGTAGCAAT GGCAACAACG TTGCGCAAAC TATTAAGTGG CGAACTACTT
 ACTCTAGCTT CCCGGCAACA ATTAATAGAC TGGATGGAGG CGGATAAAGT
 3801 TGCAGGACCA CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG
 ATAAATCTGG AGCCGGTGAG CGTGGGTCTC GCGGTATCAT TGCAGCACTG
 3901 GGGCCAGATG GTAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG
 TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCCT
 4001 CACTGATTAA GCATTGGTAA CTGTCAGACC AAGTTTACTC ATATATACTT
 TAGATTGATT TAAAACTTCA TTTTTAATTT AAAAGGATCT AGGTGAAGAT
 4101 CCTTTTTGAT AATCTCATGA CCAAATCCC TTAACGTGAG TTTTCGTTC
 ACTGAGCGTC AGACCCCGTA GAAAAGATCA AAGGATCTTC TTGAGATCCT
 4201 TTTTTTCTGC GCGTAATCTG CTGCTTGCAA ACAAAAAAAC CACCGCTACC
 AGCGGTGGTT TGTTTGCCGG ATCAAGAGCT ACCAACTCTT TTTCCGAAGG
 4301 TAACTGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCCT TCTAGTGTAG
 CCGTAGTTAG GCCACCACTT CAAGAACTCT GTAGCACCGC CTACATACCT
 4401 CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC GATAAGTCGT
 GTCTTACCGG GTTGGACTCA AGACGATAGT TACCGGATAA GGCGGAGCGG
 4501 TCGGGCTGAA CGGGGGGTTT GTGCACACAG CCCAGCTTGG AGCGAACGAC
 CTACACCGAA CTGAGATACC TACAGCGTGA GCTATGAGAA AGCGCCACGC
 4601 TTCCCGAAGG GAGAAAGGCG GACAGGTATC CGGTAAGCGG CAGGGTCGGA
 ACAGGAGAGC GCAGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT
 4701 AGTCCTGTCG GGTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG
 CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT
 4801 TACGGTTCCT GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTCCTGCG
 TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG AGTGAGCTGA
 4901 TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG
 AAGCGGAAGA GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG
 5001 ATTCATTAAT GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG
 TGAGCGCAAC GCAATTAATG TGAGTTAGCT CACTCATTAG GCACCCCAGG

5101 CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTGAGCGGA
 TAACAACAAT TTCACACAGG AAACAGCTAT GACCATGATT ACGAAT

pTZ-UTF1-Neo

```

1      TTAATACGAC TCACTATAGG GAATTCAGCG CCAGGACCGA CCCCTTAGGA
      CCCGGCGGGC GGGGCGGGGC GCGCGGGGCG GGGCGGGGAT GGTGGGAGGG
101    TCGCGTAGAC CCGGCTGGGA CGGAGGGACG GCGGCGCGGG ATGCTCAGAA
      CGCGTGTGGG CGGCCCAGGC GCGTCTGGC CCTCTCCCCA TCCTCGCGCG
201    CCGCGCCCCA GCCCCGGGGT ACCCCATGAT TGAACAAGAT GGATTGCACG
      CAGGTTCTCC GGCCGCTTGG GTGGAGAGGC TATTCGGCTA TGA CTGGGCA
301    CAACAGACAA TCGGCTGCTC TGATGCCGCC GTGTTCCGGC TGTCAGCGCA
      GGGGCGCCCC GTTCTTTTTG TCAAGACCGA CCTGTCCGGT GCCCTGAATG
401    AACTGCAAGA CGAGGCAGCG CGGCTATCGT GGCTGGCCAC GACGGGCGTT
      CCTTGCGCAG CTGTGCTCGA CGTTGTCACT GAAGCGGGAA GGGACTGGCT
501    GCTATTGGGC GAAGTGCCGG GGCAGGATCT CCTGTCATCT CACCTTGCTC
      CTGCCGAGAA AGTATCCATC ATGGCTGATG CAATGCGGCG GCTGCATACG
601    CTTGATCCGG CTACCTGCCC ATTCGACCAC CAAGCGAAAC ATCGCATCGA
      GCGAGCACGT ACTCGGATGG AAGCCGGTCT TGTCGATCAG GATGATCTGG
701    ACGAAGAGCA TCAGGGGCTC GCGCCAGCCG AACTGTTTCGC CAGGCTCAAG
      GCGAGCATGC CCGACGGCGA GGATCTCGTC GTGACCCATG GCGATGCCTG
801    CTTGCCGAAT ATCATGGTGG AAAATGGCCG CTTTTCTGGA TTCATCGACT
      GTGGCCGGCT GGGTGTGGCG GACCGCTATC AGGACATAGC GTTGGCTACC
901    CGTGATATTG CTGAAGAGCT TGGCGGCGAA TGGGCTGACC GCTTCCTCGT
      GCTTTACGGT ATCGCCGCTC CCGATTGCGA GCGCATCGCC TTCTATCGCC
1001   TTCTTGACGA GTTCTTCTGA TGTACAAGTG AGTCCGGCTG CGGCCAGTCT
      CCCCTCTCCA GGCCGAGGCC CCTCCGCCCT GACCCCTCCT GCTGCCTTCC
1101   CACCCTCGTT CTTGGGTACG TTCAATAAAA GGATTGTTTT CCTAGAGTCC
      GGGCTGTGTA GGAGTGCGGG TCTGGAGACG GGGAGGGAGT GGCGCCCTGC
1201   AGTGCTGGGG CTGCAAAGGT CAGGGCAGGG TTGCGACGCT CTCATCACAG
      AGAGGGGCTC CCACTCGCTC CTAAGGTGTA GACAGAGCTC CGTGGGGCGC
1301   ACCCAAATGT AGGCCGAGGA GGCCAGGCTC GAGGGTGGTG GGCGTGGTGC
      CTGCTGCCGG GTCTTGGTCA CAACTCGGT TCTCAAGGCC ACACCAGACT
1401   CTCCCTGCCC CTCCCGGGGA GCCCCACTGA GTAAAACGAA GTGAACGGGT
      GGGCTCCGCT GGGGGTCCGG AGAGAACCTG AGGCTGACCC CACCTCTTCA
1501   CTCAATGCAG GGGTAGGGGT TAGAGTGGTC GCCTCCACAC TGCCCAGCAG
      AGAACCCCTT TTGACCCAGG AGTAGCATGG GGGGTTCGGA AGGTCTCTCC

```

```

1601      ATGGCCTCGG GCAGCCTTGG ACCACCCGGC TGAACTCTCC CAGGTCCTGG
          GAAACAAAAG GGGAAGAGAA GCCACAGAGC TCCAGGGAGG ATGCTGGTTT
1701      TGTCTCCAGG GAGGTGGCTC TGGTCTGGGT GGGCCTAGGG AAGCGCTGAA
          CGTCTGGGGT GTGATCCTTG GGCAGAGGCT GAGGGCAGGG TCCTTGCAAT
1801      TTTGAGAAAA CCGAAGCAAA CCTGCCCAGG AAGTATTTTC CAAGTAGCAA
          TAGAGGGCCC CACCCTGGAC CCCTGAGTCA GAAGAGGGAT TGGCTGGAGC
1901      CGGAGATACA AATGTAGCCT GGAGGAGTTC CGCTAGCATA ACAATGCCTG
          CAAATTAGTG GGCCCATCCA GAAAGACAAG TGGTTTGCCC AGGGCCATGC
2001      AGCTAAATGG GGCTGGGGGG GTCCCCGCTG GCCCAGCACC CAGCAGTCCT
          GAGGGGTTC CTTGTGGGTG TACGGTTGCA GCCCCTCCTG CACATGCTTT
2101      TCTGGGGATC CAGAAGAATA GTTCCCCGTA GTGCTAAAGG CTGCTCCTGG
          GGAAATGAAC GGCTTGCCCA GGTCTGAGGG GGAGGCCTGT GAGACCACGC
2201      CTGCAAGAAC TGA CTCCAAG CCTCACCTCA TTTAAGAAAC TGAGGCGCCG
          GACGCTGTGG CTCACGCCTG TAATCCTAAC ACTTTGGGAG GCTGAGGTGG
2301      GTGGATTGCC TGAGCTCGCT CAGGAGTTCG AGACCAGCCT GGGCAACATG
          GTGAAACCCC ATCTCTACTA AAAATACAAA AATTAGCCGG GTGTGGTGGC
2401      GGGTGCCTGT AGTCCCAGTT ACTCGGGAGG CTGAGGCAGG AGAAGTGCTG
          AACCTCTAG AGTCGACCTG CAGGCATGCA AGCTTGGCAC TGGCCGTCGT
2501      TTTACAACGT CGTGA CTGGG AAAACCCTGG CGTTACCCAA CTTAATCGCC
          TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC
2601      ACCGATCGCC CTTCCCAACA GTTGCGCAGC CTGAATGGCG AATGGGAAAT
          TGTAACGTT AATATTTTGT TAATATTTTG TTAAAATTTCG CGTTAAATTT
2701      TTGTTAAATC AGCTCATTTT TTAACCAATA GGCCGAAATC GGCAAATCC
          CTTATAAATC AAAAGAATAG ACCGAGATAG GGTTGAGTGT TGTTCAGTT
2801      TGGAACAAGA GTCCACTATT AAAGAACGTG GACTCCAACG TCAAAGGGCG
          AAAAACCCTC TATCAGGGCG ATGGCCCACT ACGTGAACCA TCACCCTAAT
2901      CAAGTTTTTTT GGGGTCGAGG TGCCGTAAAG CACTAAATCG GAACCCTAAA
          GGGATGCCCC GATTTAGAGC TTGACGGGGA AAAGCCGGCG AACGTGGCGA
3001      GAAAGGAAGG GAAGAAAGCG AAAGGAGCGG GCGCTAGGGC GCTGGCAAGT
          GTAGCGGTCA CGCTGCGCGT AACCACCACA CCCGCCGCGC TTAATGCGCC
3101      GCTACAGGGC GCGTCAGGTG GCACTTTTTCG GGGAAATGTG CGCGGAACCC
          CTATTTGTTT ATTTTCTAA ATACATTCAA ATATGTATCC GCTCATGAGA
3201      CAATAACCCT GATAAATGCT TCAATAATAT TGAAAAAGGA AGAGTATGAG
          TATTCACAT TTCCGTGTCG CCCTTATTCC CTTTTTTCG GCATTTTGCC
3301      TTCCTGTTTT TGCTCACCCA GAAACGCTGG TGAAAGTAAA AGATGCTGAA

```

3401 GATCAGTTGG GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG
 TAAGATCCTT GAGAGTTTTC GCCCCGAAGA ACGTTTTCCTA ATGATGAGCA
 CTTTTTAAAGT TCTGCTATGT GGCGCGGTAT TATCCCGTAT TGACGCCGGG
 3501 CAAGAGCAAC TCGGTCGCCG CATACTAT TCTCAGAATG ACTTGTTGA
 GTACTCACCA GTCACAGAAA AGCATCTTAC GGATGGCATG ACAGTAAGAG
 3601 AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC GGCCAACTTA
 CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT TTTTGCACAA
 3701 CATGGGGGAT CATGTAATC GCCTTGATCG TTGGGAACCG GAGCTGAATG
 AAGCCATACC AAACGACGAG CGTGACACCA CGATGCCTGT AGCAATGGCA
 3801 ACAACGTTGC GCAAACATAT AACTGGCGAA CTACTTACTC TAGCTTCCCG
 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC
 3901 TGCGCTCGGC CCTTCCGGCT GGCTGGTTTA TTGCTGATAA ATCTGGAGCC
 GGTGAGCGTG GGTCTCGCGG TATCATTTGCA GCACTGGGGC CAGATGGTAA
 4001 GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG
 ATGAACGAAA TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT
 4101 TGGTAACTGT CAGACCAAGT TTACTCATAT ATACTTTAGA TTGATTTAAA
 ACTTCATTTT TAATTTAAAA GGATCTAGGT GAAGATCCTT TTTGATAATC
 4201 TCATGACCAA AATCCC'TTAA CGTGAGTTTT CGTTCCTACTG AGCGTCAGAC
 CCCGTAGAAA AGATCAAAGG ATCTTCTTGA GATCCTTTTTT TTCTGCGCGT
 4301 AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG GTGGTTTGT
 TGCCGGATCA AGAGCTACCA ACTCTTTTTTC CGAAGGTAAC TGGCTTCAGC
 4401 AGAGCGCAGA TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA
 CCACTTCAAG AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC
 4501 TGTTACCAGT GGCTGCTGCC AGTGCGGATA AGTCGTGTCT TACCGGGTTG
 GACTCAAGAC GATAGTTACC GGATAAGGCG GAGCGGTCGG GCTGAACGGG
 4601 GGGTTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC ACCGAACTGA
 GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA
 4701 AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAG
 AGGGAGCTTC CAGGGGGAAA CGCCTGGTAT CTTTATAGTC CTGTGCGGTT
 4801 TCGCCACCTC TGACTTGAGC GTCGATTTTT GTGATGCTCG TCAGGGGGGC
 GGAGCCTATG GAAAAACGCC AGCAACGCGG CCTTTTTTACG GTTCCTGGCC
 4901 TTTTGCTGGC CTTTGTCTCA CATGTTCTTT CCTGCGTTAT CCCCTGATTC
 TGTGGATAAC CGTATTACCG CCTTTGAGTG AGCTGATACC GCTCGCCGCA
 5001 GCCGAACGAC CGAGCGCAGC GAGTCAGTGA GCGAGGAAGC GGAAGAGCGC
 CCAATACGCA AACC GCCTCT CCCC GCGCGT TGGCCGATTC ATTAATGCAG

5101 CTGGCACGAC AGGTTTCCCG ACTGGAAAGC GGGCAGTGAG CGCAACGCAA
TTAATGTGAG TTAGCTCACT CATTAGGCAC CCCAGGCTTT ACACTTTATG
5201 CTTCCGGCTC GTATGTTGTG TGAATTGTG AGCGGATAAC AACAATTTCA
CACAGGAAAC AGCTATGACC ATGATTACGA AT

Appendix II Sequencing of *pTZ-UTF1-EGFP-M1mut*

Figure VIII.1. Sequencing data for *pTZ-UTF1-EGFP-M1mut*. Clone #7 was sequenced and found to have the mutated M1 sequence (GTAGTGGT compared to the original GTCTGGGT).

Appendix III Validation gels for qRT-PCR primers

Figure VIII.2. Validation gel for UTF1 transcript knockdown experiment (HUES8 hESCs). The endpoint products for the siUTF1 qRT-PCR experiment were run on an agarose gel, showing a single band for all primer pairs.

Figure VIII.3. Validation gel for qRT-PCR experiments. The endpoint products for qRT-PCR experiments were run on an agarose gel, showing a single band for all specific primer pairs.

Appendix IV Amplification curves for qRT-PCR results

Figure VIII.4. Raw qRT-PCR data for UTF1 transcript knockdown experiment. HUES8 raw data from SDS Software is displayed, showing downregulation of mainly UTF1 transcripts, and to a lesser extent, that of Nanog.

Figure VIII.5. Raw qRT-PCR data for RA and DMSO treatment comparison. HUES8 raw data from SDS Software showing the downregulation of all three pluripotency markers upon either RA (12 days) or DMSO (7 days) treated cultures. UTF1 downregulation is the most pronounced.

Figure VIII.6. Raw qRT-PCR data for pluripotent marker comparison between HUES8 and ZUN hESCs. Data from SDS Software showing the higher expression of both UTF1 and Sox2 in ZUN hESC cultures compared to parental HUES8. Oct4 and Nanog expression are relatively similar.

Figure VIII.7. Raw qRT-PCR data for modulation of G418 selection pressure experiment in ZUN hESCs. Data from SDS Software showing the downregulation of Oct4, Nanog and UTF1, but not Sox2. Pluripotency marker expression is reinstated upon re-addition of G418 antibiotic. ZUN cell lines were cultured with G418 antibiotic for 90 days (+G418), G418 was removed for 30 (-G418) or 60 (--G418) days, G418 was removed for 30 days and re-introduced for the subsequent 30 days (-+G418).

Appendix V Calculations for qRT-PCR results

Transcript knockdown

HUES8	$\Delta Ct(\text{siNT})$	$\Delta Ct(\text{siUTF1})$	$\Delta\Delta Ct$	% downregulation ($2^{-\Delta\Delta Ct} \times 100$)		
Nanog	6.13	6.52	0.39	76.18	Average %	73.67
Nanog	6.13	6.74	0.61	65.41	Standard Deviation	7.34
Nanog	6.13	6.46	0.33	79.42		
OCT4	3.68	3.88	0.20	87.26	Average %	87.13
OCT4	3.68	3.75	0.07	95.49	Standard Deviation	8.42
OCT4	3.68	4.03	0.35	78.65		
SOX2	8.64	8.51	-0.13	109.68	Average %	106.22
SOX2	8.64	8.42	-0.22	116.74	Standard Deviation	12.62
SOX2	8.64	8.76	0.12	92.23		
UTF1	8.60	11.08	2.48	17.89	Average %	15.16
UTF1	8.60	11.68	3.08	11.81	Standard Deviation	3.09
UTF1	8.60	11.26	2.66	15.79		

* Each set of ΔCt and $\Delta\Delta Ct$ values are an average of three assay replicates.*

Transcript expression after RA/DMSO treatment

HUES8	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{RA 12days})$	$\Delta\Delta Ct$	% of Undifferentiated ($2^{-\Delta\Delta Ct} \times 100$)		
Nanog	5.25	6.72	1.47	36.01	Average %	34.58
Nanog	5.18	6.77	1.59	33.14	Standard Deviation	2.03
OCT4	1.93	3.57	1.64	32.09	Average %	30.70
OCT4	1.72	3.49	1.77	29.32	Standard Deviation	1.95
UTF1	8.77	14.6	5.83	1.76	Average %	1.99
UTF1	9.07	14.56	5.49	2.23	Standard Deviation	0.33

HUES8	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{DMSO 7days})$	$\Delta\Delta Ct$	% of Undifferentiated ($2^{-\Delta\Delta Ct} \times 100$)		
Nanog	5.25	8.135	2.89	13.51	Average %	12.39
Nanog	5.18	8.325	3.15	11.28	Standard Deviation	1.57
OCT4	1.93	5.015	3.09	11.78	Average %	13.40
OCT4	1.72	4.455	2.74	15.02	Standard Deviation	2.29
UTF1	8.77	16.385	7.62	0.51	Average %	0.57
UTF1	9.07	16.375	7.31	0.63	Standard Deviation	0.09

HUES9	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{RA 12days})$	$\Delta\Delta Ct$	% of Undifferentiated ($2^{-\Delta\Delta Ct} \times 100$)		
Nanog	0.59	2.05	1.46	36.35	Average %	34.22
Nanog	0.5	2.14	1.64	32.09	Standard Deviation	3.01
OCT4	5.11	6.3	1.19	43.83	Average %	42.79
OCT4	4.96	6.22	1.26	41.75	Standard Deviation	1.47
UTF1	8.41	12.53	4.12	5.75	Average %	4.84
UTF1	8.2	12.87	4.67	3.93	Standard Deviation	1.29

HUES9	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{DMSO 7days})$	$\Delta\Delta C_t$	% of Undifferentiated ($2^{-\Delta\Delta C_t} \times 100$)		
Nanog	0.59	3.37	2.78	14.56	Average %	13.57
Nanog	0.5	3.49	2.99	12.59	Standard Deviation	1.39
OCT4	5.11	7.49	2.38	19.21	Average %	21.85
OCT4	4.96	6.99	2.03	24.49	Standard Deviation	3.73
UTF1	8.41	15.19	6.78	0.91	Average %	0.82
UTF1	8.2	15.28	7.08	0.74	Standard Deviation	0.12

hES2	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{RA 12days})$	$\Delta\Delta C_t$	% of Undifferentiated ($2^{-\Delta\Delta C_t} \times 100$)		
Nanog	4.07	7.66	3.59	8.31	Average %	9.31
Nanog	4.07	7.35	3.28	10.30	Standard Deviation	1.41
OCT4	0.36	4.19	3.83	7.03	Average %	6.82
OCT4	0.36	4.28	3.92	6.61	Standard Deviation	0.30
UTF1	9.99	15.29	5.30	2.54	Average %	2.87
UTF1	9.99	14.96	4.97	3.20	Standard Deviation	0.46

hES2	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{DMSO 7days})$	$\Delta\Delta C_t$	% of Undifferentiated ($2^{-\Delta\Delta C_t} \times 100$)		
Nanog	4.07	5.955	1.88	27.10	Average %	27.87
Nanog	4.07	5.875	1.80	28.64	Standard Deviation	1.09
OCT4	0.36	2.325	1.96	25.62	Average %	26.07
OCT4	0.36	2.275	1.91	26.52	Standard Deviation	0.64
UTF1	9.99	13.345	3.35	9.79	Average %	8.79
UTF1	9.99	13.675	3.68	7.79	Standard Deviation	1.42

* Each set of ΔC_t and $\Delta\Delta C_t$ values are an average of three assay replicates.*

Spontaneous differentiation – Pluripotency markers

Gelatin Protocol

HUES8	$\Delta\text{Ct}(\text{Undifferentiated})$	$\Delta\text{Ct}(\text{Gelatin 12days})$	$\Delta\Delta\text{Ct}$	% of Undifferentiated ($2^{-\Delta\Delta\text{Ct}} \times 100$)		
Nanog	5.21	5.25	0.04	97.11	Average %	96.57
Nanog	5.21	5.27	0.06	96.03	Standard Deviation	0.76
OCT4	1.95	3.10	1.15	45.03	Average %	45.80
OCT4	1.95	3.05	1.10	46.57	Standard Deviation	1.09
SOX2	8.27	8.18	-0.09	106.44	Average %	122.46
SOX2	8.27	7.80	-0.47	138.47	Standard Deviation	22.65
UTF1	8.36	9.46	1.10	46.60	Average %	42.51
UTF1	8.36	9.73	1.38	38.43	Standard Deviation	5.78

ZUN1	$\Delta\text{Ct}(\text{Undifferentiated})$	$\Delta\text{Ct}(\text{Gelatin 12days})$	$\Delta\Delta\text{Ct}$	% of Undifferentiated ($2^{-\Delta\Delta\text{Ct}} \times 100$)		
Nanog	5.80	5.52	-0.28	121.39	Average %	117.01
Nanog	5.80	5.62	-0.17	112.64	Standard Deviation	6.19
OCT4	1.83	2.30	0.47	71.97	Average %	72.46
OCT4	1.83	2.28	0.45	72.95	Standard Deviation	0.69
SOX2	7.85	7.34	-0.50	141.54	Average %	152.06
SOX2	7.85	7.14	-0.70	162.58	Standard Deviation	14.88
UTF1	7.24	7.51	0.27	83.13	Average %	84.71
UTF1	7.24	7.46	0.21	86.30	Standard Deviation	2.25
UTF1-Neo	8.47	8.04	-0.43	134.74	Average %	124.48
UTF1-Neo	8.47	8.27	-0.19	114.23	Standard Deviation	14.51

ZUN2	$\Delta\text{Ct}(\text{Undifferentiated})$	$\Delta\text{Ct}(\text{Gelatin 12days})$	$\Delta\Delta\text{Ct}$	% of Undifferentiated ($2^{-\Delta\Delta\text{Ct}} \times 100$)		
Nanog	5.77	5.77	0.00	99.96	Average %	92.20
Nanog	5.77	6.01	0.24	84.44	Standard Deviation	10.97
OCT4	2.01	2.24	0.24	84.85	Average %	78.85
OCT4	2.01	2.46	0.46	72.85	Standard Deviation	8.48
SOX2	7.26	7.23	-0.04	102.46	Average %	117.14
SOX2	7.26	6.86	-0.40	131.82	Standard Deviation	20.76
UTF1	7.18	7.60	0.42	74.55	Average %	83.09
UTF1	7.18	7.31	0.13	91.62	Standard Deviation	12.07
UTF1-Neo	8.67	8.38	-0.29	122.46	Average %	116.89
UTF1-Neo	8.67	8.52	-0.15	111.33	Standard Deviation	7.87

* Each set of ΔCt and $\Delta\Delta\text{Ct}$ values are an average of three assay replicates.*

Matrigel Protocol

HUES8	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{Matrigel 12days})$	$\Delta\Delta C_t$	% of Undifferentiated ($2^{-\Delta\Delta C_t} \times 100$)		
Nanog	5.21	7.46	2.24	21.10	Average %	18.53
Nanog	5.21	7.86	2.65	15.95	Standard Deviation	3.64
OCT4	1.95	4.39	2.44	18.40	Average %	19.23
OCT4	1.95	4.27	2.32	20.07	Standard Deviation	1.18
SOX2	8.27	10.25	1.97	25.45	Average %	26.32
SOX2	8.27	10.15	1.88	27.20	Standard Deviation	1.24
UTF1	8.36	10.22	1.86	27.48	Average %	20.56
UTF1	8.36	11.23	2.87	13.63	Standard Deviation	9.79

ZUN1	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{Matrigel 12days})$	$\Delta\Delta C_t$	% of Undifferentiated ($2^{-\Delta\Delta C_t} \times 100$)		
Nanog	5.80	5.62	-0.17	112.78	Average %	108.57
Nanog	5.80	5.73	-0.06	104.36	Standard Deviation	5.95
OCT4	1.83	2.51	0.68	62.42	Average %	56.75
OCT4	1.83	2.80	0.97	51.08	Standard Deviation	8.02
SOX2	7.85	7.61	-0.23	117.38	Average %	113.83
SOX2	7.85	7.70	-0.14	110.27	Standard Deviation	5.03
UTF1	7.24	7.47	0.23	85.23	Average %	92.54
UTF1	7.24	7.25	0.00	99.85	Standard Deviation	10.34
UTF1-Neo	8.47	8.47	0.00	99.90	Average %	96.92
UTF1-Neo	8.47	8.56	0.09	93.94	Standard Deviation	4.21

ZUN2	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{Matrigel 12days})$	$\Delta\Delta C_t$	% of Undifferentiated ($2^{-\Delta\Delta C_t} \times 100$)		
Nanog	5.77	6.12	0.35	78.26	Average %	77.62
Nanog	5.77	6.15	0.38	76.98	Standard Deviation	0.91
OCT4	2.01	2.88	0.88	54.43	Average %	60.09
OCT4	2.01	2.61	0.60	65.75	Standard Deviation	8.00
SOX2	7.26	7.87	0.61	65.47	Average %	76.47
SOX2	7.26	7.45	0.19	87.46	Standard Deviation	15.55
UTF1	7.18	7.79	0.61	65.72	Average %	71.81
UTF1	7.18	7.54	0.36	77.89	Standard Deviation	8.60
UTF1-Neo	9.07	9.33	0.25	83.96	Average %	92.25
UTF1-Neo	9.07	9.07	-0.01	100.54	Standard Deviation	11.72

* Each set of ΔC_t and $\Delta\Delta C_t$ values are an average of three assay replicates.*

Directed differentiation – Pluripotency markers**RA Protocol**

HUES8	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{RA 12days})$	$\Delta\Delta Ct$	% of Undifferentiated ($2^{-\Delta\Delta Ct} \times 100$)		
Nanog	5.21	6.76	1.54	34.27	Average %	42.80
Nanog	5.80	6.76	0.96	51.33	Standard Deviation	12.06
OCT4	1.95	3.47	1.52	34.76	Average %	34.82
OCT4	2.41	3.93	1.52	34.87	Standard Deviation	0.08
SOX2	8.27	10.27	1.99	25.11	Average %	24.20
SOX2	7.74	9.84	2.10	23.28	Standard Deviation	1.29
UTF1	8.93	14.58	5.65	1.99	Average %	3.42
UTF1	8.06	12.42	4.36	4.86	Standard Deviation	2.03

ZUN1	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{RA 12days})$	$\Delta\Delta Ct$	% of Undifferentiated ($2^{-\Delta\Delta Ct} \times 100$)		
Nanog	6.28	6.34	0.07	95.46	Average %	94.24
Nanog	6.87	6.98	0.10	93.01	Standard Deviation	1.73
OCT4	1.83	2.11	0.28	82.62	Average %	77.35
OCT4	2.93	3.40	0.47	72.08	Standard Deviation	7.46
SOX2	7.85	8.80	0.96	51.46	Average %	48.61
SOX2	7.99	9.11	1.13	45.76	Standard Deviation	4.03
UTF1	8.53	9.71	1.18	44.14	Average %	40.79
UTF1	8.33	9.75	1.42	37.43	Standard Deviation	4.74
UTF1-Neo	8.47	10.46	2.00	25.01	Average %	32.92
UTF1-Neo	9.02	10.31	1.29	40.82	Standard Deviation	11.18

ZUN2	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{RA 12days})$	$\Delta\Delta Ct$	% of Undifferentiated ($2^{-\Delta\Delta Ct} \times 100$)		
Nanog	5.77	5.58	-0.19	114.03	Average %	108.53
Nanog	6.91	6.86	-0.04	103.02	Standard Deviation	7.78
OCT4	2.01	2.06	0.05	96.52	Average %	87.78
OCT4	2.97	3.31	0.34	79.05	Standard Deviation	12.35
SOX2	7.26	8.25	0.98	50.53	Average %	46.15
SOX2	8.67	9.93	1.26	41.78	Standard Deviation	6.19
UTF1	9.19	10.82	1.63	32.35	Average %	25.72
UTF1	7.18	9.57	2.39	19.09	Standard Deviation	9.38
UTF1-Neo	8.67	9.49	0.82	56.74	Average %	47.98
UTF1-Neo	9.07	10.42	1.35	39.21	Standard Deviation	12.40

* Each set of ΔCt and $\Delta\Delta Ct$ values are an average of three assay replicates.*

DMSO Protocol

HUES8	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{DMSO 7days})$	$\Delta\Delta C_t$	% of Undifferentiated ($2^{-\Delta\Delta C_t} \times 100$)	Average (%)
Nanog	5.212091	7.55	2.34	19.77968256	20.46
Nanog	5.212091	7.43	2.22	21.4952706	Standard Deviation
Nanog	5.212091	7.57	2.36	19.50736982	0.96
Nanog	5.212091	7.46	2.25	21.05290433	
OCT4	1.94845	4.79	2.84	13.95109066	13.30
OCT4	1.94845	4.6	2.65	15.91489802	Standard Deviation
OCT4	1.94845	5.12	3.17	11.09860161	2.10
OCT4	1.94845	4.98	3.03	12.22960589	
SOX2	8.274448	9.98	1.71	30.66039777	34.04
SOX2	8.274448	9.65	1.38	38.54053007	Standard Deviation
SOX2	8.274448	9.72	1.45	36.71517341	4.21
SOX2	8.274448	10	1.73	30.23828701	
UTF1	7.811675	15.12	7.31	0.630921107	0.70
UTF1	7.811675	14.59	6.78	0.911005406	Standard Deviation
UTF1	7.811675	15.24	7.43	0.580565811	0.15
UTF1	7.811675	15.05	7.24	0.662288412	

ZUN1	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{DMSO 7days})$	$\Delta\Delta C_t$	% of Undifferentiated ($2^{-\Delta\Delta C_t} \times 100$)	Average (%)
Nanog	6.277209	7.75	1.47	36.02846677	34.22
Nanog	6.277209	7.89	1.61	32.69652373	Standard Deviation
Nanog	6.277209	7.815	1.54	34.44124411	1.40
Nanog	6.277209	7.845	1.57	33.73245356	
OCT4	2.751119	4	1.25	42.07744664	44.17
OCT4	2.751119	3.8	1.05	48.33429374	Standard Deviation
OCT4	2.751119	3.915	1.16	44.63103184	3.07
OCT4	2.751119	4.015	1.26	41.64222515	
SOX2	8.65319	8.79	0.14	90.95279172	87.65
SOX2	8.65319	8.85	0.20	87.24774965	Standard Deviation
SOX2	8.65319	8.955	0.30	81.12338886	4.72
SOX2	8.65319	8.785	0.13	91.26855694	
UTF1	8.581698	12.37	3.79	7.237812501	6.88
UTF1	8.581698	12.36	3.78	7.288155468	Standard Deviation
UTF1	8.581698	12.745	4.16	5.581116396	0.87
UTF1	8.581698	12.335	3.75	7.415550182	
UTF1-Neo	8.465217	11.99	3.52	8.688293185	8.10
UTF1-Neo	8.465217	12.34	3.87	6.816696671	Standard Deviation
UTF1-Neo	8.465217	12.045	3.58	8.36330274	0.87
UTF1-Neo	8.465217	12.015	3.55	8.539033508	

ZUN2	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{DMSO 7days})$	$\Delta \Delta Ct$	% of Undifferentiated ($2^{-\Delta \Delta Ct} \times 100$)	Average (%)
Nanog	6.131444	7.405	1.27	41.36389184	42.40
Nanog	6.131444	7.415	1.28	41.07817057	Standard Deviation
Nanog	6.131444	7.295	1.16	44.64108004	1.62
Nanog	6.131444	7.365	1.23	42.52678912	
OCT4	2.636464	3.795	1.16	44.79668183	45.09
OCT4	2.636464	3.905	1.27	41.50807059	Standard Deviation
OCT4	2.636464	3.695	1.06	48.01189479	2.73
OCT4	2.636464	3.755	1.12	46.05608798	
SOX2	7.29	7.505	0.22	86.15461597	95.39
SOX2	7.29	7.135	-0.16	111.3421618	Standard Deviation
SOX2	7.29	7.385	0.09	93.62722474	11.07
SOX2	7.29	7.435	0.15	90.43793776	
UTF1	9.19	12.335	3.15	11.30474222	11.39
UTF1	9.19	12.305	3.12	11.54227888	Standard Deviation
UTF1	9.19	12.165	2.98	12.71849615	1.12
UTF1	9.19	12.515	3.33	9.978729829	
UTF1-Neo	8.670849	11.945	3.27	10.33671148	9.12
UTF1-Neo	8.670849	12.195	3.52	8.692103627	Standard Deviation
UTF1-Neo	8.670849	12.235	3.56	8.454417596	0.84
UTF1-Neo	8.670849	12.145	3.47	8.998629999	

Directed differentiation – Differentiation markers

RA Protocol

HUES8	Transcript	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{RA 12days})$	$\Delta \Delta Ct$	Fold Change ($2^{-\Delta \Delta Ct}$)	Average Fold Change
Ectoderm	SOX9	9.53	4.25	-5.28	37.73	33.49
Ectoderm	SOX9	9.53	4.37	-5.16	34.64	Standard Deviation
Ectoderm	SOX9	15.90	10.90	-5.00	30.93	2.85
Ectoderm	SOX9	15.90	10.90	-5.00	30.93	
Ectoderm	SOX9	15.90	10.80	-5.10	33.22	
Ectoderm	Nestin	11.78	9.01	-2.77	5.80	5.13
Ectoderm	Nestin	11.78	8.87	-2.91	6.50	Standard Deviation
Ectoderm	Nestin	18.50	16.21	-2.29	3.89	1.03
Ectoderm	Nestin	18.50	16.04	-2.46	4.50	
Ectoderm	Nestin	18.50	15.92	-2.58	4.98	
Endoderm	ID2	8.85	5.41	-3.44	9.82	10.76
Endoderm	ID2	8.85	5.31	-3.54	10.59	Standard Deviation
Endoderm	ID2	8.85	5.34	-3.51	10.36	0.85
Endoderm	ID2	9.44	5.86	-3.58	10.96	
Endoderm	ID2	9.44	5.73	-3.71	12.09	

ZUN1	Transcript	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{RA 12days})$	$\Delta\Delta Ct$	Fold Change ($2^{-\Delta\Delta Ct}$)	Average Fold Change
Ectoderm	SOX9	9.73	7.31	-2.42	4.34	3.87
Ectoderm	SOX9	9.73	7.10	-2.63	5.19	Standard Deviation
Ectoderm	SOX9	13.78	11.65	-2.13	3.36	0.87
Ectoderm	SOX9	13.78	11.72	-2.06	3.16	
Ectoderm	SOX9	13.78	11.67	-2.11	3.30	
Ectoderm	Nestin	13.73	11.78	-1.95	2.86	2.54
Ectoderm	Nestin	13.50	11.78	-1.72	2.29	Standard Deviation
Ectoderm	Nestin	16.76	15.09	-1.67	2.17	0.33
Ectoderm	Nestin	16.76	14.79	-1.97	2.90	
Ectoderm	Nestin	16.76	14.95	-1.81	2.49	
Endoderm	ID2	8.98	6.84	-2.15	3.42	2.82
Endoderm	ID2	8.98	6.81	-2.18	3.52	Standard Deviation
Endoderm	ID2	8.75	7.01	-1.74	2.34	0.60
Endoderm	ID2	8.75	7.00	-1.75	2.37	
Endoderm	ID2	8.75	6.97	-1.78	2.44	

ZUN2	Transcript	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{RA 12days})$	$\Delta\Delta Ct$	Fold Change ($2^{-\Delta\Delta Ct}$)	Average Fold Change
Ectoderm	SOX9	9.99	6.92	-3.07	7.40	7.84
Ectoderm	SOX9	9.99	6.75	-3.24	8.47	Standard Deviation
Ectoderm	SOX9	13.13	10.09	-3.04	7.24	0.51
Ectoderm	SOX9	13.13	9.96	-3.17	8.02	
Ectoderm	SOX9	13.13	9.95	-3.18	8.08	
Ectoderm	Nestin	15.01	13.62	-1.39	1.62	1.80
Ectoderm	Nestin	15.01	13.49	-1.52	1.86	Standard Deviation
Ectoderm	Nestin	15.90	14.06	-1.84	2.58	0.48
Ectoderm	Nestin	15.90	14.71	-1.19	1.28	
Ectoderm	Nestin	15.90	14.49	-1.41	1.66	
Endoderm	ID2	8.97	7.01	-1.96	2.90	2.54
Endoderm	ID2	8.97	6.90	-2.07	3.21	Standard Deviation
Endoderm	ID2	8.50	6.72	-1.78	2.43	0.51
Endoderm	ID2	8.50	6.91	-1.59	2.02	
Endoderm	ID2	8.50	6.86	-1.64	2.12	

DMSO Protocol

HUES8	Transcript	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{DMSO 7days})$	$\Delta \Delta Ct$	Fold Change ($2^{-\Delta \Delta Ct-1}$)	Average Fold Change
Mesoderm	HAND1	11.53	7.76	-3.77	12.66480814	15.67
Mesoderm	HAND1	11.53	7.24	-4.29	18.59472335	Standard Deviation
Mesoderm	HAND1	11.53	7.44	-4.09	16.05819745	2.44
Mesoderm	HAND1	11.53	7.50	-4.03	15.36331675	
Mesoderm	IGF2	13.71188	10.94	-2.77	5.829989044	5.09
Mesoderm	IGF2	13.71188	11.07	-2.64	5.241458793	Standard Deviation
Mesoderm	IGF2	13.71188	11.33	-2.38	4.212166838	0.67
Mesoderm	IGF2	13.71188	11.11	-2.60	5.070785774	
Ectoderm	Nestin	18.50	12.02	-6.48	88.26315433	91.95
Ectoderm	Nestin	18.50	11.88	-6.62	97.35952641	Standard Deviation
Ectoderm	Nestin	18.50	12.10	-6.40	83.44808972	7.33
Ectoderm	Nestin	18.50	11.86	-6.64	98.73257423	

ZUN1	Transcript	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{DMSO 7days})$	$\Delta \Delta Ct$	Fold Change ($2^{-\Delta \Delta Ct-1}$)	Average Fold Change
Mesoderm	HAND1	11.05	8.11	-2.94	6.655800859	7.49
Mesoderm	HAND1	11.05	7.84	-3.21	8.263473787	Standard Deviation
Mesoderm	HAND1	11.05	8.00	-3.05	7.291041283	0.68
Mesoderm	HAND1	11.05	7.92	-3.13	7.733459835	
Mesoderm	IGF2	13.84	12.21	-1.63	2.085946614	1.53
Mesoderm	IGF2	13.84	12.61	-1.23	1.346829617	Standard Deviation
Mesoderm	IGF2	13.84	12.73	-1.11	1.159523632	0.40
Mesoderm	IGF2	13.84	12.49	-1.35	1.541557911	
Ectoderm	Nestin	16.76	12.42	-4.34	19.18203854	16.02
Ectoderm	Nestin	16.76	12.76	-4.00	15	Standard Deviation
Ectoderm	Nestin	16.76	12.69	-4.07	15.79546694	2.22
Ectoderm	Nestin	16.76	12.84	-3.92	14.08455257	

ZUN2	Transcript	$\Delta Ct(\text{Undifferentiated})$	$\Delta Ct(\text{DMSO 7days})$	$\Delta \Delta Ct$	Fold Change ($2^{-\Delta \Delta Ct-1}$)	Average Fold Change
Mesoderm	HAND1	10.76	8.77	-1.99	2.974748708	3.11
Mesoderm	HAND1	10.76	8.83	-1.93	2.812833819	Standard Deviation
Mesoderm	HAND1	10.76	8.67	-2.09	3.260030185	0.26
Mesoderm	HAND1	10.76	8.62	-2.14	3.395001485	
Mesoderm	IGF2	13.74	12.21	-1.54	1.898998218	2.05
Mesoderm	IGF2	13.74	11.89	-1.86	2.618908037	Standard Deviation
Mesoderm	IGF2	13.74	12.29	-1.46	1.742619432	0.39
Mesoderm	IGF2	13.74	12.18	-1.56	1.949671815	
Ectoderm	Nestin	15.90	12.50	-3.40	9.566779049	9.78
Ectoderm	Nestin	15.90	12.62	-3.28	8.723419588	Standard Deviation
Ectoderm	Nestin	15.90	12.29	-3.61	11.22246847	1.04
Ectoderm	Nestin	15.90	12.49	-3.41	9.603464248	

Long term differentiation**Embryoid body formation**

HUES8	Transcript	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{EB 21 days})$	$\Delta \Delta C_t$	Fold Change ($2^{-\Delta \Delta C_t}$)	Average Fold Change
Endoderm	α 1AT	16.26	4.10	-12.16	4576.408118	4992.07
Endoderm	α 1AT	16.26	4.01	-12.26	4887.903203	Standard Deviation
Endoderm	α 1AT	16.26	3.85	-12.42	5461.191381	368.00
Endoderm	α 1AT	16.26	3.96	-12.30	5042.767517	
Mesoderm	HAND1	11.18	3.86	-7.33	160.3410481	215.97
Mesoderm	HAND1	11.18	3.27	-7.92	241.3528412	Standard Deviation
Mesoderm	HAND1	11.18	3.22	-7.97	249.8641308	40.42
Mesoderm	HAND1	11.18	3.45	-7.73	212.3058037	
Endoderm	ID2	8.77	3.02	-5.76	54.00421095	56.54
Endoderm	ID2	8.77	2.85	-5.93	60.75789574	Standard Deviation
Endoderm	ID2	8.77	2.84	-5.94	61.18050033	5.35
Endoderm	ID2	8.77	3.12	-5.65	50.21338227	
Mesoderm	IGF2	13.75	2.24	-11.51	2916.454801	3943.14
Mesoderm	IGF2	13.75	2.01	-11.74	3408.686017	Standard Deviation
Mesoderm	IGF2	13.75	1.50	-12.25	4870.992343	931.25
Mesoderm	IGF2	13.75	1.59	-12.16	4576.408118	
Ectoderm	SOX9	10.89	7.30	-3.59	12.05589345	13.18
Ectoderm	SOX9	10.89	7.14	-3.75	13.46989469	Standard Deviation
Ectoderm	SOX9	10.89	7.23	-3.66	12.61148858	1.10
Ectoderm	SOX9	10.89	7.02	-3.87	14.58755982	

ZUN1	Transcript	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{EB 21 days})$	$\Delta \Delta C_t$	Fold Change ($2^{-\Delta \Delta C_t}$)	Average Fold Change
Endoderm	α 1AT	16.34	4.93	-11.41	2721.148547	2765.23
Endoderm	α 1AT	16.34	4.96	-11.38	2665.14812	Standard Deviation
Endoderm	α 1AT	16.34	4.90	-11.44	2778.325663	98.85
Endoderm	α 1AT	16.34	4.84	-11.50	2896.309376	
Mesoderm	HAND1	11.87	4.05	-7.82	225.9719671	300.81
Mesoderm	HAND1	11.87	3.22	-8.65	401.7070581	Standard Deviation
Mesoderm	HAND1	11.87	3.81	-8.06	266.8712348	75.27
Mesoderm	HAND1	11.87	3.6	-8.27	308.6868039	
Endoderm	ID2	7.73	2	-5.73	53.19922535	55.36
Endoderm	ID2	7.73	1.97	-5.76	54.31705416	Standard Deviation
Endoderm	ID2	7.733	1.8	-5.93	61.10986265	3.88
Endoderm	ID2	7.733	2.01	-5.72	52.83175146	
Mesoderm	IGF2	14.49	1.69	-12.80	7131.550215	8612.88
Mesoderm	IGF2	14.49	1.12	-13.37	10586.95439	Standard Deviation
Mesoderm	IGF2	14.49	1.34	-13.15	9089.593115	1555.47
Mesoderm	IGF2	14.49	1.59	-12.90	7643.406267	
Ectoderm	SOX9	9.72	6.27	-3.45	10.92832205	12.18
Ectoderm	SOX9	9.72	6.05	-3.67	12.72858374	Standard Deviation
Ectoderm	SOX9	9.72	5.99	-3.73	13.26911273	1.03
Ectoderm	SOX9	9.72	6.16	-3.56	11.79415374	

ZUN2	Transcript	$\Delta C_t(\text{Undifferentiated})$	$\Delta C_t(\text{EB 21 days})$	$\Delta \Delta C_t$	Fold Change ($2^{-\Delta \Delta C_t}$)	Average Fold Change
Endoderm	$\alpha 1 \text{AT}$	17.30	4.42	-12.88	7520.78	7218.03
Endoderm	$\alpha 1 \text{AT}$	17.30	4.51	-12.79	7065.94	Standard Deviation
Endoderm	$\alpha 1 \text{AT}$	17.30	4.33	-12.97	8004.89	733.23
Endoderm	$\alpha 1 \text{AT}$	17.30	4.68	-12.62	6280.51	
Mesoderm	HAND1	11.51	2.80	-8.71	417.32	460.37
Mesoderm	HAND1	11.51	2.36	-9.15	566.13	Standard Deviation
Mesoderm	HAND1	11.51	3.02	-8.49	358.29	91.30
Mesoderm	HAND1	11.51	2.54	-8.97	499.73	
Endoderm	ID2	7.29	1.86	-5.43	43.01	42.36
Endoderm	ID2	7.29	1.89	-5.40	42.13	Standard Deviation
Endoderm	ID2	7.29	1.93	-5.36	40.97	1.05
Endoderm	ID2	7.29	1.85	-5.44	43.31	
Mesoderm	IGF2	13.66	0.50	-13.16	9152.82	8942.10
Mesoderm	IGF2	13.66	0.61	-13.05	8480.89	Standard Deviation
Mesoderm	IGF2	13.66	0.45	-13.21	9475.59	455.27
Mesoderm	IGF2	13.66	0.58	-13.08	8659.09	
Ectoderm	SOX9	10.22	6.48	-3.74	13.39	14.48
Ectoderm	SOX9	10.22	6.35	-3.87	14.66	Standard Deviation
Ectoderm	SOX9	10.22	6.42	-3.80	13.96	1.09
Ectoderm	SOX9	10.22	6.23	-3.99	15.93	

Appendix VI List of publications

Shen Mynn Tan and Peter Dröge. 2005. Comparative analysis of sequence-specific DNA recombination systems in human embryonic stem cells, *Stem Cells* 23:868-873.

Shen Mynn Tan, Siew Tein Wang, Hannes Hentze and Peter Dröge. 2007. A *UTF1*-based selection system for stable homogeneously pluripotent human embryonic stem cell cultures, *Nucleic Acids Research* 2007 35(18):e118.

STEM CELLS®

Rapid Communication

This material is protected by U.S. Copyright law.
Unauthorized reproduction is prohibited.
For reprints contact: Reprints@AlphaMedPress.com

Comparative Analysis of Sequence-Specific DNA Recombination Systems in Human Embryonic Stem Cells

SHEN MYNN TAN, PETER DRÖGE

School of Biological Sciences, Nanyang Technological University, Singapore

Key Words. Site-specific recombination • Human embryonic stem cells • Cre recombinase
 λ integrase • Plasmid transfection • $\gamma\delta$ resolvase

ABSTRACT

The great potential of human embryonic stem cells (hESCs) in basic research, regenerative medicine, and gene therapy is widely recognized. Controlled manipulation of hESC genomes through sequence-specific DNA recombination (SSR) may play a significant role in future hESC applications. However, very little is known about the functionality of SSR systems in hESCs. We demonstrate here that mutant phage λ integrase, phage P1 Cre recombinase, and mutant $\gamma\delta$ resolvase displayed distinct activities on episomal recom-

bination substrates. Interestingly, cofactor-independent λ integrase catalyzed the integrative pathway five times more efficiently than the excisive pathway. Such a degree of directionality in hESCs could be explored for sequential gene insertions into predetermined genomic sequences. We also report an improved, easy-to-use plasmid transfection system that employs silica microspheres and, in combination with SSR, could be applied to hESC genome engineering. *STEM CELLS* 2005;23:868–873

INTRODUCTION

Human embryonic stem cell (hESC) lines are derived from the inner cell mass of the early preimplantation embryo. These cells are genetically unaltered and remain pluripotent in culture over many cell generations [1–3]. The great potential of hESCs in basic research, regenerative medicine, and gene therapy is apparent [4]. To fully explore their potential, however, certain experimental techniques need to be improved or newly developed. This includes gene transfer technologies and tools for stable genetic modifications of hESCs.

It has been noted that plasmid transfection is rather inefficient with hESCs [5]. In addition, published data comparing different transfection strategies are scarce. This is especially true for reagent-based methods. We address this issue and directly compare transfection efficiencies obtained with different reagents. We also describe an efficient three-component

plasmid transfection system for hESCs which employs silica microspheres [6].

Site-specific DNA recombination (SSR) systems derived from prokaryotic cells are valuable tools for various applications in eukaryotic cells. Notably, the phage P1 recombinase Cre is frequently used to splice out (delete) marker genes from genomes after genotargeting and for conditional mutagenesis in model organisms, particularly in the mouse [7]. Furthermore, the Cre, Flp, Φ C31, and λ integrase system can achieve targeted insertion of foreign DNA into predetermined artificial or natural genome sequences [8–10]. The latter application is especially relevant to future gene therapy approaches with hESCs because it minimizes the risk of unwanted genome alterations due to random DNA insertions, which have been reported for viral vector-based strategies [11]. A main objective, therefore, is to develop SSR systems that can be used to safely modify hESC genomes for possible clinical applications.

Correspondence: Peter Dröge, Ph.D., Nanyang Technological University, School of Biological Sciences, 60 Nanyang Drive, 637551, Singapore. Telephone: 65-6316-2809; Fax: 65-6791-3856; e-mail: pdroge@ntu.edu.sg Received January 31, 2005; accepted for publication May 11, 2005; first published online in *STEM CELLS EXPRESS* June 13, 2005. ©AlphaMed Press 1066-5099/2005/\$12.00/0 doi: 10.1634/stemcells.2005-0044

STEM CELLS 2005;23:868–873 www.StemCells.com

As a first step toward this goal, we compared the enzymatic activities of three site-specific recombinases: wild-type Cre, a modified $\gamma\delta$ resolvase ($\gamma\delta$ 102NLS) derived from bacterial transposon $\gamma\delta$ bearing the two recombination-activating mutations E102Y and E124Q, and Int-h/218, which is a mutant phage λ integrase (E174K/E218K) that functions in the absence of cofactors in mammalian cells [12–17]. It has been shown that these enzymes faithfully catalyze DNA strand transfer on their respective target sequences inside mammalian cells without adding or deleting nucleotides to or from strands in the course of the reaction, respectively [12, 13, 15, 17]. We show here that these recombinases catalyzed DNA strand transfer reactions inside hESCs on plasmid substrates after cotransfection with the respective expression vector. Whereas Cre-mediated recombination is detectable in approximately 50% of transfected cells, the integrative recombination pathway catalyzed by Int-h/218 is observed in close to 20% of transfected cells. Interestingly, the excisive recombination pathway is significantly less active in hESCs than the integrative pathway. This finding will be explored to achieve controlled site-specific gene insertions into predetermined chromosomal sequences.

MATERIALS AND METHODS

Cell Culture

hESC line hES2 (46XX; ES Cell International [ESI], Helios, Singapore, <http://www.esccellinternational.com>) was cultured on ESI murine embryonic fibroblast (MEF) feeder cells according to previously established protocols [2, 18–19]. Prior to transfection, individual colonies were manually dissociated with a 27-gauge needle (Sigma, St. Louis, <http://www.sigmaaldrich.com>) into pieces of approximately 5×10^3 cells each. Four pieces were transferred into each well of a four-well plate (Nunc, Rochester, NY, <http://www.nuncbrand.com>) coated with Matrigel (BD Biosciences, Singapore, <http://www.bdbiosciences.com>) and cultured in MEF conditioned media (CM) as described [20–22]. Briefly, stock Matrigel solution was diluted 1:20 in ice-cold Dulbecco's modified Eagle's medium (DMEM) and incubated on plates at 37°C overnight. CM was prepared by collecting day-old hESC media from inactivated MEFs and filtered through a 0.2- μ m syringe filter (BD Biosciences).

Transfection and Recombination Assays

hESCs were transfected 4 days after colony transfer. Approximately 2×10^5 cells in each well were transfected using Exgen 500 (Fermentas, Hanover, PA, <http://www.fermentas.com>), Lipofectamine 2000 (Invitrogen Corporation, Carlsbad, CA, <http://www.invitrogen.com>), Fugene 6 (Roche, Basel, Switzerland, <http://www.roche.com>), Effectene (Qiagen GmbH, Hilden, Germany, <http://www.qiagen.com>), or Effectene plus 0.15 μ M silica microspheres (Polysciences, Inc., Warrington, PA, <http://www.polysciences.com>). Transfections using Exgen 500 and

Lipofectamine 2000 were performed as described [5, 23]. For Fugene 6, hESC media without serum was mixed with either 30 μ l (for 2 μ g DNA) or 60 μ l (for 4 μ g DNA) of reagent to yield a total volume of 100 μ l and incubated at room temperature (RT) for 5 minutes. The appropriate amount of DNA was added to the mixture, incubated for 45 minutes, and subsequently added dropwise to each well. Transfection using Effectene was performed as recommended by the manufacturer at a 1:25 DNA:Effectene (μ g: μ l) ratio. The final mixture was incubated for 15 minutes at RT before application. Transfection with Effectene and silica microspheres was performed as above, but with the addition of 1×10^9 microspheres 5 minutes after the start of the final incubation period. The reagent/DNA mixtures were incubated with hESCs for 24 hours.

Transfection efficiencies were determined with 2 μ g of reporter plasmid pCMVssEGFP per well. Recombination assays were carried out with 2 μ g of substrate vectors pCH-RLNRLE, p λ IR for λ integrative, or p λ ER for λ excisive recombination. Substrate vectors were cotransfected with 2 μ g of respective recombinase vectors pPGKCre, pPGK $\gamma\delta$ 102NLS, or pCMVssInt-h/218 per well. pPGK or pCMV served as mock expression vectors. pCMVssEGFP was used as positive control for λ integrative and excisive recombination, whereas plasmid pCH-RLE, the product vector that results from recombination on pCH-RLNRLE, served as positive control for Cre and $\gamma\delta$ resolvase reactions. λ integrase recombination assays were performed using Fugene 6. Cre and $\gamma\delta$ resolvase assays were carried out with Effectene.

Immunostaining

Immunostaining of hESC colonies on Matrigel was performed using SSEA-1, SSEA-4, TRA-1-61, TRA-1-80, and OCT4 specific primary antibodies (Chemicon, Temecula, CA, <http://www.chemicon.com>). Antibodies were visualized with fluorescein isothiocyanate (FITC)-labeled secondary antibodies (Chemicon). Alkaline phosphatase (AP) activity was detected according to the manufacturer's protocol (Chemicon).

Flow Cytometry

Cells were washed 48 hours after transfection with phosphate-buffered saline (PBS) (Gibco, Invitrogen Corporation) and harvested through incubation with 0.05% trypsin/ethylenediaminetetraacetic acid (EDTA) for 5 minutes. Cells were collected, washed once with PBS, and resuspended in 200 μ l of PBS for analysis in a FACSCalibur flow cytometer (Becton, Dickinson and Company, Franklin Lakes, NJ, <http://www.bd.com>). A minimum of 20,000 total events were recorded per sample and analyzed using CELLQuest software (Becton, Dickinson and Company). All values were obtained after gating mock transfected (pCMV) hESCs to zero. Recombination efficiencies were calculated after subtracting mock control values from experimental values. The positive control values were taken as 100%.

Substrate and Expression Vectors

Plasmids pCMV, pCMVssInt-h/218, pCMVssEGFP, pCH-RLN-RLZ, p λ IR, p λ ER, pPGKCre, and pPGK γ δ 102NLS have been described [13–16]. pCH-RLNRLE was constructed by inserting the coding region for enhanced green fluorescent protein (EGFP) into pCH-RLNRLZ. The EGFP gene was derived from pCMVssEGFP via polymerase chain reaction (PCR), with a short linker (GGSGG) replacing the EGFP start codon. The following primers were used in the PCR: L1EGFP-F (5'-CGGGGTACCGGGTGGAAGCGGCGGTGTGAGCAAGGGCGAGGA-3') and L1EGFP-R (5'-CGCGGATCCGAGGCTAGAACTAGTGG-3'). Both the PCR product and vector pCH-RLNRLZ were cleaved with KpnI/BamHI. The EGFP insertion replaced a large portion of the *LacZ* gene and results in the expression of an N-terminal LacZ-Linker-EGFP fusion protein. The recombined product vector pCH-RLE served as positive control and was generated through transformation into 294-Cre cells. This *Escherichia coli* strain expresses Cre constitutively at a low level [24].

RESULTS AND DISCUSSION

hESC Pluripotency on Matrigel

It was demonstrated previously that some hESC lines can be grown on Matrigel in the presence of CM obtained from MEF cell cultures [20–22, 25]. We show that ESI cell line hES2 can also be cultured under these conditions and displays surface markers SSEA-4, TRA-1-60, and TRA-1-81. In addition, hES2 cells stain positive for AP and OCT4, but negative for SSEA-1 (Figs. 1A–1F). This indicates that they remain pluripotent under these culture conditions, which are used for plasmid transfection and recombination assays below.

hESC Transfection Efficiency

Unlike mouse ESCs, hESCs are rather difficult to transfect. Protocols using lentiviral vectors [26, 27], electroporation [28, 29], and various transfection reagents [5, 23, 30] have been developed,

but comparative analyses are scarce. Currently, lentiviral-based approaches appear to be the preferred choice to achieve high transfection efficiencies. However, these vectors are quite difficult to generate and have size limitations for transgenes [23]. Here we compare plasmid transfection efficiencies obtained with more popular reagents and demonstrate that hES2 cells exhibited a clear preference for lipid-based transfection reagents (Fig. 2A). Optimized Fugene 6, Effectene, and Lipofectamine 2000 showed similar transfection efficiencies of 5%. This number decreased to less than 1% with the polymer-based reagent Exgen 500. We also tested the addition of silica microspheres to lipid-based transfection reagents and found that it resulted in a close to threefold increase in plasmid uptake in combination with Effectene (Fig. 2A). A similar finding has been reported with COS-7 and CHO cells [6, 31]. On the other hand, addition of microspheres showed only a limited effect with Lipofectamine 2000 and none with Fugene 6 (data not shown). Improved DNA uptake through silica microspheres is based on the theory that DNA/reagent/silica complexes increase the DNA concentration at the cell surface. This dependency on local DNA concentration might also lead to rather large variations between experiments.

Figure 1. Immunostaining of hESC colonies on Matrigel. To assess pluripotency after transfer to Matrigel, hESC colonies were analyzed for expression of TRA-1-60 (A), OCT4 (B), SSEA-1 (C), TRA-1-81 (D), SSEA-4 (E), and AP (F), as described in Materials and Methods. Bars = 100 μ m. Abbreviations: AP, alkaline phosphatase; hESC, human embryonic stem cell.

Figure 2. (A): Comparative analysis of different transfection reagents. hESCs were transfected with pCMVssEGFP, using various transfection reagents as indicated. The efficiency was determined via FACS as the percentage of cells expressing EGFP after gating pCMV-transfected cells as negative control. Each experiment was repeated three times, and the mean values with SDs are shown. (B, C): EGFP-transfected hESC colonies. (B): A large EGFP-transfected colony, with many EGFP-expressing cells in the monolayered periphery, but none in the thick, clustered center. Bar = 250 μ m. (C): The smaller colony displayed has EGFP-expressing cells evenly distributed. Two-fold higher transfection efficiencies (more than 20% from preliminary results) can be obtained if the transfected colonies are monolayered and small. Bar = 100 μ m. Abbreviations: EGFP, enhanced green fluorescent protein; FACS, fluorescence-activated cell sorting; hESC, human embryonic stem cell.

In addition to being a rather efficient transfection reagent in combination with silica microspheres, Effectene also showed the least cytotoxicity based on FACS (fluorescence-activated cell sorting) flow rates (data not shown). Moreover, transfection of smaller, monolayered colonies could further enhance DNA uptake at least twofold (Figs. 2B, 2C). This is most likely due to more cells being directly exposed to DNA-bearing complexes. Our plasmid transfection efficiencies in the range of 10%–20% using the commonly available and easy-to-use Effectene in combination with silica microspheres are comparable to those achieved with more sophisticated protocols, such as nucleofection [29].

SSR in hESCs

We recently developed two new SSR systems for applications in eukaryotic cells and showed that mutant phage λ integrases and the mutant $\gamma\delta$ resolvase $\gamma\delta 102\text{NLS}$, in particular, are recombination-proficient on episomal and genomic DNA substrates [13–17]. However, because hESCs differ from other mammalian cells, including mouse ESCs, in features such as cell size, doubling time, and gene expression patterns [32, 33], we were interested in comparing the λ and $\gamma\delta$ SSR system with the widely used Cre system inside hESCs. For this, we used substrate vectors that, when cotransfected with the respective recombinase expression vector, lead to EGFP expression as an easy read-out for recombination (Fig. 3).

In the first series of recombination assays, we used recombinant product vector pCH-RLE and pPGK as positive control for recombination catalyzed by Cre and $\gamma\delta$ resolvase (taken as 100% recombination efficiency). We found that Cre recombined pCH-RLE in more than half of the transfected cells (Fig. 4). In contrast, only 5% of transfected cells displayed detectable EGFP expression due to recombination of pCH-RLE by $\gamma\delta$ resolvase. The latter finding is in agreement with previous results which showed that $\gamma\delta$ resolvase is almost as efficient as Cre in murine cells, but significantly less so in human cells [34]. We want to emphasize that we are unable to determine recombination efficiencies at plasmid level because the copy number of internalized substrate DNA molecules accessible for the recombinase is unknown. This number varies per cell, and it is reasonable to assume that a single recombined copy will suffice to produce an EGFP-positive cell. Hence, the actual percentage of recombined substrate molecules could be lower than the percentage of EGFP-expressing cells normalized to the positive control.

In a second series of assays, we used pCMVssEGFP together with pCMV mock vector as positive control and demonstrate that λ Int-h/218 recombined the substrate for integrative recombination, pAIR, in 20% of transfected cells (Fig. 4), which is similar to what we described for HeLa cells [15]. Surprisingly, this number dropped to approximately 4% with pAER, the substrate for excisive recombination. This deviates from our previous findings with HeLa cells in which this difference is less than twofold [15]. Because integrative recombination between attachment sites *attB* and *attP*

generates sites *attL* and *attR*, which are then substrates for excisive recombination, our results imply that cofactor-independent λ integrases display a significant degree of directionality inside hESCs. The molecular basis for this pronounced directionality in hESCs is not understood at the moment and deserves further investigation.

Figure 3. Recombination substrate vectors. (A): pCH-RLENRLE was used as substrate for Cre and $\gamma\delta$ resolvase. Recombination between the directly repeated 114-bp Res sites as target sequence for $\gamma\delta$ resolvase, or between the directly repeated 34-bp Lox sites as target sites for Cre, leads to the excision of the neomycin resistance gene (Neo). This deletion results in EGFP expression from the N-terminus-LacZ-EGFP-fused coding region. (B): pAIR contains directly repeated attachment sites *attB* and *attP* as target sequences for the λ integrative pathway. pAER instead contains *attL* and *attR* as direct repeats that serve as target sequences for the excisive pathway. Recombination between *attB* and *attP* or between *attL* and *attR* deletes the promoter-neomycin-transcriptional stop (PGK-Neo-TSS) cassette and results in EGFP expression then driven by the CMV promoter. Abbreviations: CMV, cytomegalovirus; EGFP, enhanced green fluorescent protein.

Figure 4. Comparative analysis of three site-specific recombination systems in human embryonic stem cells. Cells were cotransfected with substrate vector and the respective recombinase expression vector. In each case, the recombination efficiency was determined 48 hours after transfection. Experiments with $\gamma\delta$ resolvase and Cre were repeated four times. Those with λ integrase were repeated three times. The mean values with the SDs are shown. Based on the two-tailed paired Student's *t*-test, all values are significant compared with respective negative controls, and *p* values are always < .05.

Our finding that the Cre recombinase is able to recombine episomal substrates in a significant fraction of cotransfected cells indicates that this recombinase is a good candidate for future hESC genome manipulations such as removal of marker genes. However, in our direct comparison with Cre, the resolvase mutant $\gamma\delta 102\text{NLS}$ is significantly less active in hESCs than in CHO cells and, therefore, may be useful only for specific applications that do not require high recombination efficiencies. It is not clear at present what causes the reduced activity of $\gamma\delta 102\text{NLS}$ in human cells. We can exclude differences in nuclear localization because both Cre and $\gamma\delta 102\text{NLS}$ contain functional nuclear localization signals [16, 35]. One possible factor to consider may be protein modification of $\gamma\delta$ resolvase in human cells.

The λ integrative pathway catalyzed by Int-h/218 is quite efficient in hESCs. Furthermore, because of the directionality that λ Int-h/218 exhibited in hESCs, in particular, this SSR system could prove very valuable for targeted gene insertions into genomic attachment sites. These sites could be artificially introduced or occur naturally in the hESC genome. Examples for λ integrase-mediated recombination genome insertions have been reported with human Burkitt's lymphoma cells and mammalian artificial chromosomes [9, 10].

Currently, we are evaluating experimentally more than 1,000 different human genome sequence tracts as potential target sites that are homologues to the 21-bp *attB* site. These tracts contain inverted consensus λ integrase recognition sites separated by a 7-bp-long spacer. Nearly all of these genomic sequences are single copy because of a unique spacer. Based on the high fidelity of the λ system in mammalian cells (Dröge, unpublished results), it should be possible to specifically direct the insertion of vectors into a few selected sites that may be more accessible in hESC chromatin. Because a strategy of gene insertion employing the pair *attB/attP*, or derivatives thereof, will generate *attL/attR* sites in the genome, integration could be stable even in the continued presence of the recombinase. This should allow for sequential targeting events into different genomic loci.

ACKNOWLEDGMENTS

We thank Curt Davey and members of our laboratory for critical comments on the manuscript. We acknowledge the financial support of A*STAR, Singapore, and Nanyang Technological University.

REFERENCES

- Thomson JA, Itskovitz-Eldor J, Shapiro SS et al. Embryonic stem cell lines derived from human blastocysts. *Science* 1998;282:1145–1147.
- Reubinoff BE, Pera MF, Fong CY et al. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotechnol* 2000;18:399–404.
- Amit M, Carpenter MK, Inokuma MS et al. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol* 2000;227:271–278.
- Stojkovic M, Lako M, Strachan T et al. Derivation, growth and applications of human embryonic stem cells. *Reproduction* 2004;128:259–267.
- Eiges R, Schuldiner M, Drukker M et al. Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. *Curr Biol* 2001;11:514–518.
- Luo D, Saltzman WM. Enhancement of transfection by physical concentration of DNA at the cell surface. *Nat Biotechnol* 2000;18:893–895.
- Branda CS, Dymecki SM. Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. *Dev Cell* 2004;6:7–28.
- Kolb AF. Genome engineering using site-specific recombinases. *Cloning Stem Cells* 2002;4:65–80.
- Dröge P, Christ N, Lorbach E. Sequence-specific DNA recombination in eukaryotic cells. U.S. Patent Application Publication US20030027337A1.
- Lindenbaum M, Perkins E, Csonka E et al. A mammalian artificial chromosome engineering system (ACE system) applicable to biopharmaceutical protein production, transgenesis and gene-based cell therapy. *Nucleic Acids Res* 2004;32:e172.
- Hacein-Bey-Abina S, Von Kalle C, Schmidt M et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 2003;302:415–418.
- Sauer B, Henderson N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci U S A* 1988;85:5166–5170.
- Lorbach E, Christ N, Schwikardi M et al. Site-specific recombination in human cells catalyzed by phage λ integrase mutants. *J Mol Biol* 2000;296:1175–1181.
- Christ N, Corona T, Dröge P. Site-specific recombination in eukaryotic cells mediated by mutant λ integrases: implications for synaptic complex formation and the reactivity of episomal DNA segments. *J Mol Biol* 2002;319:305–314.
- Christ N, Dröge P. Genetic manipulation of mouse embryonic stem cells by mutant λ integrase. *Genesis* 2002;32:203–208.
- Schwikardi M, Dröge P. Site-specific recombination in mammalian cells catalyzed by $\gamma\delta$ resolvase mutants: implications for the topology of episomal DNA. *FEBS Lett* 2000;471:147–150.
- Schwikardi M, Dröge P. Use of site-specific recombination as a probe of nucleoprotein complex formation in chromatin. *Eur J Biochem* 2001;268:6256–6262.
- Sathananthan H, Pera M, Trounson A. The fine structure of human embryonic stem cells. *Reprod Biomed Online* 2001;4:56–61.
- Reubinoff BE, Pera MF, Vajta G et al. Effective cryopreservation of human embryonic stem cells by the open pulled straw vitrification method. *Hum Reprod* 2001;16:2187–2194.
- Xu C, Inokuma MS, Denham J et al. Feeder-free growth of undifferentiated human embryonic stem cells. *Nat Biotechnol* 2001;19:971–974.
- Carpenter MK, Rosler ES, Fisk GJ et al. Properties of four human embryonic stem cell lines maintained in a feeder-free culture system. *Dev Dyn* 2004;229:243–258.
- Rosler ES, Fisk GJ, Ares X et al. Long-term culture of human embryonic stem cells in feeder-free conditions. *Dev Dyn* 2004;229:259–274.

- 23 Vallier L, Rugg-Gunn PJ, Bouhon IA et al. Enhancing and diminishing gene function in human embryonic stem cells. *STEM CELLS* 2004;22:2–11.
- 24 Aranda M, Kanellopoulou C, Christ N et al. Altered directionality in the cre-loxP site-specific recombination pathway. *J Mol Biol* 2001;311:453–459.
- 25 Draper JS, Moore HD, Ruban LN et al. Culture and characterization of human embryonic stem cells. *Stem Cells Dev* 2004;13:325–336.
- 26 Gropp M, Itsykson P, Singer O et al. Stable genetic modification of human embryonic stem cells by lentiviral vectors. *Mol Ther* 2003;7:281–287.
- 27 Ma Y, Ramezani A, Lewis R et al. High-level sustained transgene expression in human embryonic stem cells using lentiviral vectors. *STEM CELLS* 2003;21:111–117.
- 28 Zwaka TP, Thomson JA. Homologous recombination in human embryonic stem cells. *Nat Biotechnol* 2003;21:319–321.
- 29 Lakshimpathy U, Pelacho B, Sudo K et al. Efficient transfection of embryonic and adult stem cells. *STEM CELLS* 2004;22:531–543.
- 30 Park S, Kim EY, Ghil GS et al. Genetically modified human embryonic stem cells relieve symptomatic motor behavior in a rat model of Parkinson's disease. *Neurosci Lett* 2003;353:91–94.
- 31 Luo D, Han E, Belcheva N et al. A self-assembled modular DNA delivery system mediated by silica nanoparticles. *J Control Release* 2004;95:333–341.
- 32 Sato N, Sanjuan IM, Heke M et al. Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol* 2003;260:404–413.
- 33 Rao M. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev Biol* 2004;275:269–286.
- 34 Schwikardi M. Sequenz-spezifische DNA Rekombination durch $\gamma\delta$ resolvase in eukaryotischen Zellen. Dissertation, University of Cologne, Germany, 2001.
- 35 Le Y, Gagnet S, Tombaccini D. Nuclear determinants of the phage P1 Cre DNA recombinase. *Nucleic Acids Res* 1999;27:4703–4709.

A *UTF1*-based selection system for stable homogeneously pluripotent human embryonic stem cell cultures

Shen Mynn Tan¹, Siew Tein Wang², Hannes Hentze² and Peter Dröge^{1,*}

¹School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551 and ²ES Cell International Pte Ltd, 11 Biopolis Way, #05-06 Helios, 138667 Singapore

Received March 2, 2007; Revised August 7, 2007; Accepted August 24, 2007

ABSTRACT

Undifferentiated transcription factor 1 (UTF1) was identified first in mouse embryonic stem cells and is also expressed in human embryonic and adult stem cells. *UTF1* transcription ceases at the onset of differentiation, which clearly distinguishes it from less sensitive pluripotency markers, such as *Oct4* or *Nanog*. We present here two transgenic hESC lines, named ZUN. Each line harbors one copy of the *UTF1* promoter/enhancer driving a resistance gene and yielded highly homogeneous cultures under selection pressure, with a larger proportion of *Oct4* and *Sox2* positive cells. While ZUN cultures, like parental HUES8 cultures, retained the capacity to differentiate into tissues of all three germ layers using a SICD mouse teratoma model, they surprisingly exhibited an increased refractoriness to various differentiation cues *in vitro*. Together with its small size of only 2.4 kb for the entire cassette, these features render our selection system a powerful novel tool for many stem cell applications and human somatic cell reprogramming strategies.

INTRODUCTION

The transcriptional co-activator 'undifferentiated transcription factor 1' (*UTF1*) is expressed in both mouse and human embryonic stem cells (mESCs and hESCs, respectively) (1–3). *UTF1* interacts with ATF-2 and TBP-containing complexes, and seems to be involved in cell proliferation control (4,5). *UTF1* is regarded as one of the most important pluripotency marker genes defining the hESC niche (6,7). Importantly, its transcriptional regulation is very sensitive to differentiation cues and switched off faster and with a larger magnitude than either *Oct4* or *Sox2* during embryoid body formation from both mouse and human ESCs (8–10). A similar response is seen when mouse embryonic carcinoma (EC) and mESCs were

triggered to differentiate by retinoic acid (RA) (11,12). In a comparison of hESCs cultured in medium containing either fetal calf serum or defined serum replacement, *UTF1* was found to be downregulated to a larger extent in serum medium compared to *Oct4* or *Sox2*, suggesting strongly that it is a very reliable marker for early differentiation (13). Furthermore, RNA interference (RNAi) of *Oct4* transcripts in mESCs has been shown to cause downregulation of *UTF1* transcripts (14), and more than 90% of all cells in undifferentiated mESC populations stain positively for both these markers (15). These two experiments further allude to the prevalence of *UTF1* and *Oct4* co-expression in undifferentiated ESCs.

Human embryonic stem cell cultures are always heterogeneous and usually contain a subset of less pluripotent and, hence, more differentiated cells (16–18). This is also true for mouse and medaka ESCs (19,20). It was shown that the presence of minor populations of differentiated cells compromises the pluripotency of entire stem cell cultures (21). These features represent a severe problem for many hESC applications which can be addressed either via fluorescence-activated cell sorting (FACS) in order to isolate pluripotent hESCs or by antibiotic selection as a means to eliminate differentiating cells.

FACS strategies generally utilize transgenes, such as *EGFP*, under the control of genetic elements of a pluripotency marker. For example, mouse *Rex1-EGFP* and *Oct4-EGFP* combinations have been used in hESCs to screen for undifferentiated cells via flow cytometry (17,22,23). A recent improvement in semi-automated FACS and clonal recovery utilized a mouse *Oct4-EGFP* construct to sort for pluripotent hESCs (24). However, this procedure has to be repeated routinely in order to ensure more homogeneous hESC cultures over longer periods of time.

Antibiotic selection has the advantage of being less intrusive to hESCs. Expression of the neomycin resistance (*Neo*) gene and G418 antibiotic selection, along with *EGFP* expression, have been shown to be compatible with

*To whom correspondence should be addressed. Tel: +65 6316 2809; Fax: +65 6791 3856; Email: pdroge@ntu.edu.sg or pdroge2001@yahoo.de

hESC pluripotency (23,25). In the mouse, the *Oct4-Neo* combination has been used to isolate ES and EC cell lines from embryos. (19,21,26). Thus far, only one report utilized the twin markers of *EGFP* and *Neo* driven by an *Oct4* promoter for the maintenance of pluripotency in hESCs (22). However, the ability to select against partially differentiated cells as well as the derivation of more homogeneous pluripotent hESC cultures was not established in these studies.

Here we demonstrate that *UTF1* is an exceptionally sensitive pluripotency marker. We found that *M1*, a genetic element comprising a conserved octamer sequence important for *Nanog* expression (27), is also present and functional in the *UTF1* 3' enhancer, but not in the UTR of *Oct4* or *Sox2*. As a representative example, we present two stable ZUN hESC lines in which *Neo* is under the control of *UTF1* promoter/enhancer. Application of G418 selection results in ablation of differentiated hESCs and homogeneously pluripotent ZUN cultures, containing a significantly larger number of Oct4 and Sox2 positive cells compared to parental cultures. While ZUN cultures retain the capacity to differentiate into progenitors of all three germ layers *in vivo*, they surprisingly exhibited an increased refractoriness to differentiation cues under various culture conditions.

MATERIALS AND METHODS

Plasmids

The full-length human *UTF1* gene was amplified from HeLa genomic DNA using primers UTF1-F (CCGGAATTCAGCGCCAGGACCGACCCCTTA) and UTF1-R (TGCTCTAGAGGGTTCAGCACTTCTCCTGCCTC) via PCR. The product was cleaved with EcoRI and XbaI, and ligated into phagemid *pTZ-18R*. The *UTF1* coding sequence was subsequently replaced with that of *EGFP* or *Neo* in the following way. KpnI and NcoI restriction sites were introduced between the end of the *UTF1* promoter and before the ATG codon by amplifying the promoter region of *pTZ-UTF1* via PCR with primers UTF1-F and UTF1-MR (CGGGGTACCCCGGGGCTGGGGCGCGG). The resulting fragment was cleaved with EcoRI and NcoI, and ligated to a cleaved *pCMV-EGFP* which replaces the CMV promoter. The plasmid was then digested with EcoRI and BsrGI to produce an (EcoRI) *UTF1promoter-EGFP* (BsrGI) sticky-ended fragment. *pTZ-UTF1* was used as a PCR template to amplify a segment of the *UTF1* 3' UTR region and inserting a BsrGI RE site near the end of *UTF1* transcription unit using primers UTF1-MF (AGCTGTACAAGTGAGTCCCGGCTGCGGC) and UTF1-BR (AGAATACTCAAGCTATGCATCCAAC). This produced another BsrGI-BamHI sticky-ended 1 kb fragment. *pTZ-UTF1* was then digested with EcoRI and BamHI, and this 3.2 kb fragment was ligated to the above two 1 kb fragments, yielding *pTZ-UTF1-EGFP*. The *Neo* gene was amplified with flanking KpnI and BsrGI RE sites via PCR with *pPGK-Neo* as template and Neo-F (CTGGGTACCCATGATTGAACAAGATGG) and Neo-R (ACTGTGATCATCAGAAGAACTCGTCAAGAA) primers.

The *pTZ-UTF1-EGFP* backbone was digested with KpnI and BsrGI to remove *EGFP*, and ligated with the cleaved *Neo* fragment to yield *pTZ-UTF1-Neo*. The promoter-less version of *pTZ-UTF1-EGFP* was generated using EcoRI and KpnI to remove the *UTF1* promoter. The ends were blunted using Klenow Fragment (NEB) followed by self-ligation. The enhancer-less version of *pTZ-UTF1-EGFP* was generated using PstI to remove the *UTF1* enhancer followed by self-ligation.

pTZ-UTF1-EGFP was used as template to introduce the mutant M1 sequence (GTAGTGGT; original sequence GTCTGGGT) (27). Flanking arms 200 bp from M1 were amplified using two primer pairs, UTF1-M1F (GACCCA GGAGTAGCATG) and UTF1-M1mR (CTAGGCCCA CCACTACCAGAGCCAC) (near the 5' XcmI RE site), and UTF1-M1mF (GTGGCTCTGGTAGTGGTGGGC CTAG) and UTF1-M1R (CATTGTTATGCTAGCGG) (near the 3' NheI RE site). These were combined via assembly PCR using UTF1-M1F and UTF1-M1R as primers. The resulting 370 bp fragment was digested with XcmI and NheI, and ligated to a similarly cleaved *pTZ-UTF1-EGFP* in order to replace the original M1 sequence.

Plasmids were amplified in *Escherichia coli* strain DH5 α . We used standard PCR conditions with *Pfu* polymerase (Promega). Restriction enzymes were from New England Biolabs.

Cell culture, transfection and differentiation

F3 human foreskin fibroblasts (28), A549 and HeLa carcinoma cell lines were cultured according to standard procedures. hESC line hES2 (karyotype: 46XX) was obtained from ES Cell International Pte Ltd. HUES8 (karyotype: 46XY) and HUES9 (karyotype: 46XX, inv9) was provided by the Harvard Stem Cell Institute. hESCs were cultured on G418-resistant MEFs (PMEF-N, Chemicon), as previously described (29,30).

F3, A549 and HeLa cells were transfected with 10 μ g of reporter plasmid via electroporation as described previously (31). hESCs were transfected with 2 μ g of reporter plasmid and Effectene (Qiagen) as described previously (29).

Differentiation medium was prepared by removing basic fibroblast growth factor and exchanging knockout serum replacement with fetal bovine serum (all three from Gibco). Spontaneous differentiation was conducted on matrigel (BD Biosciences) or gelatin surfaces without MEFs. Induced differentiation on MEFs was conducted either with 10 μ M (RA) or 1% (DMSO) (Sigma) which was added to differentiation medium (32). Embryoid body formation was induced by plating hESCs onto low-cell binding plates (Nunc).

SCID mouse teratoma assay

To test the differentiation potential of hESC lines *in vivo*, serially passaged hESCs were manually harvested and injected as clumps with an approximate cell dose of $2-4 \times 10^6$ cells in a volume of 50 μ l into the quadriceps of the right hind limb of a male SCID mouse (3 mice/hESC line). Mice were maintained under controlled conditions

in accordance with the National Institutes of Health (NIH) and National Advisory Committee for Laboratory Animal Research (NACLAR) guidelines, and with approval of the Biopolis Institutional Animal Care and Use Committee (Biopolis IACUC approval 050008, National University of Singapore Institutional Review Board 05-020). Teratoma formation was monitored visually using a simple grading system that was confirmed by caliper measurements (grade 0—no detectable enlargement, grade 1—enlargement just detectable, grade 2—obvious enlargement, grade 3—enlargement impedes locomotion). When teratomas reached grade 3 (after 6–8 weeks), teratomas weighing 1–2 g were excised, fixed (Bouin's solution), paraffin-embedded, sectioned and histologically analyzed following staining with hematoxylin and eosin. Additionally, teratomas were assessed by a pathologist at the Department of Pathology, National University of Singapore.

Characterization of ZUN hESC lines

ZUN cell lines were generated by transfecting HUES8 cells using Effectene with *pTZ-UTF1-Neo*. Individual colonies were mechanically expanded in the presence of G418 (500 µg/ml). Cloning efficiency was 10^{-5} , with four stable cell lines generated, of which only two contained the full-length *UTF1-Neo* cassette. These two lines were analyzed here and labeled as 'ZUN'. Karyotyping via Giemsa staining was performed by Cytogenetics Lab, KK Women's and Children's Hospital, Singapore. 50 µl of BrdU/Colcemid (50:50) mixture was added to an 80% confluent 35 mm dish containing hESCs with 2 ml of medium the night before staining and analysis. Transgenic genomic inserts in both cell lines were sequenced from correctly sized genomic PCR products amplified using standard *Taq* PCR protocols and *UTF1-F* and *UTF1-R* primers. Southern blotting was performed as previously described (33), with 10 µg of isolated genomic DNA digested with *EcoRI*. *Neo* was detected using a 32 P-labeled *Neo* probe. WST-1 cell proliferation assays were conducted using manufacturer protocols (Roche). Briefly, 3×10^4 cells were seeded onto each MEF-layered 96-well at day 0. WST-1 reagent was added to the wells for 30 min the next day, and 440 nm readings were recorded using a standard photoplate reader. This was repeated for a total of 4 days with different plates. Each cell line had five replicates per time-point. Alkaline phosphatase staining and TRA-1-60/TRA-1-81 surface marker immunostaining (visualized with FITC) were conducted using manufacturer protocols (Chemicon). Flow cytometry of surface markers, Oct4 and Sox2 was performed as previously described (29), with pre-incubation of primary and FITC-conjugated secondary antibodies (Santa Cruz).

Quantitative real-time PCR

Total RNA was recovered from cell pellets (Qiagen) and cDNA was synthesized via the SSII reverse transcriptase kit, primed with oligo dTs (Invitrogen). SYBR Green qRT-PCR (Qiagen) was conducted using the 7500 real-time PCR System and results analyzed using sequence

detection software (SDS) version 1.2.3 (Applied Biosystems). All experiments had at least biological duplicates and assay triplicates, and results were analyzed via the $\Delta\Delta C_t$ method using GAPDH as the housekeeping transcript. Primers used were as follows: GAPDH (CCCACTCCTCCACCTTTGAC and CTCCCCTCTTCAAGGGGTCT), Nanog (AGTCCCAAAGGCAAAC AACCCACTTC and ATCTGCTGGAGGCTGAGGTA TTTCTGTCTC) (6), Oct4 (CTTGCTGCAGAAGTGGG TGGAGGAA and CTGCAGTGTGGGTTTCGGGCA) (1), Sox2 (ATGCACCGCTACGACGTGA and CTTTGCACCCCTCCCATTT) (1), *UTF1* (AGCAGATCCGG AAGCTCATGGG and TCCTCGGGGATGCAGGTG), *UTF1-Neo* (GGGTACCCCATGATTGAACA and CAGGTCCGTCTTGACAAAAG), Sox9 (GAAGCTCGC GGACCAGTA and CGTTCCTTCAACCACTTCCTC), Nestin (CGTTGGAACAGAGGTGGAG and GAGC GATCTGGCTCTGTAGG), ID2 (CTGGACTCGCA TCCCACTAT and CACACAGTGCTTTGCTGTCA), HAND1 (TGCCTGAGAAAGAGAACCAG and ATGGCAGGATGAACAAACAC) (9), IGF2 (TCCTC CCTGGACAATCAGAC and AGAAGCACCAGCAT CGACTT) (9), α 1AT (AGACCCTTTGAAGTCAAG GACACCG and CCATTGCTGAAGACCTTAGTG ATGC) (34).

RESULTS

UTF1 is a sensitive pluripotency marker for hESCs

We evaluated first the sensitivity of *UTF1* expression with respect to induced differentiation using quantitative real-time PCR (qRT-PCR). Various hESC lines were triggered to differentiate via retinoic acid (RA) or dimethyl sulfoxide (DMSO) treatment. The results show that this almost completely abolished *UTF1* expression; a response that was not observed for *Nanog* and *Oct4* (Figure 1A). We next generated the following three *UTF1* transgenic constructs: *pTZ-UTF1*, *pTZ-UTF1-EGFP* and *pTZ-UTF1-Neo* (Figure 1B). In addition, three variants of *pTZ-UTF1-EGFP* were designed: one without the minimal 5' promoter, one lacking the 3' enhancer, and one containing an altered *M1* sequence. The *M1* octamer sequence was found to be conserved in the *Nanog* promoter of many species (27), and we identified it recently upstream of the Oct4/Sox2 binding site within the 3' *UTF1* enhancer.

The results demonstrate hESC-specific expression of the full-length *pTZ-UTF1-EGFP* after transfection into various human cell lines and hESC line HUES8 (Figure 1C). We next introduced the three *pTZ-UTF1-EGFP* versions into HUES8 cells and showed that the 5' promoter and 3' enhancer regions were essential for EGFP expression (Figure 1C). In addition, the altered *M1* element reproducibly led to a significantly smaller number of EGFP expressing hESCs, which is similar in magnitude to its reported effect on *Nanog* transcription (27). Together, these data confirm that *UTF1* transcription is a very sensitive pluripotency marker specific for hESCs and suggests that its regulation in hESCs involves *M1*-binding factors in addition to Oct4 and Sox2.

Figure 1. UTF1 expression in hESCs. (A) qRT-PCR Expression analysis of RA- (for 12 days) and DMSO-induced (for 7 days) differentiation in hESCs, normalized to that of respective undifferentiated hESC lines, with two biological replicates for each sample. (B) Schematic drawing of vectors. Full-length *UTF1* was cloned into phagemid *pTZ-18R* yielding *pTZ-UTF1*. Its coding region was subsequently replaced with enhanced green florescent protein (*EGFP*) or neomycin (*Neo*), yielding *pTZ-UTF1-EGFP* and *pTZ-UTF1-Neo*, respectively. Red text indicates alterations in *pTZ-UTF1-EGFP*. (C) Flow cytometry of *UTF1*-driven *EGFP* expression in various cell lines following transient transfection of *pTZ-UTF1-EGFP*. Transfection efficiencies were normalized to those determined in parallel using *pCMV-EGFP* and are the average of three biological replicates. See Supplementary Figure S7 for representative dot plots.

Characterization of transgenic *pTZ-UTF1-Neo* hESC lines

We established stable *pTZ-UTF1-Neo* cell lines from HUES8 cells via random genomic transgene insertion. Two single-copy lines with normal karyotype, dubbed ZUN1 and ZUN2, were used in subsequent experiments (Supplementary Figure S1). Careful characterization of both lines revealed that they have similar proliferation rates (Supplementary Figure S2), exhibit comparable alkaline phosphatase (AP) staining patterns (Supplementary Figure S3), and express *Nanog* at a similar level as parental HUES8 cells (Figure 2A). However, an elevated level of *Oct4*, *Sox2* and *UTF1* in ZUN cultures was reproducibly detected by qRT-PCR (Figure 2B). This correlation of expression supports the notion that the control of *UTF1* involves factors Oct4 and Sox2, but most likely not Nanog (35).

Figure 2. Characterization of ZUN hESC lines. (A) Expression analysis of ZUN hESCs. qRT-PCR comparison of pluripotency markers across the three hESC cultures normalized to HUES8 samples, performed with biological triplicates. (B) Oct4 and Sox2 expression analysis of ZUN hESCs. Percentage of hESCs expressing Oct4 and Sox2. ZUN cultures consist of significantly more Oct4- and Sox2-expressing cells compared to HUES8, based on the two-tailed paired Student's *t*-test (P -values < 0.05). (C) Surface marker analysis of ZUN hESCs. Percentage of positively stained hESCs for TRA-1-60 and TRA-1-81. Both ZUN lines consist of significantly more positively stained cells compared to parental HUES8, based on the two-tailed paired Student's *t*-test (P -values < 0.05). (D) qRT-PCR analysis of removal and restoration of G418 selection pressure on ZUN1 and ZUN2 cultures. ZUN lines were cultured in the presence of G418 for 90 days (+G418); G418 removed for 30 (−G418) or 60 (−/−G418) days; G418 removed for 30 days and re-introduced for 30 days (−/+G418). Expression levels were normalized to +G418 values and presented as an average of three experimental repeats.

Figure 3. Differentiation of ZUN hESC lines on Matrigel and Gelatin. (A) Expression analysis of spontaneous differentiation using pluripotency markers. The qRT-PCR of hESCs without MEFs plated on gelatin or matrigel for 12 days was performed with biological triplicates, and presented as expression levels relative to that of respective undifferentiated cell lines. (B) AP staining of hESCs on gelatin and matrigel. Representative phase contrast photomicrographs comparing HUES8 and ZUN cell morphologies on the 12th day and visualization of the ablative effects after 2 days of exposure to G418. Scale bar: 55 µm.

Flow cytometry revealed that the observed higher level of global *Oct4*, *Sox2* and *UTF1* expression was due to the presence of a larger fraction of cells that express these markers. A statistically significant higher percentage of ZUN cells stained positive for *Oct4* and *Sox2* compared to HUES8 cultures (Figure 2B). We could also demonstrate that ZUN cultures contained more cells which stain positive for surface markers TRA-1-60 and TRA-1-81 (Figure 2C). These results indicate that HUES8 cultures contained more hESCs with a low expression level of pluripotency markers. In contrast, such cells have been ablated in ZUN cultures via G418 selection.

We next analyzed the effect of sequential removal and restoration of G418 selection on both ZUN lines and employed qRT-PCR of pluripotency markers *Nanog*, *Oct4*, *Sox2* and *UTF1* for the detection of the onset of differentiation (9). We also included the transgene *UTF1-Neo* in this analysis because its regulation should be linked to that of the endogenous *UTF1*. The results show that spontaneous differentiation can be detected in cultures during 60 days of G418 withdrawal. This is indicated by the progressive reduction in signals obtained from all markers, with the exception of *Sox2* (Figure 2D). However, upon re-application of antibiotic selection, original expression levels could be restored in these cultures. Using a transgenic hESC line in which *Neo* is expressed from the elongation factor 1 alpha (*EF-1α*) promoter as a control, we confirmed that G418 withdrawal has no effect on the expression of this set of markers which clearly links this effect to the presence of the *UTF1-Neo* transgene in hESCs (Supplementary Figure S4). Together, these data show that ZUN cultures are more homogeneously pluripotent than the parental cultures and that G418 addition and subsequent ablation of differentiated ZUN cells is required to sustain homogeneous cultures over longer periods of time.

ZUN cultures are refractory to various differentiation cues

We demonstrated above that ZUN cultures are more homogeneously pluripotent than the parental cultures. We were interested, therefore, in investigating the differentiation behavior of ZUN cells. First, we induced differentiation by plating them on matrigel or gelatin without supporting mouse embryonic fibroblasts (MEFs) and in the absence of G418. In the matrigel protocol, HUES8 cells quickly initiated differentiation as indicated by the 5-fold downregulation of our panel of markers (Figure 3A). However, both ZUN lines maintained significantly higher overall expression levels even after 12 days of culture. Similarly, the gelatin protocol induced a 2-fold downregulation of *Oct4* and *UTF1* in HUES8, but not in ZUN populations. We observed further that under these conditions, ZUN cultures have a larger number of colonies consisting of cells which display a more compact morphology that stain stronger for alkaline phosphatase (AP), and a reduced occurrence of differentiating cells at colony edges. Upon subsequent addition of G418 for 2 days, only pluripotent cells that stained for AP survived in ZUN cultures, demonstrating again that differentiating ZUN cells are rapidly and efficiently ablated by G418 selection (Figure 3B).

We analyzed next the effects of RA or DMSO treatment on ZUN cultures in the absence of G418 selection. After 12 and 7 days, respectively, qRT-PCR showed that ZUN cultures exhibited a lower rate of differentiation in comparison to HUES8 cultures (Figure 4A). Notably, *UTF1* and *UTF1-Neo* expression diminished by the largest magnitude among the pluripotency markers. Certain differentiation markers also indicated a slower onset of differentiation in ZUN cultures. HUES8 cultures showed a greater (2- to 10-fold) upregulation of *Sox9*, *Nestin* and *ID2* due to RA treatment, and of *HAND1*, *IGF2* and *Nestin* after exposure to DMSO (Figure 4B). The expression of surface markers of pluripotency,

Figure 4. Differentiation of ZUN hESCs induced by RA and DMSO. (A) Expression analysis of induced differentiation using pluripotency markers. qRT-PCR of hESCs on MEFs induced to differentiate via RA (for 12 days) or DMSO (for 7 days) in differentiation medium. (B) Transcript analysis of induced differentiation using differentiation markers. (C) TRA-1-60 and TRA-1-81 surface marker immunostaining for induced differentiation, comparing HUES8 with ZUN cultures. The bottom row shows a control panel of undifferentiated HUES8 cells stained with secondary antibody only, TRA-1-60 or TRA-1-81 antibodies (left to right). Scale bar: 30 μ m. (D) Transcript analysis after EB formation. qRT-PCR of EBs from hESCs induced to differentiate for 21 days in suspension cultures. The qRT-PCR experiments were performed with biological duplicates, and presented as expression levels relative to that of respective undifferentiated cell lines.

TRA-1-60 and TRA-1-81, complements these transcript analyses, showing that HUES8 cells exhibited a more muted staining pattern compared to ZUN colonies (Figure 4C). These results confirm our conclusion that under various conditions, ZUN cultures exhibit a more robust phenotype against the onset of differentiation than parental cultures.

ZUN cells retain the capacity to differentiate into progenitors of all three germ layers *in vitro* and *in vivo*

The observed phenotype of increased robustness against the onset of differentiation led us to investigate ZUN cultures under long term differentiation conditions. The results show that ZUN and HUES8 cells cultured under RA and DMSO differentiation protocols for 30 days show

similar morphologies and lack of AP staining (Supplementary Figure S5). These cells were completely ablated upon addition of G418, as were those in equally long-term gelatin and matrigel cultures (data not shown). In addition, 21 days of embryoid body formation in suspension cultures caused a significant upregulation of differentiation markers from all three germ layers (Figure 4D). Both ZUN lines and HUES8 show very high expression of *HAND1*, *IGF2* and *ID2* as indicated by the large fold increase relative to their undifferentiated counterparts. They also show relatively high expression of α 1 anti-trypsin (*α1AT*) and *Sox9*. The morphological characteristics and growth rates of HUES8 and ZUN cell-derived EBs were indistinguishable (Supplementary Figure S6). Furthermore, we showed by a teratoma assay that the parental HUES8 line as well as ZUN1 and

Figure 5. Teratoma formation SCID mice studies. The hESC lines ZUN1 (A), ZUN2 (B) and the parental line HUES8 (C) were injected into the hind leg of male SCID mice, and the teratoma formation was followed by a simple grading system (see Materials and Methods section). When teratomas were full developed (1–2 g tissue weight, 6–8 weeks), they were harvested and processed by routine histological procedures, followed by hematoxylin/eosin staining. (D) Detailed histology of ectodermal (epithelium, neuroectoderm), mesodermal (bone in red, cartilage in blue) and endodermal (glandular structures) structures are shown for each hESC line.

ZUN2 formed tissue of all three germ layers (i.e. ectodermal structures, mesoderm such as cartilage and bone, as well as mesodermal glandular structures) when transplanted into immune-deficient SCID mice (Figure 5). We conclude, therefore, that the robustness which our ZUN cultures displayed against the onset of differentiation induced by various protocols does not compromise their potential to differentiate, like parental cultures, into all three germ layers *in vitro* and *in vivo*.

DISCUSSION

Expression of *UTF1* is controlled by a 5' TATA-less promoter consisting of four GC boxes. The 3' enhancer element harbors a twin octamer sequence where the synergistic binding of Oct4 and Sox2 is essential for *UTF1* expression in both mouse and human ESCs (8,35). Spl-like transcription factors that bind to GC boxes present in the *UTF1* promoter are most likely involved in the regulation of *UTF1*, since they are regarded as regulators of embryonic development in vertebrates (36) and shown to be involved in the transcriptional control of both *Oct4* and *Nanog* (37,38). We showed here that *UTF1* is indeed expressed in hESCs, but not in human primary fibroblasts and carcinoma cell lines. Further, both 5' promoter and 3' enhancer are needed for expression. A novel finding is that an octamer sequence (*M1*) which is conserved in the *Nanog* promoter

of many species (27) is also present upstream of the analogous human *Oct4/Sox2* cognate sequence in the 3' *UTF1* enhancer. The fact that *M1* mutations resulted in equally muted expression levels of *UTF1* and *Nanog*, strongly suggests that similar *M1*-binding factor(s) are involved in the control of both pluripotency marker genes. Our results also confirm previous reports showing that endogenous *UTF1* is downregulated faster than *Oct4* or *Nanog* at the onset of differentiation to nearly undetectable levels. In addition, detailed analysis of SymAtlas data from the Genomics Institute of the Novartis Research Foundation showed that both *UTF1* and *Nanog* expression are greatly diminished at day 8.5 of mouse embryo formation, unlike *Oct4* or *Sox2* (<http://symatlas.gnf.org>). Together, these results firmly establish *UTF1* as a sensitive and reliable pluripotency marker for hESCs.

Recent reports revealed that *UTF1* is also an important marker expressed in other stem cell types. For example, spermatogonial stem cells isolated from adult mouse testis were shown to exhibit features reminiscent of pluripotent ESCs. These cells express *Oct4*, *Nanog*, *Rex1* and *UTF1* and, interestingly, also downregulate *UTF1* significantly faster than these other markers upon embryoid body formation (39). Further, pluripotent stem cells derived from re-programmed primary mouse fibroblasts through co-expression of transgenes encoding Oct4, Sox2, c-Myc and Klf4 also express *UTF1*. It is noteworthy that the expression level of *UTF1* in pluripotent stem cells derived from different sources is always more tightly linked with

that of the key marker *Oct4*, and not with other markers like *Cripto* or *Nanog* (40). These data, in conjunction with results presented here, indicate that our *UTF1*-based selectivity tool should be widely applicable to select for pluripotent human stem cell lines derived from different sources.

We demonstrated that the control of both endogenous and exogenous *UTF1* in ZUN cells is linked, as expected, and that *UTF1*-driven *Neo* expression in hESCs in conjunction with G418 selection can be used to efficiently ablate differentiating cells in standard culture conditions. This resulted in ZUN cultures which display a higher global expression level of key pluripotency markers. We showed that this elevated level is most likely due to an increase in the fraction of *Oct4* and *Sox2* positive hESCs. Furthermore, the number of cells staining positive for surface markers TRA-1-60 and TRA-1-81 was also elevated which, if taken together, indicates that ZUN cultures are more homogeneously pluripotent than the parental cultures (Figure 2A–C). Upon removal of selection pressure from ZUN cultures, the overall expression level of these markers declined, which indicates loss of pluripotency and, consequently, more heterogeneous hESC cultures. Interestingly, the absence of selection pressure for 60 days did not cause widespread silencing of the exogenous *UTF1-Neo*. This enabled us to add G418 only periodically in order to select against differentiating hESCs in ZUN cultures. Hence, the *UTF1* promoter/enhancer combination could be a generally applicable control element for sustained expression of transgenes in pluripotent hESCs, even in the absence of selection.

We showed that the generation of ZUN lines harnessed the sensitivity of *UTF1* as a pluripotency marker which resulted in rapid and efficient ablation of differentiating cells. An unexpected finding was the refractoriness of ZUN cultures to the onset of differentiation. We attribute this phenotype to the increased fraction of pluripotent hESCs in homogeneous ZUN cultures compared to typical, more heterogeneous hESC cultures. Given the propensity for heterogeneous cultures to spontaneously differentiate (21), the increased stability of human stem cell cultures with respect to the uniformity in pluripotency combined with efficient ablation of differentiating cells offered by the ZUN system could significantly improve large scale growth for many future hESCs applications. In addition, ZUN cultures show a higher expression of pluripotency markers when compared to HUES8 cultures following thawing after standard slow cooling cryopreservation procedures (data not shown), thereby counteracting a reported negative effect associated with cryopreservation (41).

An intrinsic advantage of the *UTF1*-based selection system over other pluripotency marker-based strategies is its enhanced sensitivity towards many differentiation pathways. Although the well-characterized pluripotency marker *Oct4* has been a component in strategies for both mouse and human ESCs (21,22), the use of *Oct4* in this context can be problematic. Apart from being less sensitive to the onset of differentiation, its expression level is also very tightly controlled. A less than 2-fold increase in *Oct4* expression alone leads to differentiation

into primitive endodermal and mesodermal lineages in mouse ESCs (42). This implies that the *Oct4-Neo* system is potentially less sensitive to these differentiation pathways, and, thus, will not be able to efficiently ablate cells that are undergoing differentiation into these lineages. In addition, the <2 kb length of the entire human *UTF1* promoter/enhancer element represents a significant advantage for genetic manipulation of hESCs, or other human cell types, if one compares this to an unwieldy 8 kb of the *Oct4* control element.

Our ZUN system has a large range of potential applications. In its current context as a stable transgene in hESCs, it can be used for the maintenance of homogeneously pluripotent cultures. In combination with gene targeting technologies, this application can be upscaled to meet the larger cell quantities ultimately needed for stem cell-based therapies. ZUN lines can also be employed to screen and characterize synthetic factors, media components or surface matrices that permit maintenance of pluripotency, and it can be used to test 3D scaffolds or aid in the derivation of new hESC lines.

As indicated above, the ZUN system shows potential in selection protocols in order to isolate adult stem cells from primary cultures. It can also be used in hESCs in conjunction with cell differentiation strategies as a component of positive-negative selection. The system can also function as a tool in somatic cell reprogramming experiments to identify successfully reprogrammed pluripotent cells. It can also be used to improve future human somatic cell nuclear transfer protocols. With our first demonstration of its utility here and its many possible applications, we envision that our ZUN system will become a valuable selectivity tool for many facets of stem cell research.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank R.J. Zhou for help with the Southern blot analysis, C.-T. Lee for the construction of *pTZ-UTF1-EGFP(M1)* and K. Neef for hESC line EPN14. Special thanks go to C.A. Davey for critical comments on the manuscript. Funding to pay the Open Access publication charges for this article was provided by NTU special investigator grant.

Conflict of interest statement. None declared.

REFERENCES

1. Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, S., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M.K. *et al.* (2003) Differences between human and mouse embryonic stem cells. *Dev. Biol.*, **269**, 360–380.
2. Brimble, S.N., Zeng, X., Weiler, D.A., Luo, Y., Liu, Y., Lyons, I.G., Freed, W.J., Robins, A.J., Rao, M.S. *et al.* (2004) Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001. *Stem Cells Dev.*, **13**, 585–597.

3. Sun, Y., Li, H., Liu, Y., Shin, S., Mattson, M.P., Rao, M.S. and Zhan, M. (2007) Cross-species transcriptional profiles establish a functional portrait of embryonic stem cells. *Genomics*, **89**, 22–35.
4. Fukushima, A., Okuda, A., Nishimoto, M., Seki, N., Hori, T.A. and Muramatsu, M. (1998) Characterization of functional domains of an embryonic stem cell coactivator UTF1 which are conserved and essential for potentiation of ATF-2 activity. *J. Biol. Chem.*, **273**, 25840–25849.
5. Nishimoto, M., Miyagi, S., Yamagishi, T., Sakaguchi, T., Niwa, H., Muramatsu, M. and Okuda, A. (2005) Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the *UTF1* locus. *Mol. Cell. Biol.*, **25**, 5084–5094.
6. Bhattacharya, B., Miura, T., Brandenberger, R., Mejido, J., Luo, Y., Yang, A.X., Joshi, B.H., Ginis, I., Thies, R.S. *et al.* (2004) Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood*, **103**, 2956–2964.
7. Trounson, A. (2006) The production and directed differentiation of human embryonic stem cells. *Endocr. Rev.*, **27**, 208–219.
8. Nishimoto, M., Fukushima, A., Okuda, A. and Muramatsu, M. (1999) The gene for the embryonic stem cell coactivator *UTF1* carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol. Cell. Biol.*, **19**, 5453–5465.
9. Cai, J., Chen, J., Liu, Y., Miura, T., Luo, Y., Loring, J.F., Freed, W.J., Rao, M.S. and Zeng, X. (2006) Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells*, **24**, 516–530.
10. Wei, C.L., Miura, T., Robson, P., Lim, S.K., Xu, X.Q., Lee, M.Y., Gupta, S., Stanton, L., Luo, Y. *et al.* (2005) Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. *Stem. Cells*, **23**, 166–185.
11. Eifert, C., Sangster-Guity, N., Yu, L.M., Chittur, S.V., Perez, A.V., Tine, J.A. and McCormick, P.J. (2006) Global gene expression profiles associated with retinoic acid-induced differentiation of embryonal carcinoma cells. *Mol. Reprod. Dev.*, **73**, 796–824.
12. Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y. and Lemischka, I.R. (2006) Dissecting self-renewal in stem cells with RNA interference. *Nature*, **442**, 533–538.
13. Skottman, H., Stromberg, A.-M., Matilainen, E., Inzunza, J., Hovatta, O. and Lahesmaa, R. (2005) Unique gene expression signature by human embryonic stem cells cultured under serum-free conditions correlates with their enhanced and prolonged growth in an undifferentiated stage. *Stem Cells*, **24**, 151–167.
14. Sen, G., Wehrman, T.S., Myers, J.W. and Blau, H.M. (2004) Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nat. Genet.*, **36**, 183–189.
15. Hoof, D.V., Passier, R., Oostwaard, D.W.-V., Pinkse, M.W.H., Heck, A.J.R., Mummery, C.L. and Krijgsveld, J. (2006) A quest for human and mouse embryonic stem cell-specific proteins. *Mol. Cell Proteomics*, **5**, 261–273.
16. Amit, M., Carpenter, M.K., Inokuma, M.S., Chiu, C.P., Harris, C.P., Waknitz, M.A., Itskovitz-Eldor, J. and Thomson, J.A. (2000) Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev. Biol.*, **227**, 271–278.
17. Eiges, R., Schuldiner, M., Drukker, M., Yanuka, O., Itskovitz-Eldor, J. and Benvenisty, N. (2001) Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. *Curr. Biol.*, **11**, 514–518.
18. Stewart, M.H., Bosse, M., Chadwick, K., Menendez, P., Bendall, S.C. and Bhatia, M. (2006) Clonal isolation of hESCs reveals heterogeneity within the pluripotent stem cell compartment. *Nat. Methods*, **3**, 807–815.
19. McWhir, J., Schnieke, A.E., Ansell, R., Wallace, H., Colman, A., Scott, A.R. and Kind, A.J. (1996) Selective ablation of differentiated cells permits isolation of embryonic stem cell lines from murine embryos with a non-permissive genetic background. *Nat. Genet.*, **14**, 223–226.
20. Hong, Y., Winkler, C. and Scharf, M. (1996) Pluripotency and differentiation of embryonic stem cell lines from the medakafish (*Oryzias latipes*). *Mech. Dev.*, **60**, 33–44.
21. Mountford, P., Nichols, J., Zevnik, B., O'Brien, C. and Smith, A. (1998) Maintenance of pluripotential embryonic stem cells by stem cell selection. *Reprod. Fertil. Dev.*, **10**, 527–533.
22. Zwaka, T.P. and Thomson, J.A. (2003) Homologous recombination in human embryonic stem cells. *Nat. Biotechnol.*, **21**, 319–321.
23. Gerrard, L., Zhao, D., Clark, A.J. and Cui, W. (2005) Stably transfected human embryonic stem cell clones express OCT4-specific green fluorescent protein and maintain self-renewal and pluripotency. *Stem Cells*, **23**, 124–133.
24. Hewitt, Z., Forsyth, N.R., Waterfall, M., Wojtacha, D., Thomson, A.J. and McWhir, J. (2006) Fluorescence-activated single cell sorting of human embryonic stem cells. *Cloning Stem Cells*, **8**, 225–234.
25. Liu, Y.P., Dovzhenko, O.V., Garthwaite, M.A., Dambaeva, S.V., Durning, M., Pollastrini, L.M. and Golos, T.G. (2004) Maintenance of pluripotency in human embryonic stem cells stably over-expressing enhanced green fluorescent protein. *Stem Cells Dev.*, **13**, 636–645.
26. Gallagher, E.J., Lodge, P., Ansell, R. and McWhir, J. (2003) Isolation of murine embryonic stem and embryonic germ cells by selective ablation. *Transgenic Res.*, **12**, 451–460.
27. Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H. and Robson, P. (2005) Transcriptional regulation of *nanog* by OCT4 and SOX2. *J. Biol. Chem.*, **280**, 24731–24737.
28. Amit, M., Margulets, V., Segev, H., Shariki, K., Laevsky, I., Coleman, R. and Itskovitz-Eldor, J. (2003) Human feeder layers for human embryonic stem cells. *Biol. Reprod.*, **68**, 2150–2156.
29. Tan, S.M. and Dröge, P. (2005) Comparative analysis of sequence-specific DNA recombination systems in human embryonic stem cells. *Stem Cells*, **23**, 868–873.
30. Cowan, C.A., Klimanskaya, I., McMahon, J., Atienza, J., Witmyer, J., Zucker, J.P., Wang, S., Morton, C.C., McMahon, A.P. *et al.* (2004) Derivation of embryonic stem-cell lines from human blastocysts. *N. Engl. J. Med.*, **350**, 1353–1356.
31. Corona, T., Bao, Q., Christ, N., Schwartz, T., Li, J. and Dröge, P. (2003) Activation of site-specific DNA integration in human cells by a single chain integration host factor. *Nucleic Acids Res.*, **31**, 5140–5148.
32. Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R. *et al.* (2004) Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat. Biotechnol.*, **22**, 707–716.
33. Christ, N. and Dröge, P. (2002) Genetic manipulation of mouse embryonic stem cells by mutant λ integrase. *Genesis*, **32**, 203–208.
34. Schuldiner, M., Yanuka, O., Itskovitz-Eldor, J., Melton, D.A. and Benvenisty, N. (2000) Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. *Proc. Natl Acad. Sci. USA*, **97**, 11307–11312.
35. 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. *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947–956.
36. Zhao, C. and Meng, A. (2005) Sp1-like transcription factors are regulators of embryonic development in vertebrates. *Develop. Growth Differ.*, **47**, 201–211.
37. Yang, H.M., Do, H.J., Oh, J.H., Kim, J.H., Choi, S.Y., Cha, K.Y., Chung, H.M. and Kim, J.H. (2005) Characterization of putative cis-regulatory elements that control the transcriptional activity of the human *Oct4* promoter. *J. Cell Biochem.*, **96**, 821–830.
38. Wu, D.Y. and Yao, Z. (2006) Functional analysis of two Sp1/Sp3 binding sites in murine *Nanog* gene promoter. *Cell Res.*, **16**, 319–322.
39. Guan, K., Nayernia, K., Maier, L.S., Wagner, S., Dressel, R., Lee, J.H., Nolte, J., Wolf, F., Li, M. *et al.* (2006) Pluripotency of spermatogonial stem cells from adult mouse testis. *Nature*, **440**, 1199–1203.
40. Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**, 663–676.
41. Katkov, I.I., Kim, M.S., Bajpai, R., Altman, Y.S., Mercola, M., Loring, J.F., Tersikh, A.V., Snyder, E.Y. and Levine, F. (2006) Cryopreservation by slow cooling with DMSO diminished production of Oct-4 pluripotency marker in human embryonic stem cells. *Cryobiology*, **53**, 194–205.
42. Niwa, H., Miyazaki, J. and Smith, A.G. (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.*, **24**, 372–376.