Delineating the mechanisms behind fate decisions of human embryonic stem cells

Thomas François Allison

Submitted in fulfilment for the degree of doctor of philosophy

Department of Biomedical Sciences

June 2015

Acknowledgements

Firstly, I would like to thank Prof. Peter Andrews for giving me the opportunity to undertake a PhD in his laboratory. Having been able to work for someone as world-renowned, enthusiastic, thought provoking yet so approachable and down to earth has made my time in Sheffield an experience I won't forget. I want to thank Peter for his continued support; his patience and his tactful approach in convincing me certain experiments were just rubbish. Finally, I would like to thank Peter for the opportunities he has given me to travel, and the conferences he has funded me to attend. Peter has given me the best start in my research career I could have hoped for.

To my parents, Sylvie and Robert Allison, and to my sisters Laure Soeters and Claire Blowfield, thank you for your support and love throughout my education, and the years I spent away from England. The experiences I have gained have changed my life and it would not have been possible without you. Thank you for your encouragement and understanding throughout my PhD and for being there at my most difficult times. Thanks also to my brothers-in-law Matthew Blowfield and Ties Soeters for their support and banter. We are lucky to have you in our family.

A special mention to Dr. Paul. Gokhale for all his support, discussions and ideas throughout my project, and for his continued enthusiasm even when mine was wavering!

A big thanks goes to all the members of the PWA lab, of which there are too many to mention individually. I would especially like to thank Dr. Mark Jones for the ridiculous amount of cell sorts he has done for me, Dr. Ivana. Barbaric for her support and advice and Dr. Veronica Biga for patiently teaching me about stats. Thanks to my best PhD buddy, Jim Hackland, for his support in and out of work. I also want to thank the founding members of the Lad-boratory, Tom Frith, Jim Hackland, Dylan Stavish, Oliver Thompson, and David Preskey. These people have made the past four years so much fun, and at times we have even had productive discussions about science. I miss you already! I also want to thank Katie Amps (aka Ampsy) for her insane optimism and equally insane organisation of the 2012 Labolympic Games. Who thought science could be so fun?

Finally, to my best friend, Inderpal Dosanjh. Thank you for the support and encouragement you have given me throughout the years. Thank you for the countless good times we've had together, in Australia, Miami and the many many raves. We will never grow up and we will never stop raving, even if it means you moving to Cali with me!!

Abstract

Human embryonic stem (ES) cells are derived from the inner cell mass of blastocyst stage embryos. Once explanted and cultured *in vitro* under appropriate conditions, human ES cells retain pluripotency (i.e. capacity to differentiate into all somatic cell types) and acquire the ability to self-renew indefinitely. These two properties of human ES cells make them an invaluable resource for developmental biology, cell replacement therapies, drug development and toxicology screening.

In order to exploit these unique cells and to better understand human development, it is imperative to understand the mechanisms behind the specification of somatic cell types. Much work has been conducted on monolayer formats to delineate signalling and gene expression networks responsible for lineage specific differentiation, however less focus has been on the use of embryoid bodies as a more representative model of *in vivo* differentiation. In this study we develop a differentiation assay to better recapitulate embryonic development, which we also show as a useful model for predictive toxicology. Within the assay, however, we found persistently heterogeneous differentiation. To better understand how hESCs make lineage decisions, we went back to interrogate heterogeneity within the stem cell compartment, and show that discreet, but functional heterogeneity biases cells to particular fates. Finally, we shed light onto a potential mechanism through which heterogeneity arises, which could offer a platform for future work to homogenise stem cells thus resulting in uniform, controlled differentiation.

List of Abbreviations

Abbreviation	Meaning
3+6-	SSEA3(+) GATA6(-)
3+6L	SSEA3(+) GATA6(Low)
3+6H	SSEA3(+) GATA6(High)
3-6+	SSEA3(-) GATA6 (+)
APEL	Albumin Polyvinylalcohol Essential Lipids
BMP	Bone morphogenetic protein
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
EB	Embryoid body
EC cell	Embryonal carcinoma cell
ESC	Embryonic stem cell
EST	Embryonic stem cell test
FACS	Fluorescent activated cell sorting
FCS	Fetal calf serum
FGF	Fibroblast growth factor
GDF	Growth and differentiation factor
GFP	Green fluorescent protein
CHO	Chinese hamster ovary cells
GRN	Gene regulatory network
hESC	Human embryonic stem cell
hPSC	Human pluripotent stem cell
IGF	Insulin-like growth factor
iPS	Induced pluripotent stem cell
ICM	Inner cell mass
KO DMEM	Knock-out dulbecco's modified eagle medium
LIF	Leukemic inhibitory factor
MEF	Mouse embryonic fibroblast
mESC	Mouse embryonic stem cell
MEF	Mouse embryonic fibroblast
PBS	Phosphate buffered saline
PDK1	Phosphoinositide-dependent kinase-1
PDMS	Polydimethylsiloxane
PrE	Primitive endoderm
PVA	Polyvinyl Alcohol
STAT3	Signal transducers and activators of transcription 3
TGF	Transforming growth factor
UPL	Universal probe library
VEGF	Vascular endothelial growth factor
Y-27632	Rho kinase inhibitor

Table of Figures

Publications

Allison TF, Powles-Glover N, Biga V, Barbaric I. 2015. Human pluripotent stem cells as tools for high-throughput and high-content screening in drug discovery. International Journal of High Throughput Screening 5:1–13.

Jones AJ, Gokhale PJ, Allison TF, Sampson B, Athwal S, Grant S, Andrews PW, Allen ND, Patrick Case C. 2015. Evidence for bystander signalling between human trophoblast cells and human embryonic stem cells. Sci Rep 5:11694

Title Page

Acknowledgements

Abstract

List of Abbreviations

Table of Figures

Publications

Contents Page

1 General Introduction

1.1 Stem cells and pluripotency

The human body consists of hugely diverse and specialised cell types that have individual roles depending upon their location. The intricate actions and interactions of and between different cell types allows for the proper development and functioning of the human body. These markedly different cell types however, all arise from a single cell known as the zygote, which contains a single genome. Once the zygote has been established via sperm and egg fusing in sexual reproduction, it undergoes cleavage events to form the morula. The totipotent morula continues to divide and undergoes compaction before reaching the blastocyst stage, which marks the first point of lineage segregation. Here outer cells of the morula differentiate towards the trophoblastic lineage, whilst the inner cells will form the inner cell mass (ICM). It is the ICM that will go on to form all somatic and germ cell lineages, and are also the cells that can be isolated *in vitro* to generate human embryonic stem cell (hESC) lines.

hESCs have unique characteristics, specifically, they can self-renew indefinitely whilst still retaining the ability to differentiate into all cell types of the adult. For this reason, hESCs hold great potential in applications such as regenerative medicine, disease modelling and toxicology. Nevertheless, substantial progress towards these applications has only occurred relatively recently due to the ethical and practical issues of research with human embryos. The benefits of hESCs are particularly apparent within the area of toxicology and drug development. Classified as a 'near-term' use of ES cell technology by the consortium 'Stem Cells for Safer Medicine' (SC4SM) in the recent Pattison Report, it is believed that, in principle, hESCs could be used to faithfully predict drug toxicity (Rubin 2008). With an understanding of the signalling pathways that are active or inhibited in cell specification, protocols now exist for the differentiation of hESC to toxicologically relevant cell types, including cardiomyoctes and hepatocytes. These cell types are particularly important as

cardiotoxicity and hepatoxicity are often the cause of drug failure, and are difficult to source (Denning & Anderson 2008; Greenhough et al. 2010). There continues to be, however, a lack of high-throughput *in vitro* assays to predict the toxic potential of new drugs.

The use of hESCs in therapeutic and toxicological applications, however, relies on our ability to efficiently and reproducibly differentiate cells to desired cell types. Before permission was granted in 1998 for the use of human embryos for cell line derivation, some of the knowledge that was gained with respect to pluripotency, differentiation and the mechanisms behind early development were obtained from embryonal carcinoma (EC) cells, the malignant counterpart to embryonic stem cells, and animal models. The characterisation and techniques used to derive mouse EC, ESC and human EC cells *in vitro,* then laid the foundations for the successful derivation of hESCs. Since their derivation there have been a large number of reports detailing the directed differentiation of hESCs, but it is becoming increasingly apparent that the intricacy of fate decisions extends beyond our current knowledge. A reoccurring feature that continues to plague directed differentiation is the lack of efficiency and the heterogeneity of differentiated cell types that result. This points to more fundamental mechanisms through which cells respond to environmental cues and acquire appropriate fates. Accumulating evidence demonstrates that hESCs show a high degree of heterogeneity in stem cell gene expression and in signalling pathways, and this heterogeneity appears to have functional consequences for their behaviour (Blauwkamp et al. 2012; Tonge et al. 2011; Fischer et al. 2010). Although it has been reported that hESCs can also express low levels of lineage associated genes, there are very few reports on the functional consequences of this heterogeneity. Furthermore, there are no reports on the functional characterisation at the single cell level.

Within this study, we developed a differentiation assay that we show can be used as an effective tool for predictive toxicology. Within the assay, as has been previously reported, we did see underlying heterogeneity in the differentiation of hESCs. We therefore went back to investigate further the mechanisms of

differentiation, by determining whether the heterogeneity that has been observed in stem cells was having functional consequences on lineage specification. Using a reporter line for the endoderm specific gene *GATA6* we identify, characterise and provide a mechanistic insight into a sub-fraction of cells in culture that exhibit endoderm differentiation bias. Importantly, we show this functional bias at a single cell level. We conclude that hESC heterogeneity may be responsible, at least in part, for the non-uniform behaviour of hESCs during differentiation.

1.2 Research leading to the derivation of hESCs

1.2.1 Mouse embryonal carcinoma cells

Teratomas, a benign tumour, and teratocarcinomas, the malignant counterpart, have long fascinated scientists, as early as the 1800's, as these tumours exhibit features of embryonic development and consist of highly differentiated cell types from each of the three primary germ layers. Early progress within the field of development came from research in 1954 by Stevens et al who discovered a high incidence of testicular teratocarcinomas in the mouse 129 strain (Stevens & Little 1954). These teratocarcinomas contained elements of the three primary germ layers, were malignant, and also contained clonogenic, pluripotent embryonal carcinoma cells (EC), later described as cancer stem cells (Kleinsmith & Pierce 1964). The first instance of the stable maintenance of cancer stem cells or multipotent EC cell lines *in vitro* was then reported (Finch & Ephrussi 1967). The subline 402 AIII from a teratocarcinoma from the testis of a strain 129 mouse, originally passaged *in vivo,* were seeded onto a layer of irradiated mouse embryonic fibroblasts (MEFs) in DMEM and 15% fetal calf serum. It was found that MEFs appeared important in the retention of pluripotency and were presumed to provide a critical nutrient or trophic factor aiding the pluripotent state of these EC cells. This break through allowed the derivation of further EC cell lines, which were shown to have the ability to self-renew and the capacity for multi-lineage differentiation, and were thought to be analogous to cells of the inner cell mass (ICM) of the blastocyst. Indeed, these EC cells were shown to be functionally equivalent to cells of the ICM through the formation of chimeric mice following blastocyst injection (Brinster 1974). *In vitro,* their diverse differentiation capacity became apparent through embryoid body (EB) differentiation, which led to the generation of a wide variety of somatic cell types, further validating the notion that EC cells were indeed the malignant counterparts to cells of the ICM (G. R. Martin & Evans 1975). Furthermore, the EC line, F9, could be induced to differentiate through the use of retinoic acid leading to endodermal specification, showing that EC cells could

respond to exogenous signals that would bias their differentiated phenotype (Strickland & Mahdavi 1978). Subsequent characterisation of F9 EC cells identified the expression of a unique surface marker, the F9 antigen, which was also found to be expressed on cleavage stage embryos but not differentiated teratoma cell types (Artzt et al. 1973). The advent of monoclonal antibodies then allowed the identification of another surface antigen, MC480, better known as SSEA-1, which had similar expression patterns to the F9 antigen (Solter & Knowles 1978), and proved instrumental in the isolation of mouse embryonic stem cells (ESCs). The extraordinary resemblance of mouse EC cells to cells of the mouse ICM then drove interest in isolating and characterising the human counterpart.

1.2.2 Human embryonal carcinoma cells

Similarly to the 129 mouse strain, teratomas are also a feature of human disease, which are most commonly manifested as benign ovarian tumours and dermoid cysts, although in rare cases they can be present as tumours in new-borns. Using an accumulation of techniques performed to derive mouse EC cells, human EC cells were isolated and successfully propagated *in vitro* (Andrews et al. 1980; Hogan et al. 1977; Andrews 1988). Subsequent characterisation of human EC lines, however, highlighted important differences between human and mouse. Developmental differences and differences in surface antigen expression indicated that human and mouse EC cells either corresponded to different embryonic cells, or that equivalent cells differed between species **(Table 1)** (Andrews et al. 1980). Nevertheless, although differences were apparent, pluripotency associated transcription factors *OCT4*, *SOX2* and *NANOG* were later found to be expressed in both EC and ES lines of both species (Schöler et al. 1989; Gubbay et al. 1990; Mitsui et al. 2003; I. Chambers et al. 2003).

Table. 1: Comparison of mouse and human EC cell surface antigen expression profiles

Human	Mouse
SSEA1-	SSEA1+
SSEA3+	SSEA3-
SSEA4+	SSEEA4-
TRA-1-60+	TRA-1-60-
GCTM2+	GCTM2-
THY1+	THY1-
MHC+	MHC-

Subsequent work on human EC lines, notably NT2/D1, then provided insights into mechanistic events of development *in vitro*. For example, the discovery that retinoic acid, a morphogen expressed during development, induced *HOX* gene expression in a concentration dependent manner giving rise to cell types indicative of the hind-brain region at low concentrations (10^{-8}M), and more spinal cord and limb bud cell types at higher concentrations (10^{-6}M) (Simeone et al. 1990). Furthermore, the use of NT2/D1 was the first instance of the use of pluripotent cells in regenerative medicine whereby NT2/D1 derived neurones were transplanted into stroke patients (Kondziolka et al. 2000). Although useful, EC cells did have problems when studying development. Cells often showed a restricted differentiation potential, and in some cases, nullipotency. A general worry was that these cells may not truly behave in the same way as cells of the ICM and thus made it difficult to interpret the exact mechanisms responsible for self-renewal and differentiation. For these reasons, work was focussed on deriving cells directly from the ICM.

1.2.3 Mouse embryonic stem cells

The derivation of embryonic stem cells previous to 1981 had proven to be mostly unsuccessful, although cells resembling those of the ICM had been isolated transiently. The cause for the degradation of these embryo-derived cells was postulated by Kaufmann *et al* to be due to the exact embryo stage at which cells were derived, the number of cells explanted and tissue culture conditions (Evans & Kaufman 1981). Successful derivation was achieved by explanting day 2.5 mouse blastocysts into petri dishes, where they saw the generation of giant trophoblast cells and large egg cylinder-like structures deriving from cells of the ICM. These ICM cells were picked onto inactivated mouse embryonic fibroblast cells and gave rise to morphologically similar cells to established EC lines. Importantly, these proliferating ICM cells had normal 40XX or 40XY karyotypes (Evans & Kaufman 1981). Similar experiments were conducted by Gail Martin, who used teratocarcinoma conditioned media to maintain these ICM cells and who coined the term 'embryonic stem cells' due to their direct derivation from the mouse blastocyst (G. R. Martin 1981). The derivation of embryonic stem cells led to an explosion in research within the field of mammalian development, as these ICM-derived cells provided an *in vitro* tool to interrogate the genetics and signalling behind developmental processes.

The advances in molecular genetics during the 1980s allowed genetic manipulation through insertional mutagenesis via retroviral vectors (Evans et al. 1985) and the ability to target specific genes by using homologous recombination (Doetschman et al. 1987; Thomas & Capecchi 1987). The interrogation of genes involved in pluripotency subsequently identified core stem cell associated transcription factors, *OCT4* (Schöler et al. 1989), *SOX2* (Gubbay et al. 1990) and *NANOG* (I. Chambers et al. 2003), to be important for the propagation of pluripotent ES cells.

Mechanistic insight into pluripotency then identified LIF (leukemic inhibitory factor), a member of the interleukin 6 (IL6) family, as a key factor in the

maintenance of pluripotency in mESCs *in vitro* (A. G. Smith et al. 1988). In mESCs, LIF binds to a heterodimeric cell surface receptor consisting of the LIF receptor (LIFRβ) and the glycoprotein, gp130, resulting in the activation of JAKs (receptor associated janus kinases). This in turn phosphorylates STAT3 (signal transducers and activators of transcription 3), which translocates into the nucleus, resulting in target gene transcription. An important target includes GABP (GA-binding protein) which directly activates the expression of OCT3/4 thus reinforcing the pluripotent state (Kinoshita et al. 2007). Although the mainstay of experimental embryology, the mouse does show substantial differences in early developmental structures, including the placenta, extraembryonic membranes and egg cylinder. The techniques that had been learnt from the study of mouse and human EC and mouse ES cells then allowed the derivation of hESCs, which offered a better platform in which to study the development and function of these tissues within the human context.

1.3 Human embryonic stem cells

In 1998, Thomson *et al* successfully derived human embryonic stem cell (hESC) lines using MEF feeder cells. When characterised, hESCs showed the same long-term self-renewal and differentiation capacities as mESCs, but similarly to primate ESCs and human EC cells, showed morphological differences. Human cells grew as flatter colonies with well-defined borders, whereas mESCs were more tightly packed with irregular borders (Thomson et al. 1998). These hESC lines, (H1, H7, H9, H13, & H14) expressed the same markers as human EC cells as well as non-human primate ES lines, including SSEA-3, TRA-1-81, TRA-1-60, SSEA4 and ALP but similarly to human EC cells, lacked the expression of SSEA-1 found on mESCs (Thomson et al. 1998). Subsequent research then focussed on understanding the exact signalling and gene expression networks that governed self-renewal and differentiation of hESCs. This led to the finding that hESCs do not rely on the same signalling pathways for self-renewal as mESCs. hESCs do not respond to LIF and thus LIF does not confer self-renewal (Thomson et al. 1998). Additionally, BMP signalling which confers self-renewal in mESCs

actually is a potent inducer of differentiation in hESCs (Xu et al. 2002). Mechanistic differences between cells of the two species, likely due to differing embryological stages, brought into focus the relevance of using hESCs to study human development. As the field progressed, a deeper understanding of the signalling pathways involved in lineage specification brought the reality of using these cells for therapeutic and toxicological applications a step closer. In 2005, the Pattison report described ES technology in toxicology as a 'near-term' use, but despite progress, there is still a significant and urgent need for high-throughput, reliable assays for predictive toxicology.

1.4 Human embryonic stem cells in toxicology

Drug discovery is an expensive and often lengthy process which is hampered by ~90% attrition rates of new drug candidates (Hay et al. 2014). Current programmes in the identification of new drug candidates employ both *in vivo* and *in vitro* methods, using primary and transformed human cell lines and animal models prior to progression into man. Although valuable, current estimates say that the methods employed fail to detect adverse effects in up to 30% of new drugs trialled (H. Olson et al. 2000). Examples of the dangers of failed drug safety predictions include the non-steroidal anti-inflammatory (NSAID) drug Benoxaprofen which showed no adverse effects in health to rhesus monkeys, but caused severe renal and hepato-toxicity in humans (Brass 1993). An additional NSAID, Phenylbutazone, a drug routinely used in horses for pain relief also showed fatal liver and bone marrow disease in human patients (Benjamin et al. 1981). Cerivastatin, a cholesterol-lowering drug was shown to cause rhabdomyolosis in humans, which was not picked up in pre-clinical tests with rats, mice, minipigs, dogs or monkeys, unless administered at very high doses and was consequently deemed to be well tolerated in all species (Keutz & Schlüter 1998).

Reproductive toxicology is also an important aspect of pre-clinical testing to assess the potential toxic or teratogenic effects of drug candidates on the

developing foetus. Currently, studies are carried out which involve *in vivo* animal studies using a rodent (usually rat), and non-rodent species (usually rabbit) with preterm evaluation (ICH (M3), OPPTS 870.3700, OECD guideline 414). Compounds are administered during early organogenesis between implantation and the closure of the hard palate. Assessments are then made at the end of gestation for developmental external, visceral and skeletal endpoints on the foetus (Daston 2007). Although incredibly valuable, the use of model organisms does mean that observations are extrapolated to the case of the human embryo, which can show poor correlations in up to 60% of cases (Gottmann et al. 2001), the most infamous case being that of the drug Thalidomide. A drug administered to alleviate morning sickness, Thalidomide, despite not showing adverse effects on several rodent species, was responsible for extensive teratogenic effects on the human embryo during prenatal development (Brent 1964). Eventually, and rather too late, one strain of rabbit (New Zealand white) was found to be sensitive to the drug, and only upon extremely high doses were effects apparent in other species (Bailey et al. 2005).

As well as the problem of species differences for toxicological testing, the speed of data collection, the cost of such procedures and importantly the vast numbers of animals used in pre-clinical toxicity make drug development vastly expensive and time consuming (Schumann 2010). To circumvent these problems, *in vitro* studies on human cells are performed on promising drug candidates. *In vitro* studies have several benefits, for example they use human cell lines and they allow for a high-throughput approach that can be employed to screen thousands of compounds for adverse effects on cell behaviour. If employed early on within the drug development process, potential teratogenic or toxic compounds can be filtered out, reducing the amount of money spent on the development of a dangerous drug. Most *in vitro* studies to date have been performed on primary or transformed cell lines, which have significant drawbacks. Primary lines are difficult to source as they have to be taken from the tissue of origin and they have a restricted capacity for proliferation (Hayflick 1965). The development of transformed lines circumvented the problem of senescence, however these lines

often differed significantly from the cell or tissue of origin, resulting in potentially unreliable data.

To increase the speed and quantity of data, murine ESCs were employed as a model for developmental toxicology. Their ability to proliferate as an undifferentiated cell, as well as their ability to differentiate in culture led to the development of the EST (embryonic stem cell test). This method is based on the analysis of three toxicological endpoints after 10 days of exposure to the compound of interest. The first is the beating morphology of mESC derived cardiomyocytes, secondly the cytotoxicity of differentiating mESCs and finally the cytotoxicity of fully differentiated mouse fibroblasts, 3T3, using the MTT assay (Spielmann 1997). Successfully validated by the European centre for the validation of alternative methods (ECVAM) as an assay for investigating reproductive toxicology, the EST is still in use today (Genschow et al. 2004). Nevertheless, the development of humanised in vitro models is still a necessity. hESCs have generated excitement within the area of toxicology, but a lack of a complete understanding of the signalling pathways and gene regulatory networks (GRNs) in cell specification has hindered their use.

1.5 Signalling pathways in the self-renewal and differentiation of human embryonic stem cells

The maintenance of pluripotency and the differentiation of hESCs involves the activation or repression of signalling cascades resulting in the expression of transcription factors enforcing a particular lineage. The importance of signalling in hESCs was demonstrated by a study, which concluded that up to 17% of genes enriched in hESCs were involved in signal transduction and regulation. Pathways found to be important included the FGF, WNT, LIF and TGFβ pathways, with the expression of both agonists and antagonists directing cells to specific fates throughout development (Brandenberger et al. 2004). The FGF, WNT, LIF and TGFβ pathways all play important roles in pluripotency and early specification of hESCs to each of the three primary germ layers.

1.5.1 TGFβ signalling in pluripotency

The TGFβ signalling pathway is a highly conserved pathway involved in many cellular functions, including cell growth, differentiation and apoptosis as well as hPSC maintenance, and functions during both embryogenesis and within the adult organism. The TGFβ superfamily of ligands consists of 2 sub-groups. Firstly, TGFβs including Activin A, Nodal and TGFβ, which signal through SMAD2/3 proteins, via ALK4, ALK5 and ALK7 receptors (TGFBR1 & ACVR1C) (Schier 2003)(James et al. 2005); and secondly BMPs and GDFs through the SMAD1/5/8 proteins via type I receptors ALK1, ALK2, ALK3 and ALK6 (ACVERL1, ACVER1, BMPR1A and BMPR1B respectively) (Kretzschmar et al. 1997; Kingsley. 1994; James et al. 2005). Mechanistically, the TGFβ ligand binds to a type II receptor (TGF-βRII) on the cell surface, which then recruits the TGFβ type I receptor (TGF-βRI). The presence of the type II receptor is imperative as TGF-βRI is unable to bind the TGFβ ligand in its absence (Heldin et al. 1997). The activation of TGF-βRI then induces the rapid phosphorylation of SMAD2, and the closely related protein SMAD3, at the C-terminal serine residue, causing their translocation to the nucleus (Macías-Silva et al. 1996). The phosphorylated SMAD2/3 proteins synergise through functional and physical interactions with the co-factor SMAD, SMAD4. This subsequently leads to the activation of downstream targets within the nucleus (Nakao et al. 1997).

In hESCs, TGFβs of the SMAD2/3 branch have roles in maintaining the undifferentiated, pluripotent state through the sustained expression of stem cell transcription factors, for example *NANOG*, which is directly activated by TGFβ signalling (Xu et al. 2008) (Vallier et al. 2009). Studies have shown that activation of the TGFβ/Activin/Nodal pathway through SMAD2/3 with the recombinant protein Activin A is supportive of the undifferentiated state (James et al. 2005). The importance of TGFβ signalling for the pluripotent state is further demonstrated through its inhibition with chemical inhibitors, such as SB431542, and the consequential down-regulation of *NANOG* and loss of pluripotency (Xu et al. 2008). There are conflicting reports describing the ability

of Activin A stimulated TGFβ signalling alone to maintain pluripotency, or whether the presence of FGF is required (Beattie et al. 2005; Vallier et al. 2005). These conflicting reports are likely culture context dependent, as matrices and media are not consistent between reports.

1.5.2 TGFβ signalling in differentiation

Studied extensively in *xenopus*, TGFβ signalling has been found to be central to mesoderm formation and patterning. The first instance of the cloning of a TGFβ receptor was the Activin type II receptor. The cloning of a truncated dominant negative form of this receptor led to defects at the gastrula stage in the *xenopus* embryo and reduced the levels of mesoderm whilst enhancing the levels of ectoderm tissue specification (Hemmati-Brivanlou & Melton 1992). The mutations of this receptor also led to the blocking of BMP signalling, demonstrating that the same receptor is able to bind to more than one ligand (Kessler & Melton 1994). SMAD2/3 through TGFβ signalling is also important for anterior-posterior patterning of the epiblast leading to the correct formation of the primitive streak during gastrulation, in the mouse embryo (Conlon et al. 1994). Double mutant mice for both SMAD2/3 proteins exhibit severe defects with a complete failure to induce mesoderm or enter gastrulation (Dunn et al. 2004). Additionally, there was a loss of pluripotent epiblast in null mice by E7.5 however extra-embryonic ectoderm was retained, implying that SMAD2/3 is necessary for the proper formation and retention of the ICM/epiblast (James et al. 2005).

In hESCs, D'Amour et al showed that in low serum conditions, high levels of recombinant Activin A was sufficient to induce up to 80% of SOX17+ cells in culture, reasoned to be definitive endoderm-like cells (D'Amour et al. 2005). The over-expression of Lefty and Cerberus, antagonists of TGFβ signalling, or the use of the pharmacological inhibitor SB431542, which blocks ALK4/5/7 in TGFβ signalling, leads to the promotion of neuronal specification to the

detriment of mesendoderm (Smith et al. 2008). The TGFβ pathway in mesendodermal specification therefore appears key, and it's function remains well conserved throughout species.

Subsequently, a paradox is established as the same signalling pathway is involved in two very opposite cellular phenotypes. This complexity can be addressed when combining a second signalling pathway, PI3K/Akt. The levels of PI3K/Akt influence cell fate, such that at high activation levels, TGFβ signalling cooperates to enforce the pro-self-renewal phenotype, whereas at low activation levels, WNT effectors are activated leading to differentiation, and mesendodermal specification (Singh et al. 2012).

1.5.3 BMP signalling in differentiation

The second sub-group of the TGFβ family are the BMPs and GDFs. The mechanism of action for the BMP pathway is similar to that of the TGFβ, but BMP ligands are able to bind both type I and type II receptors, albeit only with low affinity. When both receptor types are present high affinity binding occurs to induce signal transduction (F. Liu et al. 1995). The binding of BMP to its receptors leads to the phosphorylation of SMAD1/5/8 proteins that, similarly to SMAD2/3, associate and form a complex with SMAD4. The phosphorylated complex then translocates to the nucleus to activate downstream targets. The activation of the BMP pathway through SMAD1/5 proteins with the supplementation of recombinant BMP4 in both serum and serum-free culture medium appears to differentiate hESCs towards the trophoblast lineage (Xu et al. 2002). Conversely, the BMP ligand inhibitor Noggin or pharmacological inhibitors such as Dorsomorphin can be used in conjunction with FGF to inhibit differentiation, and maintain pluripotency. Interestingly, the role of BMPs within the mouse context actually confers self-renewal *in vitro*, highlighting important species differences during early development, possibly due to different developmental timings of ES cell derivation (Ying et al. 2003).

In combination with the TGFβ/Activin/Nodal pathway, BMP4 is shown to be a driver of differentiation towards the mesendoderm lineage. BMP4 functions to surpass the self-renewal effects of the TGFβ/Activin/Nodal pathway by down-regulating pluripotency genes, in particular *SOX2*, which has SMAD1 binding sites, consequently pushing cells to a mesendodermal fate (Teo et al. 2012). It has also been shown that BMP4 signalling activates genes associated with the WNT signalling pathway, including WNT3, which upregulates genes associated with mesendodermal and primitive streak lineages (Kurek et al. 2015). The alternate lineages of trophectoderm and mesendoderm are thought to arise from the levels of WNT signalling activation within cells, which would explain how BMP4 can be implicated in two opposing lineages (Kurek et al. 2015).

1.5.4 FGF signalling in pluripotency

FGFs are a family of growth factors that are involved in angiogenesis, wound healing and early embryogenesis. The most commonly used FGF in hESC culture, basic FGF (FGF2), allows the maintenance of pluripotency and confers indefinite self-renewal when in combination with agonists of the TGFβ pathway (Vallier, Alexander et al. 2005). FGF signalling in hESCs functions through two distinct pathways. Firstly is the MEK/ERK pathway. Considerable confusion surrounds the role of ERK signaling in hESCs as shown by a number of conflicting reports. Reports exist which advocate ERK in the maintenance of pluripotency (Li et al. 2007; Armstrong et al. 2006), and others suggest it has roles in promoting differentiation (Na et al. 2010). Secondly is the PI3K pathway, which has important roles in pluripotency by functioning in conjunction with the TGFβ/Activin/Nodal pathway. Inhibition of PI3K using specific inhibitors such as LY294002 results in loss of stem cell transcription factor expression and consequently differentiation. Supportive of the observation that ERK signalling results in hESC differentiation, the interplay of PI3K and ERK signalling is believed to control WNT signalling and consequently self-renewal versus

differentiation. The inhibition of PI3K signalling leads to the activation of ERK and consequently the inactivation of GSK3β, the accumulation of β-catenin and mesendodermal gene expression. The inhibition of PI3K and ERK reverses this effect, maintaining pluripotency (Singh et al. 2012). The interplay of PI3K and ERK signalling therefore appears to control the pluripotent and differentiated states, whereby PI3K functions upstream of ERK and functions as a master regulator (Chen et al. 2012; Eiselleova et al. 2009). FGF2 signalling also has been shown to increase the survival of hESCs upon dissociation and in the presence of oxidative stress. The exact mechanism through which this occurs has not been elucidated (Eiselleova et al. 2009).

1.5.5 FGF signalling in differentiation

As well as its role in pluripotency, FGF signalling also plays important roles in early gastrulation. In *xenopus*, the expression of a dominant negative mutant of the FGF receptor leads to normal cleavage stage embryos, but the development of gross morphological abnormalities during gastrulation. Additionally, *xenopus* failed to fully form tails later in development (Amaya et al. 1991). Within the human context, it has been found that FGF2 is required for definitive endoderm specification and the dose of FGF affects the eventual lineage that cells will acquire. High concentrations of FGF2 appear to prevent hepatocyte differentiation, whereas moderate levels induce a pancreatic fate in hESCs. (Morrison et al. 2008, Ameri et al. 2010) FGF signalling through the MEK/ERK pathway has also been implicated in directing BMP4 induced differentiation to the mesendodermal lineage through the maintenance of *NANOG* expression (Yu et al. 2011). Genetic knock-out of the *Fgfr-1* gene in mice results in aberrant patterning of mesoderm during gastrulation, such that somites are never generated (Yamaguchi et al. 1994). Within the human context, blocking FGF signalling in the presence of BMP4 with the inhibitor SU5402 during early specification of hESCs results in the loss of mesendoderm specification in favour of the trophectoderm and primitive endoderm lineages (D'Amour et al.).

1.5.6 WNT signalling in differentiation

WNT signalling functions through 3 distinct pathways, however the most intensively researched in hESCs is the β-catenin (canonical) pathway. In the absence of WNT signalling activation, β-catenin is phosphorylated by GSK3β causing ubiquitination and degradation. However, when WNTs (glycoproteins) bind to the trans-membrane-spanning receptor frizzled, GBP/Frat-1 is recruited to displace GSK3β from axin which results in the elimination of its ability to phosphorylate β-catenin (Huelsken and Behrens 2002). Stabilised β-catenin enters the nucleus and associates with TCF-LEF, which leads to the transcription of target genes. As well as target gene transcription TCF also associates with SMAD4, possibly linking the BMP/TGFβ signalling pathways with WNT signalling (Huelsken and Behrens 2002). Despite a wealth of biochemical information on WNT signalling, its exact role in hESCs still remains elusive. The stabilisation of mESCs in serum free conditions (2i) requires inhibition of MEK1/2 and GSK3β (activation of WNT signalling), but the same conditions applied to hESCs causes rapid differentiation (Ying et al. 2008). Singly, the activation of WNT signalling in hESCs is generally found to cause differentiation (Davidson et al. 2012), which would correlate with the observation that active WNT signalling in hESCs results in higher expression levels of mesendodermal markers and consequently a bias towards mesendoderm differentiation (Blauwkamp et al. 2012). The synergistic nature of the WNT and BMP pathways has also implied a role for WNT signalling as an inducer for differentiation. In addition, the use of IWP-2, a WNT inhibitor, allows the long-term self-renewal of hESCs which actually appear more uniform (Kurek et al. 2015). In contrast, some reports have described the addition of GSK3β inhibitors resulting in self-renewal, and propagation of pluripotent cells (Sato et al. 2003). These contradictory observations are most likely culture condition and context dependent, highlighted by the necessity to inhibit GSK3β to activate WNT signalling for the derivation of naïve hESCs (Chan et al. 2013; Gafni et al. 2013; Takashima et al. 2014). Therefore, WNT signalling appears to have

different role in lineage decisions of hESCs depending upon the activation or repression status of further signalling pathways.

1.5.7 Signalling inhibition for neural specification

The generation of ectoderm also involves the aforementioned signalling pathways but neural specification requires their inhibition. Initial studies in *xenopus* demonstrated that inhibition of the BMP pathway with recombinant proteins such as Noggin were imperative for neural induction (Smith & Harland 1992, Sasai et al. 1994). Neural induction through BMP inhibition was later confirmed in mammalian cells and was found to induce neural differentiation in hESCs (Lee et al. 2007). The discovery of chemical inhibitors, including Dorsomorphin then provided a cost effective way to potently inhibit the BMP pathway and further improve neural specification in hESCs (Morizane et al. 2011). Nevertheless, the inhibition of BMP signalling alone did not prove to be that efficient for neural specification. Efficient neural differentiation also required the inhibition of the TGFβ/Activin/Nodal pathway, which was achieved through the use of the chemical inhibitor SB431542 (Smith et al. 2008). Using a combination of Noggin and SB431542 Chambers et al demonstrated more efficient generation of PAX6(+) neural cells than either inhibitor singly. The mechanism through which neural specification occurs by blocking both pathways is thought to be three-fold. Firstly, the TGFβ/Nodal/Activin pathway directly activates NANOG expression, which enforces pluripotency but also functions to block neuroectoderm and neural crest formation during early differentiation. Inhibition would therefore destabilise the NANOG mediated pluripotency network, and the block on neural cell types (Xu et al. 2008; Wang et al. 2012). Secondly, blocking BMP signalling would prevent BMP induced trophoblast differentiation (Xu et al. 2002). Finally is the suppression of endoderm and mesoderm fates through the inhibition of both TGFβ/Nodal/Activin and BMP signalling which have been shown to act in combination for mesendoderm specification (D'Amour et al. 2005; Chambers et al. 2009).

With the knowledge gained with respect to signalling pathways that induce particular lineages, differentiation protocols generally involve the addition of recombinant proteins to activate, or chemical inhibitors to block combinations of these pathways to derive cell types of interest. It is becoming increasingly apparent, however, that the uniform conditions that are applied to a population of cells to induce differentiation do not elicit uniform responses in individual hESCs. This implies, and as mounting evidence demonstrates, that hESCs cannot represent one single entity. hESCs appear to exist as a heterogeneous population of cells in culture and accumulating evidence now shows that this heterogeneity impinges on how a particular cell responds to extrinsic cues. Differences in signalling, gene expression and metabolism raise the possibility that a population of hESCs represent subtly different pluripotent cell types.

1.6 Cell states, sub-states and heterogeneity

Broadly speaking, hESCs have four options that will dictate their fates. Self-renewal, differentiation and lineage specification, death and finally quiescence (Enver et al. 2009). Furthermore, these four options are further complicated, for example in the context of differentiation and lineage specification, as hESCs have at least three choices in germ layer specification. Each of these different fates can be generally regarded as cell states and it is generally accepted that the transition to any of these states mentioned above is largely governed by transcription factors. This can be demonstrated during development when cells transition from a pluripotent state to a restricted state, passing through stable gene expression networks (Enver et al. 2009). Some of the earliest work in cellular behaviour and the depiction of cell states was introduced by the 'canalization of development' which involves cells rolling down bifurcating channels on a hillside to a final, fixed destination (Waddington 1957). This model was further developed from the idea that cells, rather than rolling smoothly down the hillside, enter stable or meta-stable states along the path of differentiation to encompass transitory intermediate cell types which appear

during development, for example progenitor cells (Andrews 2002). Within the landscape context, these states are depicted as impressions along the route of differentiation and are mathematically known as attractors **(Fig. 1.1)**. Attractors represent mathematically stable solutions to which a dynamic system gravitates to over time. In the context of a biological system, an attractor would represent the most stable gene regulatory network for a particular cell. Thus different cell states, governed by stable gene regulatory networks appear throughout development and represent equilibrium points through which a dynamical system progresses (Enver et al. 2009). There is growing evidence to suggest that more discreet states exist, known as substates, which sub-fractionate overarching cell states. In the context of the pluripotent state for example, substates represent situations in which cells are all capable of multi-lineage differentiation but the probabilities of a cell choosing a particular fate is altered, such that a cell is more or less likely to follow one particular route over another (Enver et al. 2009). Conceptually therefore, within a cell state, a cell could sit and remain at the bottom of an attractor, which would represent the most stable 'fixed point' of that state. This would correspond to the stable expression of a network of genes applicable to that attractor. On the other hand, a cell could circulate around the attractor through oscillatory states, possibly due to the stochastic input of signals, consequently generating heterogeneity within the expression levels of genes within that network (Enver et al. 2005). Between individual cells this could represent differences in growth factor responsiveness and protein expression differences, resulting in discreet yet functional changes. This concept is apparent within adult stem cell systems such as the hematopoietic system, whereby single cells have been found to express genes of both the erythroid (β-globin) and myeloid (MPO) lineages prior to exclusive commitment to either the erythroid or granulocytic lineages respectively (Hu et al. 1997).

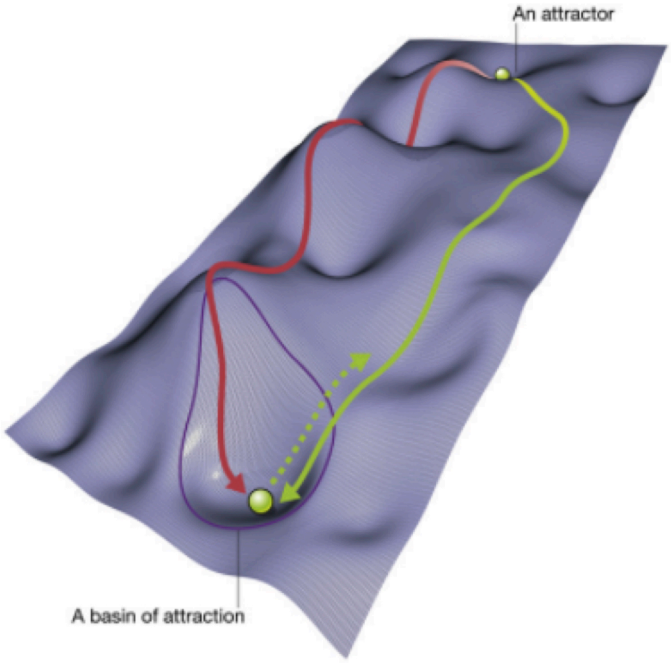
Within the pluripotent context, the most well known example of substates within the pluripotent state is that of naïve and primed ESCs. Thought to be analogous to cells of the late ICM, mouse naïve cells show increased cloning efficiencies and superior chimeric contributions than primed cells, thought to be

analogous to cells of the late epiblast (Tesar et al. 2007; Ying et al. 2008). Both subsets of cells are pluripotent and self-renew and thus can be regarded as substates of the overarching pluripotent state. Naïve-like cells have also been reported in hESCs, and exhibit similarities to their mouse counterparts (Takashima et al. 2014a; Gafni et al. 2013; Ware et al. 2014; Chan et al. 2013). Whilst the naïve and primed substates can be regarded as quite separate entities, due to the apparent need to physically reprogramme mEpi stem cells to a naïve state (Guo et al. 2009), accumulating evidence suggests the existence of more subtle forms of substates and heterogeneity that exist specifically within 'primed' hESCs that can alter the functional properties of a cell **(Fig. 1.2)**. Initially based on the heterogeneous expression patterns of hESC surface markers due to the convenience of cell staining, SSEA-3, TRA-1-60, CD9 and GCTM2 have been used to segregate functionally discreet subsets of cells in culture (Tonge et al. 2011; Enver et al. 2005; Laslett et al. 2007). These studies provided evidence that, within the stem cell compartment, there are continuous gradients of the expression of pluripotency genes. Additionally, the appearance of lineage gene expression alongside markers of pluripotency has been observed in hESCs (Laslett et al. 2007). Thus, within an attractor model cells with variable expression of pluripotency genes would represent cells in different oscillatory states, whereby the lower expression of pluripotency genes would correspond to cells residing in substates closer to the commitment barrier. This was proven functionally through clonogenic assays (Tonge et al. 2011).

Within the *in vivo* context it is also feasible that the plastic nature of ESCs is apparent within the mouse blastocyst. *NANOG* is a gene which was identified as having a central role in the pluripotent phenotype of ESCs (I. Chambers et al. 2003) but was later shown to be expressed heterogeneously within cells of the ICM of E3.5 mouse embryos (Chazaud et al. 2006). Further investigation of *NANOG in vitro* then demonstrated that it was not essential for pluripotency (I. Chambers et al. 2007), although *NANOG* negative cells did show a higher propensity to differentiate. It would therefore remain possible that some of these *NANOG*(-) cells within the ICM may still be pluripotent (Enver et al. 2009).

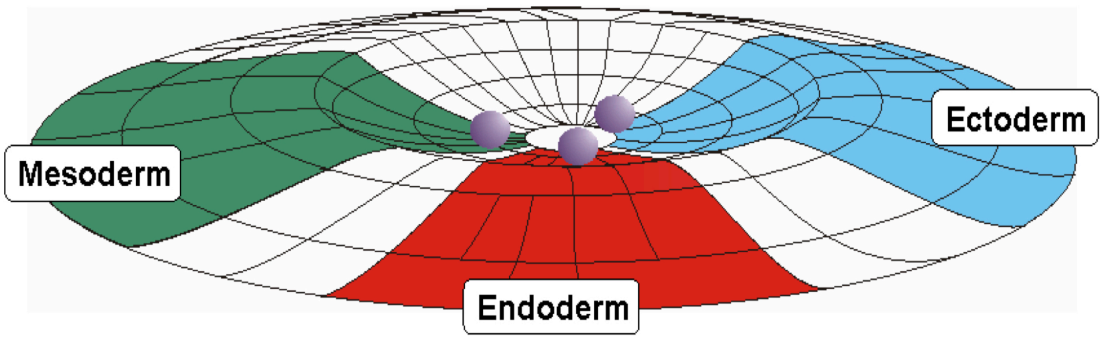
We therefore hypothesise that these oscillatory states within the stem cell compartment may be responsible, at least in part, for the non-uniform differentiation of hESCs *in vitro*. The identification and characterisation of functionally discreet substates would provide a better understanding of cell fate decisions of stem cells, and may allow for more efficient, homogeneous differentiation in culture.

Fig. 1.1: The Landscape Model of Cell Differentiation



Enver *et al* describe a model that describes the movement of cells throughout differentiation in a 3D plane. Depressions in the landscape represent observable states that cells transition through during differentiation. Mathematically, depressions represent attractors that are stable solutions to a set of mathematical equations that describe a dynamic system, such that the basin of an attractor represents the most stable point (purple line). The depth of the attractors correlates to their stability. Cells transition between attractors, not necessarily following the same pathways (red & green lines) to reach their final destination. Cells can equally move in reverse (dotted green line) for example in the case of trans-differentiation or re-programming. Furthermore, within each depression, cells can oscillate around the walls of the attractor, representing discreet substates (Enver et al. 2009).

Fig. 1.2: Sub-States Within The Pluripotent State



Conceptually, the valley represents the pluripotent state. Balls represent cells and if they reside within the valley are pluripotent. Balls can roll out of the valley and commit to differentiation, or they can reside at the basin of the attractor representing the most stable GRN of that state. Equally, through stochastic inputs of energy (signalling), balls can roll around the valley into sub-states. These sub-states may represent heightened sensitivity to signalling cues and/or gene expression changes. They therefore may represent functionally discreet sub-states altering the behaviour of cells. Furthermore, these substates may show specific biases to any of the possible lineages available to that cell.

2 Methods

2.1 Mouse embryonic fibroblast (MEF) preparation

Human ES cells were grown on mitomycin inactivated feeder cells from the CF1 mouse strain made in-house. MEFs were defrosted as P0 and cultured in DMEM/10% FCS. Cells were passaged and bulked to P4 before treatment with mitomycin C. Mitomycin C was diluted in DMEM/FCS (Sigma; M-4287) at 1µg/mL and added to MEFs for 2h. Cells were then washed in PBS, trypsinised, neutralised and counted. On average 2 x 10^6 cells were resuspended in 0.5mL of freeze media (80% DMEM, 10%FCS & 10% DMSO) and stored at -80^{O}C.

2.2 Human embryonic stem cell culture

Human embryonic stem cells (hESCs) were cultured in both feeder and feeder free formats, as specified by individual experimental procedures outlined below. In each case, hESCs were incubated at 37^{O}C in a 5% CO_2 humidified incubator.

2.2.1 MEF culture

3 mL of 0.1% Gelatin/PBS was added to T25 flasks and incubated for 30mins at room temperature. The gelatin was aspirated and MEFs were seeded at a density of 10,000 cells/cm^2 in DMEM/FCS. MEFs were incubated at 10% CO_2/37^{O}C overnight before use. Fresh MEFs were used for hESC culture wherever possible. On the day of passage, media from the MEF flasks was replaced with 2mL of hESC media and kept in a 5% CO_2/37^{O}C incubator for a minimum of 30 minutes to equilibrate. Media was aspirated from the hESCs and 1 mL/T25 of 1mg/mL collagenase IV added. Cells were incubated for 7 minutes at 37^{O}C with continual monitoring for lifting of colony edges using a microscope. After 7 min., the collagenase was removed and replaced by 3mL of fresh media.

Cells were either gently scraped using a 5mL stripette or beaded from the surface of the flask and split at a ratio of 1:3-1:4.

Table 1.1: Components of hESC culture media on MEF feeder cells

Component	Final Volume	Supplier	Cat #
Knockout DMEM	400mL	Life Technologies	10829-018
Knockout Serum Replacement	100mL	Life Technologies	10828010
1% NEAA	5mL	Life Technologies	1140-035
1mM L-Glutamine	5mL	Life Technologies	25030-081
0.1mM β-mercaptoethanol	1mL	Life Technologies	31350-010
4ng/mL human bFGF	500uL	RnD Systems	233-FB-01M

hESC medium is referred throughout this study as KO/SR, meaning the medium is made using knockout DMEM and 20% Knockout serum replacement.

2.2.2 MEF free cultures

hESCs were cultured on a combination of either Geltrex (LifeTech A1313302) & E8 (LifeTech A1517001) or Vitronectin (LifeTech A14701SA) & E8 (LifeTech A1517001). In both cases, the matrices were thawed on ice, and diluted 1:100 in either KO DMEM (Geltrex) or PBS (w/o Ca^{+}, Mg^{++} (Vitronectin)). Culture vessels were coated with 100µL/cm^2 of the diluted matrix for 1h at room temperature. These vessels could then be stored for up to 1 week at 4^0C. Cells were passaged by washing twice with PBS and then adding 100µL/cm^2 of ReleSr (StemCell Technologies). ReleSr was removed immediately so that only a thin film covered the cells. Cells were incubated at room temperature for 4 mins, and fresh E8 was added to neutralise. Cells were removed from the culture plastic by firmly rinsing with E8 using a P1000. Cells were not pipetted more than 3 times as this produced very small clumps of cells that were susceptible to death. Colonies were passaged at a ratio of 1:3-1:6. Enzymatic passaging was not used in feeder free conditions as it greatly affected cell attachment.

2.2.3 hESC culture with inhibitors

hESCs put under culture conditions with various inhibitors were grown on feeder layers in mTesr plus the appropriate inhibitors. MEF plates were prepared as described in **2.2.1**, however the density of MEF cells was increased to 40,000/cm^2. hESCs were passaged by adding 100uL/cm^2 of collagenase and incubating cells for 5 mins at 37^oC at 5% CO_2. Collagenase was removed and 100uL/cm^2 fresh media added before scraping using a 5mL p1000 tip. Cells were pipetted 2-3 times and split 1:10 every 3-4 days onto fresh MEFs. Alternatively, cells were dissociated using trypLE described in **2.5.** Cells were resuspended in fresh media and split 1:10 every 3-4 days onto fresh MEFs.

2.3 hESC freezing

hESCs were digested using collagenase IV for 7 minutes at 37^OC, 5% CO_2. The collagenase was removed and fresh hESC media or E8 added. The cells were either scraped or beaded off the vessel surface and centrifuged for 3 minutes at 1000rpm. The supernatant was aspirated and cells were resuspended gently in freeze media (60% hESC, 30% KOSR and 10% DMSO) and subsequently aliquotted into 1.5mL cryovials. Cryovials were placed into a Mr Frosty Isopropanol freezing container and placed into -80^OC for 24 hours. Vials were then transferred to liquid nitrogen after 24h for long-term storage.

2.4 hESC thawing

MEF culture vessels were prepared the night before thawing of hESCs (**See 2.1**). Fresh hESC media was added to these vessels and incubated at 37^OC at 5% CO_2

to equilibrate. Cells were removed from liquid nitrogen and transported on ice, before being placed into a 37^{O}C pre-warmed water-bath until the cell pellet was mostly defrosted. The cells were transferred to a 15mL falcon tube and centrifuged at 1000rpm for 3 minutes. The supernatant was aspirated and cells were re-suspended in fresh hESC media before seeding into the pre-equilibrated MEF flask. For improved viability, cells were seeded in 10μM Y-27632 (Abcam - ab120129).

2.5 Single cell dissociation

For experiments that required single cells, hESCs were dissociated using TrypLE (LifeTech; 12563-029). Media was aspirated from the cells and 100μL/cm^2 of 1X TrypLE added to the vessel. Cells were incubated for 2 minutes at 37^{O}C, 5% CO_2, removed and cells were dislodged by gently hitting the flask, then returned for a further minute to 37^{O}C, 5% CO_2. A 2:1 ratio of hESC media to TrypLE was added to neutralise, and cells were transferred to a 15mL falcon tube. This was then centrifuged at 1000 rpm for 3 mins, the supernatant aspirated and cells were re-suspended in hESC media for further downstream application.

2.6 Cell counting

hESCs were counted by dissociating to single cells (**see 2.5)** and resuspended in an appropriate dilution volume of hESC media. 10 μL of cell suspension was added to an improved neubauer haemocytometer and the four corner grids were counted. The cell count was determined by using the following formula

*Concentration [cell/mL] = № of cells counted ÷ № of grids counted * 10,000*
*Total cell number = Concentration * Total volume (mL)*

2.7 Screening on hESCs

In all screens, hESCs were dissociated to single cells and re-seeded on feeder free conditions. hESCs were dissociated as described in **2.5** and counted as described in **2.6.** For inhibitor screens, 10,000 cells / cm^2 were plated onto vitronectin in E8 in the presence of 10 μM Y-27632 and incubated at 37°C at 5% CO_2 overnight. The media was then changed to fresh E8 without Y-27632 but with the addition of inhibitors being screened. Media was changed on a daily basis for the duration of the assay.

2.8 Embryoid body differentiation

2.8.1 Plate set-up

Cells were either differentiated in neutral conditions that did not contain recombinant proteins or chemical inhibitors to direct differentiation, or under conditions permissive for endoderm, mesoderm and ectoderm conditions.
For neutral conditions, 50 μL of APEL media **(Table. 1.2)** (Ng et al. 2008) was pipetted into the inner 60 wells of a non-adherent U-shaped 96-well plate (Sigma Aldrich). The outer 36 wells were filled with 100μL of sterile PBS to humidify the plate and prevent evaporation. Plates were stored at 37° C at 5% CO_2 until ready to use. For directed differentiation, plates were filled with 50μL of APEL media containing the appropriate combination of human recombinant proteins and/or inhibitors at X2 concentration **(Table. 1.3)**.

2.8.2 Harvesting and seeding of cells

24h prior to EB formation, T25 flasks were coated with MEF conditioned media overnight. The next day 80-90% confluent hESCs were dissociated as described in **2.5** and resuspended in fresh hESC media without Y-27632 and seeded onto

the MEF coated flasks. Cells were incubated at 37°C for 30 mins, and the media was then changed containing 10 μM Y-27632. Cells were incubated at 37°C for a further hour before re-trypsinisation and neutralisation with APEL media, and then centrifuged. Cells were re-suspended in fresh APEL media, counted and cell dilution made such that there were 3000 cells per 50μL. 50 μL of cell suspension was then seeded into the inner 60 wells of a pre-prepared non-adherent U-shaped 96-well plate (Sigma Aldrich). Plates were centrifuged at 1000rpm for 3 mins to collect cells at the bottom of the wells. EBs were then incubated at 37°C at 5% CO_2 for the specified assay duration.

Table 1.2: Components of APEL medium

Component	Stock Solution	Component	Stock Solution
IMDM	1 x	Linoleic acid (100 ng/mL^{-1})	10,000 x
Ham's F-12	1 x	Linolenic acid (100 ng/mL^{-1})	10,000 x
Albucult	100 mg/mL^{-1}	SyntheChol (2.2 ug/mL^{-1})	7,200 x
Deionised BSA	100 mg/mL^{-1}	α-Monothioglycerol	13uL in 1mL IMDM
Polyvinylalcohol	5%	rhITS-Eth	100 x
PFHMII (5%)	1x	GlutamaxI (2mM)	200mM (100 x)

Table 1.3: Growth factors and chemical inhibitors used to differentiate hESCs in embryoid bodies

	Neutral	Endoderm	Mesoderm	Ectoderm
Activin A	X	100ng/mL	20ng/mL	X
BMP4	X	1ng/mL	20ng/mL	X
bFGF	X	X	X	100ng/mL
DMH-1	X	X	X	1uM
SB431542	X	X	X	10uM

2.9 hESC FACS antibody staining

2.9.1 FACS analysis

Media was aspirated and cells were dissociated using trypLE as described above. Cells were neutralised with FACS buffer (10% FCS/PBS), counted and centrifuged at 1000 rpm for 3 minutes. Supernatant was aspirated and cells re-suspended in FACS buffer at a cell density of 1 x 10^7 per mL. 200µL of the sample was transferred to a 5mL FACS tube and the appropriate antibodies added, as per Table 3. Cells and primary antibody were incubated at 4^0C for 30 minutes with occasional flicking to re-suspend the cells. After 30 minutes cells were washed once in FACS buffer, centrifuged at 1000 rpm for 3 minutes and re-suspended in 200µL of FACS buffer. The secondary antibody was added as per table 3 for a further 30 minutes at 4^0C with occasional flicking. After secondary staining, cells were twice washed in FACS buffer, centrifuged at 1000 rpm for 3 minutes and resuspended in 500µL for FACS analysis.

2.9.2 FACS for cell sorting

Cells for sorting were prepared in the same way as described above, however all washes and staining procedures were performed in hESC media. Single cells were sorted into FACS tubes containing 1mL of hESC media. Resulting sorted cells were seeded in the presence of 50µg/mL of gentamycin (LifeTechnologies) to prevent contamination.

Table 1.3. Antibodies used for the detection of surface antigens and transcription factor expression in hESCs

PRIMARY

Antibody	Supplier	Cat #	Working Concentration
P3-X-63-AG8*	In house	-	1:10
TRA-1-85*	In house	-	1:10
TRA-1-81*	In house	-	1:10
SSEA3*	In house	-	1:10
SSEA4*	In house	-	1:10
SSEA1*	In house	-	1:10
OCT4A**	Cell Signalling Technologies	#2890	1:100
SOX2**	Cell Signalling Technologies	#3579	1:100
SOX17***	RnD System	#AF1924	1ug/mL (1:200)
GATA6***	Santa Cruz	#sc-9055	1ug/mL (1:200)
GATA4***	RnD Systems	#AF2606	1ug/mL (1:200)

The above table details surface antigens* used to identify and purify hESCs. P3X and TRA-1-85 were used as negative and positive controls respectively. Additionally shown are stem cell associated** and lineage associated*** transcription factors.

SECONDARY

Antibody	Flurochrome	Supplier	Working Concentration
Goat anti-Mouse IgG (H+L)	Dylight 647	Stratech	1:200
Donkey anti-Rabbit IgG (H+L)	Dylight 647	Stratech	1:200
Donkey anti-Goat IgG (H+L)	Dylight 594	Stratech	1:200

The table above details the secondary antibodies used in conjunction with the primary antibodies listed in Table. 1.

2.10 Functional assays

2.10.1 Single cell deposition

24h before sorting of single cells, 96-well culture treated plates were prepared by seeding MEF cells at a density of 10,000 cells/cm^2 into each well, as described **(see 2.2.1).** MEF cells were seeded in hESC media with the addition of 12µM Y-27632 so as to condition the media. MEF cells were cultured at 37^{o}C at 10% CO_2 overnight before use. To generate clonal sublines of hESCs, cells from cultures at 80-90% confluency were dissociated to single cells using TrypLE **(see 2.5)**. Cells were resuspended in fresh hESC media and pipetting with a P1000 ensured clumps of cells were properly dissociated. Cells were then stained with the appropriate markers as previously described **(see 2.8.2)**. Using the pre-prepared MEF plate, single cells were sorted using a 'BD FACS Jazz' directly into the wells of the 96-well plate containing the conditioned media and Y-27632. The plate was then briefly centrifuged at 1000rpm for 1 min to ensure cells pierced the meniscus, and incubated at 37^{o}C at 5% CO_2 for 2 days. After 2 days, the hESC media was replaced with fresh media to remove the Y-27632, and colonies were left to develop over the next 8-12 days. Wells that contained large hESC like colonies were then passaged using the standard MEF passaging or MEF free passaging **(see 2.2.1 & 2.2.2)**.

2.10.2 Clonogenic assays

Clonogenic assays were performed after FACS cell sorting on both MEF and MEF free conditions. Cells of 80-90% confluency were dissociated with trypLE **(see 2.5)**. These cells were then stained with the appropriate antibodies **(see 2.8.2).** After staining, cells were bulk sorted using a 'BD FACS Jazz' into tubes containing 1mL of hESC media. After sorting, cells were centrifuged at 1000rpm for 3 mins and counted **(see 2.6)**. For clonongenic assays cell suspensions were made at a density of 500 cells/cm^2 in the required volume of hESC/E8 media for

MEF or feeder free conditions respectively, with the addition of 10μM Y-27632 and 50μg/mL gentamycin. Cells were left to attach for 24hrs and media was replaced to remove dead cells and the Y-27632 inhibitor. No further media changes were necessary. After four days, media was removed from each well and resulting colonies were washed once with PBS (w/o Ca^{+}, Mg^{++}), before the addition of 4% PFA. Cells were incubated for 15 mins at room temperature before PFA removal, and were washed once in PBS (w/o Ca^{+}, Mg^{++}) to remove residual PFA. Plates could then be stored at 4°C indefinitely, or stained for appropriate markers

2.11 Immunostaining

For surface marker staining *in situ*, cells were washed with twice with PBS and fixed with 4% PFA for 15 minutes at room temperature. Cells were subsequently washed in PBS, and at this stage could be stored at 4°C for later staining. Cells were blocked in blocking solution consisting of PBS, 10%FCS, 0.3M Glycine and 1% BSA for 1h at room temperature. The primary antibody was resuspended in PBS, 10% FCS and 1% BSA at the appropriate concentration, and incubated for 1h at 4°C. Cells were washed twice in PBS and the secondary was applied, again resuspended in PBS, 10%FCS &1% BSA. The secondary was incubated for 1h at 4°C, and cells were washed twice in PBS. Plates could then be stored in PBS at 4°C.

For intracellular staining, cells were fixed as above, but initial blocking included PBS, 10% serum (species specific to the animal in which the secondary antibody was raised), 0.3M Glycine, 1% BSA and 0.1% Tween for 2h at room temperature. The primary and secondary antibodies were resuspended in PBS, 10% serum (species specific), 1% BSA and 0.1% Tween. Primary antibody was incubated over night at 4°C, and secondary for 2h at 4°C. Between antibodies, cells were washed twice in PBS. Surface staining was performed before permeabilisation if combined with intracellular staining. In all cases, staining was performed under sterile conditions to allow for long-term storage.

2.12 RNA extraction

2.12.1 Extraction from monolayer cultures

Cells were washed once with sterile PBS (w/o Ca^{+}, Mg^{++}) and 100µL/cm^2 of Trizol reagent (LifeTechnologies) was added onto the surface of the culture vessel and evenly distributed by rocking. After 2 minutes, lysis was monitored under the microscope and the suspension was subsequently transferred into a 1.5mL eppendorf tube for extraction. Samples could also be stored at -20^{O}C for extraction at a later date.

2.12.2 Extraction from EB culture

EBs were transferred into a 15mL falcon tube and washed once with PBS. EBs were not centrifuged initially, but left for 2 mins to collect at the bottom of the tube through gravitational pull. Media was then aspirate and this allowed the removal of dead cells and debris. EBs were then washed in PBS, centrifuged at 1000rpm for 1 minute and resuspended in 1mL of Trizol and transferred to 1.5mL eppendorf tubes. EBs were kept in trizol for 10-12 minutes with regular vortexing to break up the structures, before being placed at -20^{O}C overnight to ensure complete cell lysis. Samples were then thawed for extraction or kept at -20^{O}C for storage.

2.12.3 Harvest from cells in suspension

Cells in suspension were pelleted by centrifugation at 1000rpm for 3 minutes, and the supernatant aspirated. The cell pellet was re-suspended and washed in sterile PBS and re-centrifuged. The cell pellet was then re-suspended in 1mL of trizol and transferred to a 1.5mL eppendorf tube for extraction or storage as above.

2.13 Total RNA isolation

Samples in trizol were thawed (if stored) at room temperature and vortexed to homogenise. 200μL of Chloroform per mL of Trizol (Sigma) was added to the sample and mixed thoroughly by vortexing for 10 seconds. The suspension was left to stand at room temperature for 10 minutes before centrifugation at 14,000rpm for 10 minutes at 4^{O}C for phase separation. The top layer containing the RNA was removed by pipetting into a new 1.5mL eppendorf tube, with care not to disturb the underlying phases. A thin layer of the RNA containing phase was always left behind to ensure minimal contamination. An 1:1 volume of 70% ethanol was then added to the extract and mixed by vortexing. Samples were then processed using the RNA clean up and concentration kit (Norgen) for total RNA. RNA quality was checked through the A260/280 and A260/230 values. RNA quality was deemed of good quality if A260/280 and A260/230 values were above 1.8.

2.14 cDNA synthesis

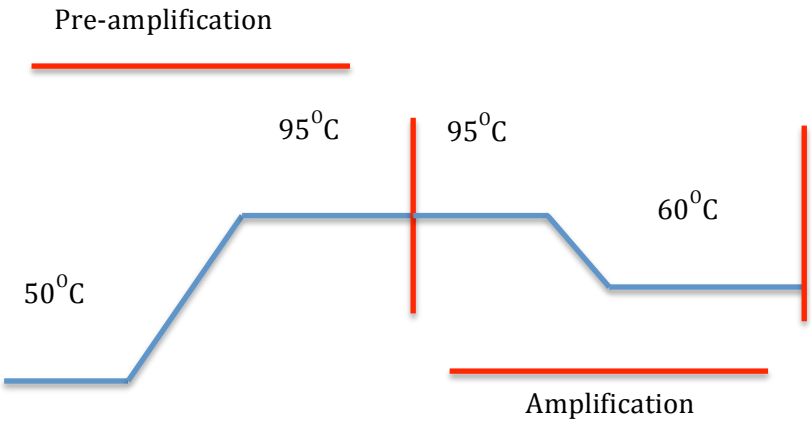
cDNA was made using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Resulting cDNA samples were assumed to have a 1:1 concentration ratio to the starting RNA.

2.15 Quantitative PCR

All preparations were done on ice and out of direct sunlight. Firstly, a master mix of 5μL Taqman Fast Universal Master Mix (LifeTechnologies), 0.2μM primer mix (containing 5μM sense-antisense oligo primers), 0.1μL Roche probe (Roche UPL) and 2.7μL of ddH_2O (Life Technologies) was made for each reaction. Secondly, a cDNA dilution of 5ng/μL was made in ddH_2O for each reaction. 8μL of the first master mix was added to wells of a Fast-96 or 384 well

plate followed by 2µL of the cDNA dilution for a total of 10µL. The plates were then spun briefly at 1000rpm. The plates were analysed using a QuantStudio 12K Flex Real-Time PCR system (LifeTechnologies) with the following parameters; 50^{O}C – 2 mins, 95^{O}C – 10 mins, 95^{O}C – 15 secs, 60^{O}C – 1 min **(Fig. 1.1).** Data analysis was performed using QuantStudio 12K Flex software provided with the machine. The raw CT values were then converted to Delta-CT by subtracting the gene of interest against the control gene (Beta-Actin in all cases for this study). Delta-CTs were then used to obtained fold change with the ΔΔCt method (Livak & Schmittgen 2001), or converted to 1/D-CT to invert values so graphs could be read in the same way as graphs using fold change.

Figure 2.1 Cycle control for quantitative PCR



The above figure details the temperature and duration of the pre-amplification steps, and the cycling steps necessary for use in quantitative PCR

2.16 List of qPCR primers

qPCR primers were designed using the Primer3 and ProbeFinder software developed by UPL Roche. Wherever possible, both forward and reverse primers were designed to contain 50% GC content and to span intron junctions. The following table describes the gene name, primer sequence and corresponding probe (UPL).

Table 1.4. Primer/Probe combinations used for the detection of mRNA levels

Gene	Sense	Anti-sense	Probe
B-Actin	ccaaccgcgagaagatga	ccagaggcgtacagggatag	#64
OCT4	agcaaaacccggaggagt	ccacatcggcctgtgtatatc	#35
NANOG	agatgcctcacacggagact	tttgcgacactcttctctgc	#31
REX1	tctgagtacatgacaggcaagaa	tctgataggtcaatgccaggt	#65
GATA4	ggaagcccaagaacctgaat	gctggagttgctggaagc	#69
GATA6	aatacttcccccacaacacaa	ctctcccgcaccagtcat	#90
FOXA2	cgccctactcgtacatctcg	agcgtcagcatcttgttgg	#9
SOX17	cgccgagttgagcaagat	ggtggtcctgcatgtgct	#13
MIXL1	gacacagatgaggggcagtt	cccgttttcagctaccattc	#6
PAX6	cgttggaactgatggagttg	Agggcaacctacgcaaga	#12
SOX1	accaggccatggatgaag	cttaattgctggggaattgg	#37
SOX7	ttcctcaccagccaggtc	atttgcgggaagttgctcta	#30
ISL1	gcagcccaatgacaaaact	ccgtcgtgtctctctggact	#83
CD34	gcgctttgcttgctgagt	Gggtagcagtaccgttgttgt	#8
DESMIN	ggagattgccacctaccg	ggtctggatggggagattg	#55
PECAM	tggaaattggaagagcacaa	Ttcaagtttcagaatatcccaatg	#37
SOX2	atgggttcggtggtcaagt	ggaggaagaggtaaccacagg	#19
TH	tcagtgacgccaaggaca	gtacgggtcgaacttcacg	#42
NEUROD1	acctcgaagccatgaacg	cttccaggtcctcatcttcg	#55
KDR	gaacatttgggaaatctcttgc	cggaagaacaatgtagtctttgc	#18
AFP	tgtactgcagagataagtttagctgac	Tccttgtaagtggcttcttgaac	#61

3 Chapter 3

3.1 Introduction

3.1.1 Development of a defined embryoid body differentiation protocol

The embryonic origin and the ability of hESCs to differentiate into all somatic cell types makes them a valuable tool for interrogating the underlying mechanisms in human development. To date, much knowledge has been gained in understanding the signalling and gene expression profiles in cellular specification using monolayer based directed differentiation. The advantages of using monolayer systems include the use of single cells that can be seeded at a known density, and the use of specific differentiation inducing factors that can be applied to interact uniformly across cells. What's more, monolayer systems have advantages when analysing differentiation endpoints, for instance *in situ* staining of cells for specific markers and flow cytometric analyses. Nevertheless, there are significant drawbacks when trying to recapitulate developmental processes *in vitro* using a 2D differentiation system. Embryonic development is a process that exhibits strict spatial and temporal control, which occurs in a 3D format. The 3D nature of the embryo means that development is in part reliant on cell-cell contacts and the establishment of polarity. For example, the development of anterior-posterior endoderm within the mouse embryo is dependent upon the surrounding mesoderm and ectoderm which signal in an instructive manner. Without these signals, endoderm fails to express more mature markers of endodermal derivatives such as *PDX1*, *SS* and *NeuroD*, which can be reversed upon co-culture with mesoderm and ectoderm of the epiblast (Wells & Melton 2000). The architecture of monolayer culture does not allow for such in-depth interactions, and consequently differentiation does not follow an embryonic-like programme. The discovery of embryoid bodies provided a new approach to ESC differentiation. Originally identified in mice, embryoid bodies are 3D aggregates of cells, which have been shown to better recapitulate *in vivo* like embryonic differentiation.

The term embryoid body (EB) was named after clumps of cells found within the ascites fluid of animals bearing intraperitoneal teratocarcinomas, and named due to the morphologically similar appearance to the 5-6 day old mouse embryo. These EBs, albeit somewhat disorganised compared to the mouse embryo, demonstrated a degree of self-organisation, which followed an embryonic program not dissimilar to the blastocyst (G. R. Martin & Evans 1975). Two morphologically different types of EBs were originally observed. The first named 'simple embryoid bodies' which showed little differentiation ability and consisted of a core of EC cells surrounded by primary extra-embryonic endodermal cell types. Second were cystic EBs that also showed endodermal differentiation on the surface of the EB, but also contained a large variety of early differentiated embryonic cell types. Within cystic EBs, advanced structures suggestive of neural tube, yolk sac, hematopoietic islands and embryonic plate were all identified (PIERCE & Dixon 1959). As a promising tool for studying mammalian development, work began to culture EBs *in vitro*. It was found that the formation of EBs could be achieved by culturing EC cells on tissue culture dishes, in the absence of feeder cells in non-adherent conditions. Single cells would form growing clumps, with a layer of endodermal differentiation surrounding an inner core of EC cells. These EBs had strikingly similar morphology to both *in vivo* EBs, and the early mouse blastocyst (G. R. Martin & Evans 1975).

After the derivation of mouse embryonic stem cells (mESCs), EBs were formed using similar techniques to that of EC cells, and similarly, ES cell derived EBs showed the ability to form cystic embryoid bodies exhibiting diverse differentiation and, if induced, complex morphogenesis such as optic cup-like structures (Evans & Kaufman 1981; Doetschman et al. 1985; Eiraku et al. 2011). The striking similarities of embryoid bodies to the early blastocyst in mouse then led to the idea that EBs from hESCs could be used to interrogate the early events of human embryogenesis and better understand the mechanisms for the derivation of somatic cell types. Although some aspects of mammalian and human development appear similar, a small number of studies demonstrate differences in placental, extra-embryonic and egg cylinder development

(Thomson et al. 1998). Immediately, it was found that the methodology used to form EBs from mESCs was not applicable to hESCs. It was not possible to generate EBs from single cells, and even multiple single cells did not readily form aggregates. The inability for single hESCs to form EBs meant that the most common method of formation was to scrape colonies from the culture flask and culture these clumps in non-adherent conditions. Itskovitz-Eldor et al demonstrated that scraped hESC colonies could aggregate on non-adherent petri dishes to form dense clusters of cells, which, over time, became cavitated and cystic (Itskovitz-Eldor et al. 2000). There were, however, large differences in the morphology and the level of organisation seen within structures derived from hESCs when compared to mESCs. Structures from these EBs did show extensive differentiation, for example cells indicative of epithelia and endoderm were apparent, and genes including αFP (endoderm) and ζ-globin (hematopoietic) were expressed (Itskovitz-Eldor et al. 2000). Furthermore, somatic cell types including cardiomyocytes, neurons, astrocytes and oligodendrocytes have been generated (Kehat et al. 2001; Zhang et al. 2001), nevertheless EBs did not show blastocyst-like structures, indicating a lack of organisation and embryonic-like differentiation programs.

3.1.2 Refining embryoid body formation from hESCs

The level of disorganisation in EB differentiation of hESCs was thought to be due to the varying sizes, geometries and cell numbers comprising each EB. Consequently, contradictory endogenous signals, the interaction of differing cell types, and random extra-cellular matrices formed within the structure resulted in non-embryonic like differentiation (Chen et al. 2007; Ungrin et al. 2008). Thus, efforts were made to identify and control parameters in EB formation with a view to better mimic the developing embryo, improve differentiation uniformity and reproducibly derive medically relevant cell types. The first parameter to be controlled was EB size. It has been shown that size has profound effects on the specification of cells within EBs. By controlling

aggregate size through microfabricated stencils, Park et al found increased mesodermal differentiation in larger cell aggregates of 300-500 μm, whereas neuronal differentiation was favoured in smaller cell aggregates of 100-300 μm diameter (Park et al. 2007). In accordance with these results, it was also shown that higher expression of cardiogenic genes were found in larger cell aggregates of 450 μm (Hwang et al. 2009). Ng et al took a slightly different approach to size control which involved counting and seeding a known density of cells per EB (Ng et al. 2005). Referred to as spin EBs, they demonstrated that erythropoietic differentiation was optimal when using 1000 cells per aggregate, and more generally, efficient blood differentiation required a minimum of 500 cells per aggregate (Ng et al. 2005).

A second parameter in controlling EB formation was spatial organisation. Ungrin et al developed PDMS casts in which the number of cells plus the shape of the EB could be tightly controlled. EBs formed in this manner showed a higher degree of self-organisation, with blastocyst-like features. EBs consisted of an inner domain of *OCT4*+ cells, an outer endodermal layer and a laminin basement membrane separating the two. These structures are reminiscent of the epiblast and visceral endoderm within the developing embryo, confirming the ability of supra-cellular order of hESCs. Additionally, this protocol allowed for the high-throughput generation of highly uniform EBs (Ungrin et al. 2008).

3.1.3 Chemically defined differentiation of embryoid bodies

Work then turned to the development of conditions in which reproducible EBs could be generated in a defined way. The development of a serum-free medium would have several advantages. Firstly, defined conditions would allow for the proper interrogation of the cues that drive embryo-like differentiation *in vitro*. Secondly, it would be permissive for the effects of exogenous growth factors for directed differentiation without conflicting signals from serum. Thirdly, it would provide neutral conditions to better understand mechanistic aspects of cellular specification. With this in mind, Albumin Polyvinylalcohol Essential Lipids

(APEL) medium was developed, which replaced animal components of conventional media with recombinant human counterparts and also included the important component polyvinyl alcohol (PVA). The addition of PVA greatly improved reproducibility of EB formation, essentially by acting as a cellular adhesive, and also allowed the generation of EBs without the addition of Y-27632 (Ng et al. 2008). Recently, more basic, defined media such as Essential 6 (E6) have been successfully used for EB formation, but the addition of PVA remains imperative (Lin & G. Chen 2014).

With methodologies to generate, in a high-throughput fashion, reproducible embryoid bodies, we wanted to investigate whether we could standardise the formation and the differentiation of EBs. Using defined conditions, we wanted to assess whether they could respond to extrinsic differentiation cues and whether we could reproducibly and efficiently direct differentiation to each primary germ layer for the derivation of medically relevant cell types. Due to the significant need for high-throughput predictive assays to predict human responses during the drug development process, and due to the improved relevance of EBs to the embryo, we also wanted to assess the use of this assay as a tool for predictive reproductive toxicology.

3.2 Results

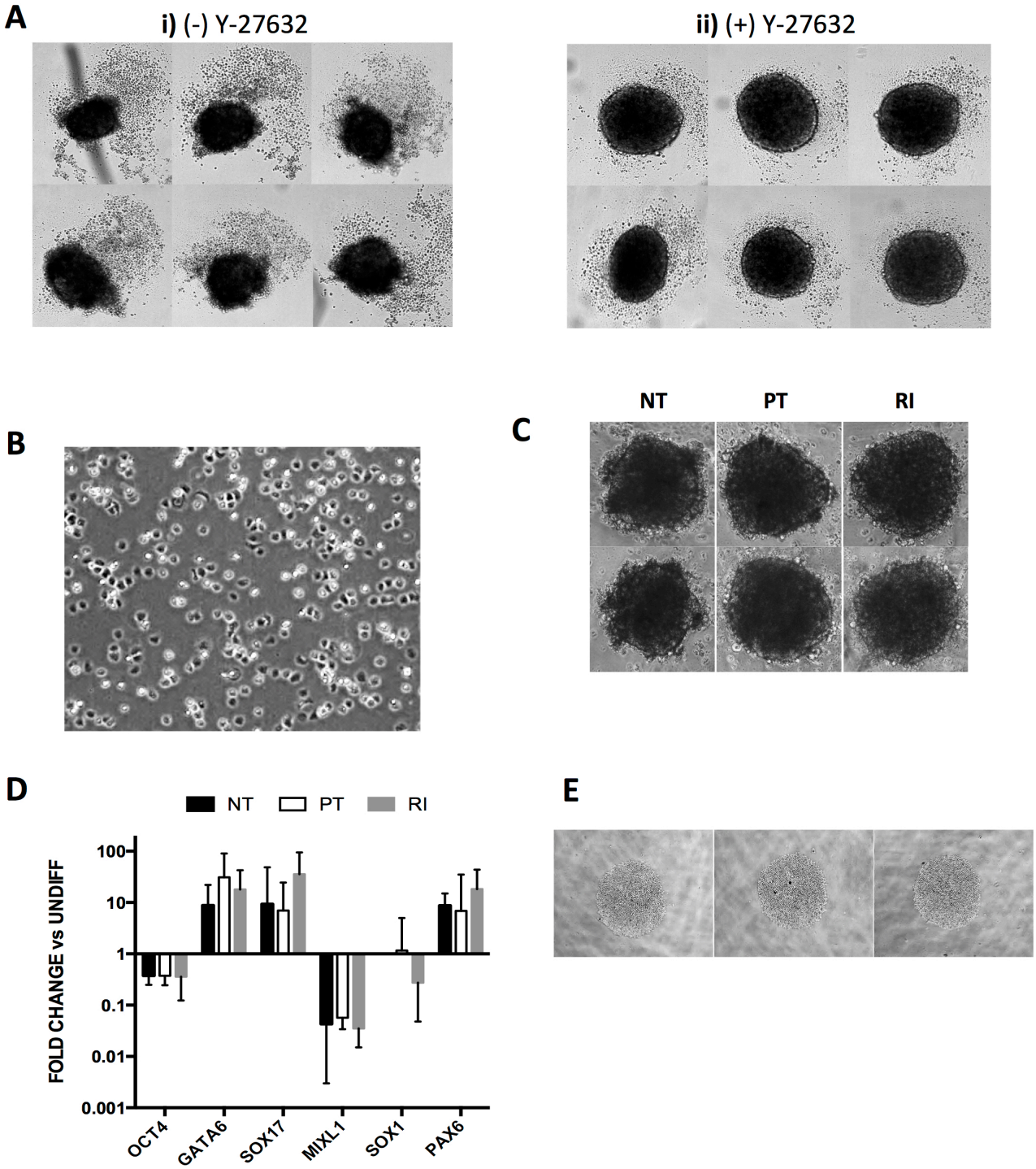
3.2.1 Development of a protocol for embryoid body formation

Although well established within the mESC field, the formation of defined EBs from single hESCs has been notoriously difficult, due to their high mortality rate upon dissociation. Firstly, we wanted to assess whether we could make EBs from single cells using the cell line H9. We found that EBs from single dissociated hESCs grown in Knockout-DMEM, 20% Knockout serum replacement (KO/SR) & MEF conditions could be formed directly from trypsinisation and seeding alone using APEL media. Although after 24h post-seeding EBs had formed, it was apparent that many cells had undergone apoptosis, and the EBs differed quite drastically in size and shape from each other **(Fig. 3.1A.i)**. This was not the case if 10 μM of the Rho Kinase inhibitor Y-27632 was added to the cells. In this case, EBs formed uniformly, and there was less evidence of cell death/debris **(Fig. 3.1A.ii)**. Thus, the inclusion of Y-27632, and consequently the prevention of cell death greatly reduced inter-well variability of EBs. Although reproducible EBs could be formed with the addition of Y-27632, we identified two problems with this approach. Firstly was the continual presence of Y-27632 in a toxicological assay. We reasoned this inhibitor may have unknown interactions with, or may mask the toxic effects of drug candidates in toxicological assays potentially through its anti-apoptotic mechanism. Secondly was the incorporation of MEF cells into EBs, which we reasoned could hinder the reproducibility of differentiation. To circumvent these issues, we included a cell 'pre-attachment step' devised by combining methodology from Ng et al as well as methodology from Andrew Elefanty at Monash University, Melbourne (Unpublished). Firstly, we coated ungelatinised culture treated plastic-ware overnight with MEF conditioned medium, referred to as MEF coated flasks (MCFs). hESCs in KO/SR & MEF culture were then trypsinised and plated onto MCFs without the addition of Y-27632. This process both eliminated MEF cells, and selected for healthy dividing hESCs **(Fig. 3.1B)**. Using this new method of pre-attachment we were able to generate more

uniform EBs than by simply trypsinising cells, although they still did not acquire the high level of uniformity seen with the continued presence of Y-27632**.** We next tested whether a media change step after the initial hESC attachment to MCFs containing 10 μM Y-27632 for 1h would improve EB uniformity. We found that the act of incubating cells with 10 μM Y-27632 for 1h and removing it before EB formation did further improve the uniformity of EBs. In fact EBs generated this way closely resembled EBs formed in the continued presence of Y-27632 **(Fig. 3.1C)**. To assess whether the act of pre-treating cells altered the differentiation of EBs, we performed qPCR analysis on day 10 EBs with no pre-treatment (NT), pre-treatment (PT) and continued presence of Y-27632 (RI). We found no differences in gene expression for the stem cell marker *OCT4*, or the lineage markers *GATA6*, *SOX17*, *MIXL1*, *SOX1* or *PAX6* between each condition **(Fig. 3.1D)**. Additionally, in the interest of high-throughput approaches, we were able to accurately seed very similar numbers of cells per well in a high-throughput way using this methodology **(Fig. 3.1E)**. Thus our modified protocol allowed for the high-throughput generation of uniform EBs without contaminating MEF cells or the continued presence of Y-27632.

The protocol thus consisted of 5 general steps. 1) Generation of MCF, 2) pre-attachment of single hESCs to MCF (without Y-27632), 3) pre-incubation of attached hESCs with 10 μM Y-27632, 4) Removal of Y-27632 and 5) EB formation **(Fig. 3.2A)**. Although robust, we did identify that certain aspects of the protocol were sensitive for the reproducibility of the assay, ultimately resulting in the loss of uniformity of EB formation, or the complete failure of cells to aggregate. Most notable was the quality of the starting cell populations, which greatly influenced aggregation of the EBs over the first 24h **(Fig. 3.2B)**.

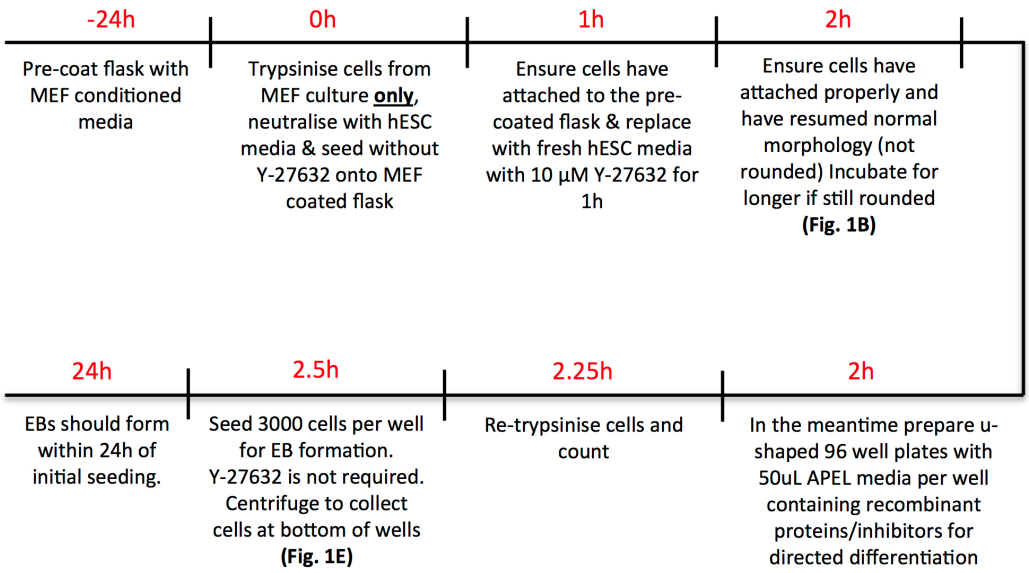
Figure 3.1: Development of a robust protocol for embryoid body formation



(A) Brightfield images at x4 magnification of 24h old EBs that have been formed in the lack of **(i)** or the presence of **(ii)** 10 µM Y-27632. **(B)** Brightfield image at x4 magnification of the morphology of cells 30 mins after seeding onto MEF coated flasks. Dead cells appear bright, and attached cells take on a rounded flat morphology. **(C)** Brightfield images at x4 magnification of 24h old EBs that had been formed using the pre-attachment protocol without Y-26732 (NT), treated with 10 µM for 1h before EB formation (PT) and in the continual presence of 10 µM Y-27632 (RI). **(D)** qPCR analysis for stem cell and lineage associated genes of day 10 EBs formed using the pre-attachment protocol without Y-26732 (NT), treated with 10 µM for 1h before EB formation (PT) and in the continual presence of 10 µM Y-27632 (RI). Data is shown as fold change against the undifferentiated starting cells, with Beta-Actin as the normalising gene. Error bars represent standard deviation from 3 technical repeats. **(E)** Brightfield images at x4 magnification of 3000 cells seeded into single wells of a u-shaped 96-well plate after centrifugation for 3 mins at 1000rpm.

Figure 3.2: Schematic of embryoid body formation and troubleshooting

A



B

Troubleshooting	
Problem	**Cause**
Cells do not attach to pre-coated plate	• Media taken from poor MEF flasks for pre-coating • Pre-coated flask not incubated for long enough • hESCs poor quality (over-confluent/ differentiation) • Incubate hESCs on MCFs at 37°C for a minimum of 20 mins
EBs do not form after 24h	• Starting quality of cells was poor (over-confluent/differentiation) • Cells were not centrifuged long enough to collect in wells **(Fig. 1E)** • Cells should not be sorted for SSEA3 • Cultures must be MEF/hESC media not feeder free

(A) Schematic of the EB formation processes with timings for each step. **(B)** Table of common problems encountered in the failure of EB formation, or the formation of non-uniform EBs.

3.2.2 Gene expression analysis of differentiating embryoid bodies

In order to investigate the early differentiation events that were occurring within these uniform EBs, we put together a small gene list that was representative of the stem cell, endoderm, mesoderm and ectoderm lineages based on existing literature **(Fig. 3.3A)**. We then generated EBs under neutral conditions (without the addition of inhibitors or recombinant proteins to direct differentiation) with endpoints at early (day 4), mid (day 10) and late (day 16) differentiation stages. The time-course was performed to determine which time-point would be most informative for the expression of the genes within our panel. By dividing genes into their associated germ layers we found that both endoderm and mesoderm associated genes showed a stepwise increase in their expression over increasing time-points **(Fig. 3.3B)**. Ectoderm associated genes were not upregulated at day 4, but saw significant upregulation at day 10 and 16 **(Fig. 3.3B)**. Day 10 therefore appeared to be the earliest and most suitable time-point to assay EBs using our gene panel.

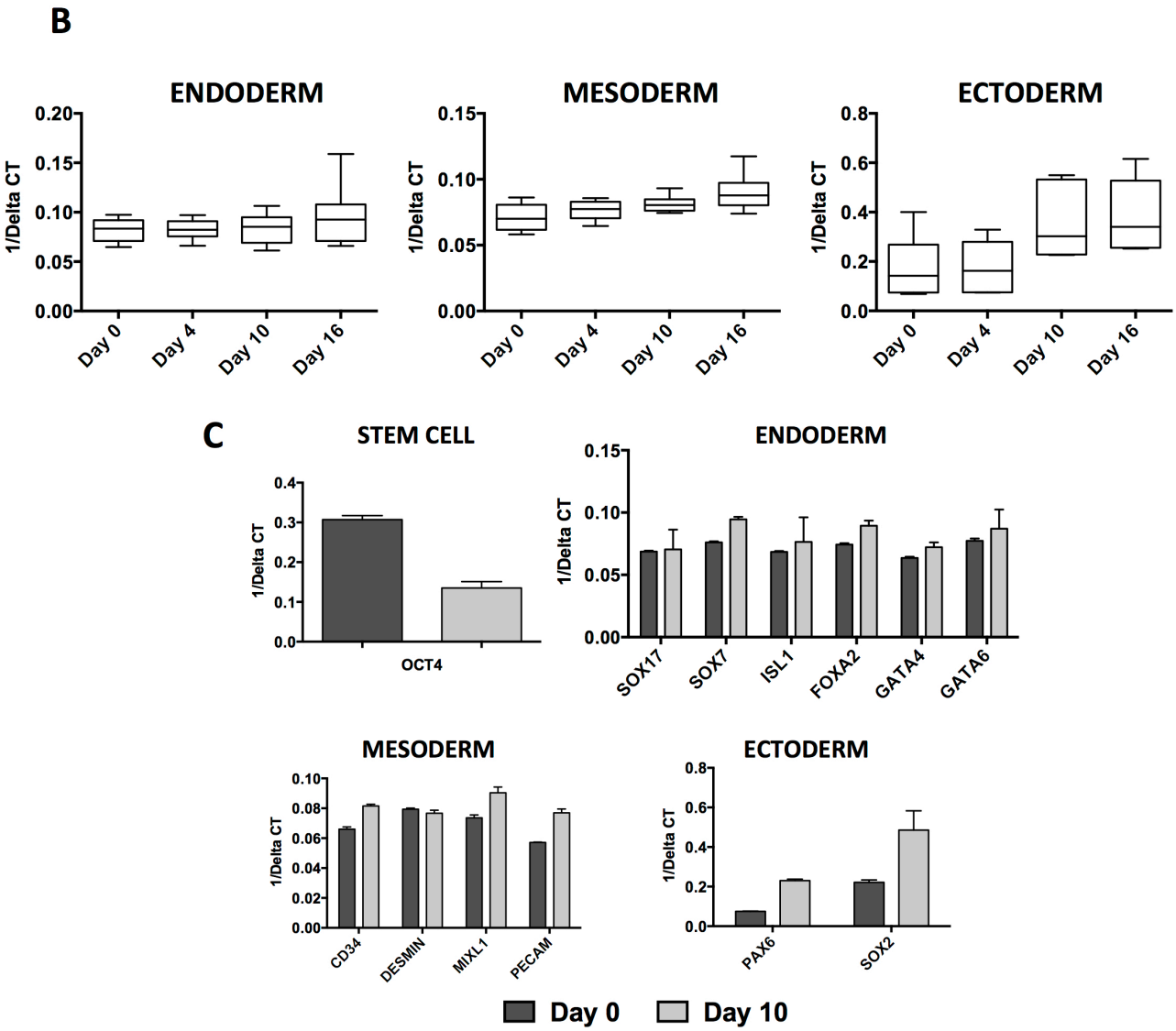
We next wanted to assess whether the protocol we had developed led to the reproducible differentiation of EBs. We performed qPCR analysis on EBs generated from three biological repeats and assessed the standard deviation of the expression of genes within our panel. The majority of genes generally showed low standard deviations, although some genes including *SOX17*, *ISL1*, *GATA6* and *SOX2* did show a higher degree of variation **(Fig. 3.3C)**. What was also striking from this data was the apparent bias in the direction of differentiation within EBs. Under the neutral conditions used, EBs expressed much higher levels of ectoderm-associated genes when compared to endoderm and mesoderm genes **(Fig. 3.3B)**. EBs were generated from hESCs grown in MEF conditions, and it is a common feature that cells undergo spontaneous differentiation. To assess whether contaminating heterogeneity in the starting population of cells was responsible for the observed differentiation bias, we sorted cells for the stem cell marker SSEA-3 before EB formation **(Fig. 3.4A)**. We found that the addition of Y-27632 was imperative for EB formation after

SSEA-3 sorting, without which no formation occurred **(Fig. 3.4B)**. After 10 days, EBs were harvested and qPCR analysis showed that the distribution of germ layers was more evenly spread. When compared to EBs from unsorted cells, SSEA-3 sorting resulted in a higher level of endoderm and mesoderm associated gene expression coupled with a reduction in ectoderm associated gene expression, although importantly, ectoderm expression levels were still high **(Fig. 3.4C)**. We also found that there was reduced variability across biological repeats in the expression of genes that did show a degree of variability when cells were sorted for SSEA-3 **(Fig. 3.4D)**. This, however, was not the case for every gene within the panel. Thus, the starting population of hESCs appeared important in the proper multi-lineage differentiation of EBs as well as the reproducibility of differentiation. The sorting of cells however was not practicable. Firstly we wanted to make it widely accessible to other laboratories, which may not have the equipment or expertise to perform fluorescent activated cell sorting. Secondly, we did not want the continual presence of Y-27632 when screening drug candidates. Having established that starting cell population quality was important, we set a threshold of 80% SSEA-3 positivity for EB formation. The threshold of 80% was chosen as it represents a high quality culture and is routinely attainable. Cells were stained with SSEA-3 and FACS analysed prior to EB formation. Cultures with ≥80% SSEA-3 were then used for subsequent EB experiments. If cultures were <80% SSEA-3 positive, cells were discarded.

Figure 3.3: Differentiation of embryoid bodies

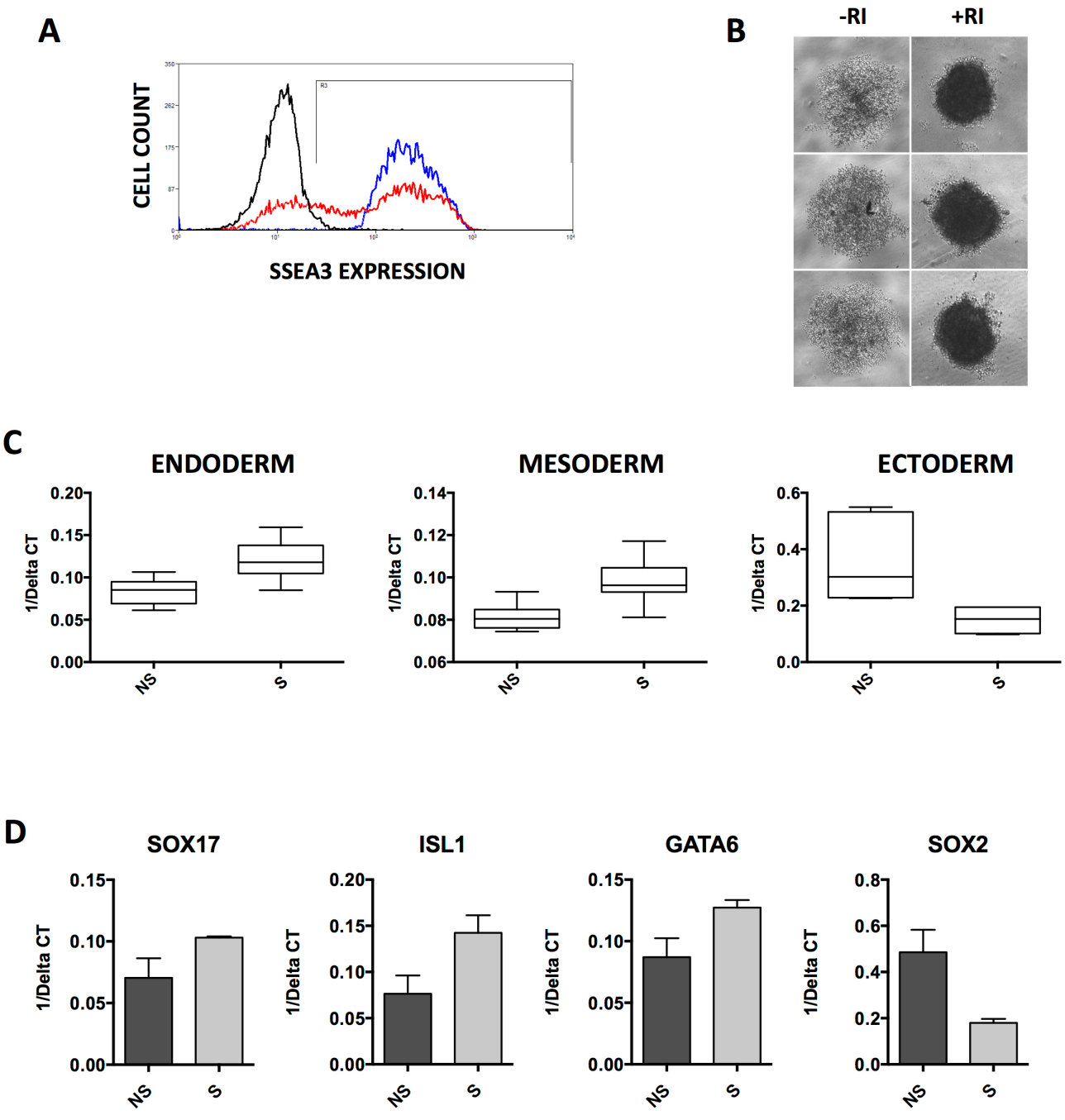
A

Stem Cell	Endoderm	Mesoderm	Ectoderm
OCT4	SOX17	CD34	PAX6
	SOX7	DESMIN	SOX2
	ISL1	MIXL1	
	FOXA2	PECAM	
	GATA4		
	GATA6		



(A) Genes used in subsequent qPCRs to mark the stem cell, endoderm, mesoderm and ectoderm lineages. **(B)** Box and whisker plots for genes specified in (A) for each primary germ layer in EBs at days 4, 10 and 16. Graph plotted as 1/delta-ct using beta-actin as the normalising gene. Box plots were generated by taking the 1/Delta-CT value for each gene in (A). Levels indicate median values, box indicates values between the 25th and 75th percentile, whisker length represents min & max values. **(C)** qPCR analysis of all genes within the panel (A) for EBs generated under neutral conditions. Data is plotted as 1/Delta-CT and error bars represent the standard deviation from three biological repeats. Day 0 is undifferentiated cells used to generate the EBs.

Figure 3.4: SSEA3 sorting improves differentiation reproducibility



(A) Representative Fluorescent Activated Cell Sorting (FACS) plot during the sorting process of SSEA3 on H9 cells. Black: P3X (negative control), Red: SSEA3 before sort, Blue: SSEA3 after sort. **(B)** Brightfield images at x4 magnification of EBs 24h post seeding after SSEA3 sort with (right panel) and without (left panel) 10 µM Y-27632 inclusion. SSEA3 sorted cells do not aggregate in the absence of Y-27632. **(C)** Box and Whisker plots of endoderm (left graph) mesoderm (middle graph) and ectoderm (right graph) associated gene expression within day 10 EBs sorted **(S)** or not sorted **(NS)** for SSEA3. Box plots were generated by taking the 1/Delta-CT value for each gene in (Fig. 3.3A). Levels indicate median values, box indicates values between the 25th and 75th percentile, whisker length represents min & max values. **(D)** qPCR analysis of genes showing a reduction in expression variation of genes from day 10 EBs upon sorting for SSEA3. Data shown as 1/Delta-CT. Error bars represent standard deviation of two biological repeats. **NS**: not sorted for SSEA3, **S**: sorted for SSEA3.

3.2.3 Directed differentiation of embryoid bodies

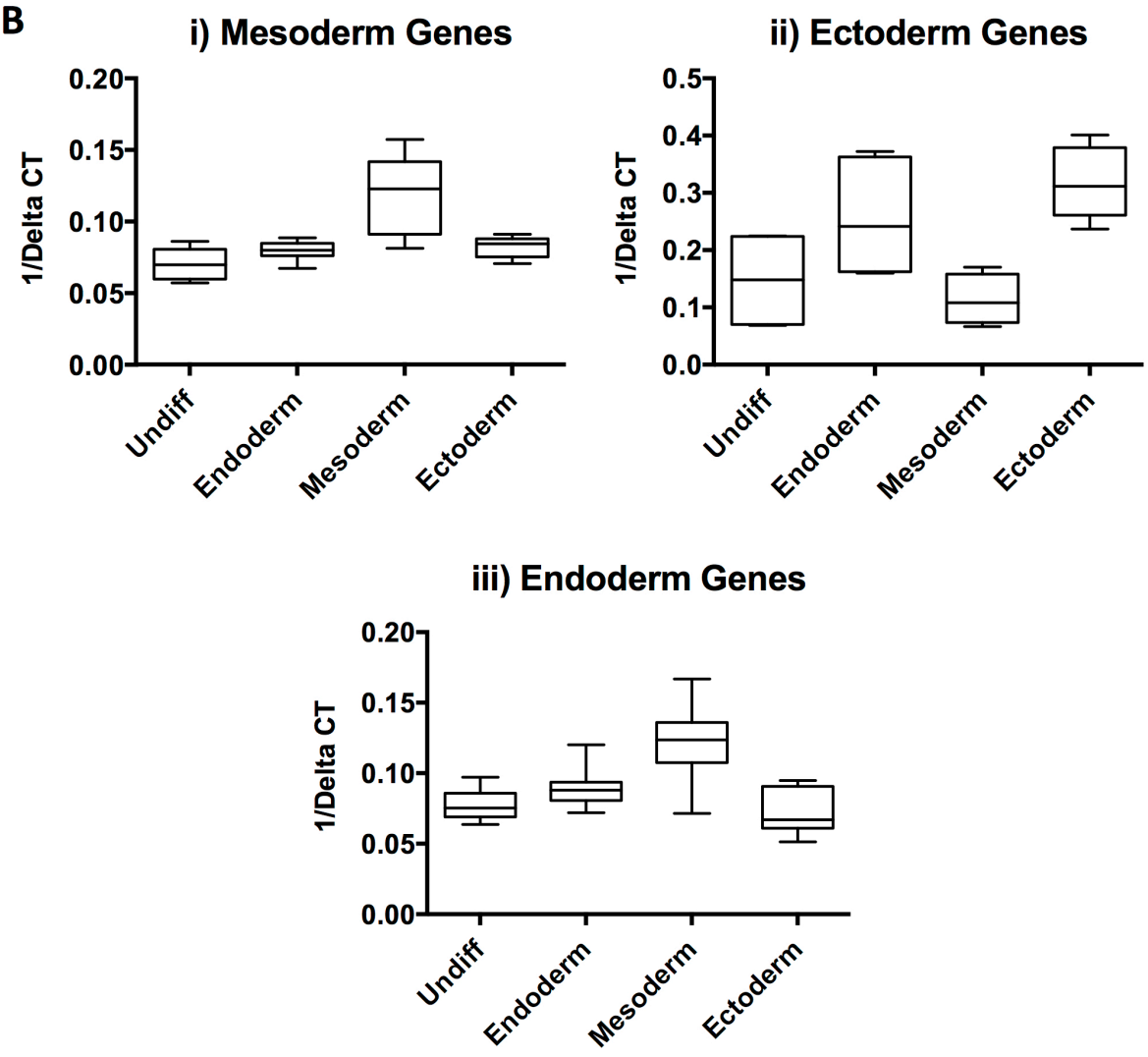
We next wanted to assess whether our assay was permissive for directed differentiation to each of the three primary germ layers; endoderm, mesoderm and ectoderm. The conditions from each germ layer were constructed from existing literature as well as unpublished data **(Fig. 3.5A)**. EBs were generated using the previously described methodology using lineage specific conditions. EBs were left to differentiate for 10 days, before harvesting and RNA extraction. Samples were analysed by qPCR for the panel of markers described previously **(Fig. 3.3A)**. We categorised genes according to their germ layer and assessed which conditions resulted in the highest expression of each. Firstly, we found that mesoderm associated gene expression was significantly higher in mesoderm inducing conditions when compared to the undifferentiated, endoderm and ectoderm conditions **(Fig. 3.5Bi)**. Secondly, ectoderm genes were strongly upregulated in ectoderm inducing conditions. Surprisingly, endoderm conditions also resulted in high expression of ectoderm genes, albeit not as efficiently as in ectoderm conditions. Mesoderm conditions did not upregulate ectodermal genes **(Fig. 3.5Bii)**. We found that endoderm conditions resulted in the upregulation of endoderm associated genes when compared to the undifferentiated control, however it was within the mesoderm inducing conditions where endoderm genes were most highly expressed **(Fig. 3.5B.iii)**. Thus it appeared that mesoderm and ectoderm differentiation conditions were effective at inducing genes of their respective lineages, but endoderm conditions were less effective at inducing endoderm genes. Additionally, it was apparent that there was not a clear segregation of the endoderm and mesoderm lineages, or the endoderm and ectoderm lineages.

Fig 3.5: Directed differentiation of embryoid bodies

A

Germ Layer	Activin A	BMP4	bFGF	SB431542	DMH-1
Endoderm	100ng/mL	1ng/mL	X	X	X
Mesoderm	20ng/mL	20ng/mL	X	X	X
Ectoderm	X	X	100ng/mL	20 µM	1 µM

B



(A) Table of conditions used to differentiate EBs to each of the three primary germ layers. X represents the absence of the recombinant protein/inhibitor. All differentiation was performed in the serum free, defined media; APEL. **(B)** qPCR analysis of day 10 EBs grown under endoderm, mesoderm and ectoderm differentiation conditions. Genes were first organised into their corresponding germ layers (Fig. 3.3A). Delta-CT values for each gene were then turned into 1/ Delta-CT and graphed as Box and Whisker plots. Horizontal bars within each box represents the median of all individual genes previously grouped into germ layers, the box shows the 25th and 75th percentile, and the whiskers represent min and max values of all genes from each germ layer. Control gene for normalisation was Beta Actin.

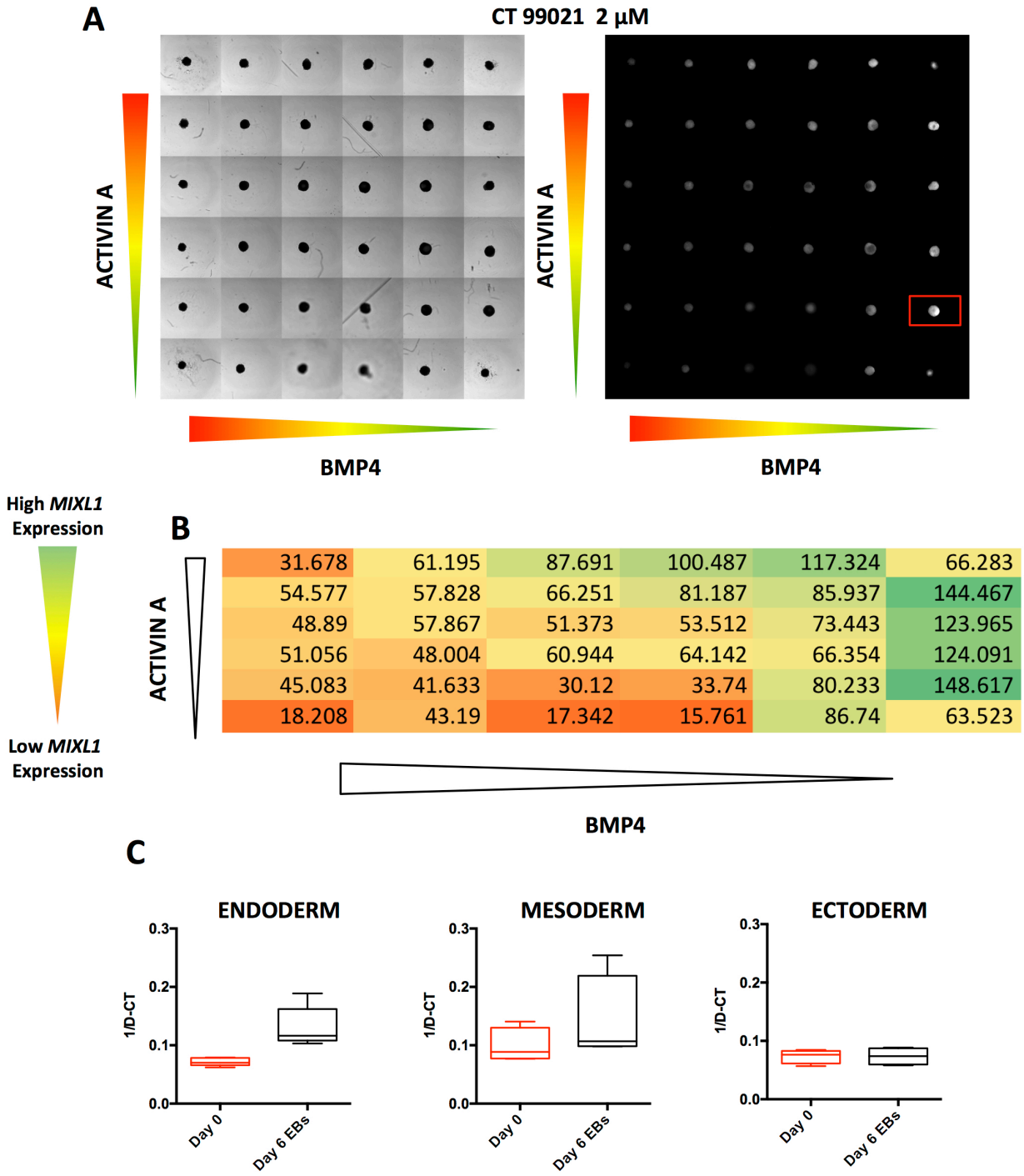
3.2.4 Optimisation of embryoid body differentiation

In an attempt to properly segregate each lineage and to find optimal conditions for endoderm differentiation, we used two separate reporter lines with GFP heterozygously knocked into the *GATA6* or *MIXL1* loci for endoderm and mesoderm read-outs respectively. Using these reporter lines we performed a 3 way titration of 2 recombinant proteins, Activin A and BMP4 as well as the GSK3β inhibitor CT 99021. Firstly, we titrated Activin A and BMP4 between 0-100ng/mL and CT 99021 between 1 μM and 3 μM, on the GFP:*MIXL1* line. We found that EBs in 2 μM CT 99021 showed the highest degree of morphological uniformity **(Fig. 3.6A)**, and were further analysed for *MIXL1* expression. *MIXL1* expression appeared highest in the absence of BMP4 but over the majority of ranges of Activin A concentrations **(Fig. 3.6A & 3.6B)**. We found that in the absence of Activin A, *MIXL1* was barely induced irrespective of BMP4 or CT 99021 concentrations, confirming the importance of TGFβ signalling through SMAD2/3 **(Fig. 3.6A)**. Activin A, however, was not able to rescue *MIXL1* expression when BMP4 concentrations were high, and in the presence of high Activin A, *MIXL1* expression increased as BMP4 was reduced. Using ImageJ analysis software, we identified that 6 ng/mL of Activin A, 0 ng/mL BMP4 and 2 μM CT 99021 resulted in the brightest expression of *MIXL1* within EBs. To ascertain whether these conditions were inductive for the expression of other mesoderm-associated genes and whether they efficiently segregate the endoderm and mesoderm lineages, we performed qPCR analysis on day 6 EBs differentiated in the new conditions. We found that mesoderm genes under these conditions were induced but we again found that EBs also showed strong expression of endoderm-associated genes. Consistently, ectoderm was not upregulated **(Fig. 3.6C)**.

Having not been able to efficiently segregate endoderm and mesoderm lineages based on *MIXL1* expression, we performed the same experiment using a reporter line for the endoderm specific marker *GATA6*. Titrating Activin A, BMP4 and CT 99021 in the same manner, we found that EBs were also most

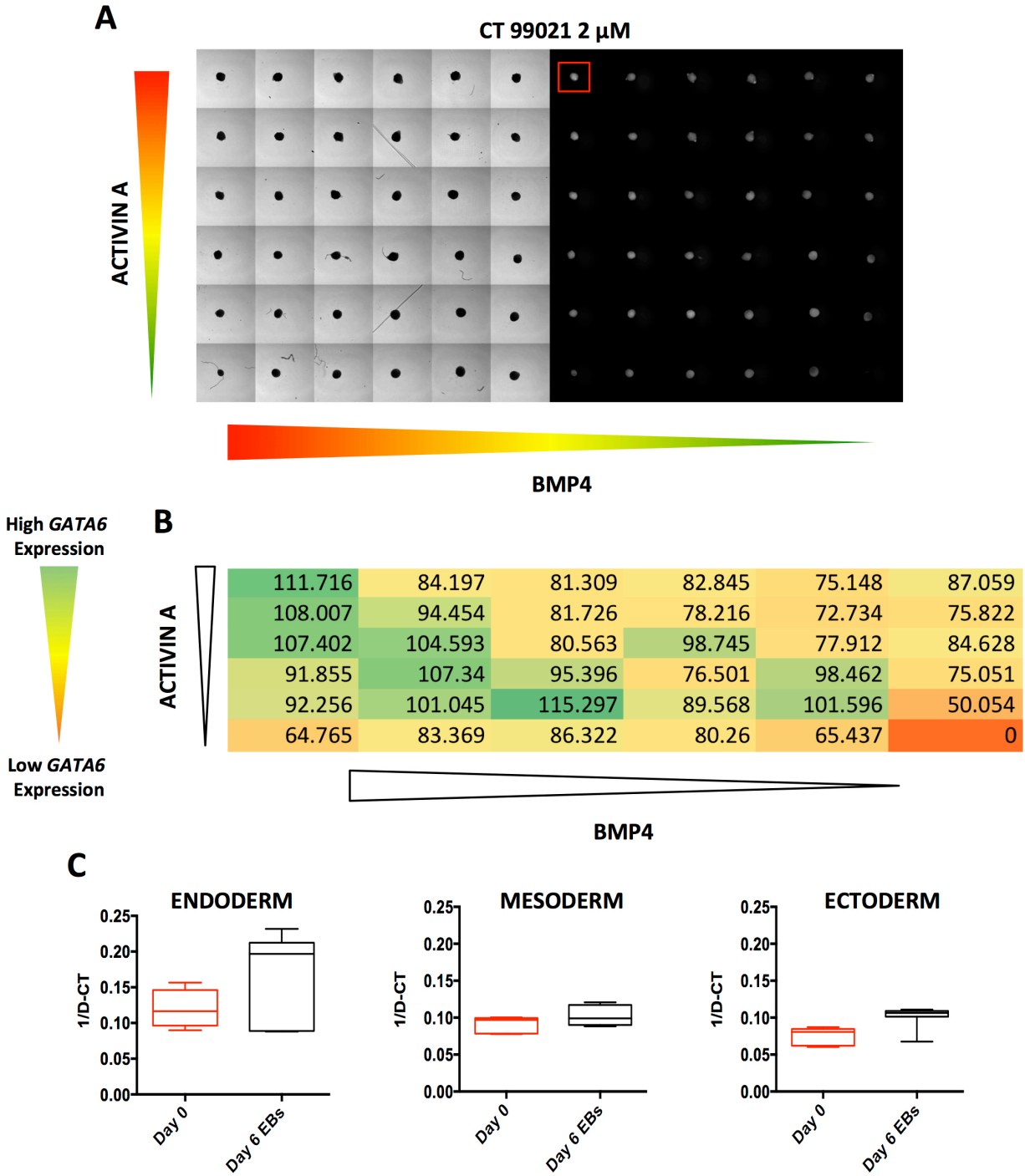
morphologically similar at 2 μM CT 99021 **(Fig. 3.7A)**, but we found that *GATA6* expression was highest in quite opposite conditions to *MIXL1* **(Fig. 3.7A & 3.7B)**. At low concentrations of BMP4 there was lower *GATA6* induction that increased as BMP4 increased. Similarly to *MIXL1*, this was also the case for Activin A, whereby *GATA6* was not strongly induced in its absence. Using ImageJ analysis software, we identified the optimal concentration of *GATA6* induction as 100 ng/mL Activin A, 100 ng/mL BMP4 and 2 μM CT 99021. Next we differentiated EBs in the presence of these optimal conditions and performed qPCR to analyse gene expression. We found gene expression for endoderm associated genes were upregulated under these conditions compared to the undifferentiated cells, but also at much higher levels than the endoderm conditions before optimisation. Furthermore, we did not see an increase in mesoderm-associated genes, and although there was a slight increase in ectoderm genes, they were not expressed at the same level as endoderm genes **(Fig. 3.7C)**. *GATA6* therefore appeared to efficiently segregate the endoderm and mesoderm lineages and its expression correlated with more efficient endoderm differentiation.

Figure 3.6: Titration of BMP4, Activin A and CT 99021 reveals conditions for optimal MIXL1 induction



(A) Brightfield (left panel) and GFP (*MIXL1*) fluorescent (right panel) images at x4 magnification of day 10 EBs grown under titrated conditions. EBs were differentiated with 2 µM CT 99021 with increasing concentrations of Activin A (100, 50, 25, 12.5 & 6.25 ng/mL top to bottom) and BMP4 (100, 50, 25, 12.5 & 6.25 ng/mL left to right). Images were put into a montage and analysed for average *MIXL1* expression using ImageJ. The red box represents the EB with the highest *MIXL1* expression. **(B)** Pixel intensity values from EB analysis using ImageJ software. Conditional formatting was done in excel to generate a heatmap of the different GFP intensities for each differentiation condition. Green represents high *MIXL1* expression, red low. **(C)** Box and Whisker plots for endoderm (left graph), mesoderm (middle graph) and ectoderm (right graph) associated genes expression from day 6 EBs under conditions resulting in brightest GFP expression (A). Box plots were generated by taking the 1/Delta-CT value for each gene in (Fig. 3.3A). Levels indicated median values, box indicates values between the 25th and 75th percentile, whisker length represents min & max values. Data is shown are 1/Delta-CT using Beta-Actin as the normalising gene.

Figure 3.7: Titration of BMP4, Activin A and CT 99021 reveals conditions for optimal GATA6 induction



(A) Brightfield (left panel) and GFP (*GATA6*) fluorescent (right panel) images at x4 magnification of day 10 EBs grown under titrated conditions. EBs were differentiated with 2 µM CT 99021 with increasing concentrations of Activin A (100, 50, 25, 12.5 & 6.25 ng/mL top to bottom) and BMP4 (100, 50, 25, 12.5 & 6.25 ng/mL left to right). Images were put into a montage and analysed for average *GATA6* expression using ImageJ. The red box represents the EB with the highest *GATA6* expression. **(B)** Pixel intensity values from EB analysis using ImageJ software. Conditional formatting was done in excel to generate a heatmap of the different GFP intensities for each differentiation condition. Green represents high *GATA6* expression, red low. **(C)** Box and Whisker plots for endoderm (left graph), mesoderm (middle graph) and ectoderm (right graph) associated genes expression from day 6 EBs under conditions resulting in brightest GFP expression (A). Box plots were generated by taking the 1/Delta-CT value for each gene in (Fig. 3.3A). Levels indicated median values, box indicates values between the 25th and 75th percentile, whisker length represents min & max values. Data is shown are 1/Delta-CT using Beta-Actin as the normalising gene.

3.2.5 An assay for predictive toxicology

3.2.5.1 Morphological analyses

Having developed and refined a robust and reproducible differentiation assay using embryoid bodies, we next wanted to assess whether the differentiation assay could be used as a tool for predictive toxicology. To do this, we generated EBs in the presence of known teratogens and known non-teratogens for 10 days and assessed whether the assay would be perturbed by and could distinguish between harmful and neutral drugs. Firstly, we assessed the morphology of exposed EBs. Images from control and test samples were analysed by 5 individuals without prior knowledge of which EBs had been exposed to which compounds or concentrations used. Control EBs were EBs that had been generated and grown in either 0.1% DMSO or 100uM Saccharin which are known non-teragogens at the specified concentrations. Individuals were asked to rate the morphology of EBs from the test samples against the control samples on a scale of 1-3. 1 indicated no morphological change; 2, noticeable morphological changes; 3, severe morphological changes. We then generated an overall 'teratogenic score' for each compound at each concentration **(Fig. 3.8A)**. This score was the sum of each morphological rating (1-3) from each of the five individuals. Therefore, compounds that were deemed to not cause any observable morphological changes (score 1) by all of the five individuals received the lowest score of 5. Compounds identified as causing severe morphological changes (score 3) by all of the five individuals received the highest score of 15 **(Fig. 3.8A)**. The rate at which teratogenic compounds were successfully identified from morphological changes ranged from 70% to 85% across the five individuals. The teratogens, Topiramate, Warfarin Dimethadione Dexamethasone, Thalidomide and Lovastatin were all successfully identified as having caused morphological changes over the 10-day time course. Fluconazole, a Category C teratogen (teratogenic) was not identified by any individual as having caused a morphological change. Conversely, the non-teratogen, Clozapine, was consistently identified by all individuals as causing severe

morphological changes at the highest concentration (100 μM). In fact, Clozapine received the highest teratogenic score of all the compounds (18). Additionally, we found that the degrees of severity of these changes were different between compounds and concentrations. Lovastatin caused severe morphological changes at low concentrations, whereas Thalidomide and Dexamethasone were found to cause only minor changes at mid to high concentrations. Warfarin and Dimethadione also caused only minor changes but at low concentrations **(Fig. 3.8A & 3.8B)**.

Figure 3.8: Teratogenic drugs induce morphological changes in hESC derived embryoid bodies

A

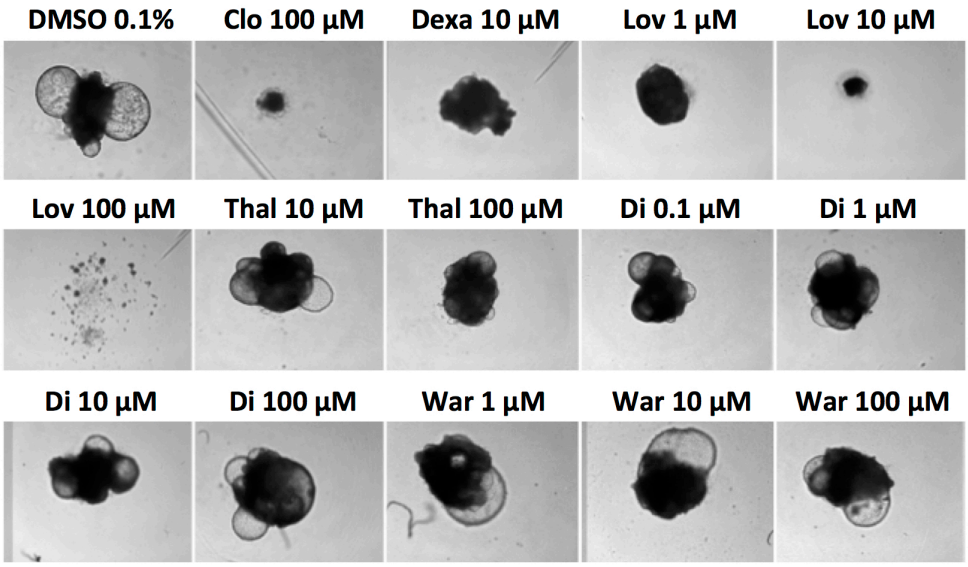
	NT Cefatoxime	NT Chlorathalideone	T Fluconazole	NT Glybenclamide	NT Hydrochlorothiazide	NT Metoclopramide
0.1uM	5	5	5	5	5	5
1uM	5	5	5	5	5	5
10uM	5	5	5	5	5	5
100uM	5	5	5	5	5	5

	NT Sulfasalazine	T Topiramate	T Warfarin	T Dimethadione	T Dexamethasone	NT Clozapine
0.1uM	5	5	5	6	5	5
1uM	5	5	6	6	5	5
10uM	5	5	6	6	5	5
100uM	5	6	6	6	10	15

	T Thalidomide	T Lovastatin
0.1uM	5	5
1uM	5	12
10uM	8	13
100uM	10	13

T - Teratogen
NT - Non-Teratogen

B



(A) Values are 'teratogenic scores' which were calculated by the summing of morphology scores (scale 1-3) of each of the five individuals who blindly analysed drug treated EBs. No change was represented by a score of 5 (1 x 5) whereas the most severe morphologies were represented by a score of 15 (5 x 3). NT – FDA classified non-teratogenic, T – FDA classified teratogenic. Red letters represent mis-classified drugs by all five individuals. **(B)** Brightfield images at x4 magnification of all drug treated day 10 EBs classified as morphologically different to control day 10 EBs (DMSO 0.1%). Each image is representative of the sample population.

3.2.5.2 Gene Expression Analyses

To further characterise EBs that had shown morphological changes when compared to the control, the lowest concentration of compounds that caused observable changes were carried forward for qPCR analysis **(Fig. 3.8B)**. In concordance with the morphological analyses, we found that the gene expression profiles of EBs grown in the presence of teratogens showed differential expression of genes within the panel when compared to control EBs **(Fig. 3.9A-C)**. Firstly, to ensure that differentiation was reproducible, we compared two control samples containing the carrier DMSO at 0.1%, which represents the highest concentration used within the assay, and the non-teratogen Saccharin at a high concentration of 100 μM. EBs showed no significant differences in the level of gene expression when compared to each other, except for the stem cell and ectoderm gene SOX2 **(Fig. 3.9B)** The ectoderm gene PAX6, although not significant compared to the 0.1% DMSO EBs, showed high variation in the Saccharin sample **(Fig. 3.9B).** Thus, both of these genes were excluded from the analysis of teratogen treated EBs. All gene expression for the following are shown in **Fig. 3.9C**.

Warfarin

Exposure of EBs to 1 μM warfarin showed significant changes in two of the genes analysed within the panel. When compared to controls, *MIXL1*, a primitive streak associated gene and *NEUROD1*, an ectoderm associated gene were significantly upregulated when compared to control EBs. The relatively minor changes in the morphology of EBs from Warfarin exposure may correlate with the expression changes in only a few genes.

Dimethadione

Exposure of EBs to 0.1 μM of Dimethadione resulted in the differential expression of three genes within the panel. Firstly, when compared to controls,

the stem cell marker *OCT4* was not downregulated to the same level of control EBs, implying a hindrance in the differentiation of hESCs. The primitive streak associated gene *MIXL1* and the ectodermal gene *NEUROD1* were both significantly upregulated. No other genes showed altered expression patterns.

Dexamethasone

EBs in the presence of 100 μM Dexamethasone showed differential expression in four genes within the panel. When compared to controls, the expression of the stem cell marker *OCT4* was downregulated at a higher rate than the control EBs, implicating Dexamethasone with a negative effect on pluripotency. The mesoderm and ectoderm genes *CD34* and *NEUROD1* respectively were upregulated, whereas the early endodermal marker *SOX17* was downregulated.

Thalidomide

Thalidomide at 10 μM induced changes in a relatively large number of genes. When compared to controls, the stem cell gene *OCT4* was not downregulated at the same rate as that of control EBs, indicating slower differentiation, similar to that of Dimethadione exposure. The endoderm and mesoderm associated genes, *AFP* and *CD34* respectively, were downregulated, but the primitive streak and ectoderm genes *MIXL1*, *NEUROD1* and *TH* respectively, were all significantly upregulated. The most significant change was in the expression of *NEUROD1*. The changes in the relatively large number of genes within the panel induced by Thalidomide may be representative of the compounds extreme teratogenic properties to the human embryo.

Lovastatin

Lovastatin also showed differential gene expression in a large number of genes within the panel. Similarly to thalidomide, lovastatin at 1 μM, when compared to controls, appeared to have reduced levels of differentiation as shown by higher levels of *OCT4* expression. The early endoderm gene *SOX17*, primitive streak

gene *MIXL1*, and ectoderm genes *NEUROD1* and *TH* were all upregulated upon lovastatin exposure. The only gene to be downregulated was the endodermal marker *AFP*. Remaining genes remained unchanged between the control and lovastatin treated EBs.

We were unable to perform qPCR on EBs treated with 100 μM Clozapine due to extensive cell death.

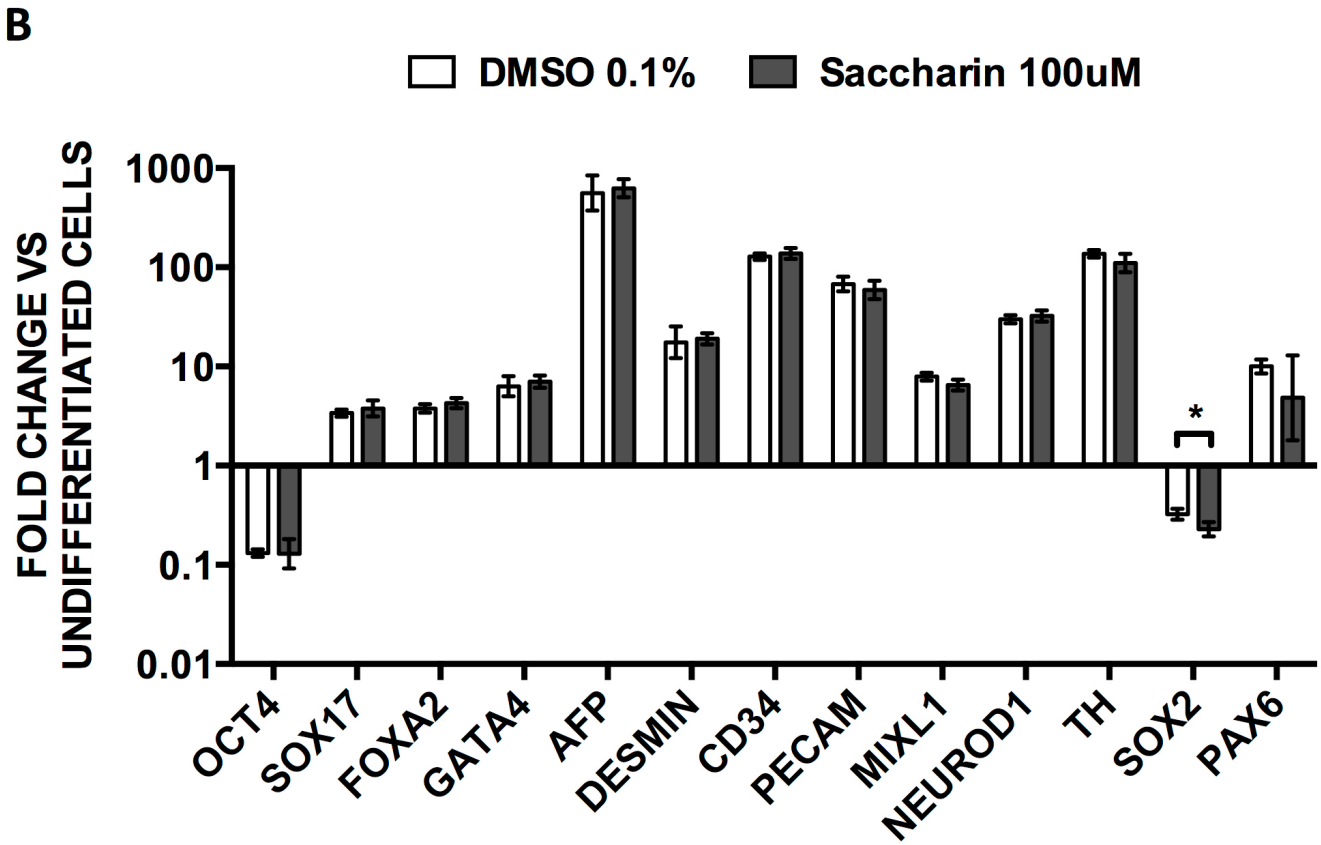
In conclusion, we found the assay resulted in the successful identification of, on average, 80% of known-teratogens based on the analysis of EB morphology by five individuals, without prior knowledge of compound or concentration, after 10 days of exposure. Analysis of EBs identified as having morphological changes by qPCR also revealed differential expression of several genes for each drug. The overall changes in gene expression between control and treated samples were quite subtle, but did show statistical significance. The assay also appeared relatively sensitive to perturbations as shown by the identification of morphological and gene expression changes at very low concentrations, for example in the case of Dimethadione at 0.1 μM, and Warfarin at 1 μM, which did not exhibit an increase in severity of morphology at high concentrations. Additionally, the severity of the teratogenic effects of compounds may be reflected by the number of genes with perturbed expression, for example, Thalidomide, which induced expression changes in many genes within the panel and is known to exhibit extensive teratogenic properties on the developing embryo.

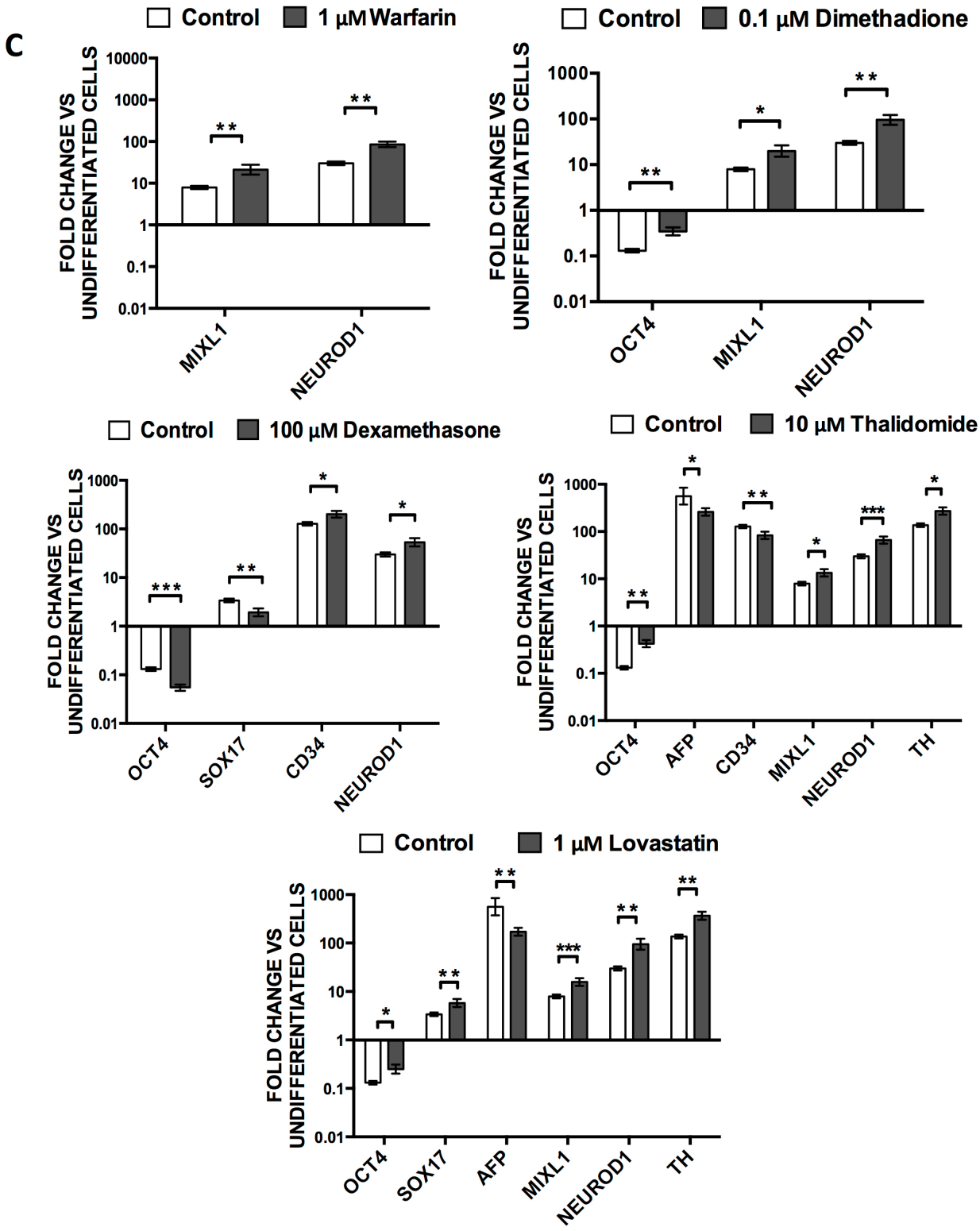
Figure 3.9: Embryoid bodies show differential gene expression upon exposure to teratogenic drugs

A

Stem Cell	Endoderm	Mesoderm	Ectoderm
OCT4	SOX17	CD34	PAX6
	FOXA2	DESMIN	SOX2
	GATA4	MIXL1	TH
	AFP	PECAM	NEUROD1

B





(A) Panel of genes selected to represent the stem cell, endoderm, mesoderm and ectoderm lineages. **(B)** qPCR analysis of two control day 10 EB samples (DMSO 0.1% and 100 µM Saccharin) used to confirm reproducible differentiation between control conditions. Data is shown as fold change against the undifferentiated control, using Beta-Actin as the normalising gene. White bars represent EBs grown in 0.1% DMSO, grey 100 µM Saccharin. Students t-test was used to assess statistical significance in genes between the two samples. Error bars are standard deviation of 3 technical repeats. **(C)** qPCR analysis of day 10 EBs grown in the presence of 0.1% DMSO (control – white bars) and the drug of interest (grey bars). Lowest drug concentration causing morphological changes were analysed. Data is shown as fold change against the undifferentiated control, using Beta-Actin as the normalising gene. Error bars are standard deviation of three technical repeats and significance was calculated using student's t-test. Genes within the panel not shown on the graphs were not significantly different to the control sample.

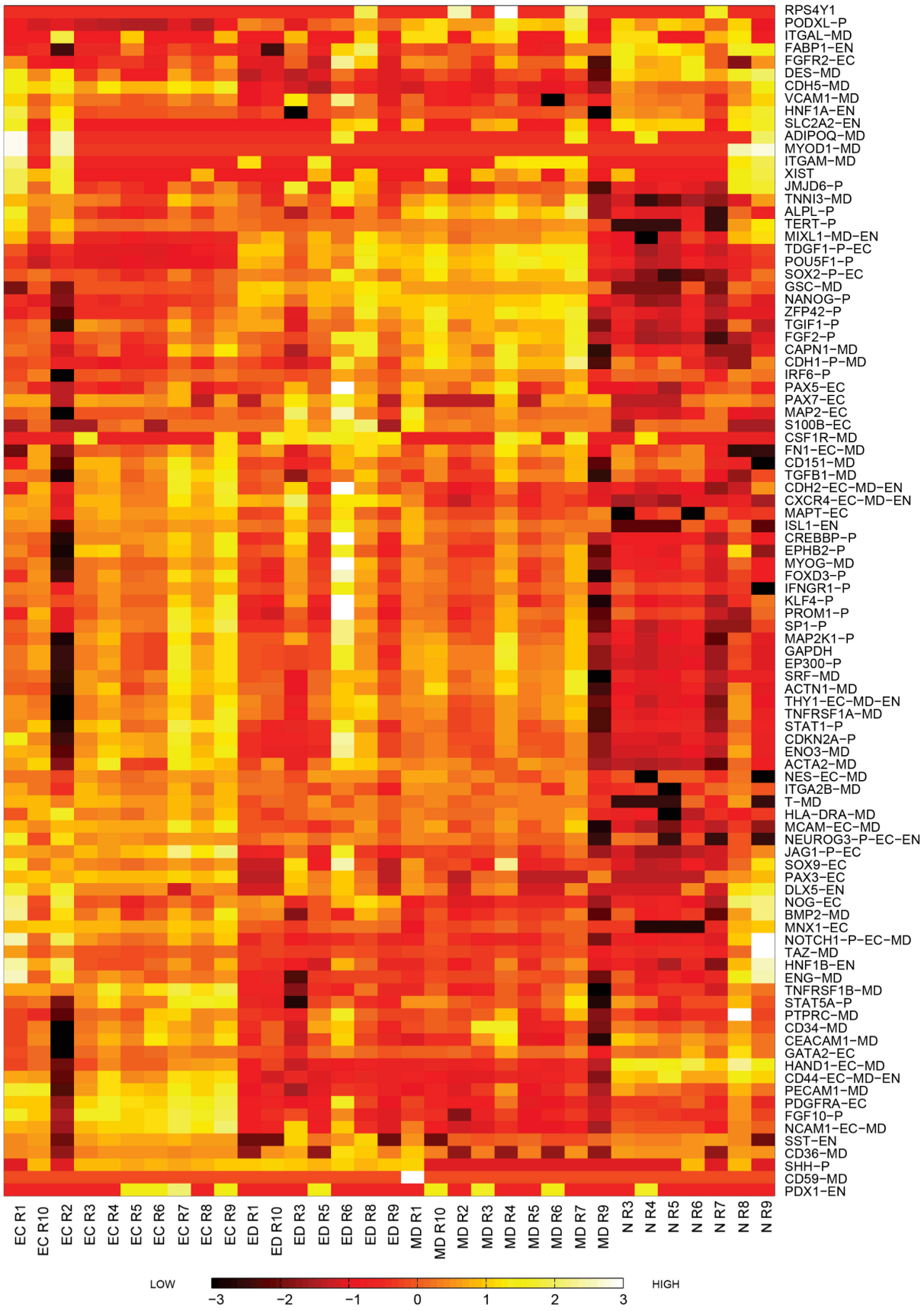
3.3 Differentiation heterogeneity of embryoid bodies

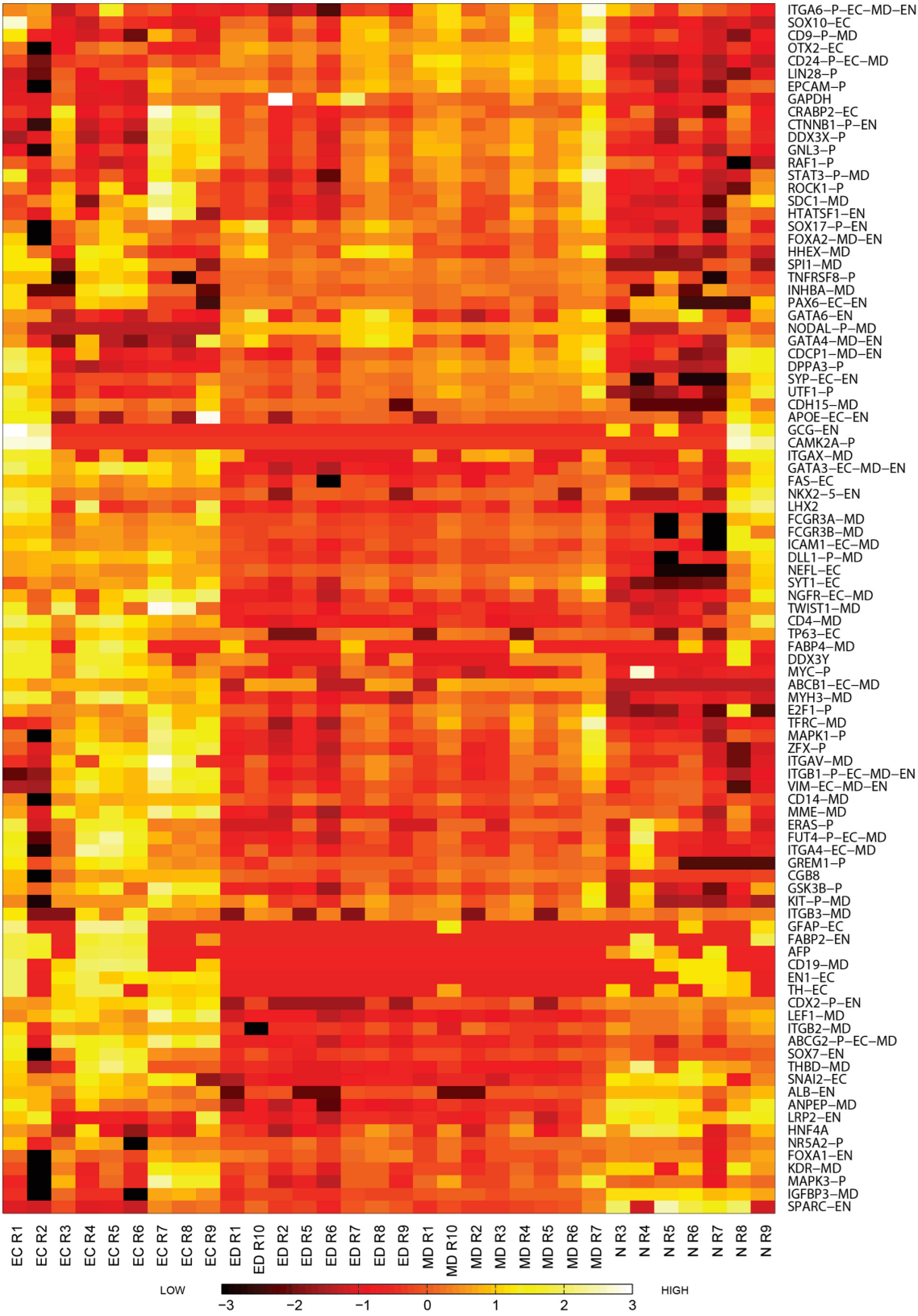
Upon further interrogation of the morphology of EBs in the screen, we did notice that single EBs exposed to the same conditions did vary somewhat from each other. We also looked at the morphology of EBs that had been formed in neutral conditions as well as EBs that had been under directed differentiation, and found the same trend. By performing qPCR on pooled EBs, the gene expression levels become averaged, and consequently standard deviations are generally reduced. We therefore wanted to assess in single EBs how variable gene expression really was and whether directed differentiation was specifying cells in a uniform manner between EBs.

To investigate the heterogeneity between single EBs, we performed qPCR analysis on EBs under neutral conditions for 196 genes using the Fluidigm system. We found that single EBs did show substantial differences in the expression of the majority of the genes within the panel **(Fig. 3.10)**. It is known that the differentiation of hESCs in EBs under neutral conditions can occur in a relatively chaotic, stochastic manner, so in an attempt to reduce the random nature of differentiation, we generated EBs under directed differentiation conditions for endoderm, mesoderm and ectoderm. We found that EBs even under directed conditions showed variation in gene expression **(Fig. 3.10)**.

Thus, the variation seen between single EBs in neutral and in directed differentiation conditions demonstrated that by just applying uniform conditions does not necessarily result in uniform differentiation. We reasoned that as the single EBs were all generated from the same starting population, there must be a more inherent heterogeneity within stem cells that was generating the variable differentiation. In order to further refine this differentiation assay, we became interested in investigating the more fundamental aspects of stem cell lineage decisions.

Figure 3.10: Embryoid bodies under uniform differentiation conditions display gene expression heterogeneity





qPCR analysis of 192 genes performed on single EBs under neutral **(N)**, endoderm **(ED)**, mesoderm **(MD)** and ectoderm **(EC)** specific conditions. Each column represents an individual EB, such that 10 were analysed for each condition. Arrays were performed on the Fluidigm system and data is shown as 1/Ct. Data was analysed using Mat Lab whereby the expression of each single gene from all of the single EB samples were firstly normalised in order to compare expression across samples. All genes in all samples were then normalised and labelled with a scale between -3 (low expression_ and +3 (high expression). This scale was arbitrarily set by Mat Lab. Colour coding was such that black represents low expression, white high expression. Gene names are followed by their lineage classification (**P** – Pluripotency, **EN** – Endoderm, **MD** – Mesoderm, **EC** – Ectoderm).

3.4 Discussion

Within this study, we have developed a reproducible differentiation assay using embryoid bodies, which shows potential as a tool for predictive toxicology. Reproductive toxicology currently consists of *in vitro* and *in vivo* assays, which involve the exposure of drugs to pregnant animal models, and the use of cell lines in culture, late on in the drug development process. Although valuable, data obtained from *in vivo* animal studies are extrapolated to the human context, which can have disastrous consequences, as shown by Thalidomide. Here we have developed an assay with hESCs, in embryoid body format so to better mimic embryonic development, which correctly identified 80% of known teratogenic drugs. This assay showed high concordance with existing toxicological data, and existing assays.

The development of the assay was initially hindered by the high sensitivity of hESCs to single cell dissociation, which has been previously reported (Watanabe et al. 2007). Myosin hyperactivation induced by single cell dissociation is a direct cause of apoptosis in hESCs, which can be effectively blocked through the use of Rho-associated protein kinase inhibitors such as Y-27632 (Ohgushi et al. 2010). Although not reported, the inclusion of Y-27632 within the assay had potential to either increase the survival of cells in the presence of toxic compounds, due to its anti-apoptotic effects (Watanabe et al. 2007), or interact with drugs through unknown mechanisms, enhancing or dampening their effects leading to false positives or more concerning, false negatives respectively. To circumvent the potential issues of Y-26732 being present for the duration of the assay, we found that initial pre-treatment with the inhibitor noticeably improved the morphological reproducibility of EBs by significantly reducing the level of cell death after dissociation. Moreover, we did not notice a difference in differentiation potential of hESCs pre-treated with Y-27632 when compared to un-treated cells, in concordance with existing reports (Watanabe et al. 2007).

As MEF cells are known to secrete growth factors (Greber et al. 2007), involved in self-renewal and specification of hESCs, and as our assay used cells specifically from MEF cultures, we devised a method in which MEF cells could be eliminated for more defined differentiation. The elimination of MEF cells was achieved by seeding dissociated hESCs onto flasks pre-incubated with MEF conditioned media. Components in the conditioned media allowed the attachment of hESCs although the mechanisms of exactly how are unknown. It has been reported that MEF cells secrete a large array of matrix forming proteins, including collagens, which may provide a suitable environment for hESCs to adhere (J. W. E. Lim & Bodnar 2002). Thus, the pre-attachment step plus pre-treatment with Y-27632 appeared to efficiently circumvent the need for Y-27632 throughout differentiation and eliminate MEF cells.

After having established a protocol for the formation and differentiation of EBs, we found that differentiation under neutral conditions (no addition of recombinant proteins or chemical inhibitors to direct differentiation) showed good reproducibility between biological experiments. EBs in these conditions seemed to have a bias strongly in favour of the ectoderm lineage, which was eradicated upon sorting for the stem cell marker SSEA-3. It is known that hESC cultures tend to be contaminated by spontaneous differentiation, but stem cells themselves can also exhibit a high level of heterogeneity in gene expression which has been shown to alter the functional behaviour of cells (Toyooka et al. 2008; Chambers et al. 2007). For example, hESCs in culture show graded expression of the stem cell surface marker SSEA-3. Upon fractionation of hESCs on SSEA-3 it has been shown that high SSEA-3 expression correlates with clonogenic stem cells, whereas SSEA-3 low and SSEA-3 negative do not (Tonge et al. 2011). By sorting cells for high SSEA-3 expression before EB formation to reduce heterogeneity, we found that differentiation was not only more evenly distributed between each primary germ layer, thereby eradicating ectoderm bias, but also led to a decrease in the standard deviation of gene expression between biological replicates. Yet we did find that cells, which had been sorted for SSEA-3, did not aggregate without the addition of Y-27632. This observation has also been reported in cells sorted for uniformly high *OCT4* expression.

Similar to our findings, this could be rectified using Y-27632 (Ungrin et al. 2008). The reasons for why highly pure populations of stem cells do not aggregate are currently unknown, but may be in part related to the absence of cell surface proteins involved in cell adherence.

Having established a reproducible protocol for the formation of EBs from single hESCs, we investigated whether the assay could be perturbed using established protocols for specific lineage differentiation through the addition of recombinant proteins and/or chemical inhibitors. Although highly effective for ectoderm specification, we were unable to efficiently segregate the endoderm and mesoderm lineages. We used two reporter lines for the *GATA6* and *MIXL1* genes in an attempt to optimise differentiation conditions for endoderm and mesoderm respectively. The action and gradients of Activin A, BMP4 and WNT in endoderm and mesoderm specification has been well established, as reviewed by (Wells & Melton 1999), therefore we titrated these morphogens in an attempt to segregate the two lineages. We found that Activin A was vital for the induction of both *MIXL1* and *GATA6*, supported by its use in mesoderm and endoderm differentiation protocols (Evseenko et al. 2010; D'Amour et al. 2005). This is also consistent with the findings that TGFβ signalling is involved in directing mesendoderm destined cells to the primitive streak *in vivo* (Arnold & Robertson 2009). It was BMP4, however, that appeared to induce a lineage switch in EBs such that lower levels induced *MIXL1* and higher levels *GATA6*.

qPCR analysis on day 10 EBs grown in optimised mesoderm conditions did show upregulation of mesodermal genes, however endodermal genes also remained highly expressed. Although heavily involved in mesodermal progenitors, such as haematopoietic cells (Ng et al. 2005), *MIXL1* has been shown to activate the promoter of endoderm associated genes *SOX17 and GSC (Goosecoid)* (Lim et al. 2009). Additionally, *MIXL1*-null mouse embryos show deficiencies in definitive endoderm formation, and null ES cells show a reduced potency to colonise the embryonic gut, an endoderm derived structure (Lim et al. 2009; Hart et al. 2002). Together, this suggests an important role of *MIXL1* in endoderm specification, whether direct or indirect, which may explain the

appearance of high endodermal gene expression in *MIXL1*(+) EBs. We also titrated Activin A, BMP4 and CT 99021 on a reporter line harbouring GFP at the *GATA6* locus. We found that, unlike *MIXL1*, *GATA6* was efficient at segregating the endoderm and mesoderm lineages, and also correlated with improved endoderm differentiation. Although *GATA6* is involved in mesodermally derived cells such as cardiomyocytes later on in development (Koutsourakis et al. 1999), it would appear at this early differentiation time-point it is almost exclusively involved in endoderm specification.

Using the assay that we developed, we wanted to assess its use as a tool for predictive toxicology. In order to assess how sensitive the EB differentiation assay was to known teratogenic compounds, we tested a variety of drugs that have previously been classed as teratogenic or non-teratogenic by the FDA. Drugs were selected from an existing library of compounds to represent wide chemical diversity, including physiochemical properties, molecular weight and polar surface areas (Gustafson et al. 2012). Five individuals were asked to blindly analyse the morphology of EBs after exposure to control and test samples, which resulted in the identification of, on average, 80% of teratogenic drugs. One drug, Clozapine, was identified by each individual at 100 μM to cause severe morphological changes of EBs, even though it is classed as non-teratogenic and falls into Category B of FDA classifications. Category B means that there have been no animal reproduction studies demonstrating a risk to the foetus, however, there are equally no adequately well-controlled studies in pregnant women. Our data from this particular assay would classify Clozapine as potentially teratogenic and/or embryotoxic to the human embryo at high concentrations. Although FDA guidelines state no adverse effects of Clozapine in animal models, it should be recalled that the highly teratogenic drug, Thalidomide, also showed no adverse effects in animal models until administered at extremely high concentrations (Bailey et al. 2005). This highlights the importance of humanised assays, as model organisms do not always accurately predict human responses to drugs. Conversely, the drug Fluconazole was not picked up by any individual as causing morphological changes in EBs, even though Fluconazole falls into Category C of FDA

classifications. Category C means that animal models have shown adverse effects, but there are no adequately controlled studies in humans. One poorly controlled clinical study does exist which describes fluconazole exposure during early gestation (weeks 4-9) with further exposure at later gestation (week 22) resulting in child abnormalities (Aleck & Bartley 1997). Conversely a study into brief fluconazole intake at the time of conception from 239 women reported no abnormalities to the resulting child (Inman et al. 1994). Similarly to Thalidomide, Fluconazole may have a tight window in which it exhibits teratogenic behaviour, which according to the existing case studies, would not appear to be early on in development. Making the assumption that our assay correlates to the earliest stages of embryonic development, these reports would support the absence of morphological change within the EB assay. A parallel study in zebrafish using the drugs Clozapine and Fluconazole also saw that Clozapine showed teratogenic behaviour, whereas Fluconazole did not. They found that Fluconazole had a very low uptake within zebrafish embryos, whereas Clozapine was very high, reasoned to have contributed to the observed phenotypes (Gustafson et al. 2012). It should therefore be emphasised that this assay is probably most relevant to the earliest stages of human development; therefore care should be taken when extrapolating data from this assay for the entire human gestation period. We believe this assay shows potential to be used, specifically, in conjunction with existing *in vitro* and *in vivo* assays to identify teratogenic drugs.

Nevertheless, we do postulate that this assay may be predictive of potential teratogenic effects later in gestation if the particular drug interferes with pathways also involved in hESC self-renewal and/or differentiation. The compound Thalidomide, which is known to have very specific teratogenic effects in later development (Miller & Strömland 1999) did show morphological and gene expression changes within this assay. Although many mechanisms of action have been published, as reviewed by (Stephens 1988), the mechanism which may explain our observed effect is through Thalidomide binding to and blocking the IGF-1 and FGF-2 promoters, both of which are involved in angiogenesis during limb formation and are involved in maintaining

pluripotency and early lineage specification in hESCs (Stephens et al. 2000; Bendall et al. 2007).

Although highly successful in identifying teratogenic drugs, upon analysis of individual EBs generated from the same starting cell population under uniform conditions, different morphologies were apparent. In concordance with these morphologies, the gene expression heterogeneity within EBs was quite striking. Of the 192 genes analysed, many showed differential expression even in conditions directing differentiation. Relating back to our data showing that the variation in gene expression could be reduced upon sorting cells for SSEA-3, it would appear that the starting cell heterogeneity might be, in part, responsible for the observed differentiation heterogeneity. In order to further refine this assay, we wanted to ascertain whether heterogeneity was indeed an aspect of stem cell culture, and whether this heterogeneity had functional consequences on the differentiation of hESCs.

4 Chapter 4

4.1 Introduction

4.1.1 GATA6 Heterogeneity in hESCs

The differentiation of lineage specific cells *in vivo* is an intricate, yet precise process through which cells of the ICM firstly enter gastrulation deriving each of the three primary germ layers, and later, further specification to terminal somatic cell types. A step-wise process of events allows cells to choose a particular fate whereby gene regulatory networks are established and cells eventually become fixed to a particular cell type, residing within a stable state, for example fibroblasts. This step-wise process involves the cell receiving a signal, signal transduction, and subsequently gene activation or repression. These signalling pathways have been the subjects of much research in order to try and recapitulate *in vitro*, decision-making that occurs during embryonic development *in vivo.* This has proven relatively successful, for example in endodermal, mesodermal and ectodermal specification (D'Amour et al. 2005; Laflamme et al. 2007; Chambers et al. 2009), but the efficiency with which cells arrive at the same destination at the same time varies dramatically. Following on from our data showing the heterogeneous differentiation of EBs, and the importance of the starting quality of cells, an outstanding question within the field is why hESCs show a large disparity in lineage specification under uniform conditions. A large body of work describes functional heterogeneity within populations of mESCs and hESCs with respect to stem cell associated markers, but there are gaps in our knowledge with respect to the role of lineage specific genes which have been shown to be expressed in hESCs (Gokhale et al. 2015). We believe that heterogeneity in the stem cell compartment may account at least in part for differentiation heterogeneity; therefore we investigated the role of the expression of the transcription factor *GATA6* in hESCs, which displays extensive expression heterogeneity in culture.

4.1.2 Heterogeneity of human embryonic stem cells

The apparent ability to maintain hESCs in states that perhaps represent varying stages of embryonic development *in vitro* has brought into question whether states may exist in which cells are biased to particular fates. Previous studies have reported that transcription factors, surface markers and signalling pathways, which are strongly correlated with the stem cell state, are expressed in a heterogeneous fashion (Stewart et al. 2006; Fischer et al. 2010; Bhatia et al. 2013; Blauwkamp et al. 2012; Tonge et al. 2011).

One example is that of *Nanog* expression in both mESCs as well as hESCs, whereby 10-20% of cells in culture actually do not express the gene (Chambers et al. 2007; Fischer et al. 2010). These *Nanog*(-) cells do not necessarily represent a differentiated cell type, as they can convert to a *Nanog*(+) state and remain pluripotent. There are, however, functional differences between these two states, such that *Nanog*(-) cells differentiate at a higher rate than *Nanog*(+) (Chambers et al. 2007). Similar observations have also been made in mESCs and the expression of the stem cell transcription factor *Rex1*, whereby two separate functionally different states were identified. The first whereby cells expressed *Rex1* and *Oct4* which were postulated to be analogues of cells within the ICM and the second where cells only expressed *Oct4* but not *Rex1*, which were deemed to be primitive endoderm like (Toyooka et al. 2008). Similar to the observations made with *Nanog*, when purified into *Rex1*(-) *Oct4*(+) and *Rex1*(+) *Oct4*(+), both states could regenerate each other resulting in the re-establishment of the original culture heterogeneity. Interestingly, when performed on hESCs, it was found that *REX1*(+) cells were able to generate *REX1*(-) cells, but not vice versa, even under prolonged culture highlighting potential differences in state residence between the two species (Bhatia et al. 2013).

Additionally, the stem cell surface markers PECAM-1 and SSEA-1 have been shown to exhibit heterogeneous expression patterns on mESCs. The expression

of PECAM-1 and SSEA-1 correlates strongly with the pluripotency of cells and their ability to form chimerae in the epiblast of mouse embryos. The loss of SSEA-1 or more importantly PECAM-1 results in an increase in lineage associated genes within cells, but does not mean the loss of their ability to self-renew. Nevertheless, cells did exhibit a shift in the probabilities of self-renewal versus differentiation such that differentiation was favoured (Furusawa et al. 2004). What is more, the expression of PECAM-1 showed near correlation to *Rex1* providing a useful link between surface marker expression and core stem cell transcription factor expression (Toyooka et al. 2008). Heterogeneity was also found to exist in the expression of the stem cell and germ cell specific gene *Stella*. *Stella* is strongly down regulated and methylated in epiblast-like cells, but highly expressed alongside PECAM-1 and NANOG in ICM-like cells. Most striking from this study was the re-equilibration of *Stella*(+) cells at a stable 20-30% of total culture even upon purification and re-seeding of *Stella*(-) PECAM-1(-) and *Stella*(+) PECAM-1(+) cells (Hayashi et al. 2008). From our lab, it has been shown that hESCs also show heterogeneity of SSEA-3, similar to that of mouse and SSEA-1. SSEA-3(+) and SSEA-3(-) cells were found to interconvert and thus regenerate each other, but the SSEA-3(-) cells showed impaired self-renewal capacity. Nevertheless, a small proportion of SSEA-3(-) cells were able to self-renew (Tonge et al. 2011).

Thus, there would appear to be a large overlap between the network of transcription factors governing self-renewal and with stem cell surface markers, which ultimately correlates with the functional status and developmental stage of ESCs. These reports specifically looked at the heterogeneity in stem cell associated markers, which then prompted a study to assess lineage marker expression in hESCs. The examination of single hESCs fractionated by their expression of SSEA-3 revealed extensive heterogeneity of lineage marker expression including genes associated with endoderm such as *GATA6*, *GATA4* and *SOX17* (Gokhale et al. 2015). The co-expression of *OCT* and *NANOG* alongside lineage genes provided initial evidence of discreet substates within the stem cell compartment that are potentially lineage specific. The early

endoderm gene *GATA6* was particularly interesting due to the high degree of heterogeneity observed in SSEA-3(+) cells.

4.1.3 The transcription factor *GATA6*

GATA6 is a zinc finger transcription factor that has DNA binding capabilities to the consensus sequence A/T**GATA**/G, (Koutsourakis et al. 1999). It plays crucial roles in the specification and development of tissue types during various stages of development, in particular the formation of the primitive endoderm during the mid to late blastocyst. In mouse development, zygotic *GATA6* gene expression has been detected as early as the 4-cell embryo (Guo et al. 2010), with clear protein expression at the 8-cell stage (Plusa et al. 2008). The endoderm-specific marker GATA6 increases in expression in the majority of cells up to the 32-cell stage alongside the opposing epiblast lineage specific marker NANOG. This co-expression of opposing lineages continues throughout the morula into the early blastocyst, but by the mid blastocyst, GATA6 and NANOG become increasingly mutually exclusive, creating a salted and peppered expression phenotype within cells of the ICM (Rossant et al. 2003; Chazaud et al. 2006). At this point in development, cells expressing GATA6 appear destined for a primitive endoderm fate (PrE) by showing a down regulation of NANOG alongside an up-regulation of other PrE associated markers, including GATA4, SOX17, Dab2 and Pdgfrα (Artus et al. 2012). By the time of GATA4 induction, very few cells within the ICM are co-expressing PrE markers and NANOG (Plusa et al. 2008). Initially, it is believed that GATA6 is induced through a signal-transduction pathway involving the signal adaptor protein, growth receptor bound protein 2 (Grb2) (Cheng et al. 1998). Grb2 functions to link receptor tyrosine kinase activation to the downstream Ras-MAP kinase-signalling pathway. Indeed, both *in vivo* and *in vitro*, the Ras/MAP kinase pathway has been shown to direct mESCs towards the endodermal lineage (Yamanaka et al. 2010). It is likely that FGF signalling works up-stream of Grb2, and is therefore important in PrE specification, as seen by PrE defects in FGF4 mutant embryos

(Feldman et al. 1995). Furthermore, the ICM of Grb2 mutant embryos completely lacks the expression of GATA6 (Rappolee et al. 1994). A critical role of GATA6 at this stage of development is for the correct formation of the primitive endoderm, which is required for proper patterning, as well as nutrient exchange in the developing embryo (Koutsourakis et al. 1999; Rossant & Tam 2009). By E4.5 *GATA6* positive cells relocate onto the surface of the ICM, forming a basement membrane between the ICM and blastocoel (Rossant et al. 2003). This relocation arises from GATA6 activated Dab2 which is required for proper epithelial organisation (Fujikura et al. 2002), as well as the roles of the extracellular matrix protein Laminin C1 and Pdgfrα (Plusa et al. 2008). Mutant forms of Dab2 and Laminin C1 allow the normal expression of *GATA6* within cells of the ICM, but spatial organisation does not occur so that *GATA6* positive cells remain mixed throughout the ICM (Smyth et al. 1998; Yang et al. 2002). Additionally, at E4.5, a new site of *GATA6* expression appears within the parietal endoderm on Reichert's membrane. This expression however is lost by E5.5, demonstrating strict spatial and temporal control of the gene (Koutsourakis et al. 1999). As development continues, *GATA6* becomes crucial in mesodermal cell types such as cardiomyocytes. By E8.0, *GATA6* is expressed in lateral plate mesoderm, which houses the cardiogenic plate, and remains present during the differentiation and migration of cardiomyocytes that form the heart tube (Koutsourakis et al. 1999). *GATA6* is then strongly expressed within the developing intestine, stomach, aorta and bladder which are all derived from endodermal and mesodermal precursors (Morrisey et al. 1996).

In humans very little is known about exactly how the early blastocyst develops, and how the expression of *GATA6* comes about, due to the lack of access and the ethical concerns over the use of human embryos. Although the early segregation events are thought to progress in a similar fashion to that of the mouse, the timing of events appears delayed (Cockburn & Rossant 2010). Even less is known about EPI and PrE segregation events. There is no strong evidence to support exactly when *GATA6* expression is initiated, or even whether the PrE formation during gastrulation is the same as the mouse. For these reasons, the majority of our understanding about this early stage of human development has

been investigated using *in vitro* hESC lines. Certainly, results from hESC studies indicate at a basic level that the mechanisms of PrE and EPI segregation are similar between the two species. For example, the knockdown of *NANOG* in hESCs induces extraembryonic endoderm, and the upregulation of *GATA6* and *GATA4* (Hyslop et al. 2005). Conversely, if *NANOG* is over expressed, hESCs follow a primitive ectodermal fate at the expense of endoderm (Darr et al. 2006). Additionally, the derivation of stable hESCs, with close resemblance to mESCs, and with high expression of *NANOG* and *GATA6*, indicates the stable co-expression of these opposing lineage specific markers, as is found within the early mouse blastocyst but within the human context (Chan et al. 2013).

With the observation that *GATA6* is expressed alongside genes of pluripotency, and given the role of *GATA6* in early endoderm specification, we wanted to follow the dynamics of *GATA6* expression in routine hESC culture to better understand its role during development and also to investigate whether we could identify functionally discreet substates within the stem cell compartment that could bias cell fate decisions.

4.2 Results

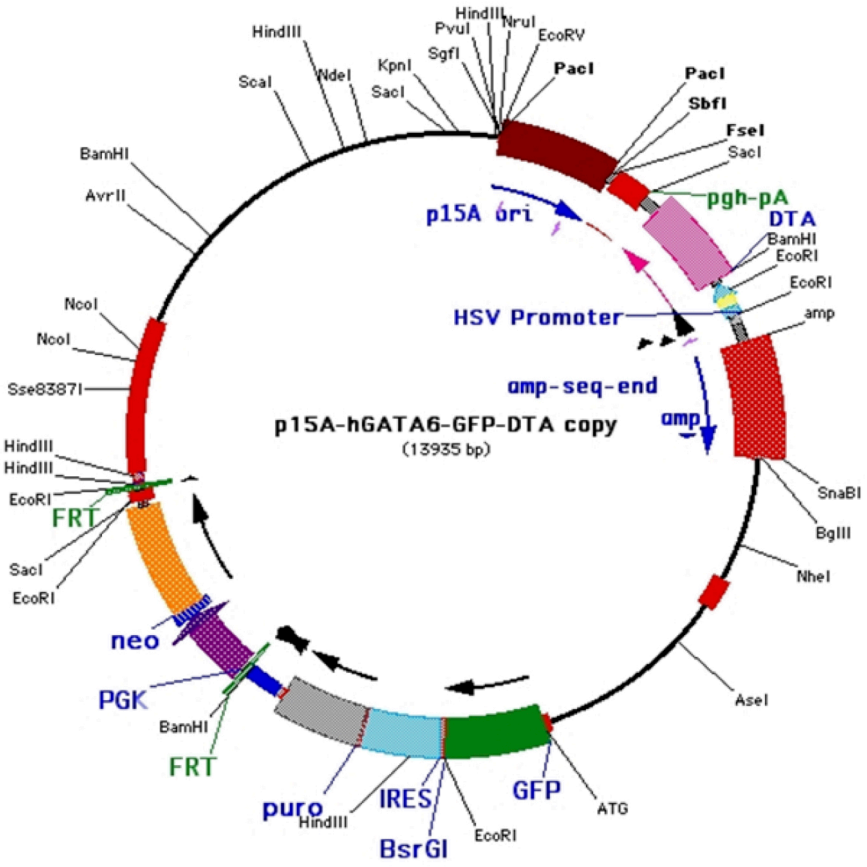
4.2.1 Generation and characterisation of a *GATA6* reporter line

In order to track the expression of the early endodermal marker *GATA6 in vitro* and in real-time, we used a reporter line whereby a GFP construct was knocked into the ATG site of the second exon of one allele at the *GATA6* locus, creating a heterozygous *GATA6*:GFP cell line. The targeting of the ATG site had been performed using zinc finger nucleases by Dr Andrew Smith (University of Edinburgh/Oxford) to create a double stranded break at the specific integration site. The GFP cassette **(Fig. 4.1)** was electroporated into the Shef4 line with induced DSBs, and the process of integration was dependent upon homologous recombination by the internal cell machinery. Cells with a successfully integrated GFP cassette had been selected for neomycin resistance, and the resulting cells sub-cloned to make clonal lines. Two clonal lines were used during this project, the first named S4G6 4/F-9, and a further sub-clone of S4G6 4/F-9, S4G6 4/-F9 A3. The A3 subclone had the neomycin resistance removed and was used for all functional experiments.

To ensure that the new S4G6 4/F-9 reporter line behaved similarly to the parental Shef4 line, the reporter was screened against a panel of stem cell markers after several passages of the cell line with GFP integration. The line expressed high levels of the stem cell markers SSEA-3, TRA-1-60, SSEA-4 and TRA-1-81, but low levels of the early differentiation markers SSEA-1, similar to that of the parental Shef4 line **(Fig. 4.2A).** Secondly, we tested whether the reporter line had still retained the capacity for multi-lineage differentiation. The formation of EBs from single hESCs has been shown to allow the specification of cells from each of the three primary germ layers (Itskovitz-Eldor et al. 2000), therefore S4G6 4/F-9 were dissociated and re-aggregated in ultra-non-adherent plates for EB formation. After 10 days of differentiation, both the Shef4 parental and the S4G6 4/F-9 line showed expression of genes associated with endoderm

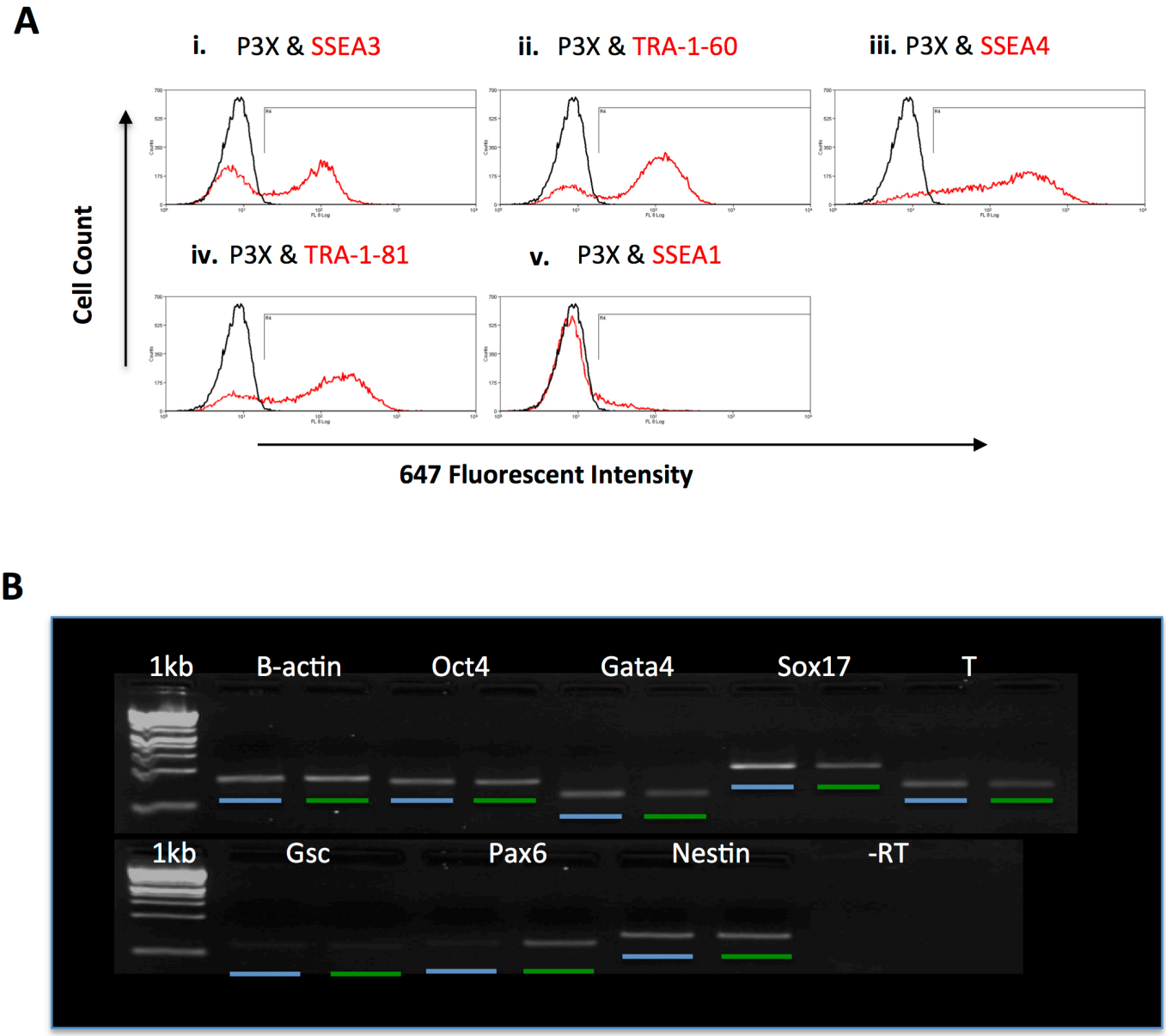
(*GATA4 & SOX17*), mesoderm (*T & GOOSECOID*) and ectoderm (*PAX6 & NESTIN*) providing evidence of pluripotency. **(Fig. 4.2B)**

Figure 4.1: GFP construct for the production of a GATA6:GFP reporter line



The GFP cassette consists of the wild-type GFP gene bolted to puromycin resistance through an IRES. Constitutive expression of neomycin allowed for the selection of successfully transfected cells. Within the S4G6 4/F-9 A3 clone, neomycin was removed using TAT-FLP recombinase.

Figure 4.2: Characterisation of the Shef4-GATA6:GFP reporter line

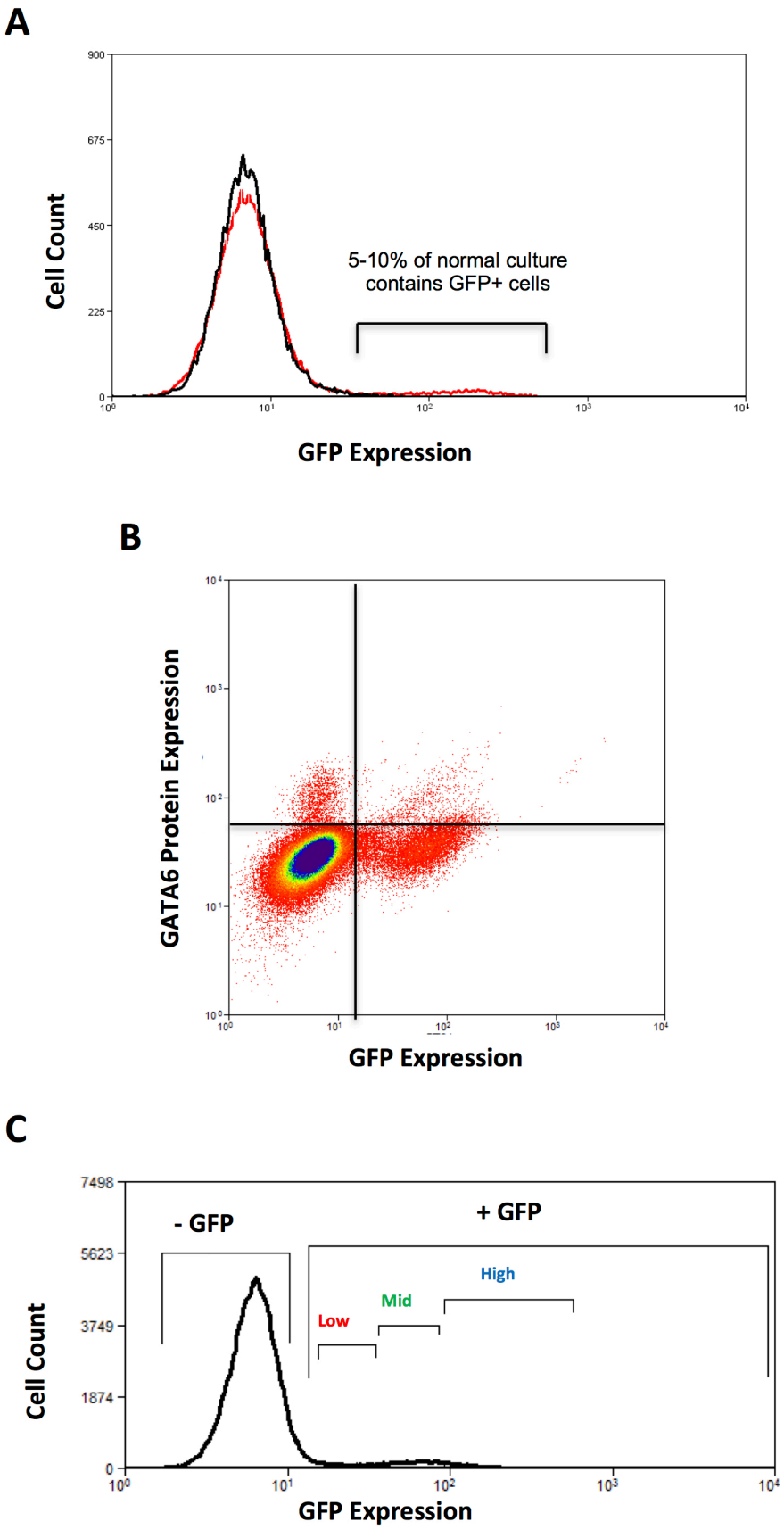


(A) The Shef4-GATA6-GFP reporter line was analysed with FACS for the markers SSEA3, TRA-1-60, SSEA4, TRA-1-81. **(i-iv)** stem cell markers and **(v)** early differentiation marker shows the absence of differentiation and the retention of pluripotency. **(B)** The Shef4-GATA6-GFP reporter line was differentiated for 10 days in an EB format. RNA was extracted and RT-PCR performed for genes of each primary germ layer. Both the parental and the reporter line showed upregulation of markers for each germ layer. B-Actin: positive control, –RT: negative control. Markers were indicative for the following lineages. *OCT4* – Stem cell; *GATA4* & *SOX17* – Endoderm; *T* & *GOOSECOID* (*GSC*) – Mesoderm; *PAX6* & *NESTIN* (Ectoderm).

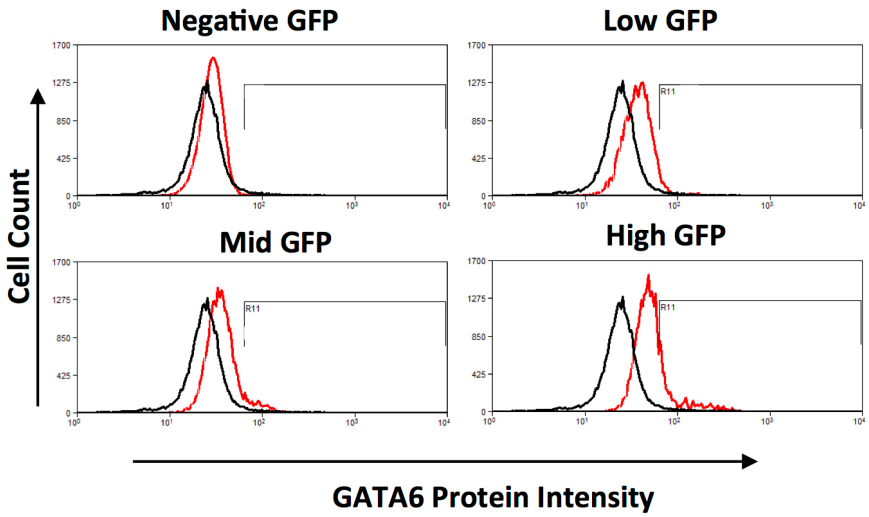
4.2.2 Detection of *GATA6* in hESCs in routine culture

We performed FACS analysis on the reporter line and found that a small subset of cells did appear to be expressing *GATA6* (**Fig 4.3A)**. As our reporter line was measuring the expression levels of GATA6 mRNA with GFP as the read-out, we sought to assess the relationship of GFP to that of GATA6 protein levels. We found that there was no linear correlation between the expression of GFP and the expression of GATA6 protein. Thus, GFP(+) cells were not necessarily translating the protein (**Fig. 4.3B)**. We then categorised the GFP expressing cells into low, mid and high expression levels to assess whether increasing GFP intensity showed an increase in the expression of GATA6 protein, even if this trend was not linear **(Fig 4.3C).** As expected, cells that did not have GFP expression showed negligible expression of GATA6 protein. As cells began to express low-level GFP, GATA6 protein did appear, with 5.5% of this population being protein positive. At mid-level GFP, a further increase of GATA6 protein was found, as 8.3% of cells became positive, however the largest increase, and the highest levels of protein were found within the high GFP fraction, at 23.2% of the population **(Fig. 4.3D & 4.3E).** Therefore, a stepwise increase in GATA6 protein was apparent as GFP intensity increased. The expression of GFP will be referred to herein as *GATA6*.

Figure 4.3: *GATA6* expression in KO/SR and MEF culture does not show linear correlation with GATA6 protein levels



D



E

Gated sample	% of Total Population	% GATA6 protein in gated sample
-ve GFP	91%	0.9%
Low GFP	2%	5.5%
Mid-GFP	3.2%	8.3%
High-GFP	1.5%	23.2%

(A) FACS plot of the *GATA6*:GFP reporter line showing low level *GATA6* expression under KO/SR & MEF culture conditions. Black represents the untransfected parental control line, and red the *GATA6*:GFP reporter line. **(B)** FACS analysis of the *GATA6* reporter line reading mRNA levels (x-axis) vs. *GATA6* antibody reading protein levels (y-axis). Gating was set off the parental Shef4 line as a negative control, and secondary only for antibody staining. There is no linear correlation between mRNA and protein levels of *GATA6*. **(C)** FACS analysis of *GATA6* mRNA expression in KO/SR & MEF culture conditions. Gates were set for low, mid & high GFP expression for GATA6 protein analysis. As GFP increases, the level of GATA6 protein also increases. **(D)** FACS plots showing the expression of GATA6 protein (red) within the GFP fractions from (C). Black represents the negative control, red GATA6 protein levels. As GATA6 mRNA increases, GATA6 protein also increases. This is quantitated in **(E)**.

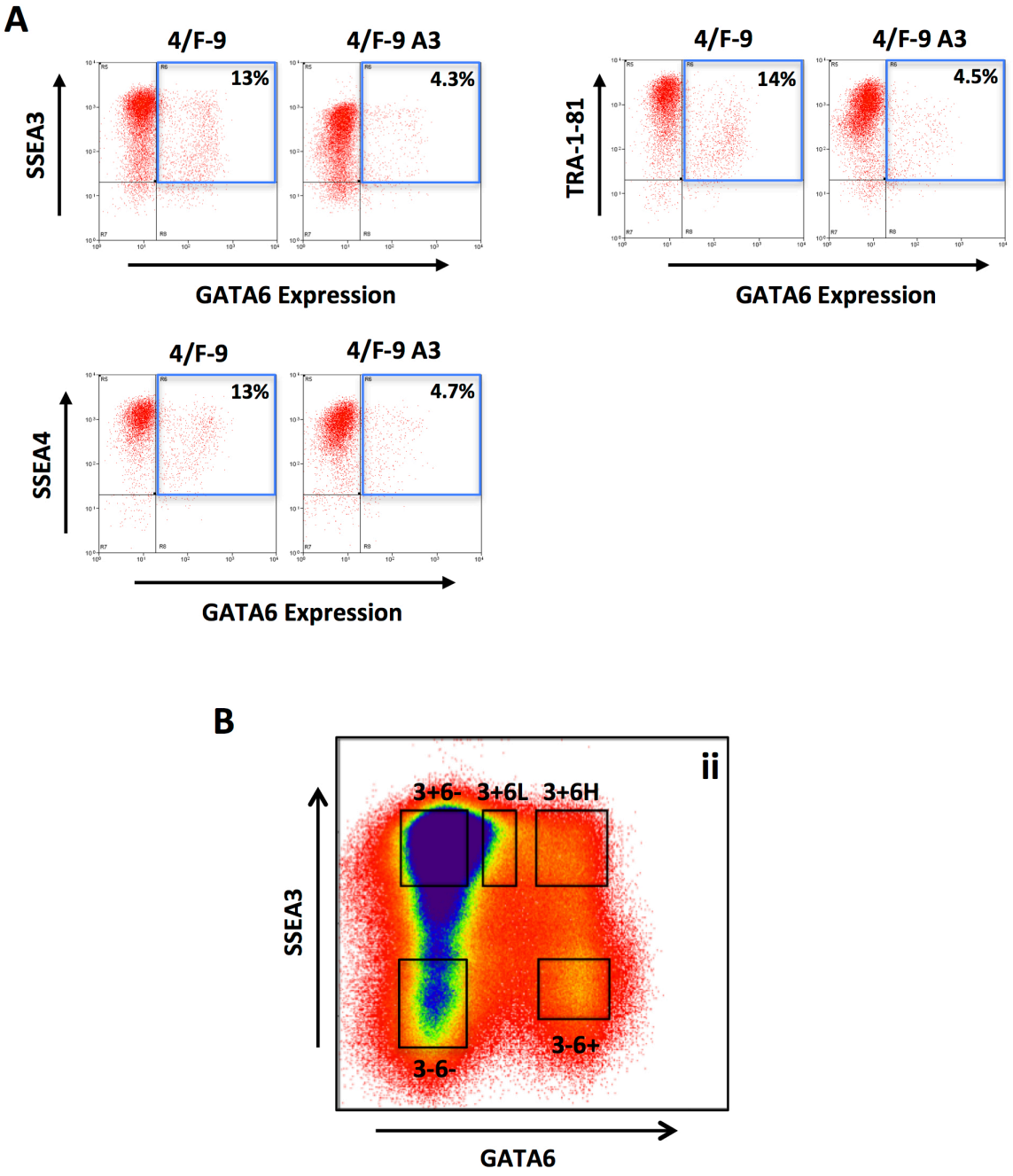
4.2.3 Stem Cell marker and *GATA6* co-expression

Having established that *GATA6* is expressed by a subset of cells within the reporter culture, we went on to investigate whether these cells were simply spontaneously differentiating following an endodermal lineage program, or whether all or a proportion of the cells were *bona fide* stem cells. Firstly, we wanted to assess whether cells of the S4G6 4/F-9 and S4G6 4/F-9 A3 clones co-expressed *GATA6* alongside characterised surface stem cell markers. We used a panel of markers consisting of SSEA-3, SSEA-4 and TRA-1-81 using two clones of the reporter line. As expected, both clones were positive for each stem cell marker tested and in both clones we found a subset of cells co-expressing each stem cell marker tested with *GATA6* **(Fig. 4.4A)**. The percentage of co-expressing cells did not appear to change depending upon the surface marker used, but there was a higher percentage of co-expressing cells in the 4/F-9 clone than the A3 clone. With previous reports describing the interference of antibiotic resistance in reporter lines we continued with the A3 clone that had the neomycin construct removed. We used the stem cell marker SSEA-3 for subsequent experiments as it has been characterised as one of the first markers to be lost upon differentiation, demonstrating its high specificity for the stem cell compartment, but also because SSEA-3 fractionated cells have previously been shown to exhibit heterogeneity in *GATA6* expression (Enver et al. 2005; Gokhale et al. 2015).

Next, we stained the reporter line with SSEA-3 and ran large cell numbers through FACS analysis to identify distinct regions where potential stable cell populations existed. We found a new degree of heterogeneity within the cultures based on SSEA-3 and *GATA6* expression and through visual analysis of the plot identified several distinct, high-density cell populations **(Fig. 4.4B)**. We found that the area with the highest density of cells was the SSEA-3(+) *GATA6*(-) fraction (3+6-). We hypothesised that these cells were the most stable state for cells to reside in under self-renewal conditions as the highest percentage of cells resided within this fraction. Secondly a distinct population of SSEA-3(-),

GATA6(-), cells (3-6-) existed which we classed as cells that had differentiated away from the endoderm lineage. Thirdly, a clear population of SSEA-3(-), *GATA6*(+) cells (3-6+) existed, again cells we assumed to have differentiated, but this time towards an endodermal lineage **(Fig. 4.4B)**. Finally, we found two populations of cells that were co-expressing SSEA-3 and *GATA6*. Firstly, a shoulder off the main SSEA-3(+), *GATA6*(-) population, which we classified as 3+6L, due to low *GATA6* expression, and secondly a separate, distinct population classified as 3+6H due to high *GATA6* expression **(Fig. 4.4B).**

Figure 4.4: *GATA6* is co-expressed with surface stem cell markers on hESCs



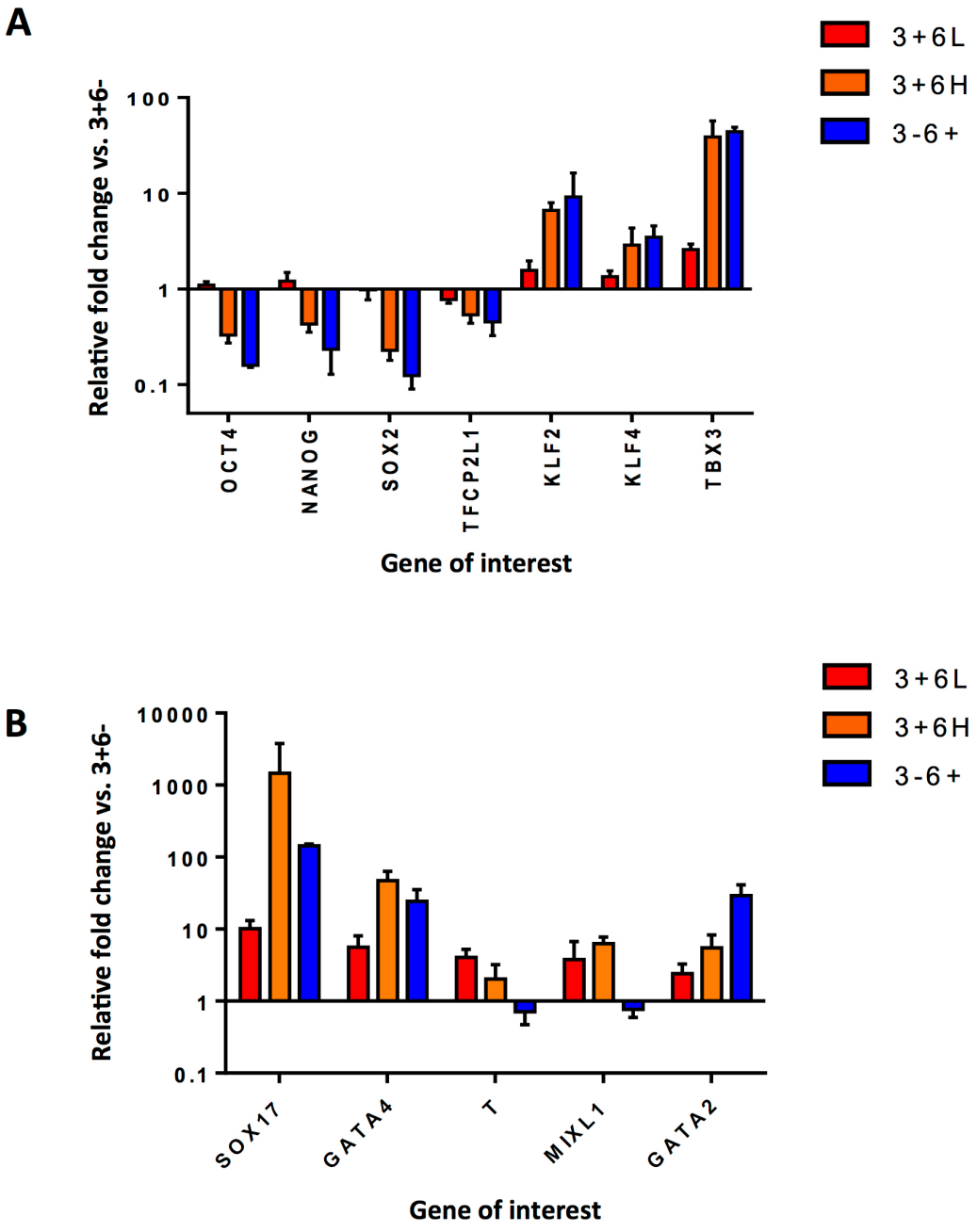
A) FACS analysis of two clones (4/F-9 & 4/F-9 A3) harbouring the GATA6:GFP construct. 4/F-9 A3 represents the clone with the neomycin gene excised. Stem cell & *GATA6* co-expressing cells are found irrespective of the stem cell marker used. **(B)** FACS analysis of the expression patterns of SSEA3 and GATA6 of the 4/D-9 A3 clone cultured in hESC media on MEFs before cell sorting The plot identified distinct cell populations in culture. 3+6- (SSEA3 high GATA6 negative); 3+6L (SSEA3 high GATA6 low); 3+6H (SSEA3 high GATA6 high) and 3-6+ (SSEA3 negative GATA6 high). Blue areas represent high cell density, red areas low cell density.

4.2.4 *GATA6* is co-expressed alongside stem cell genes

Due to our interest in GATA6 expression in stem cells, we took the 3+6-, 3+6L, 3+6H and 3-6+ fractions forward for further analysis. We wanted to investigate whether, across the four cell fractions, there was uniformly high expression of stem cell associated genes, even with increasing *GATA6* expression. We performed qPCR for the core stem cell transcription factors, *OCT4*, *NANOG* and *SOX2*. We found that compared to the 3+6- fraction, the 3+6L fraction had very similar levels of all core transcription factors **(Fig. 4.5A)**. The 3+6H and 3-6+ fractions, however, saw a significant reduction in all genes, most notable within the 3-6+ fraction **(Fig. 4.5A)**. Thus low *GATA6* expression did not appear to correlate with a reduction in stem cell gene expression, however, high *GATA6* expression did. As *GATA6* is expressed in cells of the morula and early blastocyst, we also looked at genes that are expressed at this stage of development within our *GATA6* fractions. Although the 3+6L fraction didn't show a reduction in core stem cell gene expression, this fraction did show a slight reduction in the naïve marker *TFCP2L1* **(Fig. 4.5A)**. *TFCP2L1* was also reduced in the 3+6H and 3-6+ fractions.

Counter-intuitively, the remaining 'naïve' markers, *KFL2*, *KFL4* and *TBX3* were all upregulated in the 3+6L and 3+6H and 3-6+ fractions **(Fig. 4.5A).** To confirm whether GATA6(+) fractions represented more primed-like cells, we then looked at several markers associated with early differentiation. We found that the 3+6L and 3+6H fractions both showed upregulation of *SOX17* and *GATA4* (endoderm), *T* and *MIXL1* (primitive streak) and *GATA2* (mesoderm). The 3-6+ fraction also showed upregulation of *SOX17*, *GATA4* and *GATA2*, but not *T* or *MIXL1*. Endoderm genes were most highly expressed in the *GATA6*(+) fractions. Thus, the 3+6L fraction, although having similar levels of stem cell genes compared to the 3+6- fraction, did show an increase in lineage associated genes. The 3+6H fraction saw an increase in lineage associated genes, but also a decrease of stem cell genes **(Fig. 4.5B).** Thus, the readout of SSEA-3 and *GATA6* appeared to identify discreet, incremental stages of cells transitioning through early differentiation.

Figure 4.5: *GATA6* expressing cells downregulate core stem cell genes and upregulate lineage genes



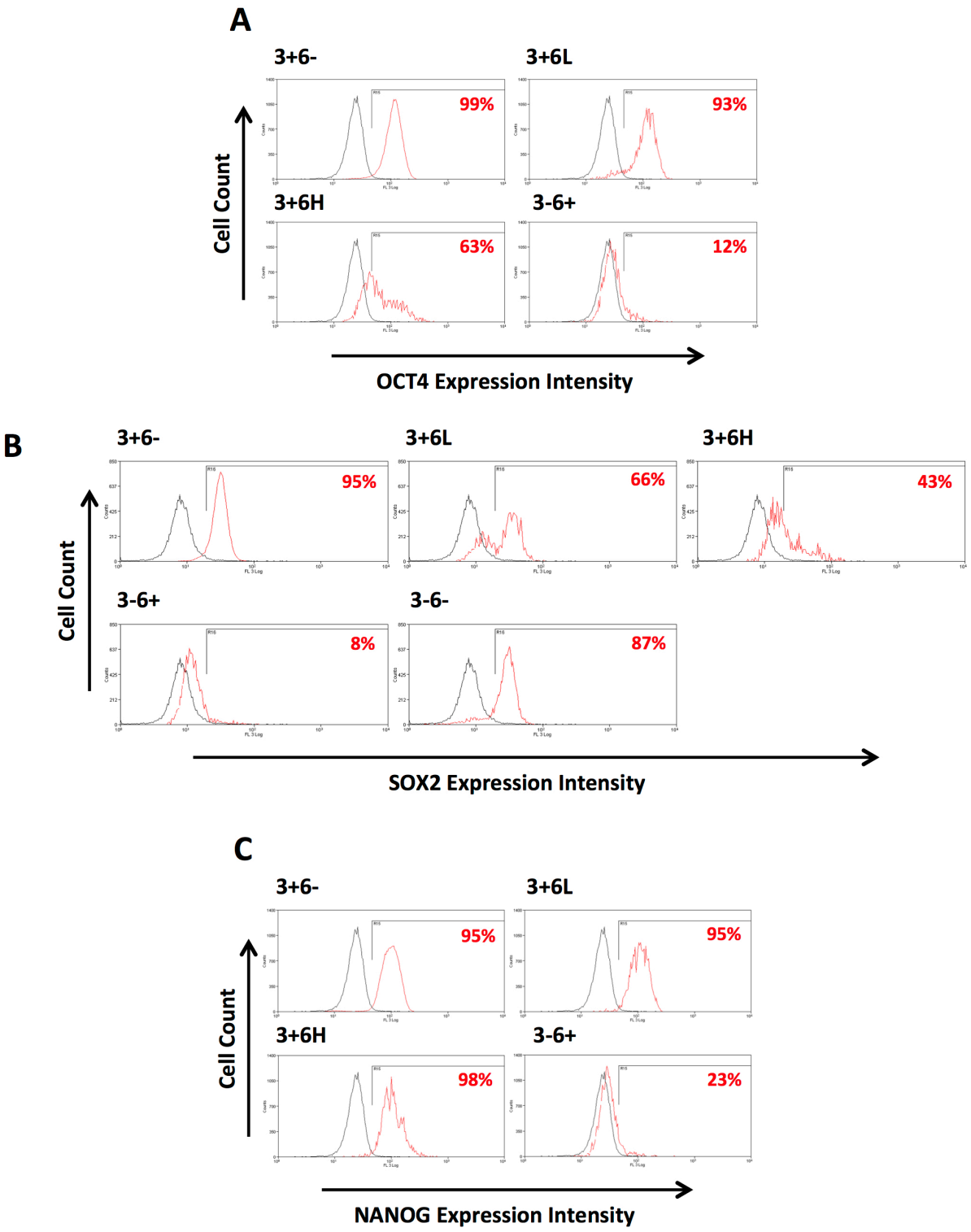
(A) qPCR analysis showing fold change of the 3+6L (red), 3+6H (orange) and 3-6+ (blue) fractions versus 3+6- for core stem cell and naïve associated genes, as well as lineage specific genes **(B)**. Beta Actin was the normalising gene and error bars represent standard deviation of three technical repeats.

4.2.5 Stem cell protein analysis of four fractions

To corroborate our findings from the qPCR, we performed FACS analysis on the protein levels of the core stem cell transcription factors, OCT4, NANOG and SOX2 within each fraction. We found that as *GATA6* intensity increased within cells, the levels of OCT4 protein decreased. The OCT4 profiles of the 3+6- and 3+6L fractions were very similar, with 98% and 93% of cells expressing OCT4A protein respectively, but within the 3+6H fraction, this level dropped to 63% of the cells. This was further reduced in the 3-6+ fraction to only 12% of cells, consistent with the quantified mRNA levels **(Fig. 4.6A)**. We found a slightly different trend with the marker SOX2. The 3+6- fraction again was highly positive at 95%, but the 3+6L was significantly reduced to 66%. This reduction was more apparent in the 3+6H fraction at 43%, and SOX2 was effectively switched off within the 3-6+ fraction with only 8% positive cells **(Fig. 4.6B)**. We also analysed the 3-6- fraction and found that the majority of cells (87%) were expressing SOX2. As well as a pluripotent stem cell marker, SOX2 is expressed in neural stem cells and thus implied the 3-6- fraction to be neural-like cell types **(Fig. 4.6B)**. Therefore it would appear that *GATA6* is an effective marker to efficiently segregate neural and endodermal destined cells in culture. Finally, we analysed the expression of NANOG. It has been reported that GATA6 and NANOG are mutually exclusive and therefore we expected to see a similar downregulation of protein levels as *GATA6* expression increased. We actually saw the opposite such that NANOG protein levels were not downregulated at all until SSEA-3 expression was lost (3+6-: 95%, 3+6L: 95%, 3+6H: 98%, 3-6+: 23%) **(Fig. 4.6C)**.

In conclusion, the quantification of the protein levels of the core stem cell genes correlated with qPCR data, with the exception of NANOG, which appeared to remain within all fractions even when mRNA was reduced.

Figure 4.6: OCT4 & SOX2 protein levels are reduced in *GATA6* expressing cells, but NANOG is maintained



FACS plots for intracellular staining for OCT4 **(A)**, NANOG **(B)** and SOX2 **(C)**. Both OCT4 and SOX2 protein levels are reduced in *GATA6* expressing cells, however NANOG is maintained until the loss of SSEA3. Cells not expressing SSEA3 or *GATA6* are SOX2 positive, suggesting a neural subpopulation in culture.

4.2.6 Lineage specific gene expression within the four fractions

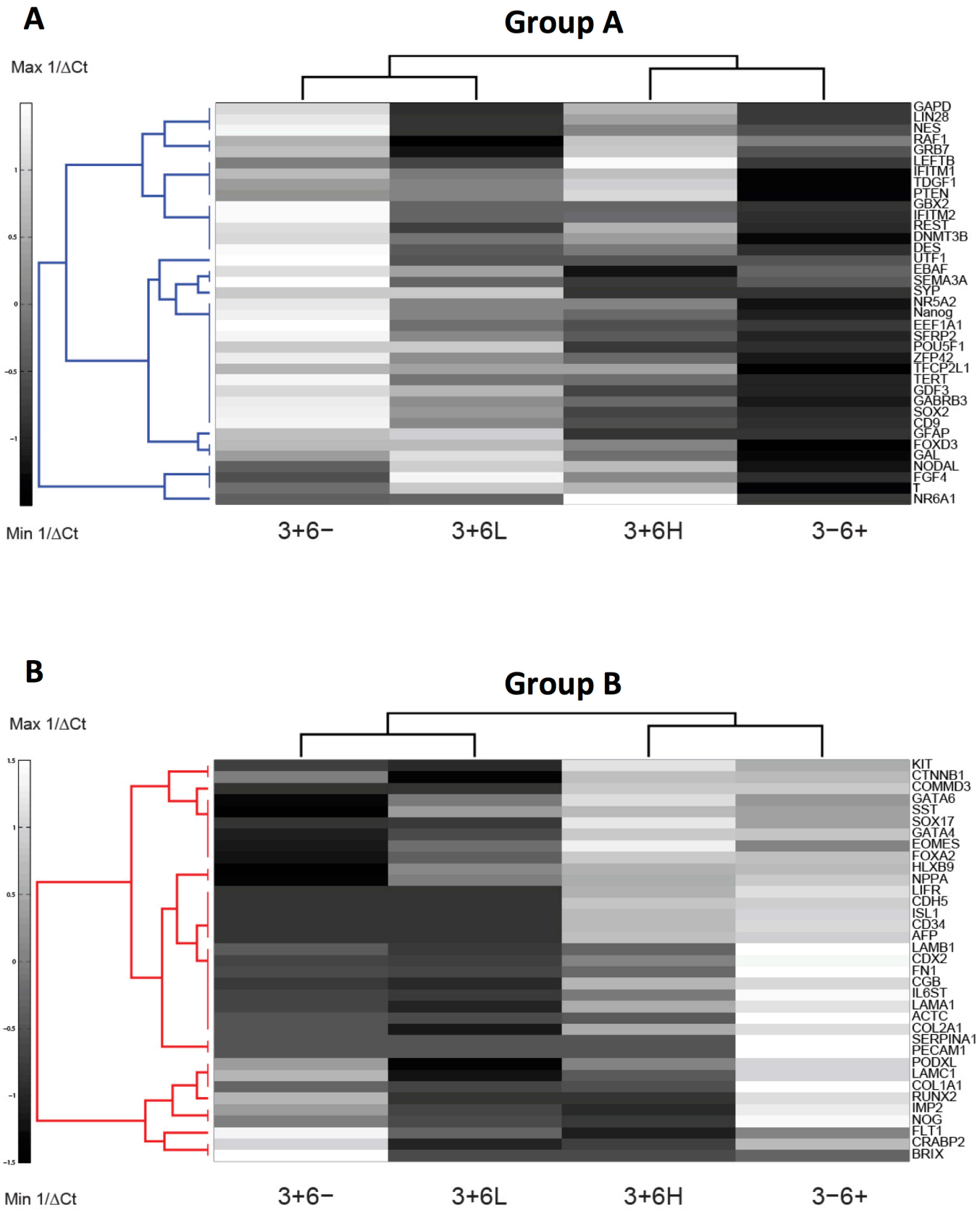
To investigate further the gene expression profiles of each fraction, we ran a low-density TLDA array for stem cell, early and late lineage specific markers. Hierarchical analysis indicated that gene expression within the fractions clustered strongly into two groups. Group A contained predominantly stem cell specific markers **(Fig. 4.7A)**; and group B, predominantly consisting of lineage associated markers (**Fig. 4.7B)**. We found that the 3+6- fraction predominantly showed high expression of group A and low expression of group B and that these expression patterns were reversed in the 3-6+ fraction. This correlated with the notion that the 3+6- fraction represented a more pristine like cell population, and the 3-6+ a differentiated population. Upon analysis of SSEA-3 and *GATA6* co-expressing cells, we found that as *GATA6* expression increased, there was an increased loss of stem cell marker expression, coupled with an increased gain in lineage marker expression. The hierarchical clustering by gene expression analysis across individual fractions showed that the 3+6- and 3+6L fractions clustered together, whilst the 3+6H and 3-6+ fractions clustered together (**Fig. 4.7A**). Therefore, the 3+6L fraction had a gene expression profile more similar to pristine stem cells, whereas the 3+6H fraction has a profile more similar to differentiated cells.

With the knowledge that *GATA6* is involved in endodermal specification during lineage segregation of the ICM, we next wanted to investigate whether the *GATA6*(+) cells were also showing increased expression of other lineage genes and whether these genes were also endoderm associated. Genes were grouped according to their associated germ layers (endoderm, mesoderm or ectoderm) and each fraction was analysed for individual germ layer expression profiles (**Fig. 4.7C**). We found that the 3+6- and 3+6L fractions did not show any notable upregulation in genes associated with any one specific germ layer. However, within the 3+6H population we noted a significant upregulation of genes associated with the endoderm lineage but not genes indicative of mesoderm and ectoderm. Similarly, the 3-6+ fraction showed increased endoderm lineage gene

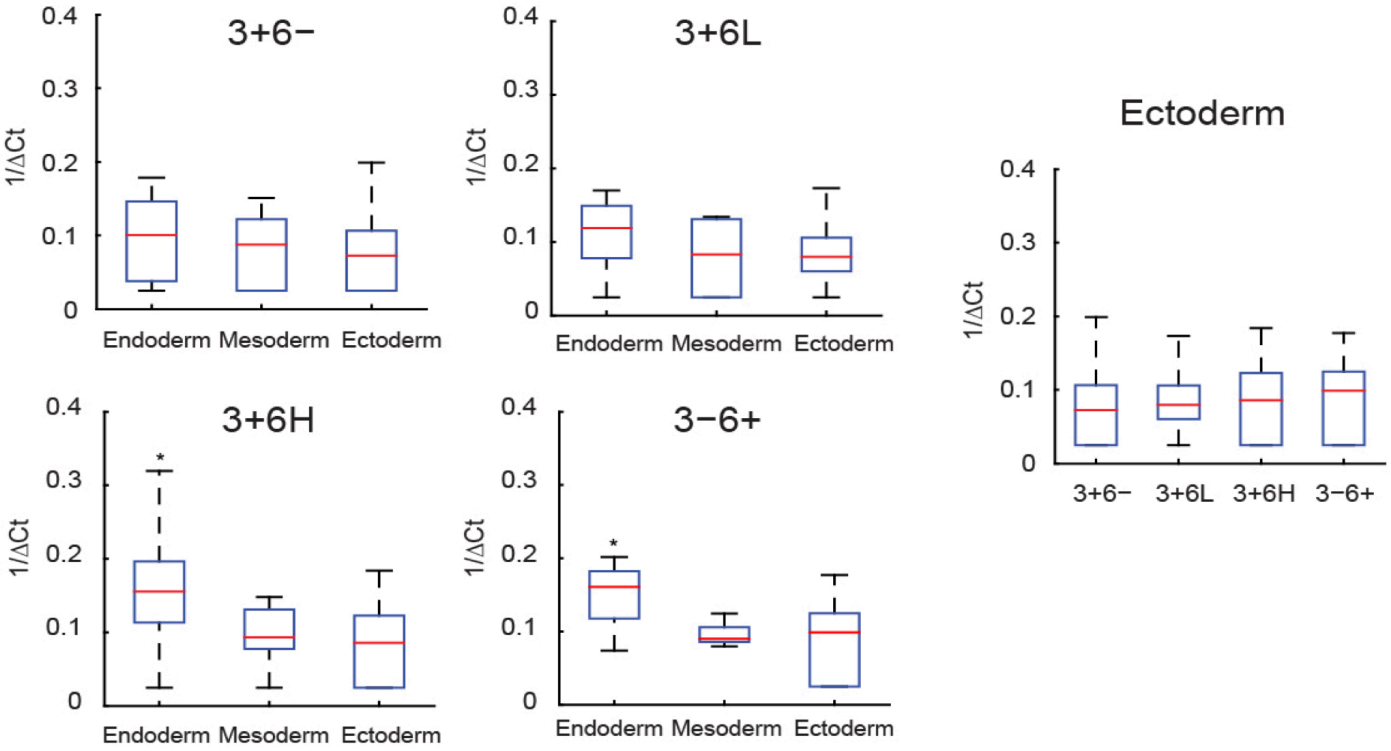
expression. Ectoderm associated activity was unchanged across fractions (**Fig. 4.7C**).

Overall, these findings indicate that GATA6 appears to correlate with a gene expression profile more representative of a differentiated cell type, which is exemplified as *GATA6* expression increases. Nevertheless, stem cell genes were still expressed in *GATA6*(+) cell fractions, which would indicate that *GATA6* expression may not necessarily correlate with differentiation. Additionally, cells expressing *GATA6* appeared to preferentially upregulate other genes associated with the endoderm lineage only.

Figure 4.7: *GATA6*+ cells upregulate other endoderm, but not mesoderm or ectoderm genes



C



(A-B) 3+6-, 3+6L, 3+6H and 3-6+ fractions were run on low density TLDA array card to assess their gene expression profiles. Expression is based on Delta-CT values using Beta-Actin as the control. Overall the data clustered in to 2 main subsets of genes. **'Group A'** representing mostly stem cell specific genes, and **'Group B'** representing mostly lineage genes. As GATA6 expression increases, profiles shift to a more differentiated expression profile. **(C)** Each fraction was analysed for upregulation of genes associated with each primary germ layer. Box plots were generated by taking the 1/Delta-CT value for genes associated with either endoderm, mesoderm or ectoderm. Levels indicated median values, box indicates values between the 25th and 75th percentile, whisker length represents min & max values. Kruskal-Wallis statistical test results are indicated for p values <0.05. Only the 3+6H fraction showed significant upregulation of endodermally associated genes. Genes in other germ layers were not upregulated.

4.2.7 Functional testing of GFP expressing cells

Having established that a subset of cells in normal culture conditions exist that co-express the surface stem cell marker SSEA-3 and the lineage marker *GATA6*, and that endodermal associated genes are significantly upregulated within the 3+6H and 3-6+ fractions, we wanted to investigate whether functional differences existed across each of the fractions in terms of their self-renewal potential. We performed functional testing of the four cell fractions (3+6-, 3+6L, 3+6H & 3-6+) to assess differences in self-renewal capacities of the sub-populations using a quantitative single cell clonogenic assay. FACS sorting obtained pure cell populations, and single cells were seeded at a clonogenic density of 500 cells/cm^2 in the presence of 10 μM Y-27632 for 24h. After 24h, media was changed to remove the inhibitor. Colonies were left to develop over 4 days, and resulting colonies were fixed and stained using *OCT4* as a marker for stem cells. If the resulting colonies contained at least one *OCT4* positive cell, we classed that colony as having originated from a stem cell. *OCT4*(+) colonies resulted from each of the four fractions, including the 3-6+, which we previously classed as a differentiated population **(Fig. 4.8A).** The efficiency of cloning however was influenced as *GATA6* expression increased. The 3+6- fraction was the most clonogenic with a cloning efficiency of 6.1% (± 0.8%). Consistent with the clustering of 3+6- with 3+6L from the qPCR analysis, the 3+6L fraction cloned with a very similar efficiency of 6.3% (±0.5%). We saw a statistically significant reduction in cloning efficiency of the 3+6H fraction and the 3-6+ fraction at 2.5% (±0.3%) and 0.2% (±0.07%) respectively **(Fig. 4.8B).** In order not to rely solely on the use of OCT4 as a stem cell marker, we repeated the clonogenic assays in exactly the same way, but this time stained with SOX2. Firstly, as we saw with OCT4, each fraction was able to generate SOX2(+) colonies **(Fig. 4.8A).** We found very similar patterns of cloning efficiencies of our 4 fractions when using SOX2 as the stem cell marker. The 3+6- fraction proved most clonogenic at 7.2% (±2.2%), and the 3+6L fraction cloned at 6.2% (±0.5%). As in the case of OCT4, the cloning efficiency of the 3+6H fraction dropped dramatically to 2.2% (±0.1%) and the 3-6+ cells remained the least

clonogenic at 0.4% (±0.2%), both of which were statistically significant to the 3+6- fraction **(Fig. 4.8B)**. All analyses were automated to prevent any bias in colony classification.

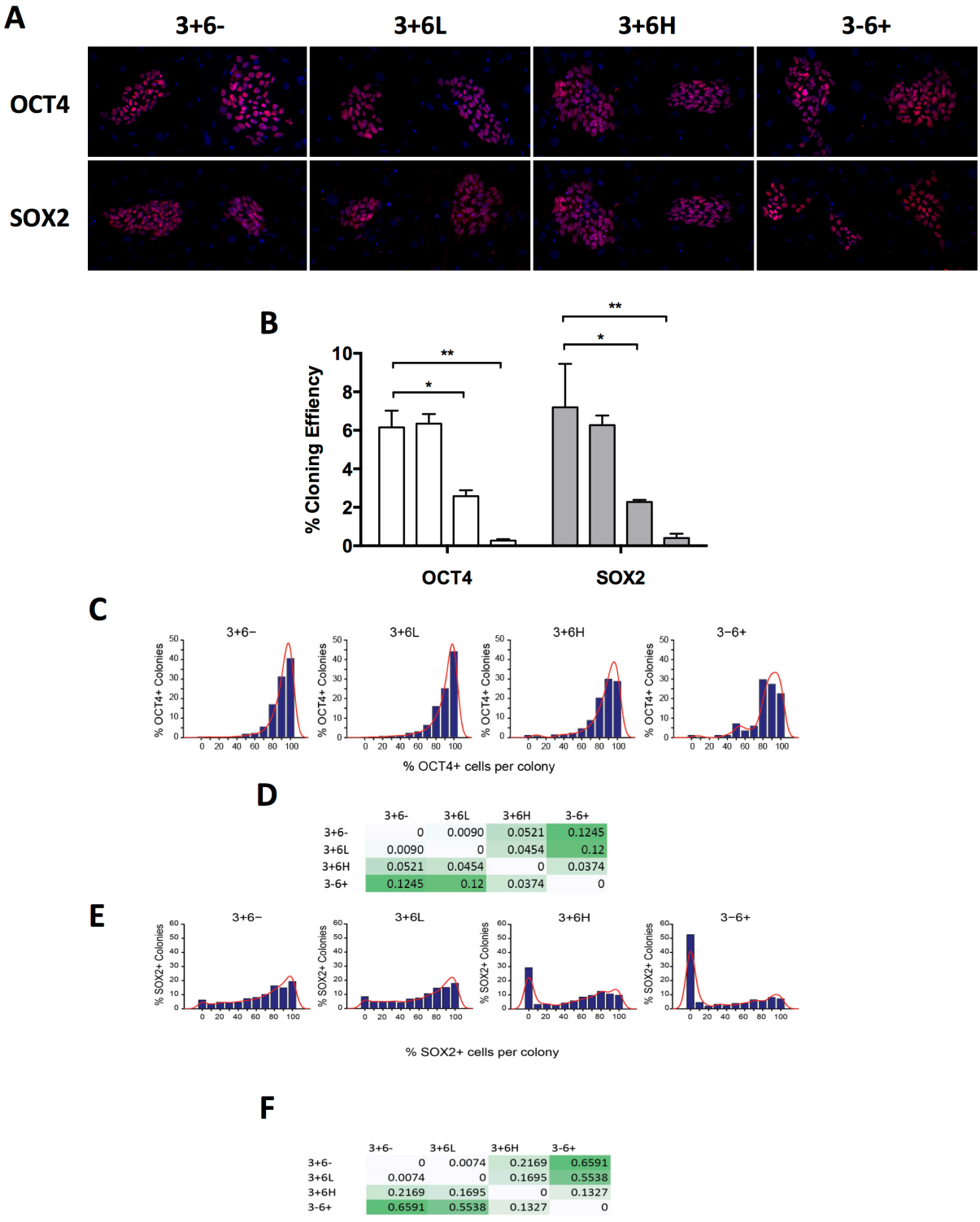
Importantly, this assay showed that a large proportion of cells that expressed SSEA-3 together with the lineage marker *GATA6* at low levels, and a smaller proportion of cells expressing *GATA6* at high levels, were clonogenic stem cells capable of self-renewal. The increasing expression of *GATA6* did alter the functional behaviour of the cells, as the percentage of stem cells within the 3+6H fraction decreased. The loss of the surface marker SSEA-3 then correlated with, in the most part, the inability to self-renew. We found that the functional data correlated well with the hierarchical clustering of the qPCR data for each fraction in that the 3+6- and 3+6L behaved similarly, whilst the 3+6H and 3-6+ behaved similarly.

We then performed a detailed analysis for the percentage of OCT4+ cells within stem cell colonies derived from each fraction. We noticed the appearance of colonies that had a decrease in the overall percentage of OCT4+ cells in 3+6H and 3-6+ fractions **(Fig. 4.8C)**. We quantified changes in the distribution of OCT4+ cells using Kullback-Leibler symmetric divergence. This measure is low for similar distributions and high for divergent distributions. As expected, the level of change was most prominent in 3+6H and 3-6+ fractions compared to 3+6- and 3+6L cells in agreement with the reduced cloning efficiency levels (**Fig. 4.8D)**. We performed the same analysis on colonies stained with SOX2 from each of the four fractions, where we saw the same trend as OCT4. There was an increase in the number of colonies with a reduction in SOX2 expressing cells in the 3+6H and 3-6+ compared to 3+6- and 3+6L (**Fig. 4.8E**). The Kullback-Leiber symmetric divergence showed that the most distinct fractions were again the 3+6H and 3-6+ when compared to 3+6- and 3+6L (**Fig. 4.8F**), which again correlated well with decreasing cloning efficiency levels **(Fig. 4.8B)**.

In conclusion, we found that the 3+6H and 3-6+ fractions generated more colonies that consisted of fewer *OCT4*(+) and *SOX2*(+) cells than the 3+6- and

3+6L fractions. Thus, the stem cells within these fractions showed a higher propensity to differentiate.

Figure 4.8: *GATA6* expressing cells generate OCT4 & SOX2 positive colonies



(A) 3+6-, 3+6L, 3+6H and 3-6+ cells were FACS sorted and seeded at clonogenic density. After 4 days, resulting colonies were fixed and stained with Hoechst (blue) and mABs to OCT4A and SOX2 (red). Single cells from each fraction gave rise to both OCT4+ and SOX2+ cell containing colonies with hESC morphology. Images were take using an InCell Analyzer. **(B)** Quantification of the absolute cloning efficiencies on OCT4 or SOX2 positive cell containing colonies shows that only the 3+6H & 3-6+ fractions, clone significantly worse than the 3+6- fraction. Secondary only controls were used to set a threshold for OCT4 and SOX2 positivity. **(C)** Proportion of *OCT4*(+) cells in *OCT4*(+) colonies Counts are shown as a bar plot (blue) with superimposed estimated nonparametric distribution (red). **(D)** Kullback-Leibler pairwise symmetric divergence between OCT4-associated distributions shown in (C). **(E, F)** Same analyses as in (C) and (D) for SOX2. Green indicates higher divergence between fractions.

4.2.8 Sub-Cloning of GATA6 expressing cells

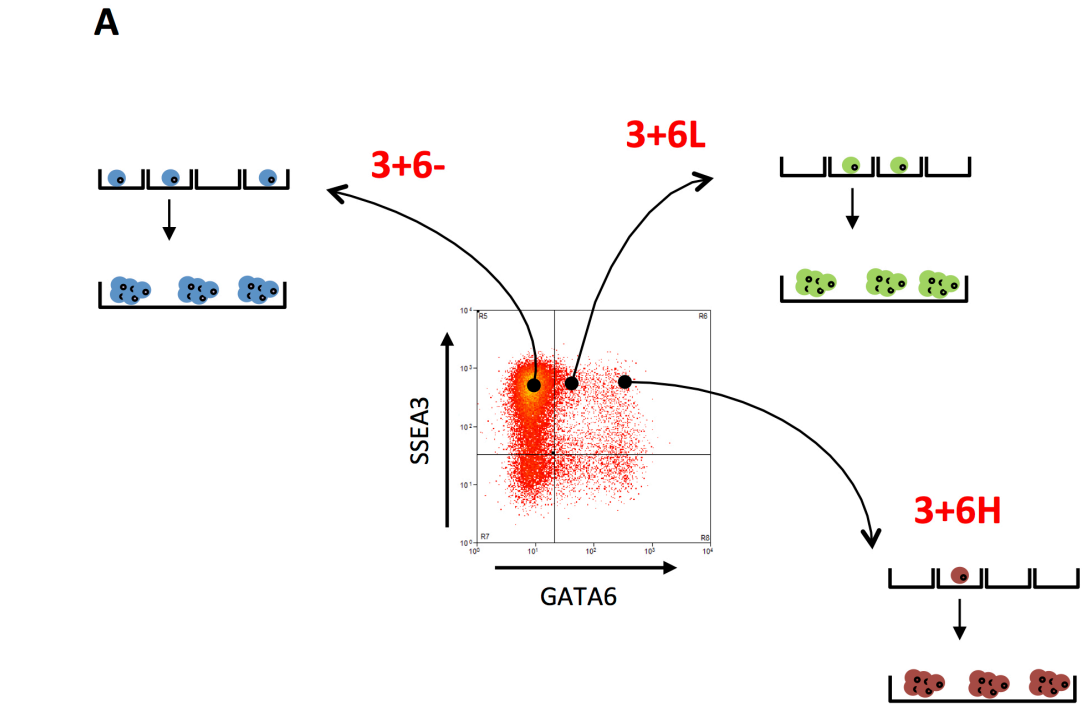
Having demonstrated through the use of a single cell quantitative clonogenic assay that a proportion of *GATA6* expressing cells had the ability to produce colonies positive for *OCT4* and *SOX2*, we then wanted to investigate the long-term functional potential of these cells. We wanted to ascertain whether *GATA6*(+) hESCs retained the ability for efficient multi-germ layer differentiation, or whether cells represented a more restricted cell type, and whether *GATA6* expressing cells showed culture reconstitution capacity. To address these points, we sub-cloned the S4G6 4/F-9 A3 line from cells of the 3+6-, 3+6L and 3+6H fractions, using high stringency single cell deposition **(Fig. 4.9A)**. Due to the very low cloning efficiency of 3-6+ cells, we excluded this fraction from sub-cloning experiments.

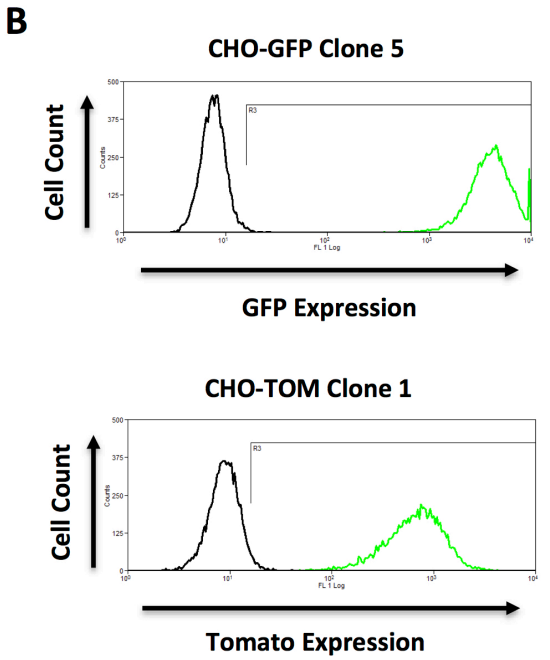
Achieving a high sorting purity in this assay was imperative to ensure that the sub-clones were truly derived from the said cell type of origin. We therefore developed a control assay to run alongside the sorting of the four hESC fractions to track any mis-classification during the sorting procedure. By using CHO cells, which have a high plating efficiency, we were able to over-estimate any miss-classification that occurred during the sorting process, and accurately predict the level of error during hESC single cell deposition. CHO cells were transfected with either pCAG-TOMATO or pCAG-GFP plasmids so cells could be easily tracked. Clonal CHO lines harbouring the GFP or Tomato cassette were then sub-cloned and the brightest clones were picked and maintained for experimental use. **(Fig. 4.9B)**. Single CHO-GFP and CHO-TOM cells were mixed together and then CHO-GFP cells were sorted into half of a 96-well plate and CHO-TOM cells into the second half of the plate. These cells were fixed 3 days after sorting and imaged. We then quantitated the number of colonies per well, and also checked that the correct CHO reporter had been sorted. We found that the vast majority of the wells contained single colonies with no instances of mis-classification. One well was found to contain the wrong CHO cell giving a mis-classification of 0.6% overall. Doublets were also found in 5.6% of the wells, but

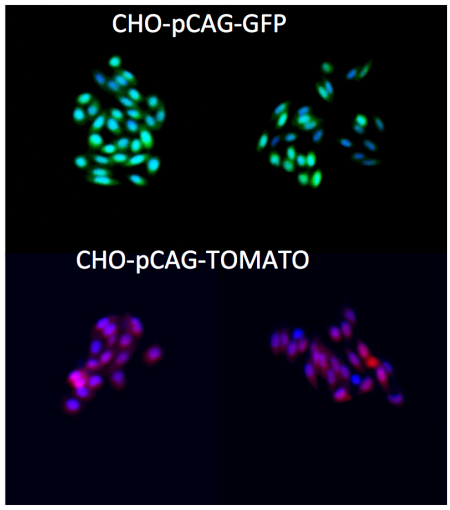
each colony was derived from the correct CHO cell type. We were therefore confident that hESC clones derived from the four fractions were truly from the said fraction. **(Fig 4.9C).**

Single cells from the 3+6-, 3+6L and 3+6H subsets were sorted into single wells of a 96-well plate in several conditions. Firstly, we tried to derive sub-clones on feeders and under feeder-free conditions. Using MEFs and hESC media, we were able to derive sub-clones with cloning efficiencies of 15%, 8% and 2.5% for the 3+6-, 3+6L and 3+6H fractions respectively **(Fig. 4.10A)**. We were unable to derive any sub-clones on feeder-free conditions irrespective of the conditions tried **(Fig. 4.10B)**. The number of cells sorted and the resulting clones, based on colony morphology are quantified in **Fig. 4.10C**. Six clones that had hESC colony-like morphology from each fraction were randomly picked and carried forward for continued culture. Clones in continued culture showed typical hESC morphology with dense, compacted colonies, and individual cells showing a high nuclear to cytoplasmic ratio **(Fig. 4.10D)**.

Figure 4.9: FACS analysis of CHO clones with GFP or Tomato integrated plasmids demonstrate accurate cell sorting





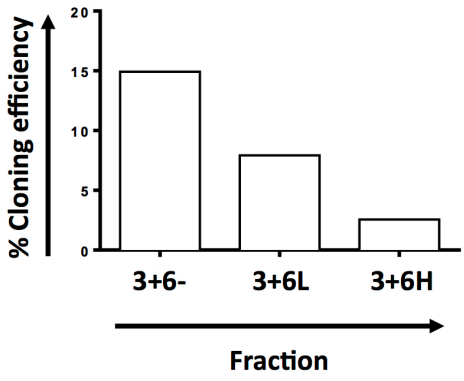


	Miss Class	Singlets	Doublets	No cells
Number	1	105	9	46
Percentage	0.6	65.2	5.6	28.6

(A) Schematic representing the single cell deposition process to derive subclones from 3+6-, 3+6L and 3+6H fractions. **(B)** FACS analysis of two CHO clones harbouring the pCAG-GFP plasmid (top) or pCAG-Tomato plasmid (bottom). CHO clones were derived by single cell deposition and brightest clones were maintained for experimental use. **(C)** Fluorescent images of colonies from the CHO-GFP and CHO-Tomato lines 3 days after single cell sorting. Quantification of the misclassification shows only 0.6% rate of error in the sorting process.

Figure 4.10: Cloning efficiency of 3+6-, 3+6L and 3+6H after single cell deposition

A



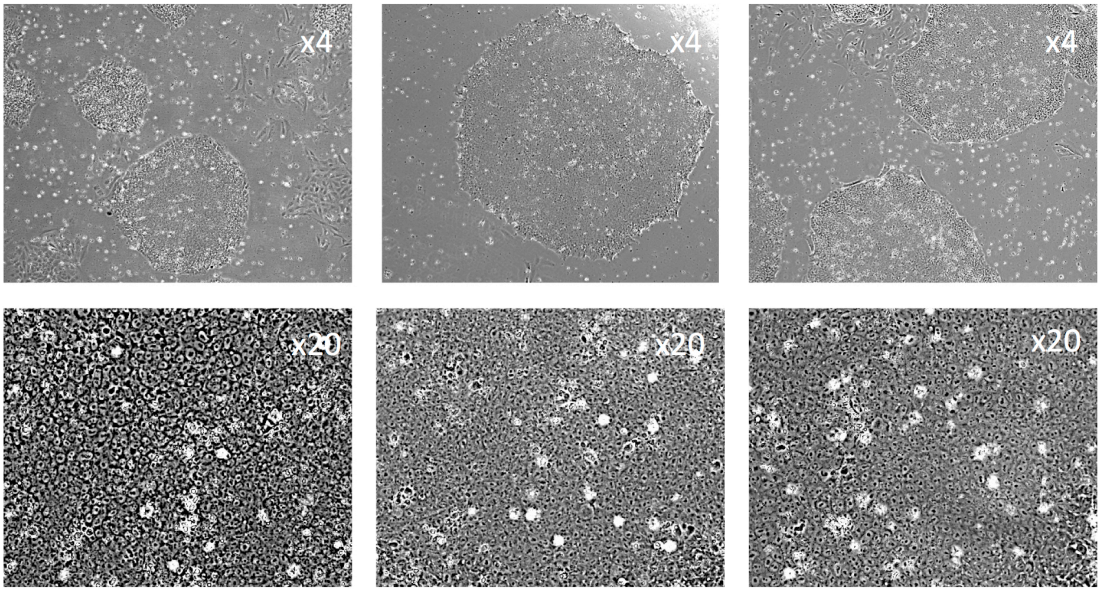
B

	Media	Matrix
Feeder	hESC media	MEFs
Feeder free	mTesr	Matrigel
	E8	Geltrex
	mTesr conditioned	Matrigel

C

Fraction	Number of Cells sorted	Number of Clones	% Cloning Efficiency
3+6-	288	43	15
3+6L	960	76	8
3+6H	1920	49	3

D



(A) Single cells were sorted into single wells of a 96-well plate and resulting colonies were imaged. Morphologically resembling colonies from hESCs were counted and the graph shows absolute cloning efficiency. 3+6- cloned at 15%, 3+6L at 8% and 3+6H at 2.5%. **(B)** Conditions used to assess the ability for subcloning of the GATA6 reporter line. Only the feeder conditions gave rise to sub-clones whereas no survival was seen on feeder free conditions. **(C)** Quantification of the number of cells sorted and the subsequent number of hESC-like colonies after 10 days. 6 from each fraction were randomly picked and maintained. **(D)** Brightfield images at x4 magnification (top) and x20 magnification (bottom) of a representative clone from each of the three fractions after single cell deposition.

4.2.9 Characterisation of sub-clones

We then characterised these sub-cloned lines to investigate whether they behaved in a similar manner to the unsorted parent population. Firstly, the expression of a series of stem cell-associated surface markers were analysed on the sub-clones and compared to the parental line. Four clones irrespective of the cell of origin expressed SSEA-3, TRA-1-81 and SSEA4, at similar levels to the unsorted parental population **(Fig. 4.11A)**. Subtle differences in the levels of expression for each marker, in particular, SSEA-3 were apparent between clones and the parental line, but were more likely differences in culture conditions as opposed to something more intrinsic to clones, as not all of the four clones analysed from the same fraction showed such changes from the parental line.

We then looked at the expression levels of core stem cell transcription factors *OCT4, NANOG, SOX2* and *REX1* in two clones from each fraction. We found that all clones irrespective of their cell of origin, expressed similarly high levels of all stem cell genes analysed, and looked similar to the unsorted parental line **(Fig. 4.12A).** Additionally, we analysed the levels of lineage specific genes to determine whether sub-clones from *GATA6*(+) cells were showing higher rates of spontaneous differentiation than the parental lines. We found that all clones had low expression levels for genes indicative of each primary germ layer, and were again similar to the unsorted parental line **(Fig. 4.12B)**.

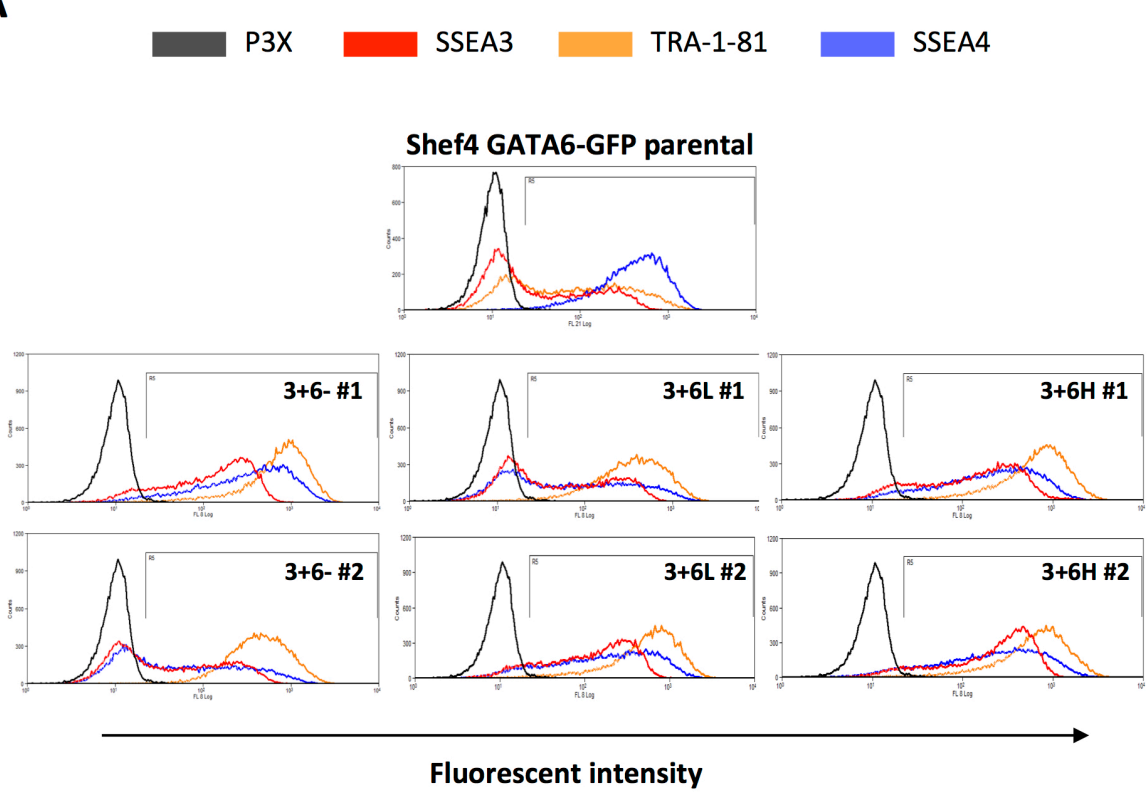
We then wanted to ensure that our clones had remained pluripotent and that they did not represent a more restricted endodermal progenitor cell type. To test pluripotency, we made EBs from two clones of each fraction under a defined EB protocol (see chapter 1). Resulting EBs were morphologically similar between all clones and the parental line and formed round, dense cellular aggregates **(Fig. 4.13A)**. Upon qPCR analysis EBs from the parental and all sub-cloned lines from each fraction had downregulated stem cell associated genes **(Fig. 4.13Bi)** and had upregulated genes associated with both the mesoderm **(Fig. 4.13Bii)** and ectoderm lineages **(Fig. 4.13Biii)** with high efficiency. Clonal

lines from each cell fraction were therefore pluripotent cell lines, and did not represent endoderm restricted cell types.

Having established that clonal lines could be derived from *GATA6* expressing cells, and that these clones were indeed pluripotent and resembled closely the parental line, we sought to assess the ability of SSEA-3(+) *GATA6*(+) expressing cells to convert back to an SSEA-3(+) *GATA6*(-) state. We stained our clonal lines with SSEA-3 to assess the redistribution of the original heterogeneity within the FACS plots of the parental line before sub-cloning. We found that clones, irrespective of their initial *GATA6* status, or the conditions in which they were maintained, were able to redistribute entirely the starting culture heterogeneity **(Fig. 4.14A).** To confirm the FACS analysis, we performed qPCR for the *GATA6* gene comparing two clones from each fraction directly after sorting with established clones from each fraction. We found negligible *GATA6* expression in the 3+6- fraction directly after sorting, but in clones the expression of *GATA6* had increased, indicating a re-establishment of *GATA6* expressing cells in culture. In the 3+6H fraction directly after sorting, *GATA6* levels were higher than in all clones analysed. Cells from this fraction had therefore inter-converted to a *GATA6*(-) state upon culture **(Fig. 4.14B).** The 3+6L fraction showed the same levels of GATA6 directly after sorting and within clones. Analysis of GATA6 gene expression therefore corroborated the re-expression or reduction in GATA6 of clones from each fraction within the FACS analysis.

Figure 4.11: Sub-Clones derived from the 3+6-, 3+6L and 3+6H fractions express stem cell surface markers

A

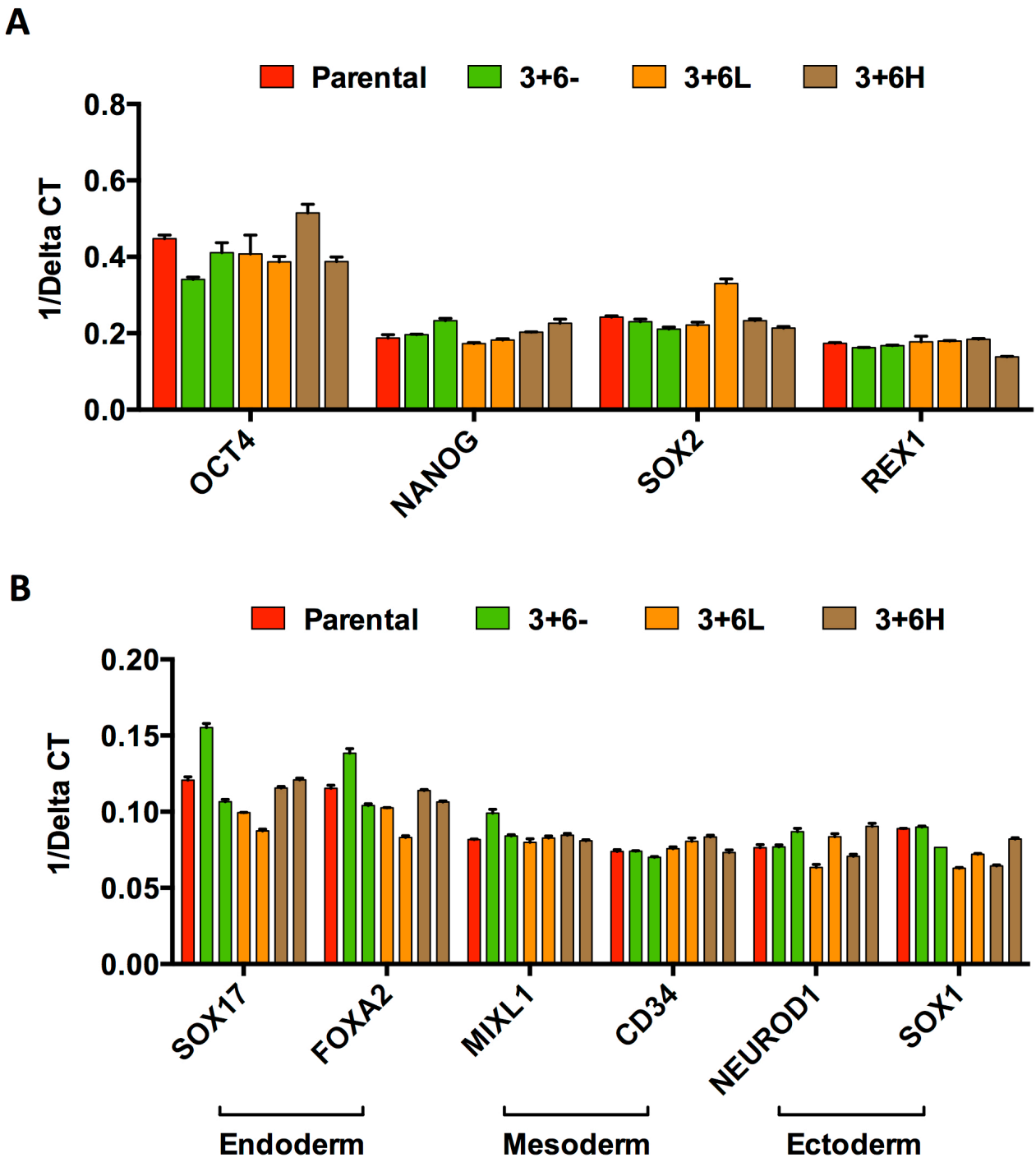


B

Clone	% SSEA3+	% SSEA4+	% TRA-1-81+
3+6- #1	86	99	95
3+6- #2	51	99	95
3+6L #1	53	99	57
3+6L #2	86	99	88
3+6H #1	86	99	92
3+6H #2	92	99	91

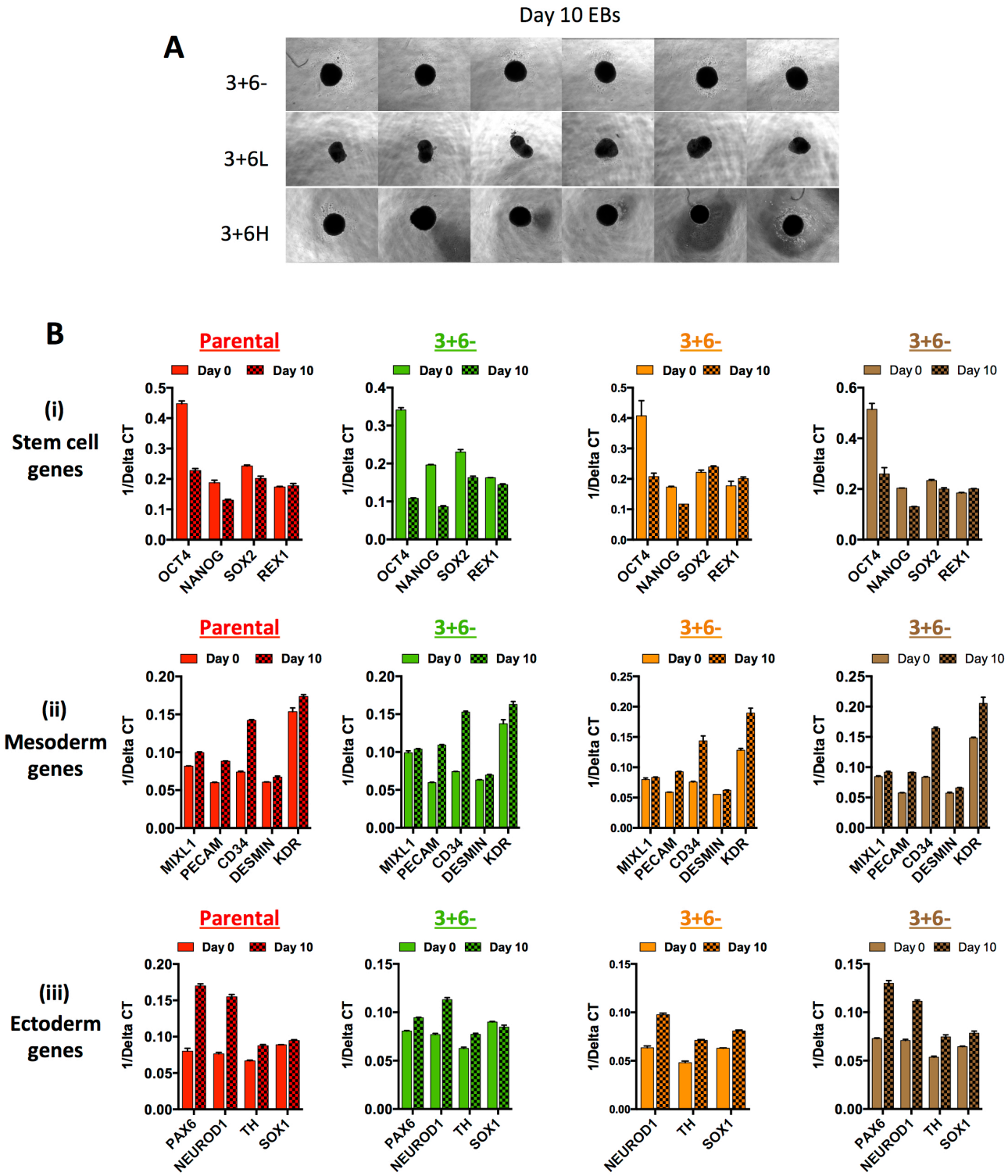
(A) FACS analysis of sub-clones from each fraction stained with the stem cell surface markers SSEA3 (red), TRA-1-81 (orange) and SSEA4 (blue). Black plots represent the negative control P3-X. **(B)** Quantification of the percentage positive cells for each of the surface markers for clones grown in feeder and feeder free conditions.

Figure 4.12: Clones derived from the 3+6-, 3+6L and 3+6H fractions show high expression of stem cell genes



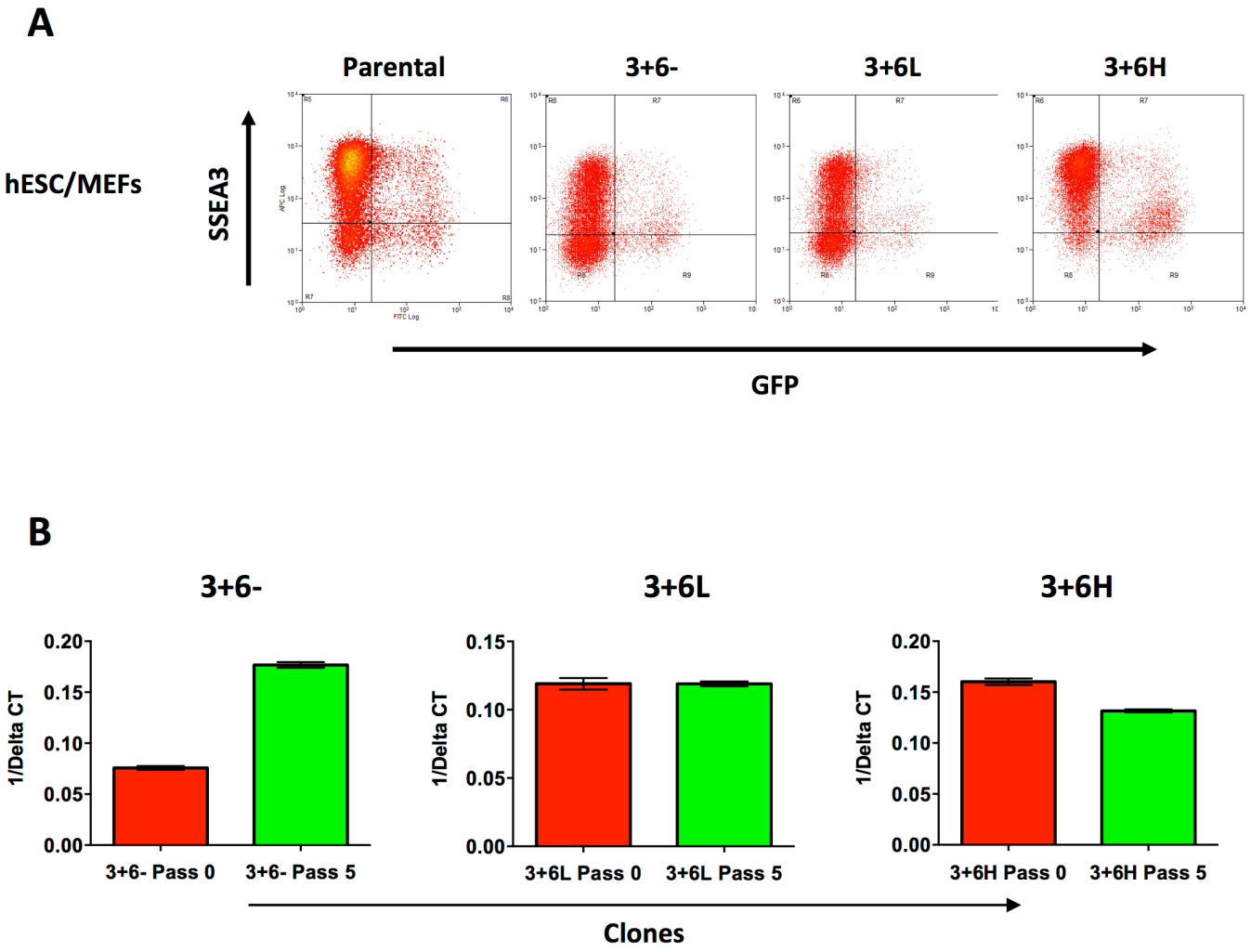
(A) qPCR analysis of the core stem cell genes *OCT4*, *NANOG*, *SOX2* and *REX1* for the parental (red bar) and 2 clones from the 3+6- (green bars), 3+6L (orange bars) and 3+6H (brown bars) fractions. Data is shown as Delta-CT and error bars are from 3 technical replicates. **(B)** qPCR analysis performed in the same manner as (A), except looking at lineage specific genes. Values are represented as 1/Delta CT and error bars represent standard deviation of three technical repeats. Clones are colour coded according to fraction of origin as in (A).

Figure 4.13: Clones from the 3+6-, 3+6L and 3+6H fractions are pluripotent



(A) Brightfield images at x4 magnification of day 10 EBs generated from cells of the 3+6-, 3+6L and 3+6H fractions. **(B)** qPCR analysis for stem cell, **(i)** mesoderm and **(ii)** ectoderm **(iii)** associated genes in undifferentiated control cells (Day 0) and day 10 EBs differentiated for 10 days (day 10) from clones of each fraction colour coded as red (parental), green (3+6-), orange (3+6L) and brown (3+6H). Amplification is shown as 1/Delta-CT and the normalising gene was beta-actin. Error bars represent standard deviation of three technical repeats.

Figure 4.14: FACS analysis demonstrates substate interconversion



(A) FACS analysis of sub-clones from the 3+6-, 3+6L and 3+6H showing inter-conversion of SSEA3/*GATA6* sub-states. *GATA6*(+) cells convert to *GATA6(-)* states, and *GATA6(-)* cells convert to *GATA6*(+). FACS plots show 1 representative example of a clone from each cell fraction. **(B)** qPCR analysis of the GATA6 gene in clones derived from the 3+6- (left), 3+6L (middle) and 3+6H (right) fractions. Red bars indicate 1/Delta-CT for GATA6 expression in cells from each fraction directly after sorting and green represent sub-clones after a minimum of 5 passages in culture. Data is shown as 1/Delta-CT using beta-actin as the normalising gene. Error bars represent standard deviation of 3 technical repeats.

4.2.10 Discussion

A large amount of published data is now convincingly pointing towards the idea that populations of hESCs are heterogeneous and are composed of cells that differ, at times quite drastically, in terms of their gene expression profiles, surface marker expression patterns and their response to signalling activation/repression. This has evolved the idea of the existence of discreet substates within the stem cell compartment that functionally alter a single cell's interpretation of extrinsic inputs and consequently functional behaviour. However, most of the work that has been done on hESC heterogeneity has mainly focussed on exploring surface stem cell markers and transcription factors. Here we provide evidence that the lineage marker *GATA6* also shows a large degree of expression heterogeneity within hESCs in culture, and this heterogeneity proves to have functional consequences at the single cell level. We were interested in looking at the early endodermal marker *GATA6,* as this marker has previously shown to be co-expressed at high levels together with the stem cell markers *OCT4* and *NANOG* (Gokhale et al. 2015). Our results show that *GATA6* does appear to be expressed in culture and its co-expression with SSEA-3, typically within the range of 5-10% of the total population, gave preliminary evidence of functionally discreet *GATA6*(+) substates within the stem cell compartment. We did see that the 4/F-9 clone had higher levels of co-expressing cells than the neo-excised clone (4/F-9 A3). Although the differences between the clones could have been due to slightly different culture conditions at the time of analysis, it has been shown that PGKneo can directly alter targeted gene transcription as well as closely associated genes within knock-out or reporter lines (Olson et al. 1996). As a precautionary measure, we therefore proceeded with the clone whereby PGKneo had been removed (4/F-9 A3).

Firstly, we assessed the correlation between *GATA6* mRNA and GATA6 protein expression. We found that there was no linear correlation with *GATA6* mRNA and GATA6 protein expression, although protein did increase as GFP increased.

Cells were therefore transcribing from the GATA6 locus, but translation of these transcripts was low.

We then looked at the gene expression profiles of SSEA-3 and *GATA6* co-expressing cells. During very early mouse development, GATA6 has been detected from the 4-8 cell stage of the embryo (Plusa et al. 2008), alongside the pluripotency marker NANOG, and this co-expression continues until the early ICM where GATA6 cells lose NANOG expression and become destined for a primitive endodermal fate (Koutsourakis et al. 1999). A recent study by Chan et al reported conditions in which 'naïve-like' hESCs expressing *GATA6* could be stabilized and it was possible therefore that our *GATA6(*+) cells could represent a naïve cell type. According to our qPCR data, the 3+6L and 3+6H fractions did not appear naïve in nature, due to the downregulation of core stem cell associated genes and the naïve marker *TFCP2L1*, coupled with the upregulation of lineage associated genes. Although *KLF2*, *KLF4* and *TBX3*, all naïve associated markers, were upregulated in *GATA6*(+) fractions, this is probably attributable to the role of these transcription factors in early mesoderm and extraembryonic endoderm specification respectively. Furthermore, all of these lineages involve *GATA6* (Chiplunkar et al. 2013; Aksoy et al. 2014; Teo et al. 2011). SSEA-3(+) *GATA6*(+) cells therefore appear to represent primed cells, which are mediating the transition between the 3+6- and 3-6+ profiles. This is consistent with reports of hESCs in KO/SR and MEF conditions exhibiting a functional hierarchy of pluripotent cells in culture, whereby cells can be fractionated according to their expression of the surface markers CD9 and GCTM2, whereby high expression correlates with more pristine, clonogenic stem cell states. (Gokhale et al. 2015; Laslett et al. 2007).

We also looked at the levels of core stem cell transcription factors at the protein level through FACS analysis, which further corroborated our qPCR data, but also provided supportive evidence of the importance of stem cell associated genes in lineage decision processes (Wang et al. 2012). The seemingly endodermal nature of *GATA6*(+) cells correlated with the rapid reduction in the embryonic stem cell and neural marker *SOX2* (Pevny & Nicolis 2010), whereas all these

fractions retained equal levels of NANOG protein. NANOG has been reported to be involved in the repression of neuroectodermal and neural crest differentiation, thereby promoting the differentiation of cells towards the endodermal lineage (Wang et al. 2012). Our data supports this concept. We did find that although the levels of *NANOG* mRNA had decreased within the 3+6H fraction, the protein levels hadn't. *NANOG* can be post-translationally modified to increase the stability of the protein (Moretto-Zita et al. 2010), and although speculative, this may be a cellular mechanism involved to enforce early mesendoderm differentiation in hESCs. Further work is required to validate this hypothesis. Evidence supportive of the conclusion that *GATA6*(+) cells were beginning to show endoderm bias came from studying the expression of the stem cell and neural marker *SOX2* within the SSEA-3(-)*GATA6*(-) fraction. Within this fraction, cells were found to have high *SOX2* expression, implying neural differentiation. *GATA6* therefore appears to be an efficient marker to dissect two opposing lineages in culture.

Upon comparing gene expression changes the four fractions demonstrated a transition from a relatively pure stem cell state to one of a more differentiated state as *GATA6* expression increased. The overall changes in gene expression of stem cell associated markers, however, did not differ hugely between the four fractions. This could explain the ability for a proportion of cells from all fractions to remain within the stem cell compartment. In conclusion, these results point to the existence of discreet pluripotent substates in hESCs, which span from a more pristine stem cell, to a more differentiated cell type.

Upon analysis of which genes showed the greatest increase in expression between *GATA6*(+) and *GATA6*(-) hESCs, we found that only genes of the endoderm lineage, including *SOX17*, *GATA4*, *FOXA2* and *AFP* were significantly upregulated. Genes associated with mesodermal lineages were also upregulated, including *CD34* and *NPPA*, but not to the extent of endoderm genes. As cells transition through early differentiation, there is evidence that they first pass through a bipotent mesendodermal stage, which then segregates to form either endoderm or mesoderm precursors (Tada et al. 2005). It is unsurprising

therefore that genes involved in both these lineages were upregulated in *GATA6*(+) cells, as these *GATA6*(+) substates all appear to relate to very early differentiation. Furthermore, *GATA6* is heavily involved in cardiac differentiation (Koutsourakis et al. 1999), a mesodermal cell type, and so we cannot not rule out that at least a proportion of these cells are mesodermal in nature. The use of genes as markers of specific cell types is often very difficult, especially in the case of endoderm and mesoderm, as they can overlap.

Following on from the gene expression data, we then went on to investigate the functional behaviour of each of the four fractions. We hypothesised two scenarios. Firstly, these fractions represent cells co-expressing a stem cell marker and *GATA6* that have already committed to differentiate towards the endodermal lineage. This co-expression may represent very early time points in commitment, such that cells have not had time to switch off and shed SSEA-3 from the cell surface. This observation would be indicative of a substate residing outside of the stem cell compartment resulting in non-clonogenic cells. The second is that the co-expression we see represents a sub-state that does reside within the stem cell compartment, and these cells retain the capacity to self-renew and the expression of *GATA6* does not necessarily correlate with differentiation commitment.

We found that each of the four fractions generated stem cell colonies providing evidence for the second scenario outlined above. We found that functional changes were only apparent when *GATA6* expression was high, such that the 3+6L fraction did not show significant differences in self-renewal potential. Therefore, it would appear that the functional effects of *GATA6* is dose dependent and must be present above a certain expression threshold in order to overtly alter cell functional behaviour. This may be explained from our data on the protein levels of GATA6 within the fractions. The 3+6L fraction did not show noticeable expression of GATA6 protein, whereas 3+6H did begin to translate the protein. As GATA6 is proposed to be one of the first transcription factors involved in endoderm specification (Schrode et al. 2014), the absence of GATA6 protein in the 3+6L fraction would mean a lack of enforcement of the chain of

events to induce this lineage. Conversely, the 3+6H fraction did have detectable levels of *GATA6* protein in some cells, which would correlate in the reduced cloning efficiency.

Upon further analysis of the colonies derived from each of the four fractions, we found that within the 3+6H and 3-6+ fractions there were an increased percentage of colonies containing fewer *OCT4*(+) cells than the 3+6- and 3+6L fraction. This would imply that stem cells within the 3+6H and 3-6+ fractions show higher probabilities of differentiation than stem cells within the 3+6- and 3+6L fractions. This supports the notion that differentiation is a probabilistic rather than a deterministic process. The probabilistic nature of differentiation is also supported in that the 3-6+ fraction, which we assumed to consist of differentiated cells, was capable of producing stem cell colonies, albeit at very low levels. Thus the probabilities in cellular behaviour within this fraction are largely shifted towards differentiation, although some cells can self-renew. The existence of clonogenic SSEA-3(-) cells has also been reported elsewhere (Enver et al. 2005).

Furthermore, from the clonogenic data, we found that although the 3+6H and 3-6+ fractions did have a proportion of cells that were capable of long-term self-renewal, equally there were cells that were not. The existence of clonogenic and non-clonogenic cells within these fractions indicate further degrees of functional heterogeneity within the same fraction and the probable existence of further substates. This has been previously described in stem cells of the hematopoietic system. By sorting cells for $Sca1^{lo}$ and $Sca1^{hi}$, erythroid biased and myeloid biased states can be captured respectively. The $Sca1^{lo}$ fraction was found to exhibit culture reconstitution capacity and also showed high expression of the erythroid gene GATA1. Upon analysis of this fraction at single cell level, it became apparent the $Sca1^{lo}$ population was heterogeneous for GATA1. Further fractionation of the $Sca1^{lo}$ state using CD34 revealed two distinct compartments. $Sca1^{lo}$ CD34- containing almost all the GATA1 protein expression with no culture reconstitution capacity, and $Sca1^{lo}$ CD34+ containing minimal GATA1 protein and culture reconstitution capacity (Pina et al. 2012). It

is possible therefore that an appropriate third dimension within the 3+6H and 3-6+ fractions would allow for further fractionation of these cell populations that may led to the complete dissection of functionally discreet compartments.

The quantitative clonogenic assay, although particularly useful to quantify the short-term self-renewal potential of GATA6 expressing cells was restricted in that the long-term functional ability of *GATA6*(+) cells could not be assessed. For this reason we sub-cloned three fractions, 3+6-, 3+6L, 3+6H and maintained clones for analysis after several passages. The only conditions in which sub-clones could be derived were KO/SR and MEF culture conditions. This indicates that either a component of KO serum or something that the MEFs were secreting was responsible for the survival of single hESCs that was lacking in serum free/MEF free conditions. Nevertheless, sub-clones were obtained from the three fractions, in concordance with results from the quantitative clonogenic assay. Analyses of clones were performed between 5-8 passages of the initial seeding, and we found that clones derived from *GATA6* expressing cells showed long-term propagation, stem cell surface marker expression, and pluripotency, comparable to that of the original, unsorted parental line. As has been demonstrated in cells of the hematopoietic stem cell system as well as in mESCs (Chang et al. 2008; Pina et al. 2012; Hayashi et al. 2008), these clonal lines also demonstrated interconversion. Single *GATA6*(+) cells were able to revert back to a *GATA6*(-) state to repopulate each of the original fractions in culture. Thus, it would seem that single hESCs have the capacity to activate genes associated with a lineage program, whilst retaining a high degree of plasticity so that pluripotency is retained. Furthermore, the expression of these lineage programs can be reversed.

Having shown the existence of *GATA6*(+) substates within the stem cell compartment, and that *GATA6* correlates with alterations in self-renewal potential, the next question was whether the 3+6H and 3-6+ substates showed any lineage bias upon differentiation commitment, or whether cells were rather just biased for general non-directional differentiation. We hypothesised that, as

GATA6 is a marker of endoderm, cells would have a higher probability of moving towards the endodermal lineage.

5 Chapter 5

5.1 Introduction

5.1.1 Lineage Bias of GATA6+ cells

The mechanisms by which hESCs balance self-renewal and differentiation remain unanswered questions. A growing body of research has convincingly shown that hESCs *in vitro* are heterogeneous and this heterogeneity has functional significance for cell behaviour (Hayashi et al. 2008; Canham et al. 2010; Tonge et al. 2011). The concept that cells can express lineage specific markers that bias their eventual fate decisions, commonly referred to as lineage priming, has been well documented within several systems, including the haematopoietic stem cell system. One such well established model is that of the inhibitory feedback loop of a cell population based model for the transcription factors *GATA1* and *PU.1* involved in erythrocyte and myeloid specification respectively (Zhang et al. 1999). Multipotency is achieved by the negative regulation of one lineage specific factor to another (*GATA1* to *PU.1* and vice versa) through protein-protein and protein-DNA interactions. The expression of one protein can therefore poise cells for lineage specification without actually losing multipotency if the other is expressed. Only when signalling cues are received, and one transcription factor accumulates do cells transition to the differentiated state (Zhang et al. 1999). This concept of multilineage priming was subsequently shown at the single cell level which identified that transcripts for the genes globin and *MPO* (erythroid and granulocytic lineages respectively) as well as other key lineage regulators, could be expressed within the same cell. Thus, multipotent cells appeared to be expressing key markers for multiple lineages within a stem cell phenotype (Hu et al. 1997; Delassus et al. 1999). Functional evidence of lineage priming was then shown using the leukemic line HL60 and DMSO induced neutrophil differentiation. These findings identified a 'primed' state in which cells expressed the neutrophil marker CD11b, which became more sensitive to further differentiation stimuli without necessarily

losing the stem cell phenotype. This developed a model in which cells exist along a path of differentiation, whereby priming is both closer to differentiation but also reversible away from differentiation such that cells have not necessarily made a permanent switch to the neutrophilic lineage (Chang et al. 2006).

Within the pluripotent context, the concept of biased states was shown in the embryonal carcinoma line NTERA2 whereby pro-neural and non-neural fates were shown to exist at the single cell level. Single NTERA2 cells were shown to behave in a non-uniform manner when exposed to retinoic acid, where some cells formed homogeneous TuJ1+ neuronal colonies, and others homogeneous TuJ1- colonies. Retinoic acid was then delayed to allow one round of cell division, resulting in two cells each with the potential to reside within a pro or non-neural substate. This time a proportion of NTERA2 cells generated heterogeneous TuJ1(+)/TuJ1(-) colonies demonstrating pro and non-neural substate interconversion (Tonge et al. 2010). Thus cells appeared to have made lineage decisions even before commitment, indicative of discreet functional interconvertible substates within the stem cell compartment.

In concordance with the existence of lineage directing substates within NTERA2, studies examining the expression of lineage genes in mESCs have also pointed to the idea that the pluripotency of stem cells spans a broad state space divided into substates that does not simply represent cells with equal lineage probabilities. A study by Hayashi et al found that a subset of mESCs in culture conditions containing serum expressed the gene *STELLA* under self-renewing conditions. *STELLA* is a germ cell specific marker expressed in the pre-implantation embryo but repressed within the epiblast. Upon fractionation, *STELLA*(-) cells were shown to have an increased propensity to form differentiating EBs, enhanced neural specification, and an increased ability to form trophectoderm. *STELLA*(-) cells, although found to be more permissive for differentiation were not necessarily committed cells (Hayashi et al. 2008). Furthermore, the gene *HEX*, a marker for anterior visceral endoderm was also found to be expressed at low levels in mESC cultures (Canham et al. 2010).

HEX(+) cells showed a PrE like transcriptome, through the expression of *GATA4*, Dab2, *SOX7* and Hnfα, and were found to have a significant impairment in chimeric embryo contribution when compared to *HEX*(-) cells. Furthermore, *HEX*(+) cells upon chimeric EB formation were found to segregate to the surface of EBs and expressed endodermal markers such as *GATA6*, *FOXA2* and *SOX17* (Canham et al. 2010). This pattern of endoderm formation is consistent with *in vitro* cultured ICMs of the mouse blastocyst, which form a layer of endoderm on the surface (Cockburn & Rossant 2010). Similarly to *Stella* expression, HEX(+) cells were not necessarily committed to differentiation, but showed directional lineage specification if they did commit. Lineage priming has also been observed within the hESC context. Blauwkamp et al demonstrated that different levels of Wnt signalling conferred distinct lineage-specific differentiation propensities. Cells with low WNT activation showed an enhanced ability to differentiate into neural cells using a directed differentiation protocol, whereas cells with high WNT activation acquired primitive streak-like characteristics and were found to rapidly differentiate into mesodermal and endodermal cells. Furthermore, stabilization of hESCs with high WNT activation differentiated into mesodermal and endodermal cells with greater speed and efficiency than heterogeneous populations (Blauwkamp et al. 2012). Although this study provided strong evidence that heterogeneity of WNT signaling in hESCs did cause directed lineage bias, it was only performed at a population level.

The functional consequences of lineage priming may also be pertinent when trying to explain why there is such disparity in the behaviour between hESC lines. Subtle differences in lineage marker expression between cell lines were found upon the characterisation of 59 established cell lines from 17 labs worldwide. The lines H14, H7, H13 and H9, showed higher levels of endoderm specific gene expression, in particular *AFP*, than other lines tested (International Stem Cell Initiative et al. 2007). Reports also describe substantial differences between cell lines in directed differentiation. For example, it was found that the HSF1 line was much more efficient at producing functional forebrain cell types, primarily GABAergic synaptic networks than the HSF6 line, which produced primarily glutamergic networks under a standard differentiation assay. The cell

lines exhibited distinctly different differentiation potentials that led to the speculation that each cell line may be 'pre-programmed' for lineage specification (Wu et al. 2007). A further study assessed the differentiation propensities of 17 HUES lines, resulting in the finding that HUES 8 was most efficient at pancreatic differentiation, and HUES 3 for cardiac differentiation. Gene expression differences after differentiation were often quite large at >100 fold between cell lines directed to the same lineage. The conclusion here was that the disparity between lines was most likely reflective of the considerable genetic variation between hESC lines as well as variable epigenetic statuses (Osafune et al. 2008). This altered propensity for HUES lines to preferentially differentiate to a specific lineage has also been shown in blood differentiation. HuES 8, 14, 15 and a non-HuES line, H1, were more efficient at producing CD34+CD45+ hematopoietic precursors than other lines tested, even using multiple differentiation protocols (Melichar et al. 2011). Functional heterogeneity has also been observed at the protein level of hESCs. It has been reported that the cell line HES2, when induced to different to cardiomyocytes, produces a significantly higher level of ventricular-like cells than atrial-like and pacemaker-like cells. In contrast, the H1 line generated a much more even distribution of these cells types. Upon protein analysis of undifferentiated HES2 and H1 cells, it was found that HES3 had higher levels of troponin and annexin II, both cardiac specific proteins (Moore et al. 2008). These observations raise the possibility that the expression of certain cardiac proteins in pluripotent hESCs may bias their cardiogenic outcomes.

From the existing reports detailing lineage directed differentiation in haemotopoetic stem cells, embryonal carcinoma cells, mESCs and hESCs, and with our data providing evidence that *GATA6* can be expressed in *bona fide* stem cells, we wanted to ascertain whether *GATA6* expression caused a lineage specific differentiation bias. As *GATA6* is primarily involved in the endodermal lineage during early mouse development, and as our data demonstrated a significant increase in other genes of the endodermal lineage but not mesoderm or ectoderm, we hypothesised that GATA6(+) cells, if they committed to differentiate, would preferentially choose the endodermal lineage. In this study,

we wanted to perform functional analysis of *GATA6* expression on both the population and the single cell level. The single cell bias was important as interpretation from population data relies on the assumption that single stem cells exhibit both self-renewal and differentiation biased properties.

5.2 Results

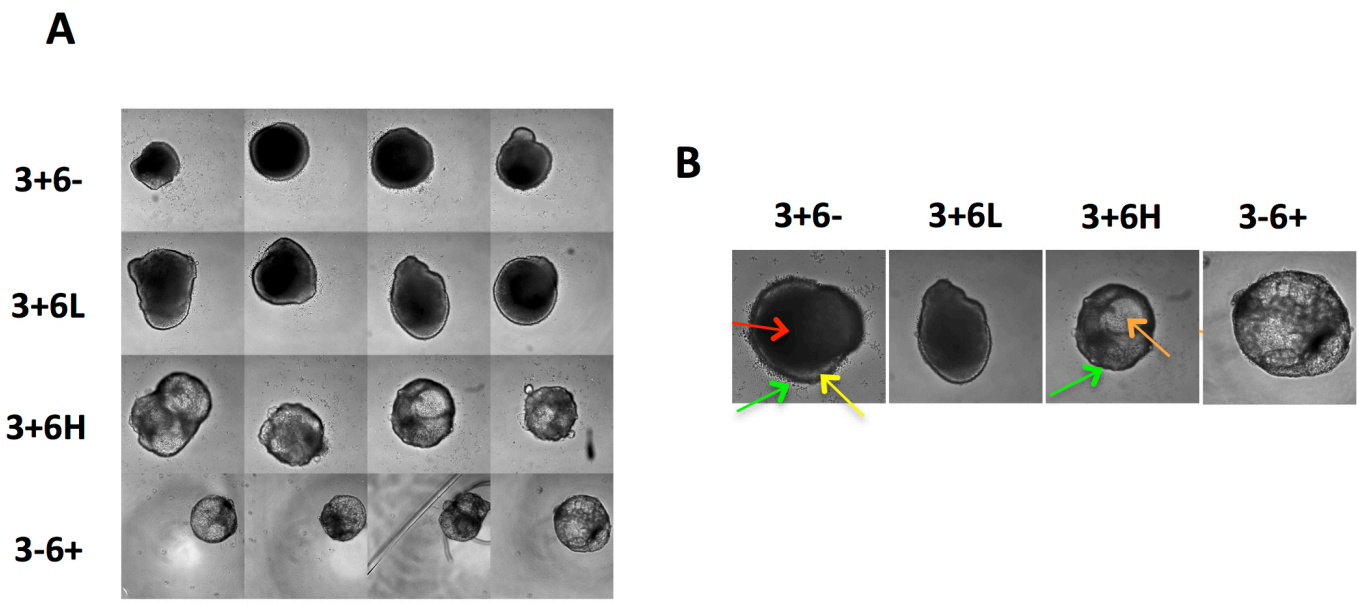
5.2.1 GATA6 expressing cells exhibit endoderm population bias

To assess whether each of the four fractions of cells showed similar differentiation propensities at the population level, we used FACS sorting to isolate each fraction (3+6-, 3+6L, 3+6H & 3-6+) for seeding into a defined, neutral spin-EB protocol (Ng et al. 2008). Neutral within this context meant without the addition of exogenous growth factors or chemical inhibitors to direct differentiation. Cells were aggregated over a 24h period and were left to differentiate for 10 days. We found differences in the morphology of EBs depending upon their fraction of origin **(Fig. 5.1A)**. EBs made from the 3+6- fraction showed a dense, compacted morphology with a degree of structural organisation consisting of a dense inner core, a distinct outer border and a less dense middle layer of cells separating the two **(Fig. 5.1B)**. EBs from the 3-6+ fraction were much more cystic and showed less structural organisation, such that the three layers seen within the 3+6- EBs were not present **(Fig. 5.1B)**. The 3+6L fraction showed morphology more similar to that of the 3+6- fraction, and the 3+6H cells showed morphology more similar to that of the 3-6+ fraction.

We then analysed gene expression of EBs from the 3+6-, 3+6L, 3+6H and 3-6+ fractions. Although the morphology of the 3+6L EBs were similar to 3+6- EBs, when the gene expression was compared, we found an upregulation of genes associated with endoderm. Similarly, we found that EBs made from the 3+6H and 3-6+ fractions also showed more efficient endoderm differentiation when compared to 3+6- **(Fig. 5.2)**. Genes including *GATA6*, *AFP*, *SOX17* and *FOXA2* were all upregulated in the 3+6L, 3+6H and 3-6+ fractions. Also upregulated, but not to the same extent as the endoderm genes were mesoderm genes. Genes including *PECAM*, *KDR*, and *DESMIN* were all upregulated in the 3+6L, 3+6H and 3-6+ fractions. We also looked at the expression of ectodermally associated markers. The 3+6L, 3+6H and 3-6+ EBs all showed downregulation of ectoderm associated markers, which was most striking in EBs from the 3+6H and 3-6+

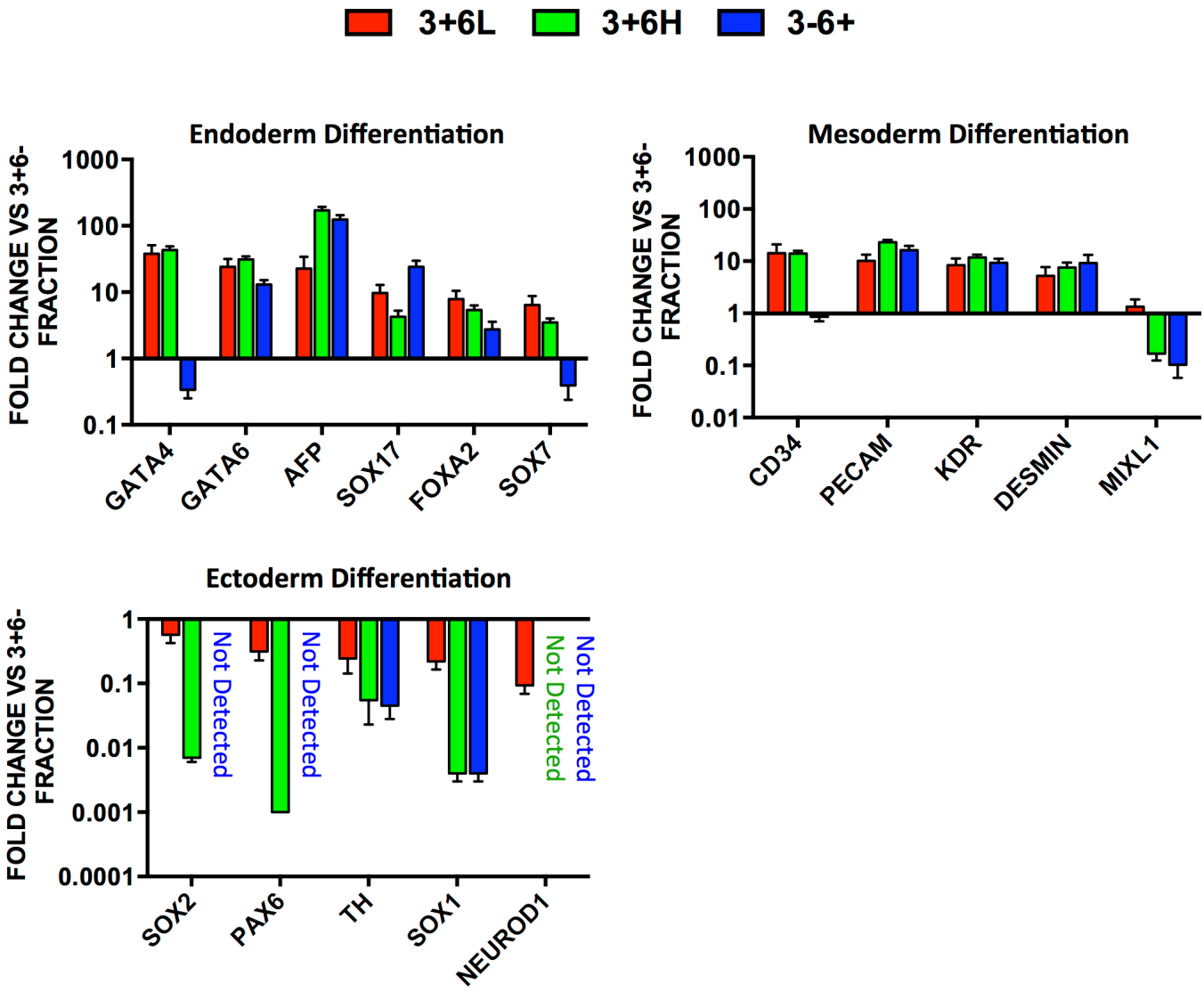
fractions **(Fig. 5.2)**. EBs from the 3+6H and 3-6+ fractions did not show any detectable expression of *NEUROD1*, and 3-6+ EBs also did not show detectable expression of *SOX2* or *PAX6* **(Fig. 5.2)**. We therefore concluded that under differentiation conditions, EBs from the 3+6L, 3+6H and 3-6+ fractions showed a strong endoderm differentiation bias when compared to EBs from the 3+6- fraction. These fractions also showed a moderate upregulation of mesoderm genes. Conversely, ectoderm genes were strongly downregulated within *GATA6*(+) fractions and this bias away from ectoderm appeared graded, so that increasing *GATA6* expression resulted in a stronger bias away from ectoderm.

Figure 5.1: High *GATA6* leads to morphological changes in EBs



(A) Representative brightfield images of EBs derived from the 3+6-, 3+6L, 3+6H and 3-6+ fractions. Images were take on an InCell Analyzer 2000 at x4 magnification at day 10 of differentiation. **(B)** Individual EBs show structural organisation from the 3+6 and 3+6L, but not the 3+6H or 3-6+ fractions. Arrows represent different structures: Red, a dense inner core of cells; Green: outer border of cells; Yellow: less dense layer of cells separating inner and outer layers ; Orange: cystic cavity.

Figure 5.2: *GATA6* expressing fractions display endoderm differentiation bias



qPCR analysis for Endoderm (top left), Mesoderm (top right) and Ectoderm (bottom left) genes from day 10 EBs for 3+6L, 3+6H and 3-6+ fractions. Values shown as fold change when compared to day 10 EBs from the 3+6- control fraction. Beta-actin was used as a normalising gene. Not detected indicates no amplification of target gene over 40 cycles. Coloured bars represent each fraction; red: 3+6L, Green: 3+6H, Blue: 3-6+. Error bars indicate standard deviation of three technical replicates and graphs are representative of two biological repeats.

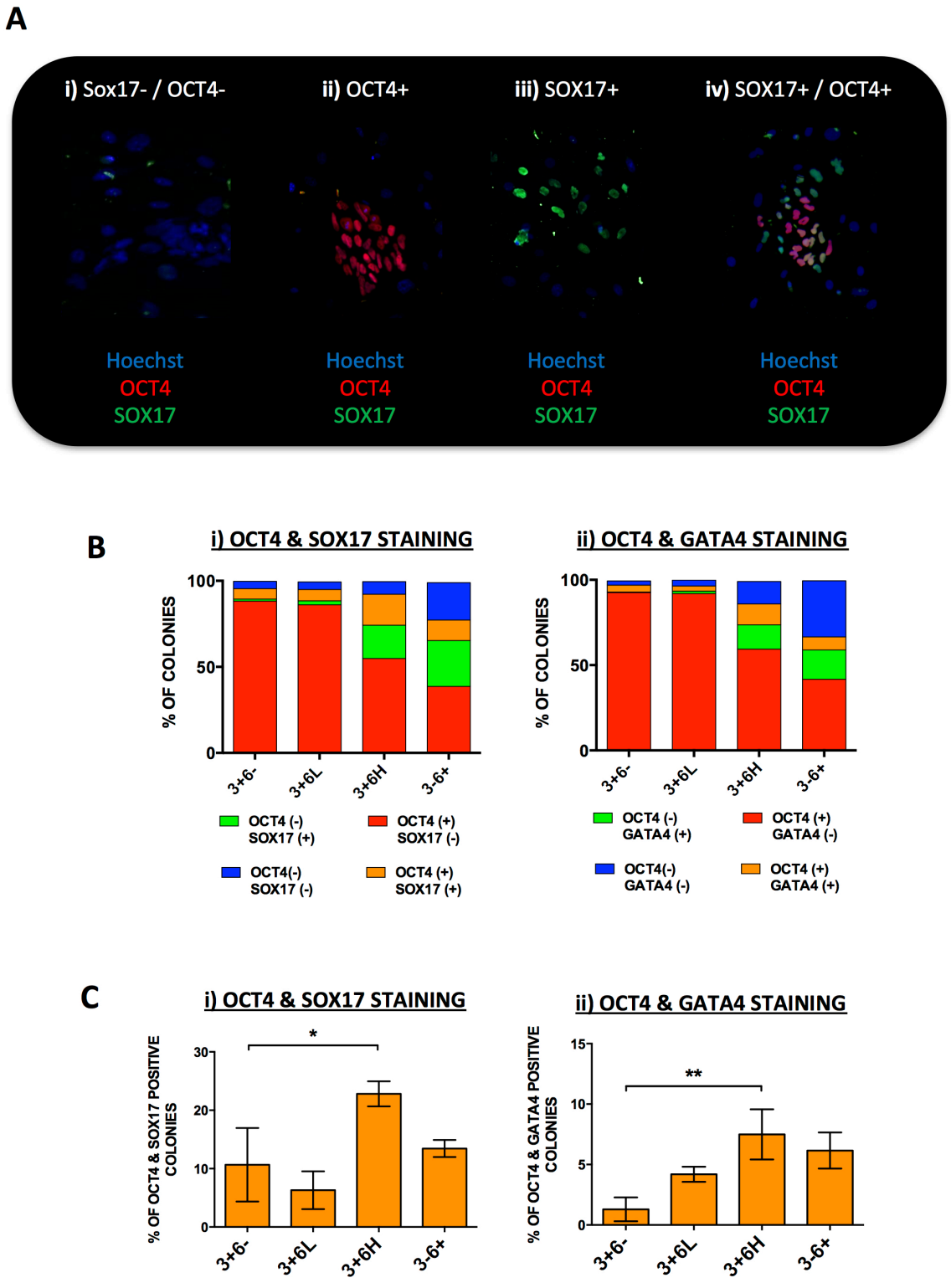
5.2.2 Stem cells of the 3+6H fraction show single cell endoderm bias

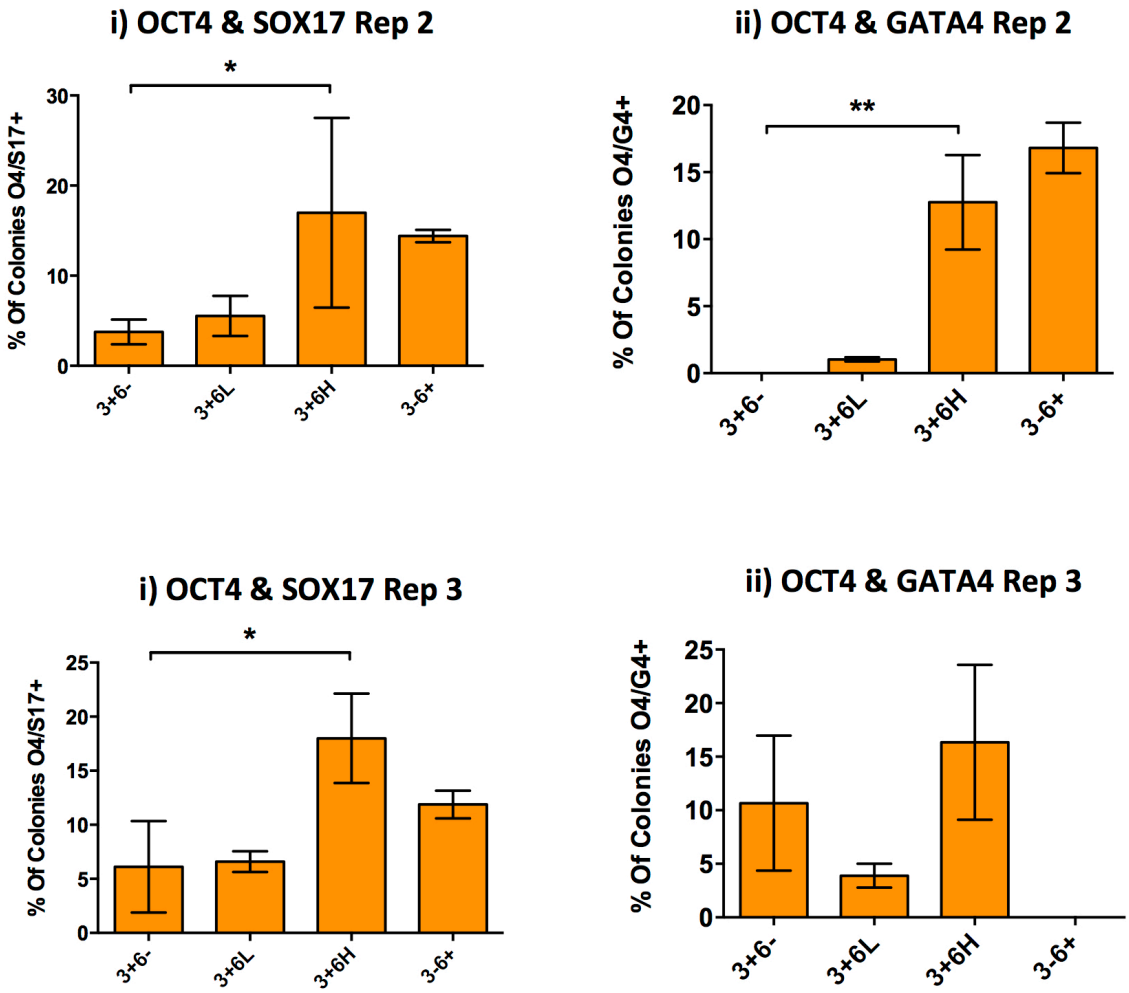
From the experiments performed on EBs, it became apparent that cells expressing *GATA6* did show a bias towards endoderm and although less pronounced mesoderm at the expense of ectoderm. What we were unable to establish from these experiments, however, is whether the differentiation bias was a result of already committed cells or was due to biased stem cells. In order to answer this question, we devised an assay in which we analysed colonies resulting from single cells of the 3+6-, 3+6L, 3+6H and 3-6+ fractions, looking specifically for colonies containing cells positive for the stem cell marker OCT4, and the endodermal markers, SOX17 or GATA4. These colonies would provide evidence that the initial cell was a stem cell, through the functional generation of OCT4(+) colonies, but would also allow the detection of spontaneous differentiation to the endoderm lineage. We reasoned that if *GATA6* expressing cells showed a higher percentage of OCT4(+)/SOX17(+) and OCT4(+)/GATA4(+) colonies, then stem cells within these fractions were endoderm biased. The four fractions of cells were FACS sorted and seeded at clonogenic density. Colonies were left to develop over four days and cells were then fixed and stained in OCT4/SOX17 or OCT4/GATA4 combinations.

There were 4 types of colonies that resulted after staining in these combinations **(Fig. 5.3A)**. Colonies either did not express OCT4 or SOX17 **(Fig. 5.3Ai)**, expressed either OCT4 **(Fig. 5.3A.ii)** or SOX17 **(Fig. 5.3A.iii)**, or expressed both **(Fig. 5.3A.iv)**. The percentage of each colony type was then quantified for each fraction **(Fig. 5.3B)**. We found that only when *GATA6* expression was high, did percentages of each colony type change from the 3+6- fraction. Firstly, we analysed the percentage of colonies across the four fractions showing OCT4(+) staining only. Colonies derived from the 3+6L fraction, showed no noticeable decrease in the percentage of OCT4(+) only colonies when compared to the 3+6- fraction **(Fig. 5.3B.i)**. This is consistent with our previous clonogenic data that the 3+6L fraction clones as efficiently as the 3+6- fraction. Within the 3+6H and 3-6+ fractions, there was a decrease in the percentage of colonies that contained

OCT4(+) cells only **(Fig. 5.3B.i)**. Again this correlates with previous clonogenic data that high *GATA6* expression resulted in a lower cloning efficiency. Next, we analysed colonies with SOX17(+) staining only. We saw very little change in the percentages of OCT4(-) SOX17(+) colonies from the 3+6L fraction when compared to the 3+6- fraction, but the percentage of these colonies did increase within the 3+6H fraction and more so in 3-6+ fraction when compared to 3+6-. This implied a higher rate of endoderm differentiation within these fractions **(Fig. 5.3B.i)**. Thirdly, we analysed colonies, which did not contain OCT4(+) or SOX17(+) cells. In this instance, we found no real differences between the 3+6-, 3+6L or 3+6H fractions, however this colony type did increase in the 3-6+ fractions. Finally we found that *OCT4*(*+) SOX17*(+) colonies only increased within the 3+6H and 3-6+ fractions when compared to the 3+6-. The 3+6H fraction consistently showed statistically significant higher percentages of this colony type for all biological replicates **(Fig. 5.3C.i)**. The replacement of SOX17 with GATA4 as a marker of endoderm also showed the same pattern of percentage colony types within each fraction, although one biological repeat did not show statistical significance **(Fig. 5.3B.ii, Fig. 5.3C.ii)**. This demonstrates that the 3+6H fraction, specifically, consists of the largest percentage of endoderm biased stem cells

Figure 5.3: High *GATA6* expression results in endoderm biased stem cells





(A) Cells from each fraction (3+6-, 3+6L, 3+6H & 3-6+) were sorted and seeded at clonogenic density and fixed after 4 days. Resulting colonies were co-stained with OCT4 and SOX17 or OCT4 and GATA4. 4 colony types were apparent with colonies containing no OCT4, SOX17 or GATA4 **(i)**, OCT4 only **(ii),** SOX17 or GATA4+ only **(iii)**, and OCT4 & SOX17 or OCT4 & GATA4+ cells **(iv). (B)** Graphs representing the percentage of each colony type from (A) using SOX17 **(i)** or GATA4 **(ii)** as the endoderm marker from each of the four cell fractions**.** Only colonies that contained at least 1 OCT4+ cell and 10% SOX17+ cells were classed as double positive. Graphs represent three biological repeats each with three technical repeats. **(C)** The percentage of OCT4+ & SOX17+ **(i)** or OCT4+ & GATA4+ **(ii)** colonies were plotted for each fraction. Each graph represents three technical repeats for each biological repeat. Error bars represent standard deviation of three technical repeats and significance was calculated using student's t-test.

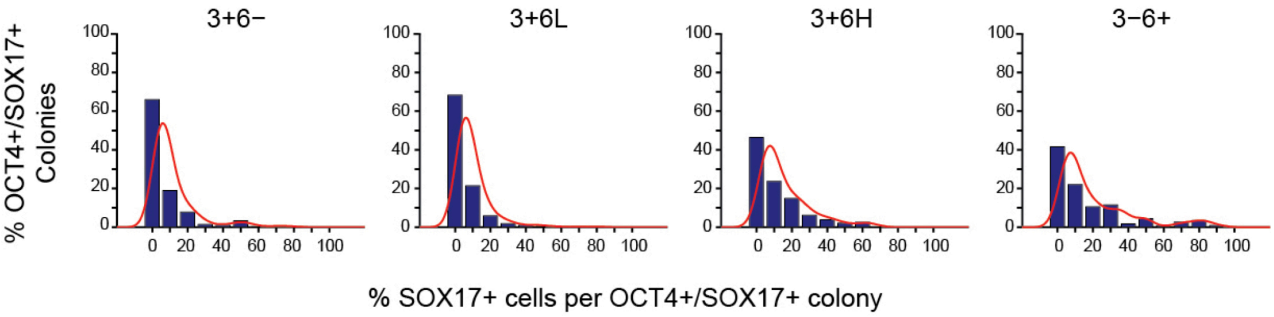
5.2.3 3+6H and 3-6+ fractions generate more colonies with a higher percentage of *SOX17*(+) and *GATA4*(+) cells

We then performed a detailed analysis of the distribution of SOX17(+) cells in OCT4(+)/SOX17(+) colonies by histogram counts **(Fig. 5.4A)**. By quantitating the changes in the shape of the frequency distribution graphs using the Kullback-Leibler measure, we noted that although overall values were low implying subtle changes, differences were evident between fractions **(Fig. 5.4B)**. The trend in divergence was most apparent in the 3+6H and 3-6+ fractions when compared to the 3+6- fraction. This indicated that a higher percentage of cells per stem cell colony within the 3+6H and 3-6+ fractions were SOX17 positive. Similarly, we noted a similar trend when analysing the distributions of OCT4(+)/GATA4(+) colonies **(Fig. 5.4C)**. In the context of GATA4 bias, the Kullback-Leibler measure indicated that the most prominent differences occurred again in colonies arising from the 3+6H and 3-6+ fractions when compared to 3+6- **(Fig. 5.4D)**. For both SOX17 and GATA4, the 3+6L fraction resembled 3+6-, consistent with the finding that these fractions were functionally similar.

In conclusion, the 3+6L fraction under self-renewal conditions did not show functionally different behaviour in terms of self-renewal capacity or differentiation bias from the 3+6- fraction. In contrast, the 3+6H and 3-6+ fractions did show a decrease in self-renewal potential coupled with an increase in endoderm differentiation. Most interestingly, the 3+6H and 3-6+ fractions also showed an increase in endoderm differentiation bias within stem cell colonies. This was most apparent within the 3+6H fraction.

Figure 5.4: Colony analysis from each fraction for OCT4 & SOX17 or GATA4 expressing cells

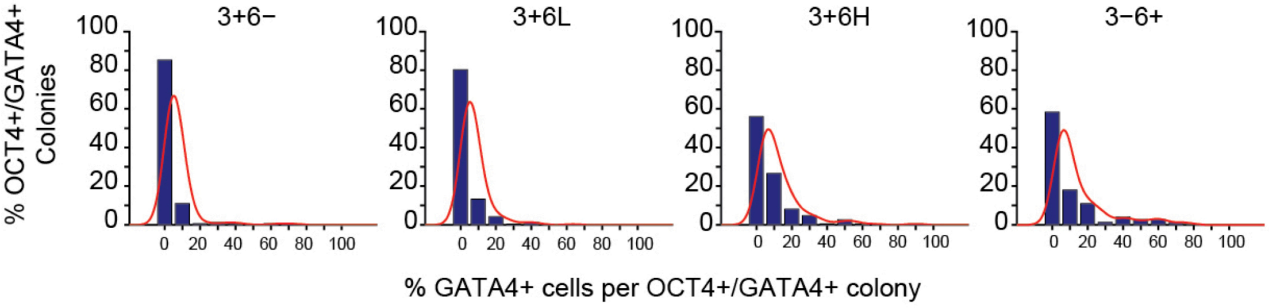
A



B

	3+6-	3+6L	3+6H	3-6+
3+6-	0	0.0159	0.1088	0.1918
3+6L	0.0159	0	0.1483	0.2515
3+6H	0.1088	0.1483	0	0.2087
3-6+	0.1918	0.2515	0.2087	0

C



D

	3+6-	3+6L	3+6H	3-6+
3+6-	0	0.0251	0.2253	0.2411
3+6L	0.0251	0	0.1613	0.1940
3+6H	0.2253	0.1613	0	0.0430
3-6+	0.2411	0.1940	0.0430	0

(A) Proportion of OCT4+/SOX17+ cells in double positive colonies resulting from plated single cells from fractions 3+6-, 3+6L, 3+6H and 3-6+. Double positive denotes colonies with min one OCT4+ cells and min 10% SOX17+ cells. Counts are shown as a bar plot (blue) with superimposed estimated nonparametric distribution (red). **(B, D)** Kullback-Leibler pairwise symmetric divergence between distributions shown in A and C respectively. Measure compares fractions on X axis to fractions on Y axis. This measure increases with reduced similarity between distributions; zero indicates identical distributions. Green bars represent largest divergences. **(C)** Proportion of OCT4+/GATA4+ cells in double positive colonies resulting from plated single cells from fractions 3+6-, 3+6L, 3+6H and 3-6+. Double positive denotes colonies with min one OCT4+ cells and min 10% SOX17+ cells. Counts are shown as a bar plot (blue) with superimposed estimated nonparametric distribution (red).

5.2.4 Identification of signalling pathways that induce GATA6 expression

From single cell clonogenic data we found that a proportion of stem cells that were expressing *GATA6* at high levels had a differentiation bias towards the endodermal lineage. The actual percentage of stem cells in the 3+6H fraction in culture however is relatively low at around 5% of the total population, and the percentage of biased stem cells within this fraction is even smaller. We therefore wanted to understand the mechanism through which these cells appear in routine culture, to develop strategies in order to enrich and stabilise these biased stem cells. We set up a screen using chemical inhibitors targeting various pathways active in hESC, to ascertain which signalling pathways were important in the expression of *GATA6* **(Fig. 5.5A)**. The screen was performed in fully defined conditions (Vitronectin & E8) on single cells to eliminate unknown parameters present in KO/SR and MEF conditions, or matrigel cultures. We screened each inhibitor at two concentrations (1 μM & 2 μM) for two days on SSEA-3(+) sorted cells and performed FACS analysis for the induction of GFP against the untreated control sample. Firstly, we noticed that the expression of *GATA6* was reduced, and almost lost in feeder free conditions before the addition of any inhibitors **(Fig. 5.5B)**. After screening, we found that the induction of WNT, inhibition of MEK/ERK and inhibition of BMP pathways were effective at inducing *GATA6* expression **(Fig. 5.5C)**. Firstly, CT 99021, a GSK3β inhibitor, caused a large increase at 43% of *GATA6* expressing cells. The effect of CT 99021 was dose dependent, as lower concentrations resulted in fewer *GATA6*(+) cells **(Fig. 5.5D)**. Secondly, PD 0325901, a MEK/ERK inhibitor, induced 40% of cells to express *GATA6* **(Fig. 5.5D)**. PD 0325901 did not seem to function in a dose dependent manner, at least at the concentrations we tested. Finally, Dorsomorphin also induced *GATA6* and similarly to CT 99021 appeared dose dependent **(Fig. 5.5D)**. In order to validate these results, we again treated SSEA-3(+) sorted cells with CT 99021, PD 0325901 and Dorsomorphin at optimal concentrations (2 μM, 0.5 μM and 2 μM respectively) and analysed *GATA6* induction using flow cytometry **(Fig. 5.6A)**. Both CT 99021 and PD 0325901 showed robust induction of *GATA6* expressing cells, whereas

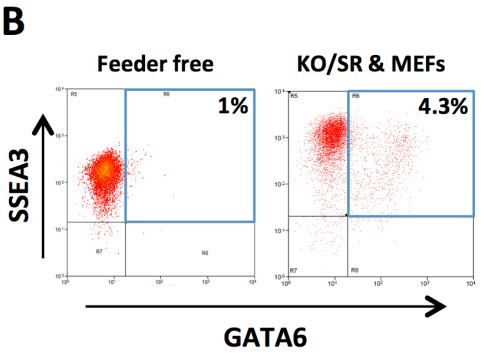
Dorsomorphin did not **(Fig. 5.6B)**. In both repeats of the validation assay, Dorsomorphin failed to induce GATA6 above control levels and was therefore treated as a false positive and eliminated from further study **(Fig. 5.6B)**. The loss of SSEA-3(+)/*GATA6*(+) expression in fully defined conditions then prompted the question of the mechanism behind how these cells arise in KO/SR and MEF conditions.

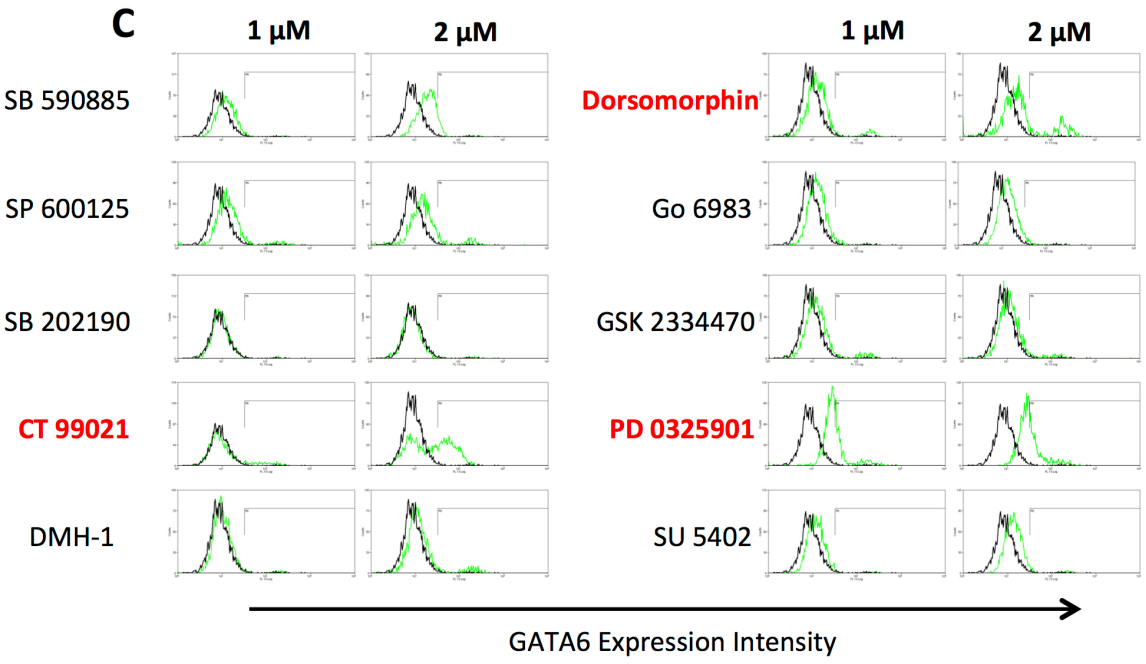
Figure 5.5: Activation of WNT and inhibition of BMP and MEK signalling pathways induce *GATA6*

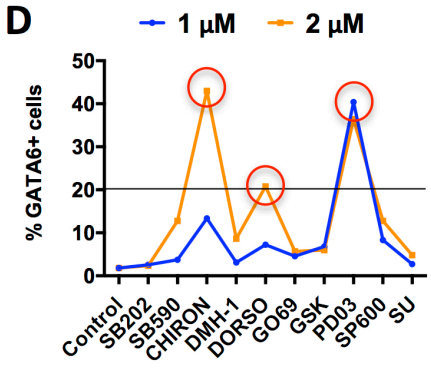
A

Inhibitor	Pathway
SB 202190	P38 MAPK
SB 590885	B-Raf
CT 99021	WNT
DMH-1	BMP
Dorsomorphin	BMP

Inhibitor	Pathway
Go 6983	PKC
GSK 2334470	PDK1
PD 0325901	MEK1/2
SP 600125	JNK
SU 5402	FGFR & VEGFR

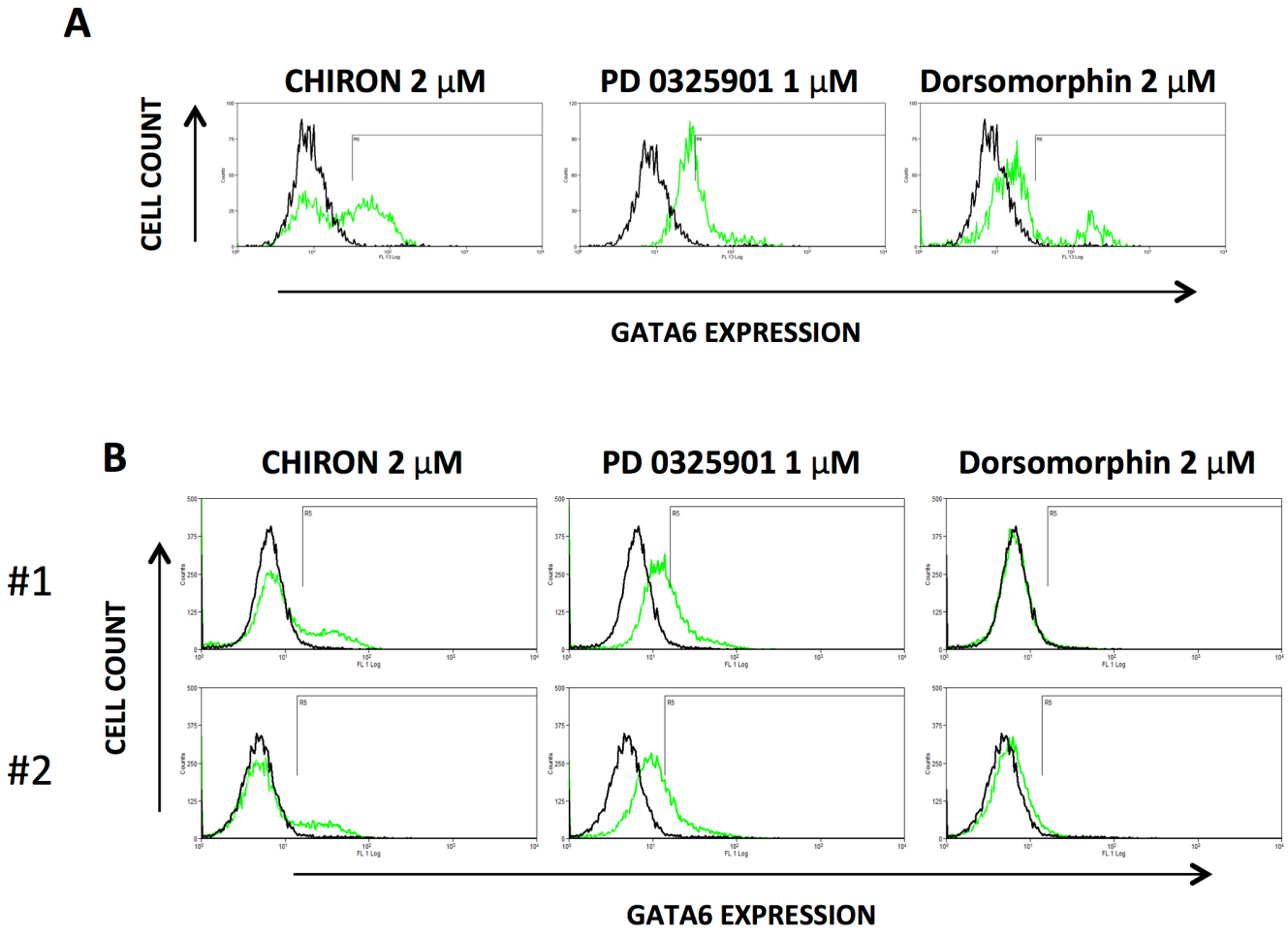






(A) Table of signalling pathways assessed for their ability to regulate *GATA6*, and their corresponding commercially available inhibitors. **(B)** FACS comparison of SSEA3+ *GATA6*+ cells in feeder free (left panel) and feeder (right panel) conditions. **(C)** Fluorescent cytometric analyses of the Shef4-GATA6:GFP reporter line after 2 days exposure to 1 µM and 2 µM of individual inhibitors in defined conditions. Black histogram: reporter line in standard self-renewal conditions; green histogram: reporter line after 2 days of exposure. Inhibitors were compared to back-ground control GFP levels. **(D)** Graph of individual inhibitors and the percentage of GATA6 induction at 1 µM (blue) and 2 µM (yellow) concentrations. ≥20% induction were further analysed.

Figure 5.6: WNT and MEK pathways reproducibly alter GATA6 expression



(A) Individual signalling pathways identified to induce GATA6 from the initial screen. **(B)** Validation studies were performed and analysed by flow cytometery by re-exposing SSEA3 sorted cells to individual inhibitors at optimal concentrations. Black: reporter line without inhibitors; Green: reporter line with corresponding inhibitors for 2 days.

5.2.5 WNT signalling induces SSEA-3(+) *GATA6*(+) fractions in defined conditions

To better understand the mechanism of how the four fractions are generated in KO/SR & MEF conditions, we assessed whether the addition of the *GATA6* inducing inhibitors identified previously (CT 99021 for WNT activation and PD 0325901 for MEK/ERK inhibition) could recapitulate each fraction in defined conditions. Furthermore, as we have previously shown that the 3+6H fraction specifically exhibited differentiation bias, we wanted to assess whether CT 99021 and/or PD 0325901 could induce this particular fraction. We performed FACS analysis of *GATA6* against SSEA-3, and found that cells exposed for two days to CT 99021 generated each of the four fractions originally identified in KO/SR and MEF culture (3+6-, 3+6L, 3+6H & 3-6+), and efficiently induced the 3+6H fraction **(Fig. 5.7A)**. We also confirmed that *GATA6* induction through CT 99021 was dose dependent, and that the optimum concentration was 2 μM as found within the initial screen **(Fig. 5.7B)**. PD 0325901 also induced *GATA6* as expected, but did not recapitulate feeder conditions. The addition of this inhibitor only induced the 3+6L fraction but not the 3+6H fraction, irrespective of the concentration used **(Fig. 5.7C)**.

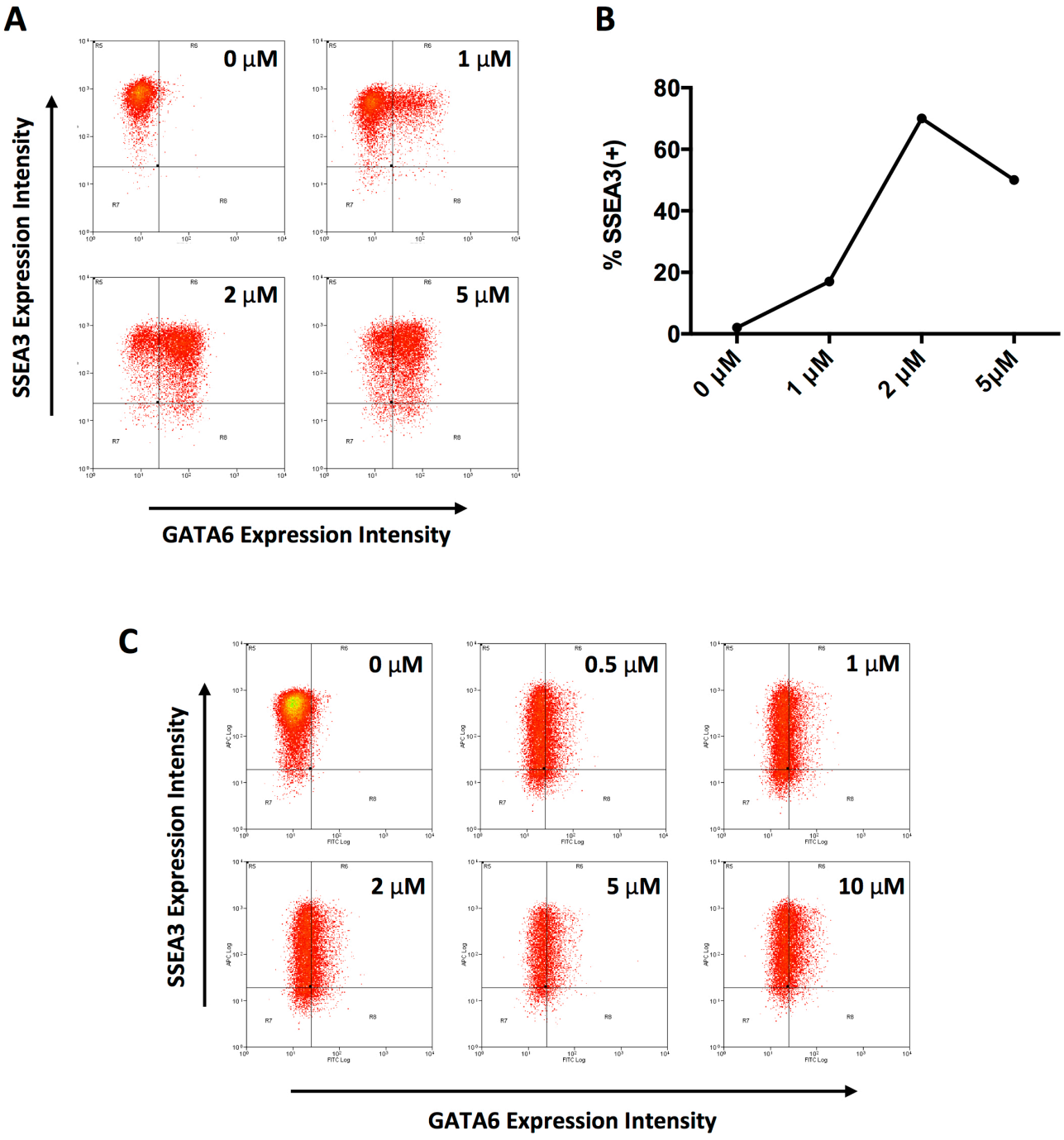
We then performed qPCR analysis to investigate the gene expression profiles of cells that had been treated with CT 99021 or PD 0325901, and whether the expression profiles matched with the fractions from KO/SR & MEF conditions. Firstly, cells treated with CT 99021 showed a downregulation in stem cell associated genes which became more apparent with increasing concentrations **(Fig. 5.8A.i)**. This was coupled with a strong upregulation in endoderm and primitive streak associated genes such as *SOX17, GATA4, FOXA2, BRACHYURY* & *MIXL1*, as was seen in the fractions from KO/SR & MEF conditions **(Fig. 5.8A.ii & Fig. 5.8A.iii)**. This upregulation was apparent for all concentrations tested, however, we noticed that at higher concentrations of CT 99021, SOX17 and FOXA2 were not expressed at levels higher than the control sample **(Fig. 5.8A.ii)**. In concordance with the upregulation of genes associated with

endoderm and the primitive streak, the expression of genes indicative of ectoderm, *PAX6* and *SOX2* were downregulated at all concentrations. Conversely, *SOX1* expression showed no change between CT 99021 treated and non-treated cells **(Fig. 5.8A.iv)**. Thus, CT 99021, in a dose dependent manner, induced *GATA6* expression and directed cells to an endodermal / primitive streak like fate. Furthermore, CT 99021 exposure induced the 4 fractions (3+6-, 3+6L, 3+6H & 3-6+), which are ordinarily lost in defined conditions, similar to that of KO/SR & MEF culture.

Cells that had been exposed to PD 0325901 showed a different lineage gene expression profile to that of CT 99021. Similarly to CT 99021 treatment, stem cell associated genes were downregulated. The endoderm genes *GATA6* and *GATA4* were upregulated but other, markers of endoderm (*SOX17*, *FOXA2*) and primitive streak (*MIXL1*) were not **(Fig. 5.8B.i)**. *T* expression did not show a change between treated and untreated cells **(Fig. 5.8B.ii & Fig. 5.8B.iii)**. The definitive endoderm marker *SOX17* was undetected at all concentrations **(Fig. 5.8B.ii)**, but the ectodermal marker *PAX6* was increased in all concentrations. This trend was not apparent for other ectodermal markers including *SOX1* and *SOX2* **(Fig. 5.8B.iv)**. Thus, PD 0325901 appeared to induce differentiation but cells were directed to an alternative lineage when compared to CT 99021 treatment, which did not seem to correlate with endoderm differentiation. Furthermore, *GATA6* was not highly induced by PD 0325901 and did not re-establish the four fractions in defined conditions.

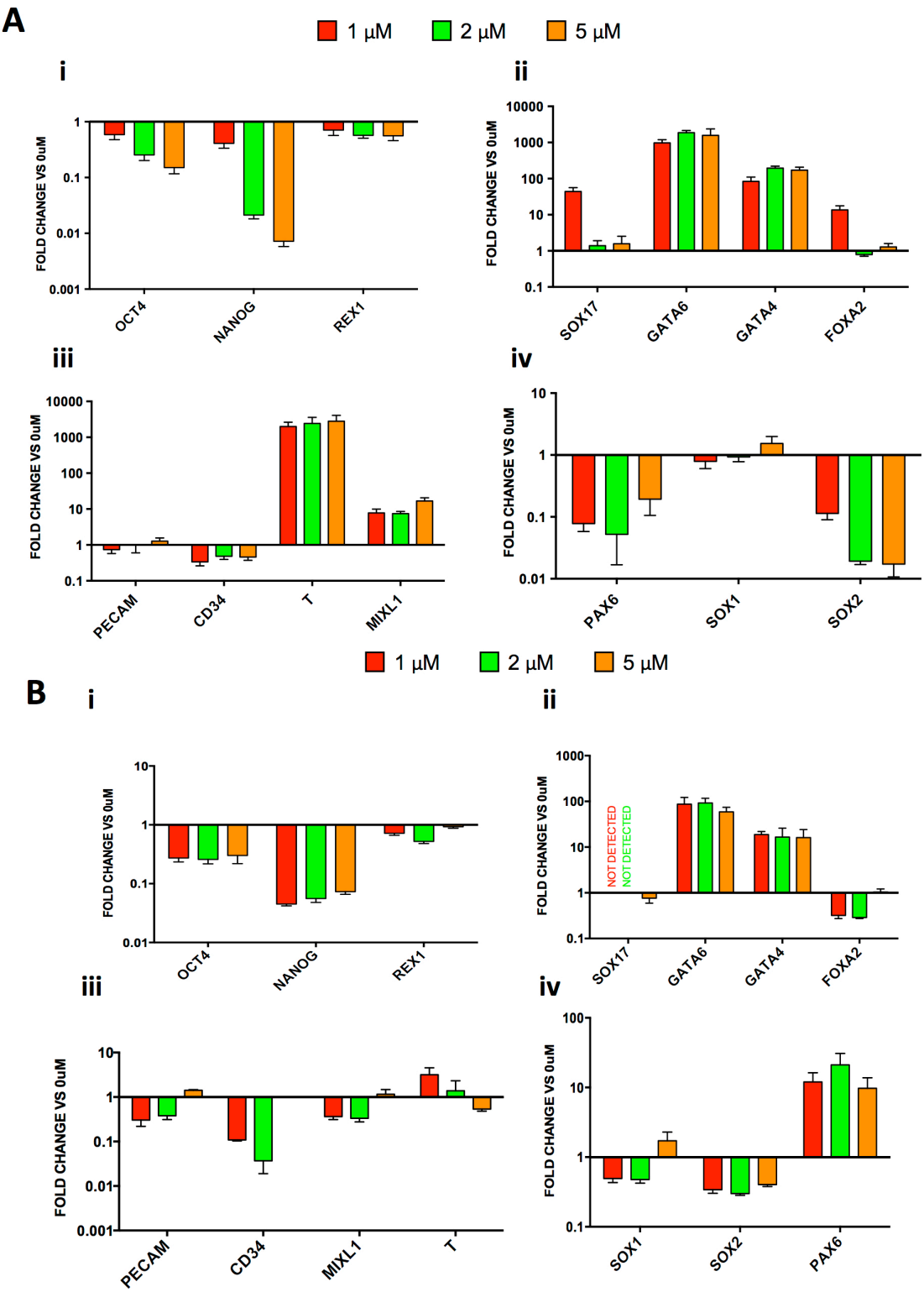
In conclusion, WNT activation through the GSK3β inhibitor efficiently induced *GATA6* expression and generated the four cell fractions identified in KO/SR & MEF culture that were otherwise missing in defined culture. WNT signalling therefore appears to play, at least in part, a role in the generation of *GATA6*(+) cells in culture.

Figure 5.7: WNT signalling induces the four cell fractions in defined conditions



(A) Flow cytometric analysis of Shef4 GATA6:GFP line after 2 days exposure with 2 µM CT 99021 on single cells in E8 media. Varying concentrations of CT 99021 were used to induce GATA6 to determine a dose response. **(B)** Quantification of the percentage of the SSEA3+/*GATA6*+ cells induced by varying concentration of CT 99021 after 2 days exposure. **(C)** Flow cytometric analysis of GATA6 induction upon treatment with varying concentrations of the MEK1/2 inhibitor PD 0325901 after 2 days exposure.

Figure 5.8: WNT signalling and MEK inhibition induce differentiation of hESCs towards different cell types



(A) qPCR analysis of CT 99021 treated cells at 1 µM (red), 2 µM (green) and 5µM (orange) and **(B)** PD 0325901 treated cells at 0.5 µM (red), 1µM (green) and 2µM (orange) concentrations for stem cell **(i)**, endoderm **(ii)**, mesoderm **(iii)** and ectoderm **(iv)** associated genes. Shown as fold change against untreated control samples using beta-actin as the normalising gene. Error bars represent standard deviation of three technical repeats.

5.2.6 Functional characterisation of CT 99021 induced GATA6(+) cells

To examine whether CT 99021 induced *GATA6*(+) cells were functionally equivalent to *GATA6*(+) cells in KO/SR & MEF culture, we performed functional assays on cells pre-treated with 2 μM of CT 99021 for two days. Consistent with previous findings, we were able to generate the four cell fractions similar to the KO/SR & MEF conditions **(Fig. 5.9A),** which we went on to FACS sort. To assess the self-renewing potential of each fraction, single cells were sorted and seeded at clonogenic density of 500 cells/cm^2 and colonies were left to develop for 4 days. Colonies were fixed and stained for OCT4 or SOX2.

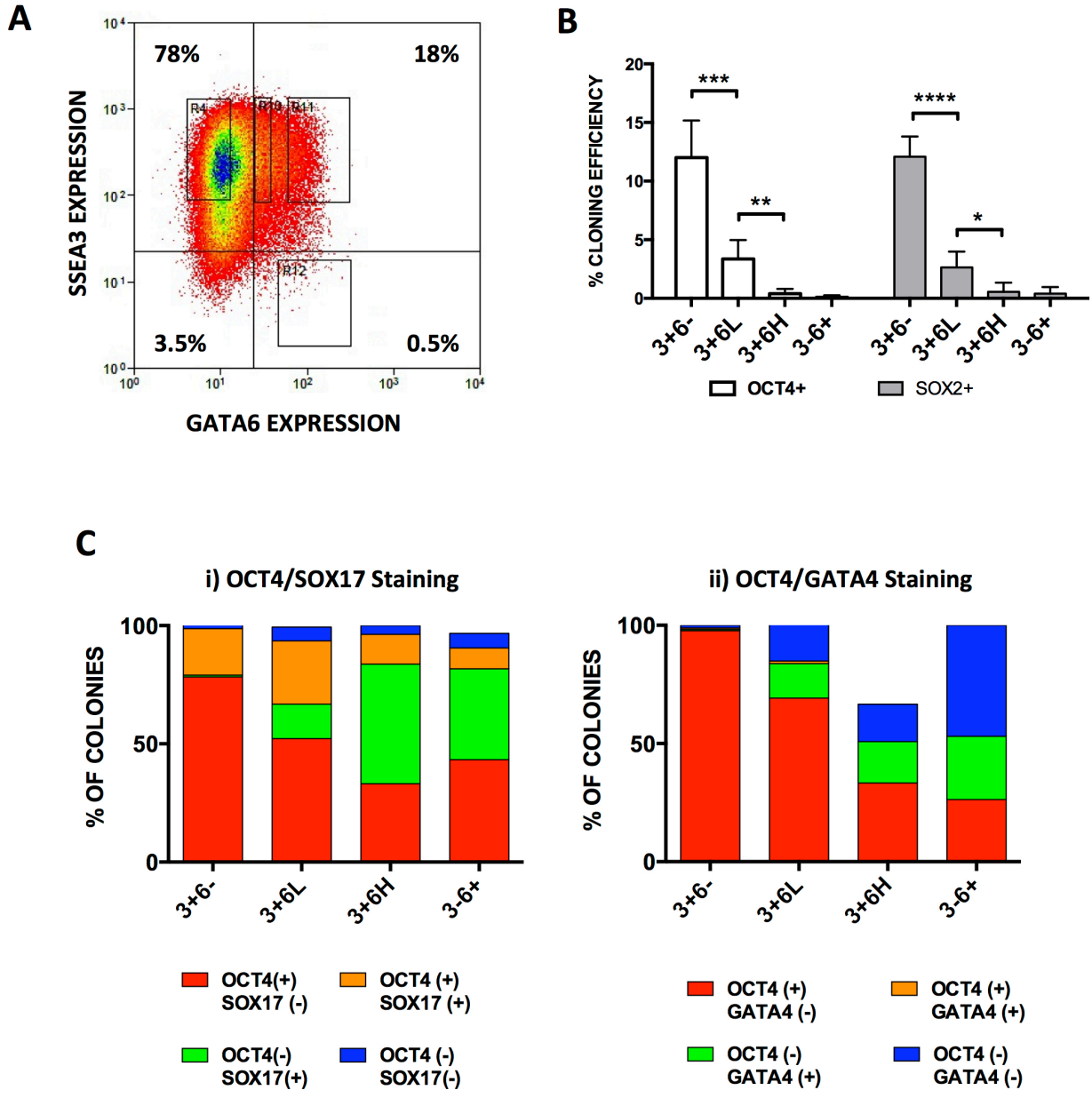
Firstly, we found that each fraction (3+6-, 3+6L, 3+6H & 3-6+) on vitronectin/E8 treated with 2 μM CT 99021 for 2 days, similarly to KO/SR & MEF conditions, generated both OCT4 and SOX2 positive stem cell colonies **(Fig. 5.9B),** but we found that the cloning efficiencies between fractions were different. Most notably, the 3+6L fraction cloned at a significantly lower efficiency than the 3+6- fraction, which was not observed between the two fractions on KO/SR & MEFs **(Fig. 5.9B)**. The 3+6H fraction had a further reduction in cloning efficiency, as did the 3-6+ fraction when compared to 3+6- **(Fig. 5.9B)**. Although the reduction in cloning efficiency of the 3+6H and 3-6+ fractions was consistent with what we observed on KO/SR & MEF conditions when compared to 3+6-, the reduced cloning efficiency in this case was more dramatic. Thus, when comparing the CT 99021 treated and KO/SR & MEF conditions, all *GATA6*(+) fractions had a reduction in cloning efficiency, implying a higher rate of differentiation and/or cell death from CT 99021 treatment.

Upon analysis of each fraction with respect to single cell differentiation bias, we found that the 3+6L fraction in the CT 99021 pre-treated cells as opposed to the 3+6H fraction in KO/SR & MEF conditions showed greatest endodermal bias **(Fig. 5.9C)**. Additionally, we saw a relatively high percentage of colonies within the 3+6- fraction that were double positive for *OCT4* and *SOX17*, which was not observed in KO/SR & MEF conditions to the same extent. The majority of

colonies from the 3+6H fraction contained only *SOX17*+ cells and likely represent endoderm commitment within this fraction. Curiously, the same was not true when using *GATA4* as the endodermal marker **(Fig. 5.9C.ii)**. With CT 99021 pre-treatment, almost no colonies showed co-staining of *OCT4* and *GATA4* in any fraction, even though there were colonies containing *GATA4* only.

Having identified and characterised the existence of a *GATA6*(+) substate within the stem cell compartment, that biases cells to the endoderm lineage in both KO/SR / MEF and in fully defined conditions, and with some knowledge of the mechanism through which these biased stem cells arise, we sought to develop culture conditions in which these cells could be stabilised and propagated.

Figure 5.9: WNT activation induces the *GATA6*(+) cell fractions at the expense of cloning efficiency



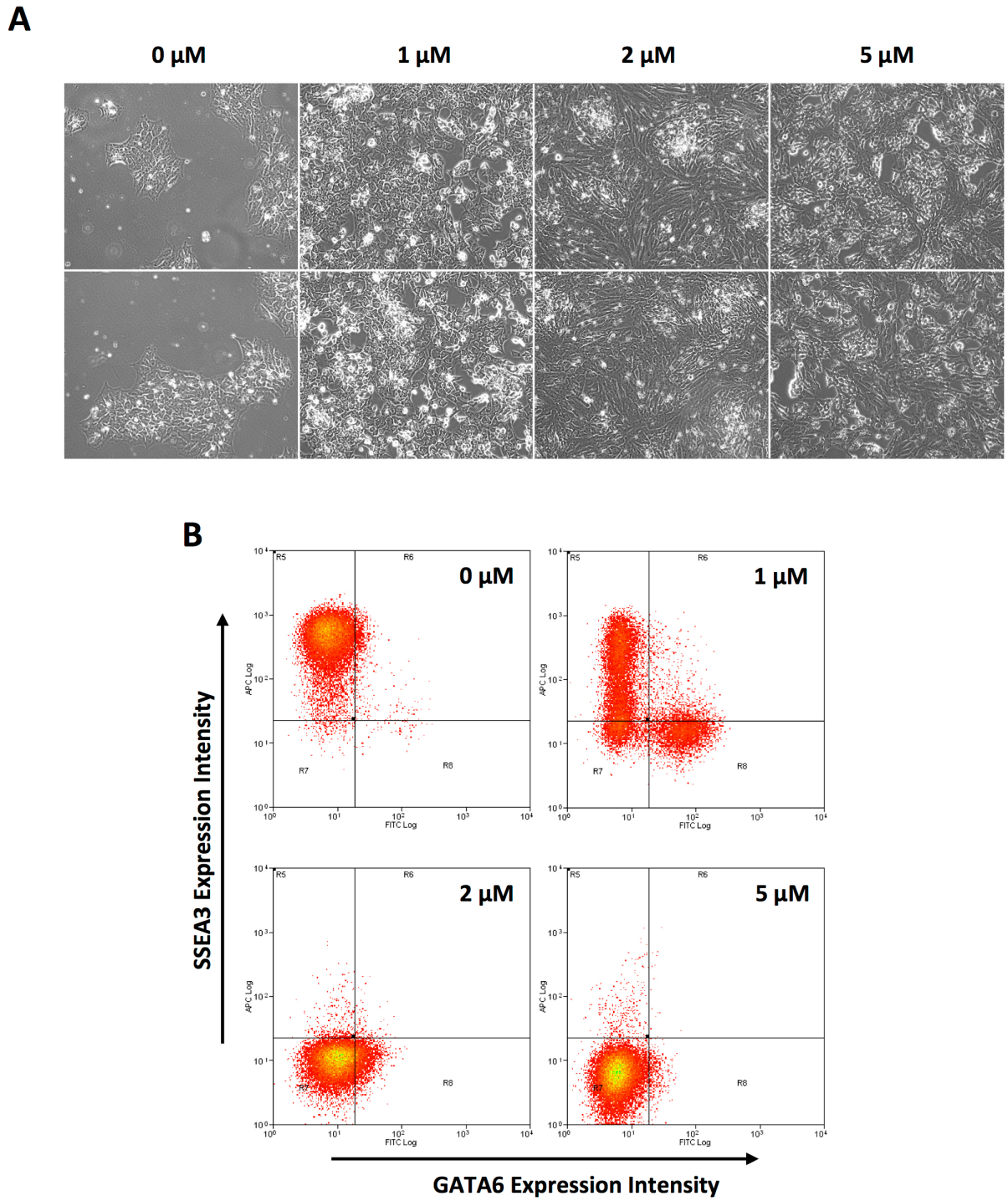
(A) FACS plots of cells pre-treated with 2 µM CT 99021 for 2 days prior to cell sorting and gating of cells for clonogenic assays. **(B)** Quantitative clonogenic assay based on OCT4 (left) or SOX2 (right) from the four cells fractions 4 days post sort. Cells were seeded at clonogenic density of 500 cells/cm^2, and stem cell colonies were classified by having at least 1 OCT4 or SOX2 positive cell. Graphs are representative of 3 biological repeats, error bars indicate standard deviation from 5 technical repeats. Significance was calculated with students t-test. **(C)** Quantitative single cell clonogenic assay to assess differentiation bias of fractions pre-treated with 2 µM CT 99021 for 2 days using SOX17 **(i)** and GATA4 **(i)** as the eodnderm marker. Only colonies with 10% SOX17 within OCT4+ SOX17+ colonies were counted as double positive. Graphs represent three biological repeats each with three technical repeats

5.2.7 Development of media conditions to stabilise the *GATA6*(+), endodermally biased substate

5.2.7.1 Prolonged CT 99021 exposure results in hESC differentiation

As CT 99021 was effective at inducing *GATA6* expression, we wanted assess if the exposure of the inhibitor could propagate *GATA6*(+) stem cells long-term. We therefore cultured cells in the presence 1 μM, 2 μM and 5 μM of the inhibitor. After five days exposure, cells under each concentration tested had a very different morphology to untreated cells. Cells had lost the compacted colony morphology and had proliferated to form larger and longer cells **(Fig. 5.10A)**. FACS analysis showed that cells exposed to high concentrations of CT 99021 had almost completely lost the expression of SSEA-3 as well as *GATA6* **(Fig. 5.10B)**. Cells at 1 μM had distinct differentiated SSEA-3(-) *GATA6*(+) but also had a population of SSEA-3(+) *GATA6*(-) cells, which had resisted differentiation even upon CT 99021 exposure **(Fig. 5.10B)**. Prolonged exposure to CT 99021 therefore appeared to induce complete differentiation at 2 μM and above and was unable to stabilise the 3+6H fraction in the long-term. At 1 μM, remaining stem cells were not *GATA6*(+).

Figure 5.10: 5 day exposure to WNT agonist results in hESC differentiation

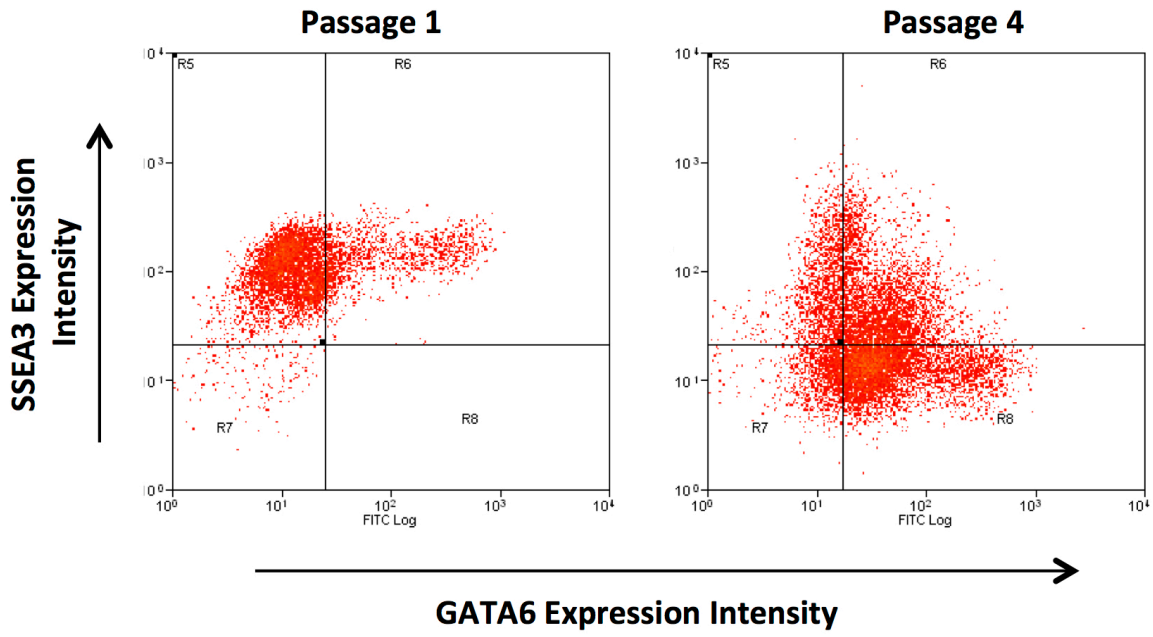


(A) Brightfield images at x4 magnification of cells treated with 1 µM, 2 µM and 5 µM CT 99021 for 5 days. 0 µM represents untreated control. **(B)** FACS analysis of the reporter line after exposure to 0 µM (top left), 1 µM (top right), 2 µM (bottom left) and 5 µM (bottom right) CT 99021 for 5 days in defined conditions. 50,000 cells were seeded with 10 µM Y-27632 24h before exposure. Gates were set against P3X (for SSEA3) and the untransfected Shef4 line (-GFP).

5.2.7.2 GATA6 expressing hESCs are not stable in culture

The inability for a single inhibitor of GSK3β to both induce *GATA6* and hold cells within the stem cell compartment then led to the idea that a combination of inhibitors may be required to block CT 99021 induced hESC differentiation. Upon cross comparison of the inhibitors identified within our screen to induce *GATA6*, we found that both CT 99021 and PD 0325901 were integral components of all media conditions in which 'naïve' hESCs had been derived. (Gafni et al. 2013; Ware et al. 2014; Valamehr et al. 2014; Theunissen et al. 2014). Strikingly, one particular set of conditions reported the derivation of pre-implantation epiblast like hESCs that expressed *GATA6* (Chan et al. 2013). This medium used mTeSR1 as the base with the addition of BIO, PD 0325901, Dorsomorphin and rhLIF, and was called 3iL (3 inhibitors + LIF). The targeted pathways in 3iL were also the pathways that were found to induce *GATA6* expression in our initial inhibitor screen. We therefore exposed our *GATA6* reporter line to 3iL to ascertain whether these conditions could stabilise the *GATA6*(+) endodermally biased substate. Although the SSEA-3(+)/*GATA6*(+) fraction was induced after one passage, cells began to lose SSEA-3 over the next three passages. *GATA6* expression did remain throughout, but after four passages, cells were difficult to maintain and had lost entirely their hESC morphology **(Fig. 5.11)**.

Figure 5.11: 3iL conditions result in hESC differentiation



(A) FACS analysis of cells cultured in 3iL after 1 passage (left) and 4 passages (right) in 3iL conditions. Gates were set using P3X for SSEA3(-) control, and untransfected Shef4 line for GFP(-) control. Cells were passaged as single cells without 10 µM Y-27632 addition. *GATA6* is induced, but SSEA3 is lost over time.

5.2.7.3 Off-target effect of chemical inhibitors results in hESC differentiation

By cross-referencing both BIO and Dorsomorphin with the MRC Kinase Screen database we found that both these inhibitors strongly inhibit kinases that were classed as off target effects. We hypothesised that the inability to stabilise stem cells in these conditions were due to the off-target effects, therefore we titrated each inhibitor as well as LIF in an attempt to stabilise hESCs.

Upon titrating BIO and Dorsomorphin as well as rhLIF **(Fig. 5.12A)**, we were still unable to maintain the SSEA-3(+)/*GATA6*(+) cell phenotype for more than two passages. Cells rapidly lost the expression of SSEA3, and were difficult to maintain in culture due to differentiation **(Fig. 5.12B)**. We therefore used more specific inhibitors to GSK3β and inhibitors to BMP signalling that did not show the extensive off-target effects of BIO and Dorsomorphin. We termed this new media formulation '3iL Sheffield' (2 μM CT 99021, 1 μM PD 0325901, 2 μM DMH-1 & 10 ng/mL rhLIF). Under these conditions, we were able to successfully propagate hESCs that retained normal hESC morphology similar to the parental line **(Fig. 5.13A)**, which remained SSEA-3(+) over a prolonged period of time **(Fig. 5.13B)**. Cells under these modified 3iL conditions were passaged up to 10 passages with no apparent loss of hESC morphology or the stem cell marker SSEA-3 **(Fig. 5.13B)**. All media conditions for the derivation of naive-like cells thus far, with the exception of a report by Takashima et al require feeders in order to stabilise hESCs within the naïve state. Using our new 3iL conditions we wanted to assess whether hESCs could be stabilised in feeder free conditions. Similarly to previous reports, however, we were not able to maintain SSEA-3(+) hESCs in feeder free conditions in our improved 3iL media. Morphologically, cells began differentiating after the first passage **(Fig. 5.13C)** and FACS analysis demonstrated the loss of SSEA-3 **(Fig. 5.13D).**

We then assessed whether the 3iL Sheffield conditions were inducing *GATA6* in hESCs. In contrast to the published 3iL conditions, the optimised 3iL Shef

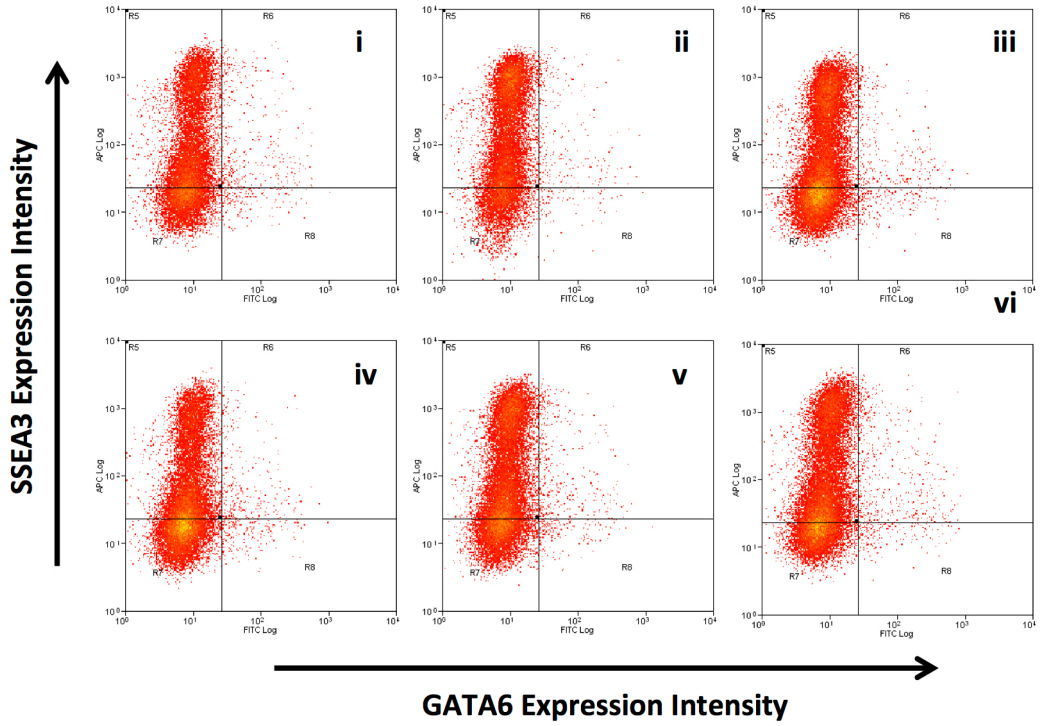
conditions did not lead to the upregulation in expression of *GATA6* at levels any higher than in KO/SR & MEF culture conditions **(Fig. 5.13B)**.

Figure 5.12: Titration of 3iL does not result in the stabilisation of *GATA6* expressing hESCs

A

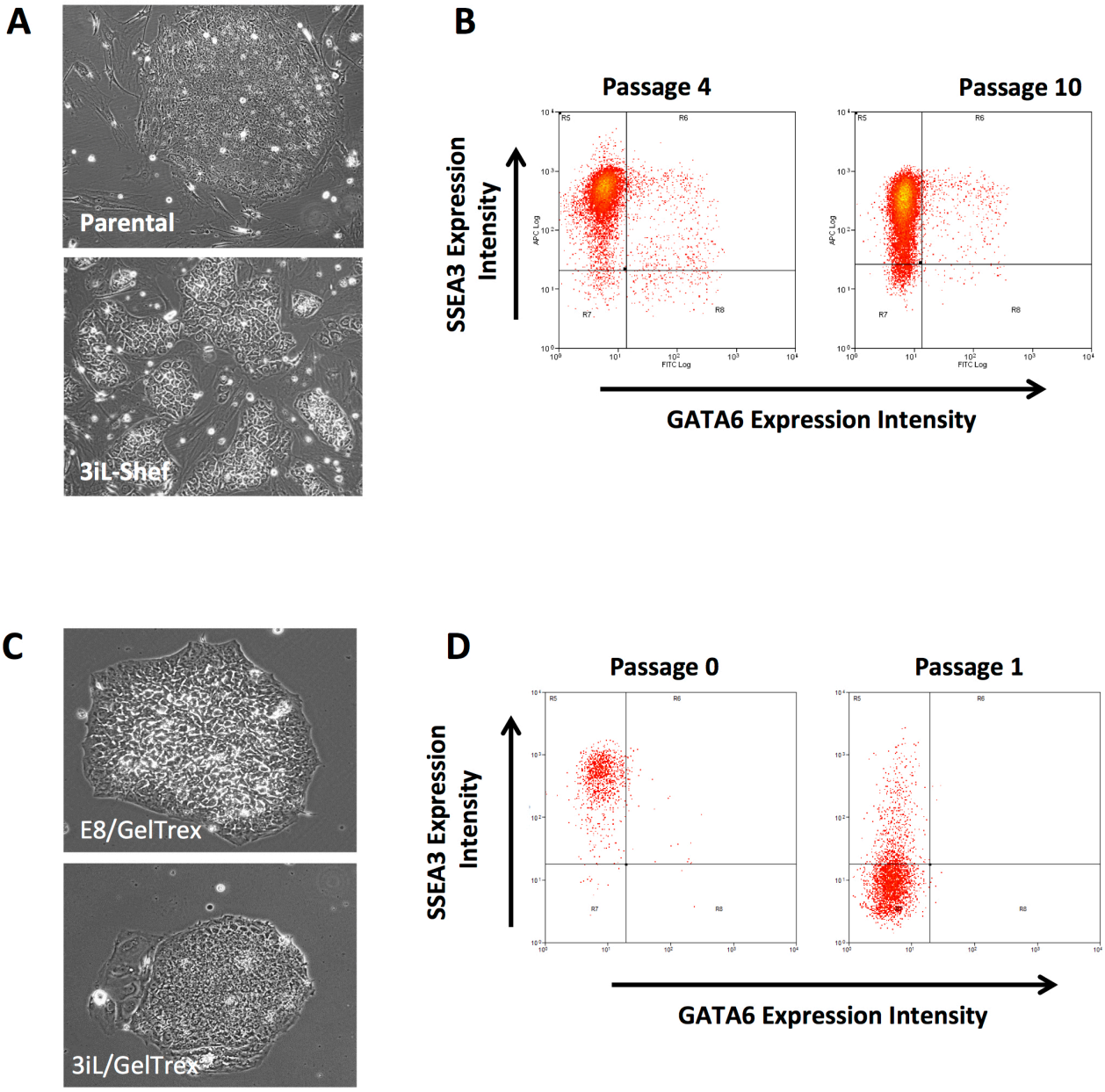
	PD 0325901	BIO	Dorsomorphin	LIF
Media i	1 µM	2 µM	2 µM	10ng/mL
Media ii	1 µM	1 µM	2 µM	10ng/mL
Media iii	1 µM	2 µM	1 µM	10ng/mL
Media iv	1 µM	2 µM	2 µM	15ng/mL
Media v	1 µM	1 µM	2 µM	15ng/mL
Media vi	1 µM	2 µM	1 µM	15ng/mL

B



(A) Table describing the different concentrations and combination of BIO, Dorsomorphin and rhLIF used to try to stabilise GATA6 expressing hESCs. PD 0325901 was kept constant due to minimal off target effects. **(B)** FACS analysis of cells treated for 2 passages with the titrated concentrations of inhibitors in 3iL media. Gates were set using P3X for SSEA3(-) control, and untransfected Shef4 line for GFP(-) control. Titrating 3iL did not result in *GATA6* expressing hESCs.

Figure 5.13: Optimised 3iL conditions result in stabilisation of hESCs



(A) Brightfield image at x4 magnification of the reporter line under standard KO/SR & MEF culture (top) and in 3iL Shef conditions (bottom) after 10 passages. Corresponding FACS plots for the cells in 3iL at passage four (left) and ten (right) are shown in **(B)**. **(C)** Brightfield images at x4 magnification of the reporter line under feeder free conditions (E8 / Geltrex) (top) and the reporter line also in feeder free conditions but in the 3iL Shef media (bottom). **(D)** Corresponding FACS plots of SSEA3 expression of cells in feeder free conditions with 3iL Shef media for passage zero (left) and passage one (right).

5.2.7.4 Further optimisation could not yield GATA6 expressing hESCs

To investigate whether the induction of GATA6 was due to an off-target effect of inhibitors used in the original 3iL formulation, we identified several key pathways involved in hESC specification and self-renewal which were also being targeted by the action of BIO and Dorsomorphin, but not by CT 99021 or DMH-1. Protein kinase C (PKC), FGF, VEGF, Src and PDK1 were all strongly inhibited by BIO and/or Dorsomorphin and all have roles in hESC self-renewal and specification **(Fig. 5.14A)**. Several pathways other than those listed are also involved in hESC self-renewal or specification but specific inhibitors were not commercially available. Inhibitors to each pathway were then added to the optimised 3iL Sheffield, generating what we called 7iL (3iL + Go 6893 - PKC inhibitor; Su 5402 - FGF 7 VEGF inhibitor; WH-4-023 - Src inhibitor and GSK 2334470 - PDK1 inhibitor). Cells were exposed to the new 7iL media formulation to assess hESC propagation and *GATA6* expression. Cells exposed to the new 7iL conditions did propagate well initially, and did retain hESC like morphology **(Fig. 5.14B)**. Cells were maintained over four passages, but over time cells did begin to lose SSEA-3 expression **(Fig. 5.14C)**. Although cells in 7iL could be maintained for up to four passages, the level of *GATA6* was not upregulated at any point under these conditions, and therefore these targeted pathways, at least in combination, did not promote *GATA6* expression. We were therefore unable to stabilise cells residing within the 3+6H substate.

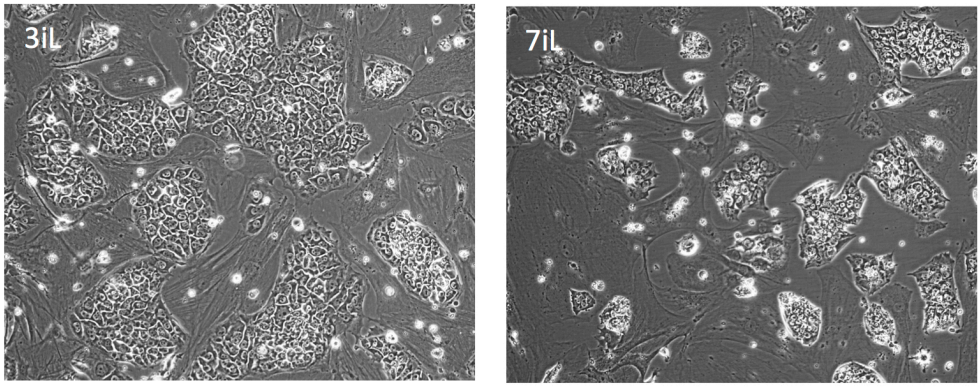
In conclusion, we were able to demonstrate that high *GATA6* expression leads to endoderm biased stem cells, and that WNT signalling is likely to be, at least in part, responsible for the appearance of these cells in culture. We unfortunately were unable to stabilise *GATA6* expression hESCs.

Figure 5.14: Inhibition of off-target effects of BIO do not induce *GATA6* expression

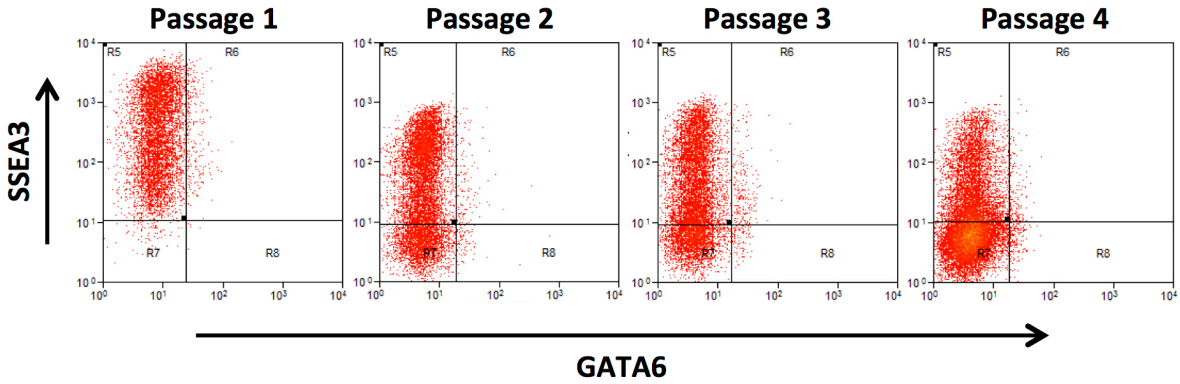
A

PATHWAY	% REMAINING ACTIVITY
PDK1	1
FGF-R1	3
Src	5
VEGFR1	5
PKC alpha	37

B



C



(A) The identification of off-target signalling pathways from 1 µM BIO and/or Dorsomorphin. Data was obtained from the MRC Kinase Screen. Values represent the percentage of remaining activity of kinases after 1 µM BIO treatment. **(B)** Brightfield images at x4 magnification of cells cultured for two passages in 3iL Sheffield (left) and 7iL Sheffield (right) with corresponding FACS plots for SSEA3 and *GATA6* expression after 3 passages **(C)**. Cells lose SSEA3 over time and do not express *GATA6*. Gates were set using the Shef4 parental line for –GATA6 and P3X for SSEA3- for all FACS plots.

5.3 Discussion

Following on from our data demonstrating the existence of stem cells that express the early endodermal marker *GATA6*, we wanted to investigate whether these cells showed directed differentiation bias. Firstly, we investigated at the population level whether *GATA6* conferred a lineage bias under differentiation conditions. Using a defined, neutral EB system, we demonstrated that cell fractions expressing *GATA6* at both low and high levels showed differentiation bias towards predominantly the endoderm, but also the mesoderm lineages at the expense of ectoderm when compared to the *GATA6* negative fraction. The morphology of EBs from the 3+6H and 3-6+ fractions resembled that of 'cystic EBs' originally classified in EC cell derived EBs that contained a wide range of endodermal cell types (Itskovitz-Eldor et al. 2000). Consistent with previous reports that *GATA6* is a key regulator and is fundamental in the specification of endoderm lineages (Koutsourakis et al. 1999; Fujikura et al. 2002; Morrisey et al. 1996; Morrisey et al. 1998) we found within our assay that *GATA6* also appeared to be a positive and sensitive readout for lineage specification, as even at low expression levels, was able to identify cells with an endodermal fate. Although the strongest upregulation was in genes for the endodermal germ layer, we did find that mesoderm associated genes were also upregulated within EBs derived from the GATA6 positive fractions. GATA6 is implicated specifically in the endoderm lineage during the segregation of the ICM and extra-embryonic lineages in the mouse blastocyst, but it is possible that a small proportion of SSEA-3(+) *GATA6*(+) cells are representative of mesodermal lineages as *GATA6* is also expressed in cells of the lateral plate mesoderm which gives rise to structures such as the cardiac tract (Koutsourakis et al. 1999). Additionally, as we have shown, inter-converting, pluripotent stem cells do exist within all of the *GATA6*(+) fractions (3+6-, 3+6L, 3+6H, 3-6+) and these stem cells may be accountable for the appearance of the multi-lineage differentiation in the 3+6L and 3+6H fraction. The random nature of EB differentiation would be permissive for the appearance of the three primary germ layers (Itskovitz-Eldor et al. 2000). The step-wise decrease in ectoderm specification within the 3+6H

and 3-6+ fractions would also correlate with the reduction of the pluripotent stem cell pool capable of inter-converting and specifying neural lineages, which are generally not associated with *GATA6* at this stage of development. The population differentiation, however, was not informative as to whether biased stem cells, or a population of differentiated cells was responsible for the observed bias.

We therefore devised an assay in which the self-renewal potential and differentiation bias could be assessed at single cell level. This assay allowed us to analyse OCT4 positive stem cell colonies positive for the appearance of endoderm differentiation using SOX17 and GATA4. We classified a colony as being OCT4 and SOX17 or GATA4 positive if that colony had at least 1 OCT4+ cell, and at least 10% SOX17 or GATA4 positive cells. We chose 10% for SOX17 and GATA4 as the culture conditions that cells were seeded into (KO/SR & MEFs) can be inductive for mesendoderm differentiation due to the secretion of WNT protein by MEFs (Hao et al. 2006). We therefore wanted to try and eliminate colonies that were showing stochastic differentiation as opposed to actual stem cell bias. Increasing the percentage past 10% resulted in very few colonies being double positive, which implies the effect of the differentiation bias observed within stem cells is subtle.

By quantifying the percentage of these colonies within each of the cell fractions (3+6-, 3+6L, 3+6H & 3-6+), we found that the 3+6H fraction consisted of significantly more stem cell colonies with SOX17 or GATA4 positive cells, indicative of a higher percentage of stem cells possessing endoderm bias within this fraction. Although SOX17 and GATA4 have been implicated in lineages from an non-endodermal origin at later stages of development, in particular cardiac cells (Masino et al. 2004; Martin et al. 2004), published reports strongly link the expression of SOX17 and GATA4 during the first differentiation events of hESCs into endodermal lineages (D'Amour et al. 2005; Wang et al. 2011; Fujikura et al. 2002).

Although the 3+6L fraction at population level under differentiation conditions did show differentiation bias, under self-renewing conditions, this bias was eradicated. Thus, there would appear to be discreet substates within the stem cell compartment that represent different functional behaviours in individual cells, which are context dependent and characterise different stages of lineage bias. Within the *GATA6* scenario, cells that express *GATA6* at low levels would represent a state in which cells become more sensitive to differentiation cues, when compared to *GATA6* negative cells. This functionally manifests as lineage biased differentiation within the EB context. In the presence of self-renewal conditions and the absence of differentiation cues, however, this substate retains functional equivalence to *GATA6* negative cells with respect to self-renewal potential. *GATA6* at high levels represents a second substate in which cells also show a heightened sensitivity to differentiate, but under self-renewing conditions, retains differentiation bias, even within stem cell colonies. The observation of multiple stages of cell bias would confirm our previous data that differentiation commitment is not a deterministic, but is a probabilistic event that is context dependent.

We have therefore shown that high *GATA6* expression shifts self-renewal vs. differentiation probabilities in such a way as to favour not only differentiation, but also lineage specific as opposed to stochastic differentiation, in this case, towards endoderm. 3+6-, 3+6L, 3+6H and 3-6+ therefore appear to represent incremental stages along the self-renewal versus differentiation pathway, whereby endodermal differentiation is the most probable route. Although few, the appearance of colonies within all of the SSEA-3(+) fractions containing cells that did not have either OCT4, SOX17 or GATA4 staining may represent differentiation to a lineage other than endoderm. This proves important in the context of lineage bias versus lineage priming. Lineage bias implies that cells are capable of choosing lineage fates other than endoderm, although the probability they do choose endoderm is highly favoured. Lineage priming would imply more that cells are moving solely towards an endodermal trajectory without necessarily committing at the expense of other lineages.

To better understand the mechanism through which biased cells arise in KO/SR & MEF culture, we performed a screen in fully defined conditions on inhibitors to several pathways known to impinge on hESC behaviour (Chan et al. 2013; Gafni et al. 2013; Takashima et al. 2013; Theunissen et al. 2014). The action of CT 99021, an inhibitor of GSK3β which activates canonical WNT signalling, was capable of activating the expression of *GATA6* and recapitulating the 4 fractions seen in KO/SR & MEF conditions. qPCR analysis demonstrated that the treatment of cells with CT 99021 pushed cells to a mesendodermal like cell type through upregulation of endoderm and mesoderm and the downregulation of ectoderm specific genes, consistent with the role of WNT signalling during development (Bakre et al. 2007; Sumi et al. 2008; Aubert et al. 2002). Importantly, this lineage gene expression profile was similar to the fractions in KO/SR & MEF conditions. Although the exact role of WNT signalling in hESCs is still in dispute (Sato et al. 2004; Davidson et al. 2012) WNT signalling does have roles during gastrulation by setting up the primitive streak, directing axis formation and subsequently giving rise to migrating mesoderm and embryonic endoderm, both of which involve *GATA6* (Liu et al. 1999; Tam & Behringer 1997; Huggon et al. 1997; Zhao et al. 2005). It has been shown that mouse embryonic fibroblasts (MEFs) secrete detectable levels of many WNT proteins including WNT3a (Hao et al. 2006), which may in part explain the origins of *GATA6*(+) stem cells in KO/SR & MEF conditions. We did find that the functional behaviour of *GATA6*(+) cells from KO/SR & MEF compared to CT 99021 treated cells was different, in that *GATA6*(+) fractions from CT 99021 treated cells were less clonogenic than the equivalent fractions in KO/SR & MEF conditions. The substantial downregulation of stem cell associated genes after short-term CT 99021 exposure, that is not apparent in KO/SR & MEF culture, would indicate a significantly higher level of WNT signalling by CT 99021 addition, impairing self-renewal. This data also implies that the levels of *GATA6* expression do not necessarily correlate with differentiation, as similar levels of *GATA6* expression between CT 99021 treated and cells in KO/SR & MEF ultimately results in different functional outcomes.

Having established that cells of the 3+6H fraction in KO/SR & MEF conditions, and 3+6-/3+6L in CT 99021 treated cells show fate bias at the single cell level, it became apparent that if these cells could be stabilised and propagated as biased stem cells, it may provide a platform for better controlled, more uniform endodermal differentiation. To better understand the mechanism and signalling behind the *GATA6* biased substate we attempted to trap cells within the *SSEA-3*(+) / *GATA6*(+) fraction. Work by Ying et al demonstrated that the capacity for self-renewal and the retention of pluripotency of ground state mESCs involves the prevention of cells leaving that state, rather than the active process of maintaining it (Ying et al. 2008). We therefore postulated that *GATA6*(+) hESCs could be stabilised not simply through the active process of inducing *GATA6*, but by also blocking differentiation pathways. This idea became feasible after the publication of conditions known as 3iL (3 inhibitors + LIF), which also contained the WNT activator BIO and which gave rise to *GATA6* expressing hESCs (Chan et al. 2013). We were unable, however, to repeat these findings with our reporter line as 3iL conditions rapidly resulted in differentiation. We did find that through the formulation of an optimised medium using more specific inhibitors, we could stabilise hESCs. We found that the off-target effects of inhibitors originally used by Chan et al were potentially responsible for differentiation as replacement of BIO with CT 99021 and Dorsomorphin with DMH-1, which show much less off-target effects, stabilised long-term hESC self-renewal. Off-target pathways including IGF-1R, FGF-R1 and PDK1 (to a lesser extent) are know to be heavily involved in self-renewal, thus the destabilisation of self-renewal is not surprising upon their inhibition (Bendall et al. 2007; Burdon et al. 2002).

Even though we were able to stabilise pluripotent stem cells using a combination of more specific inhibitors to the same pathways identified by Chan et al, we did not see the induction of *GATA6* expression within these cells, as we saw using the original 3iL conditions. Again we hypothesised that off-target effects of inhibitors used within the original 3iL conditions were responsible for *GATA6* activation either directly or indirectly. Other existing media formulations that use CT 99021 for WNT activation in combination with

other pathway inhibitors for the propagation of hESCs have not reported the elevated expression of *GATA6*. (Takashima et al. 2014; Gafni et al. 2013; Ware et al. 2014). This indicates that the use of BIO as opposed to CT 99021 for WNT activation may play a part in *GATA6* induction, and that it is likely not WNT signalling alone, at least within this context, which is responsible. Further work is required to dissect specific pathways that relate to *GATA6* expression within the 3iL context.

In conclusion, we have shown that *GATA6*(+) cells in KO/SR & MEF culture show a heightened sensitivity to differentiate and exhibit endoderm bias at the population level under differentiation conditions. We found that it is specifically the high expression of *GATA6* in KO/SR & MEF conditions that results in endoderm biased stem cells under self-renewing conditions. Care should therefore be taken when describing hESCs as being 'primed' or 'biased' as our data shows that lineage gene expression levels are important in the functional translation of a cell, whereby, at least in the case of *GATA6* within our cell line, low expression levels do not translate to behavioural alterations in self-renewing conditions.

Although we were not able to stabilise the 3+6H state in culture, our work provides evidence that one or more off-target effect of the GSK3β inhibitor BIO may result in *GATA6* expression. The ability to assign particular signalling pathways with a molecular phenotype in a particular context is currently very difficult due to the nature of off-target effects of chemical inhibitors (Karaman et al. 2008).

6 Final Discussion

6.1 Development of a defined, reproducible differentiation assay

Our results demonstrate that by controlling the input cell number and the quality of the starting population of hESCs, highly reproducible differentiation of pooled EBs can be achieved. The use of EBs to better mimic embryonic development and for the generation of cell types of interest have been gaining substantial interest, and substantial progress has been made (Ungrin et al. 2008; Ng 2005). We wanted to build upon existing knowledge to develop a reproducible differentiation assay that could be efficiently manipulated and controlled through the use of exogenous cues. The reproducibility of this assay subsequently led to its use as a model for predictive reproductive toxicology. Although we only used a small panel of genes to assess drug induced differential gene expression, we demonstrated that the controlled differentiation of EBs did lead to a high success rate in the identification of known teratogens in concordance with existing data. The reproducibility and the high-throughput format of this assay has the potential to accurately screen for the on-going generation of thousands of new compounds by academia and industry. As therapeutics progress, however, it is emerging that personalised medicine is becoming a new requirement for healthcare, as it is very much apparent that individual patients respond differently to drugs. Within this assay, we use a single hESC line, H9, which almost certainly is not representative of the genetic make-up of the general population. A more powerful approach to this assay would be to incorporate induced pluripotent stem cell (iPS) technology so that the genetic, and epigenetic make-up of individual patients is reflected (Rubin 2008; Davila et al. 2004). The incorporation of iPS technology in predictive toxicology would also allow for a high-throughput approach to fully understand the mechanisms behind drug induced toxicity, which could aid in the development of both general and personalised medicine. Furthermore, the use of iPS technology would allow for detailed modelling and subsequently, a

thorough understanding of the mechanistic cause for disease (Allison et al. 2015). A more systematic approach could then be applied to the drug development process.

6.2 Differentiation of hESCs is heterogeneous

During the development of the differentiation assay, we noticed that single embryoid bodies, even under uniform differentiation conditions, did not show exactly the same morphologies, and showed differential gene expression patterns. Even though this assay was designed to eliminate, as much as possible, random differentiation, it is possible that the discrepancy between EB differentiation was due to further uncontrolled parameters during the formation process, for example, the lack of spatial restriction, previously shown to be beneficial in differentiation reproducibility (Ungrin et al. 2008). We found that the spatial organisation of EB formation, although important, was unlikely to be the only parameter in the hindrance of reproducibility as by purifying the quality of the starting population of hESCs using SSEA-3, we were able to reduce gene expression variability, thereby improving reproducibility. This raised the possibility that the heterogeneity in differentiation of hESCs may be more inherent to the properties of individual stem cells, and the dynamics of the pluripotent stem cell compartment. Supportive of this theory is the finding that single cells of the embryonal carcinoma cell line NTERA2/D1 appear to make lineage decisions before commitment (Tonge et al. 2010). The uniform application of exogenous differentiation agents therefore may not be sufficient to override any intrinsic decisions that have been already made by the cells, thus resulting in heterogeneity in the resulting cell population. Heterogeneity of surface markers and gene expression patterns within the embryonic stem cell compartment have been documented and may be indicative of stem cells with these pre-made commitment decisions (Graf & Stadtfeld 2008). Furthermore, the observation that lineage associated genes can be highly expressed with

pluripotency associated genes in single hESCs implies certain lineages may be favoured if a cell then commits to differentiate (Gokhale et al. 2015). Thus, the heterogeneous differentiation of hESCs even when exposed to uniform differentiation conditions may be, in part, accounted for by the use of a heterogeneous starting population of stem cells. The mechanisms of lineage decisions thus appears more complex than the general activation or inhibition of signalling pathways and indicates that subtle forms of heterogeneity within the stem cell compartment can alter specification trajectories.

6.3 Functionally discreet substates in the stem cell compartment

Following on from our observations and existing reports that hESC differentiation is not a uniform process, we went back to the basic biology of fate decision mechanisms in hESCs. We were interested in the expression of the early endodermal marker *GATA6* due to the high level of co-expression with genes associated with pluripotency (Gokhale et al. 2015). In concordance with this report, we identified cells in culture that co-expressed the stem cell marker SSEA-3 and *GATA6*, whereby a subset of these cells retained self-renewal capacity over extended periods of time. The identification of a novel substates within the stem cell compartment, characterised by *GATA6* expression supports the concept that hierarchies of stem cells exist in culture (Laslett et al. 2007). We have shown at the single cell level that the *GATA6* substate does confer functional alterations on the behaviour of cells, such that differentiation is biased towards the endoderm lineage. The finding that cells within the haematopoietic stem cell system transition through stages of priming before commitment to differentiate also translates to our findings within the pluripotent stem cell scenario (Chang et al. 2006). Cells with low *GATA6* eaxpression under differentiation conditions show lineage bias, but under self-renewal conditions do not, whereas cells with high *GATA6* expression exhibit differentiation bias in both conditions, indicative of incremental stages towards

the committed phenotype. The concept of substates may also be applicable to adult systems, where evidence exists, for example in stem cells of the intestinal crypt, that the stem cell compartment shows degrees of heterogeneity in differentiation gene expression. Cells moving up the crypt progressively take on a more differentiated phenotype, which eventually becomes terminal. Some cells, however, can also move back down the crypt and still have the capacity for multi-lineage differentiation and self-renewal (Booth & Potten 2000). The question as to why functional heterogeneity exists within the stem cell compartment remains an unanswered question, but hypotheses propose that heterogeneity allows the retention of self-renewal capacity but provides windows of opportunity to respond to extrinsic environmental cues for specific differentiation (Graf & Stadtfeld 2008). The extensive heterogeneity seen within hESCs may also be the reason for their pluripotency, and it is the gradual suppression of heterogeneity as cells transition through differentiation to multipotent cell types, which eventually results in a terminally differentiated specialised cell. In regards to the translation of *in vitro* heterogeneity to the *in vivo* context, there is evidence that heterogeneity does exist within pluripotent cells of the ICM (Chazaud et al. 2006). These findings raise the question of whether all individual cells of the ICM are truly bound for one fate, or rather primed with the capacity to revert to an alternate cell type should the embryo require. Furthermore, the few examples of *in vivo* heterogeneity do not negate its importance in development. The tightly controlled environment of the embryo may mean that the substates representing heterogeneity are transitory, and the forced retention of the stem cell state *in vitro* allows for these states to be manifested. The mechanisms behind the generation of these fluctuations in transcription factor expression and sensitisation to extrinsic signalling within stem cells remain important questions in stem cell biology.

6.4 WNT signalling generates GATA6 expressing hESCs

The underlying mechanism of what drives the expression of lineage genes in hESCs, without necessarily resulting in their commitment to differentiate is unknown. With an understanding of how cells can balance two opposing lineages would not only provide information as to how the embryo regulates heterogeneity, but it would also offer potential strategies to develop more efficient and uniform differentiation *in vitro*. Our results indicate that the activation of WNT signalling, at least in part, is responsible for the generation of *GATA6* expressing, endodermally biased stem cells in culture. A lack of mechanistic understanding of this substate meant, however, that we were unable to stabilise the *GATA6* substate within the stem cell compartment. Further work is required to find stable conditions for the propagation of these cells, for further analysis and for improved directed differentiation. The observation that the ground state in mESCs is actually maintained by the prevention of cells leaving the state through signalling inhibition as opposed to the active process of maintaining it may be relevant within this context (Ying et al. 2008). The long-term maintenance of *GATA6* expressing stem cells may require the blocking of signalling pathways inducing commitment, but further work is required to elucidate which pathways are important. Finally, we found that we could induce endoderm-biased stem cells in feeder free conditions by inducing WNT signalling. The generation of these cells in feeder free conditions has provided a step forward to capturing these endoderm biased stem cells in a clinically relevant context. Although we were unable to stabilise GATA6 expressing stem cells, the act of pre-treatment with WNT agonists (specifically CT 99021 in this study) to induce biased cells, and to sensitise them to further differentiation may aid in homogenising differentiation whilst increasing the efficiency at which endoderm progenitors are generated. Pre-treatment of cells using chemical inhibitors has been used to improve the efficiencies of differentiation, for example rapamycin and the specification of definitive endoderm (Tahamtani et al. 2013), but the underlying mechanisms have not been studied. Our observation of discreet functional substates may well be

pertinent in explaining these studies. Equally, the resulting homogeneity of differentiated cell types has not been fully investigated in this context. Finally, our study provides a paradigm that could be used with other gene reporters to identify and characterise further substates within the stem cell compartment that may bias cells towards both mesoderm and ectoderm. This study may also provide a paradigm to adult stem cells, which have also found to be heterogeneous (Graft & Stadtfeld, 2008).

7 Concluding remarks

In this study, we sought to delineate lineage specific differentiation of hESCs, with a view to improving our understanding of human development. We developed a reproducible differentiation assay using embryoid bodies to better recapitulate embryonic development, which also proved useful as a model for predictive toxicology. Within this assay, it was apparent that the differentiation of stem cells proceeded in a non-uniform and heterogeneous fashion. We therefore went back to interrogate the cause of this heterogeneity. Consequently, we found that discreet, yet functional substates existed within the stem cell compartment, and these states represented alterations in the probabilities of cell fate decisions. Using a *GATA6* reporter line, we were able to identify and characterise an endoderm biased state in uncommitted stem cells. Furthermore, we have shed light onto a potential mechanism through which these biased stem cells arise. With a deeper knowledge of the causes and consequences of heterogeneity within the stem cell compartment, we have developed a platform in which the differentiation of hESCs *in vitro* could be more homogeneous and thus more controllable. Finally, the assays we have developed could be applied in the search for other functionally discrete substates, not only within the pluripotent stem cell compartment, but also more committed cell types.

References

Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P., Beighton, G., Bello, P., Benvenisty, N., Berry, L., Bevan, S., Blum, B., Brooking, J., Chen, K., Choo, A., Churchill, G., Corbel, M., Damjanov, I., Draper, J., Dvorak, P., Emanuelsson, K., Fleck, R., Ford, A., Gertow, K., Gertsenstein, M., Gokhale, P., Hamilton, R., Hampl, A., Healy, L., Hovatta, O., Hyllner, J., Imreh, M., Itskovitz-Eldor, J., Jackson, J., Johnson, J., Jones, M., Kee, K., King, B., Knowles, B., Lako, M., Lebrin, F., Mallon, B., Manning, D., Mayshar, Y., McKay, R., Michalska, A., Mikkola, M., Mileikovsky, M., Minger, S., Moore, H., Mummery, C., Nagy, A., Nakatsuji, N.,

O'Brien, C., Oh, S., Olsson, C., Otonkoski, T., Park, K.-Y., Passier, R., Patel, H., Patel, M., Pedersen, R., Pera, M., Piekarczyk, M., Pera, R., Reubinoff, B., Robins, A., Rossant, J., Rugg-Gunn, P., Schulz, T., Semb, H., Sherrer, E., Siemen, H., Stacey, G., Stojkovic, M., Suemori, H., Szatkiewicz, J., Turetsky, T., Tuuri, T., Brink, S., Vintersten, K., Vuoristo, S., Ward, D., Weaver, T., Young, L., Zhang, W., 2007. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nature Biotechnology 25, 803–816.

Aksoy, I., Giudice, V., Delahaye, E., Wianny, F., Aubry, M., Mure, M., Chen, J., Jauch, R., Bogu, G., Nolden, T., Himmelbauer, H., Doss, M., Sachinidis, A., Schulz, H., Hummel, O., Martinelli, P., Hübner, N., Stanton, L., Real, F., Bourillot, P.-Y., Savatier, P., 2014. Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells. Nature Communications 5, 3719.

Aleck, K., Bartley, D., 1997. Multiple malformation syndrome following fluconazole use in pregnancy: report of an additional patient. American journal of medical genetics 72, 253–256.

Allison, T.F., Powles-Glover, N., Biga, V., Barbaric, I., 2015. Human pluripotent stem cells as tools for high-throughput and high-content screening in drug discovery. International Journal of High Throughput Screening 5, 1–13.

Amaya, E., Musci, T., Kirschner, M., 1991. Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in xenopus embryos. Cell 66, 257–270.

Ameri J, Ståhlberg A, Pedersen J, Johansson J, Johannesson M, Artner I, Semb H. 2010. FGF2 specifies hESC-derived definitive endoderm into foregut/midgut cell lineages in a concentration-dependent manner. Stem cells 28:45–56.

Andrews, PW, 2002. From teratocarcinomas to embryonic stem cells. Philosophical Transactions of the Royal Society B: Biological Sciences.

Andrews, P., 1988. Human teratocarcinomas. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 948, 17–36.

Andrews, P., Bronson, D., Benham, F., Strickland, S., Knowles, B., 1980. A comparative study of eight cell lines derived from human testicular teratocarcinoma. International journal of cancer. Journal international du cancer 26, 269–280.

Armstrong, L., Hughes, O., Yung, S., Hyslop, L., Stewart, R., Wappler, I., Peters, H., Walter, T., Stojkovic, P., Evans, J., Stojkovic, M., Lako, M., 2006. The role of PI3K/AKT, MAPK/ERK and NFo2 signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. Human Molecular Genetics 15, 1894–1913.

Arnold, S., Robertson, E., 2009. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nature Reviews Molecular Cell Biology 10, 91–103.

Artus, J., Chazaud, C., 2014. A close look at the mammalian blastocyst: epiblast and primitive endoderm formation. Cellular and molecular life sciences : CMLS 71, 3327–3338.

Artzt, K., Dubois, P., Bennett, D., Condamine, H., Babinet, C., Jacob, F., 1973. Surface antigens common to mouse cleavage embryos and primitive teratocarcinoma cells in culture. Proceedings of the National Academy of Sciences of the United States of America 70, 2988–2992.

Aubert, J., Dunstan, H., Chambers, I., Smith, A., 2002. Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. Nature Biotechnology 20, 1240–1245.

Bailey, J., Knight, A., Balcombe, J., 2005. The future of teratology research is in vitro. Biogenic Amines 19, 97–145.

Bakre, M., Hoi, A., Mong, J., Koh, Y., Wong, K., Stanton, L., 2007. Generation of multipotential mesendodermal progenitors from mouse embryonic stem cells via sustained Wnt pathway activation. Journal of Biological Chemistry 282, 31703–31712.

Beattie, G.M., Lopez, A.D., Bucay, N., Hinton, A., Firpo, M.T., KING, C.C., Hayek, A., 2005. Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers 23, 489–495.

Bendall, S., Stewart, M., Menendez, P., George, D., Vijayaragavan, K., Werbowetski-Ogilvie, T., Ramos-Mejia, V., Rouleau, A., Yang, J., Bossé, M., Lajoie, G., Bhatia, M., 2007. IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. Nature 448, 1015–1021.

Benjamin, S., Ishak, K., Zimmerman, H., Grushka, A., 1981. Phenylbutazone liver injury: a clinical-pathologic survey of 23 cases and review of the literature. 1, 255–263.

Bhatia, S., Pilquil, C., Roth-Albin, I., Draper, J., n.d. Demarcation of stable subpopulations within the pluripotent hESC compartment. PLoS ONE 8, e57276.

Blauwkamp, T., Nigam, S., Ardehali, R., Weissman, I., Nusse, R., 2012. Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. Nature Communications 3, 1070.

Booth, C, Potten, CS, 2000. Gut instincts: thoughts on intestinal epithelial stem cells 105, 1493–1499.

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M., Mandalam, R., Lebkowski, J., Stanton, L., 2004. Transcriptome characterization elucidates signaling networks that control

human ES cell growth and differentiation. Nature Biotechnology 22, 707–716.

Brass, E., 1993. Hepatic toxicity of antirheumatic drugs. 60, 466–472.

Brent, R., 1964. Drug testing in animals for teratogenic effects: thalidomide in the pregnant rat. The Journal of Pediatrics 64, 762–770.

Brinster, R., 1974. The effect of cells transferred into the mouse blastocyst on subsequent development. The Journal of Experimental Medicine 140, 1049–1056.

Burdon, T., Smith, A., Savatier, P., 2002. Signalling, cell cycle and pluripotency in embryonic stem cells. Trends in cell biology 12, 432–438.

Canham, M., Sharov, A., Ko, M., Brickman, J., 2010. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. PLoS Biology 8, e1000379.

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

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., Smith, A., 2007. Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230–1234.

Chambers, S., Fasano, C., Papapetrou, E., Tomishima, M., Sadelain, M., Studer, L., 2009. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nature Biotechnology 27, 275–280.

Chan, Y.-S., Göke, J., Ng, J.-H., Lu, X., Gonzales, K., Tan, C.-P., Tng, W.-Q., Hong, Z.-Z., Lim, Y.-S., Ng, H.-H., 2013. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. Cell Stem Cell 13,

663–675.

Chang, H., Hemberg, M., Barahona, M., Ingber, D., 2008. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells 453, 544–547.

Chang, H., Oh, P., Ingber, D., Huang, S., n.d. Multistable and multistep dynamics in neutrophil differentiation. BMC cell biology 7, 11.

Chazaud, C., 2006. Lineage allocation and asymmetries in the early mouse embryo. Philosophical Transactions of the Royal Society B: Biological Sciences 358, 1341–1349.

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

Chen, S., Fitzgerald, W., Zimmerberg, J., Kleinman, H., Margolis, L., 2007. Cell-cell and cell-extracellular matrix interactions regulate embryonic stem cell differentiation. Stem cells (Dayton, Ohio) 25, 553–561.

Cheng, A., Saxton, T., Sakai, R., Kulkarni, S., 1998. Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. Cell 95, 793–803.

Chiplunkar, A, Lung, T, Alhashem, Y, 2013. Krüppel-like factor 2 is required for normal mouse cardiac development 8, e54891.

Cockburn, K., Rossant, J., 2010. Making the blastocyst: lessons from the mouse. The Journal of clinical investigation 120, 995–1003.

Conlon F, Lyons K, Takaesu N, Barth K, Kispert A, Herrmann B, Robertson E. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development 120, 1919–1928.

D'Amour, K., Agulnick, A., Eliazer, S., Kelly, O., Kroon, E., Baetge, E., 2005. Efficient differentiation of human embryonic stem cells to definitive endoderm. Nature Biotechnology 23, 1534–1541.

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

Daston, G.P., 2007. Genomics and developmental risk assessment. Birth defects research. Part A, Clinical and molecular teratology 79, 1–7.

Davidson, K., Adams, A., Goodson, J., McDonald, C., Potter, J., Berndt, J., Biechele, T., Taylor, R., Moon, R., 2012. Wnt/^{2}-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. Proceedings of the National Academy of Sciences 109, 4485–4490.

Davila, J., Cezar, G., Thiede, M., Strom, S., Miki, T., Trosko, J., 2004. Use and application of stem cells in toxicology. Toxicological Sciences 79, 214–223.

Delassus, S., Titley, I., Enver, T., 1999. Functional and molecular analysis of hematopoietic progenitors derived from the aorta-gonad-mesonephros region of the mouse embryo. Blood 94, 1495–1503.

Denning, C., Anderson, D., 2008. Cardiomyocytes from human embryonic stem cells as predictors of cardiotoxicity. Drug Discovery Today: Therapeutic Strategies 5, 223–232.

Doetschman, T., Eistetter, H., Katz, M., Schmidt, W., Kemler, R., 1985. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. Journal of embryology and experimental morphology 87, 27–45.

Doetschman, T., Gregg, R., Maeda, N., Hooper, M., Melton, D., Thompson, S., Smithies, O., 1987. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature 330, 576–578.

Dunn, N., Vincent, S., Oxburgh, L., Robertson, E., Bikoff, E., 2004. Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. Development 131, 1717–1728.

Eiraku, M., Takata, N., Ishibashi, H., Kawada, M., Sakakura, E., Okuda, S., Sekiguchi, K., Adachi, T., Sasai, Y., 2011. Self-organizing optic-cup morphogenesis in three-dimensional culture. Nature 472, 51–6.

Eiselleova L, Matulka K, Kriz V, Kunova M, Schmidtova Z, Neradil J, Tichy B, Dvorakova D, Pospisilova S, Hampl A, et al. 2009. A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells. Stem Cells 27:1847–57.

Enver, T., Pera, M., Peterson, C., Andrews, P., 2009. Stem Cell States, Fates, and the Rules of Attraction. Cell Stem Cell 4, 387–397.

Enver, T., Soneji, S., Joshi, C., Brown, J., Iborra, F., Orntoft, T., Thykjaer, T., Maltby, E., Smith, K., Dawud, R., 2005. Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. Human Molecular Genetics 14, 3129–3140.

Evans, M., Bradley, A., Kuehn, M., Robertson, E., 1985. The ability of EK cells to form chimeras after selection of clones in G418 and some observations on the integration of retroviral vector proviral DNA into EK cells. Cold Spring Harb Symp Quant Biol 50, 685–689.

Evans, M.J., Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154–6.

Evseenko, D., Zhu, Y., Schenke-Layland, K., Kuo, J., Latour, B., Ge, S., Scholes, J., Dravid, G., Li, X., MacLellan, W., Crooks, G., 2010. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. Proceedings of the National Academy of Sciences 107, 13742–13747.

Feldman, B., Poueymirou, W., Papaioannou, V., DeChiara, T., Goldfarb, M., 1995. Requirement of FGF-4 for postimplantation mouse development. Science 267, 246–249.

Finch, B., Ephrussi, B., 1967. Retention of multiple developmental potentialities by cells of a mouse testicular teratocarcinoma during prolonged culture in vitro and their extinction upon hybridization with cells of permanent lines. Proceedings of the National Academy of Sciences of the United States of America 57, 615–621.

Fischer, Y., Ganic, E., Ameri, J., Xian, X., Johannesson, M., Semb, H., 2010. NANOG reporter cell lines generated by gene targeting in human embryonic stem cells. PLoS ONE 5, e12533.

Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Ji, J., Niwa, H., 2002. Differentiation of embryonic stem cells is induced by GATA factors. Genes & Development 16, 784–789.

Furusawa, T., Honda, C., Tokunaga, T., Takahashi, S., Ohkoshi, K., 2004. Embryonic Stem Cells Expressing Both Platelet Endothelial Cell Adhesion Molecule-1 and Stage-Specific Embryonic Antigen-1 Differentiate Predominantly into Epiblast Cells in a Chimeric Embryo. Biology of Reproduction 70, 1452–1457.

Gafni, O., Weinberger, L., Mansour, A., Manor, Y., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., Rais, Y., Shipony, Z., Mukamel, Z., Krupalnik, V., Zerbib, M., Geula, S., Caspi, I., Schneir, D., Shwartz, T., Gilad, S., Amann-Zalcenstein, D., Benjamin, S., Amit, I., Tanay, A., Massarwa, R.,

Novershtern, N., Hanna, J., 2013. Derivation of novel human ground state naive pluripotent stem cells. Nature 504, 282–286.

Genschow, E., Spielmann, H., Scholz, G., Pohl, I., Seiler, A., Clemann, N., Bremer, S., Becker, K., 2004. Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. Alternatives to laboratory animals 32, 209–244.

Gokhale, P., Au-Young, J., Dadi, S., Keys, D., Harrison, N., Jones, M., Soneji, S., Enver, T., Sherlock, J., Andrews, P., 2015. Culture Adaptation Alters Transcriptional Hierarchies among Single Human Embryonic Stem Cells Reflecting Altered Patterns of Differentiation. PloS one 10, e0123467.

Gottmann, E., Kramer, S., Pfahringer, B., Helma, C., 2001. Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. Environmental health perspectives 109, 509–514.

Graf, T., Stadtfeld, M., 2008. Heterogeneity of Embryonic and Adult Stem Cells. Cell Stem Cell 3, 480–483.

Greber, B., Lehrach, H., Adjaye, J., 2007. Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal. Stem cells 25, 455–464.

Greenhough, S, Medine, CN, Hay, DC, 2010. Pluripotent stem cell derived hepatocyte like cells and their potential in toxicity screening. Toxicology 278, 250–255.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Vivian, N., Goodfellow, P., Lovell-Badge, R., 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature 346, 245–250.

Guo, G., Huss, M., Tong, G., Wang, C., Sun, L., Clarke, N., Robson, P., 2010. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Developmental Cell 18, 675–685.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., 2009. Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development 135, 1063-1069.

Gustafson, A., Stedman, D., Ball, J., Hillegass, J., Flood, A., Zhang, C., Panzica-Kelly, J., Cao, J., Coburn, A., Enright, B., Tornesi, M., Hetheridge, M., Augustine-Rauch, K., 2012. Inter-laboratory assessment of a harmonized zebrafish developmental toxicology assay - progress report on phase I. Reproductive Toxicology 33, 155–164.

Hao, J., Li, T.-G., Qi, X., Zhao, D.-F., Zhao, G.-Q., 2006. WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. Developmental Biology 290, 81–91.

Hart, A., Hartley, L., Sourris, K., Stadler, E., Li, R., Stanley, E., Tam, P., Elefanty, A., Robb, L., 2002. Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo. Development 129, 3597–3608.

Hay, M., Thomas, D., Craighead, J., Economides, C., Rosenthal, J., 2014. Clinical development success rates for investigational drugs. Nature biotechnology 32, 40–51.

Hayashi, K., Lopes, S., Tang, F., Surani, M., 2008. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell stem cell 3, 391–401.

Hayflick, L., 1965. The limited in vitro lifetime of human diploid cell strains. Experimental Cell Research 37, 614–636.

Heldin, C., Miyazono, K., Dijke, P., n.d. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. Nature 390, 465–471.

Hemmati-Brivanlou, A., Melton, D., 1992. A truncated activin receptor inhibits mesoderm induction and formation of axial structures in Xenopus embryos. Nature 359, 609–614.

Hogan, B., Fellous, M., Avner, P., Jacob, F., 1977. Isolation of a human teratoma cell line which expresses F9 antigen. Nature 270, 515–518.

Hu, M., Krause, D., Greaves, M., Sharkis, S., 1997. Multilineage gene expression precedes commitment in the hemopoietic system. Genes & Development 11, 774-785

Huggon, I., Davies, A., Gove, C., Moscoso, G., Moniz, C., Foss, Y., Farzaneh, F., Towner, P., 1997. Molecular cloning of human GATA-6 DNA binding protein: high levels of expression in heart and gut. Biochimica et biophysica acta 1353, 98–102.

Hwang, Y.-S., Chung, B., Ortmann, D., Hattori, N., Moeller, H.-C., Khademhosseini, A., 2009. Microwell-mediated control of embryoid body size regulates embryonic stem cell fate via differential expression of WNT5a and WNT11. Proceedings of the National Academy of Sciences 106, 16978–16983.

Hyslop, L., Stojkovic, M., Armstrong, L., Walter, T., Stojkovic, P., Przyborski, S., Herbert, M., Murdoch, A., Strachan, T., Lako, M., 2005. Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. Stem cells 23, 1035–1043.

Inman, W., Pearce, G., Wilton, L., 1994. Safety of fluconazole in the treatment of vaginal candidiasis. A prescription-event monitoring study, with special reference to the outcome of pregnancy. European journal of clinical

pharmacology 46, 115–118.

Itskovitz-Eldor, J., Schuldiner, M., Karsenti, D., Eden, A., Yanuka, O., Amit, M., Soreq, H., Benvenisty, N., 2000. Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. Molecular medicine 6, 88.

James, D., Levine, A., Besser, D., Hemmati-Brivanlou, A., 2005. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development 132, 1273–1282.

Karaman, M., Herrgard, S., Treiber, D., Gallant, P., Atteridge, C., Campbell, B., Chan, K., Ciceri, P., Davis, M., Edeen, P., Faraoni, R., Floyd, M., Hunt, J., Lockhart, D., Milanov, Z., Morrison, M., Pallares, G., Patel, H., Pritchard, S., Wodicka, L., Zarrinkar, P., 2008. A quantitative analysis of kinase inhibitor selectivity. Nature Biotechnology 26, 127–132.

Kehat, I, Kenyagin-Karsenti, D, Snir, M, 2001. Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. Journal of clinical investigation 108, 407–414.

Kessler, D., Melton, D., 1994. Vertebrate embryonic induction: mesodermal and neural patterning. Science 266, 596–604.

Keutz, E., Schlüter, G., 1998. Preclinical safety evaluation of cerivastatin, a novel HMG-CoA reductase inhibitor. The American journal of cardiology 82, 11J–17J.

Kinoshita, K, Ura, H, Akagi, T, Usuda, M, Koide, H, 2007. GABPα regulates Oct-3/4 expression in mouse embryonic stem cells. Biochemical and Biophysical Research Communications 353, 686–691.

Kleinsmith, L., Pierce, G., 1964. Multipotency of single embryonal carcinoma

cells. Cancer research 24, 1544–1551.

Kondziolka, D., Wechsler, L., Goldstein, S., Meltzer, C., Thulborn, K., Gebel, J., Jannetta, P., DeCesare, S., Elder, E., McGrogan, M., Reitman, M., Bynum, L., 2000. Transplantation of cultured human neuronal cells for patients with stroke. Neurology 55, 565–569.

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

Kurek, D., Neagu, A., Tastemel, M., Tüysüz, N., Lehmann, J., Werken, H., Philipsen, S., Linden, R., Maas, A., IJcken, W., Drukker, M., Berge, D., 2015. Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. Stem Cell Reports 4, 114–128.

Laflamme, M., Chen, K., Naumova, A., Muskheli, V., Fugate, J., Dupras, S., Reinecke, H., Xu, C., Hassanipour, M., Police, S., O'Sullivan, C., Collins, L., Chen, Y., Minami, E., Gill, E., Ueno, S., Yuan, C., Gold, J., Murry, C., 2007. Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. Nature Biotechnology 25, 1015–1024.

Laslett, A., Grimmond, S., Gardiner, B., Stamp, L., Lin, A., Hawes, S., Wormald, S., Nikolic-Paterson, D., Haylock, D., Pera, M., 2007. Transcriptional analysis of early lineage commitment in human embryonic stem cells. BMC Developmental Biology 7, 12.

Lee, H., Shamy, G., Elkabetz, Y., Schofield, C., Harrsion, N., Panagiotakos, G., Socci, N., Tabar, V., Studer, L., 2007. Directed differentiation and transplantation of human embryonic stem cell-derived motoneurons. Stem cells 25, 1931–1939.

Li, J., Wang, G., Wang, C., Zhao, Y., Zhang, H., Tan, Z., Song, Z., Ding, M., Deng, H.,

2007. MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. Differentiation; research in biological diversity 75, 299–307.

Lim, J., Bodnar, A., 2002. Proteome analysis of conditioned medium from mouse embryonic fibroblast feeder layers which support the growth of human embryonic stem cells. Proteomics 2, 1187–1203.

Lim, S., Pereira, L., Wong, M., Hirst, C., Vranken, B., Pick, M., Trounson, A., Elefanty, A., Stanley, E., 2009. Enforced expression of Mixl1 during mouse ES cell differentiation suppresses hematopoietic mesoderm and promotes endoderm formation. Stem cells 27, 363–374.

Lin, Y., Chen, G., 2014. Embryoid body formation from human pluripotent stem cells in chemically defined E8 media. StemBook 1.

Liu, F., Ventura, F., Doody, J., Massagué, J., 1995. Human type II receptor for bone morphogenic proteins (BMPs): extension of the two-kinase receptor model to the BMPs. Molecular and cellular biology 15, 3479–3486.

Liu, P., Wakamiya, M., Shea, M., Albrecht, U., Behringer, R., Bradley, A., 1999. Requirement for Wnt3 in vertebrate axis formation. Nature Genetics 22, 361–365.

Livak, K., Schmittgen, T., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408.

Maci as-Silva, M., Abdollah, S., Hoodless, P., Pirone, R., 1996. MADR2 is a substrate of the TGF^2 receptor and its phosphorylation is required for nuclear accumulation and signaling. Cell 87, 1215–1224.

Martin, C., Meeson, A., Robertson, S., Hawke, T., 2004. Persistent expression of

the ATP-binding cassette transporter, Abcg2, identifies cardiac SP cells in the developing and adult heart. Developmental Biology 265, 262–275.

Martin, G., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United State of America 78, 7634–7638.

Martin, G., Evans, M., 1975. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. Proceedings of the National Academy of Sciences of the United States of America 72, 1441–1445.

Masino, A., Gallardo, T., Wilcox, C., Olson, E., Williams, R., Garry, D., 2004. Transcriptional regulation of cardiac progenitor cell populations. Circulation Research 95, 389–397.

Melichar, H., Li, O., Ross, J., Haber, H., Cado, D., Nolla, H., Robey, E., Winoto, A., 2011. Comparative study of hematopoietic differentiation between human embryonic stem cell lines. PLoS ONE 6, e19854.

Miller, M., Strömland, K., 1999. Teratogen update: thalidomide: a review, with a focus on ocular findings and new potential uses. Teratology 60, 306–321.

Moore, J., Fu, J., Chan, Y.-C., Lin, D., Tran, H., Tse, H.-F., Li, R., 2008. Distinct cardiogenic preferences of two human embryonic stem cell (hESC) lines are imprinted in their proteomes in the pluripotent state. Biochemical and biophysical research communications 372, 553–558.

Moretto-Zita, M., Jin, H., Shen, Z., Zhao, T., Briggs, S., Xu, Y., 2010. Phosphorylation stabilizes Nanog by promoting its interaction with Pin1. Proceedings of the National Academy of Sciences 107, 13312–13317.

Morizane, A., Doi, D., Kikuchi, T., Nishimura, K., Takahashi, J., 2011. Small-molecule inhibitors of bone morphogenic protein and activin/nodal signals

promote highly efficient neural induction from human pluripotent stem cells. Journal of Neuroscience Research 89, 117–126.

Morrisey, E., Ip, H., Lu, M., Parmacek, M., 1996. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. Developmental Biology 177, 309–322.

Morrisey, E., Tang, Z., Sigrist, K., Lu, M., Jiang, F., Ip, H., Parmacek, M., 1998. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes & Development 12, 3579–3590.

Morrison GM, Oikonomopoulou I, Migueles RP, Soneji S, Livigni A, Enver T, Brickman JM. 2008. Anterior definitive endoderm from ESCs reveals a role for FGF signaling. Cell Stem Cell 3:402–15.

Na, J., Furue, M., Andrews, P., 2010. Inhibition of ERK1/2 prevents neural and mesendodermal differentiation and promotes human embryonic stem cell self-renewal. Stem Cell Research 5, 157–169.

Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C., Miyazono, K., Dijke, P., 1997. TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. The EMBO Journal 16, 5353–5362.

Ng, E., Azzola, L., Sourris, K., Robb, L., Stanley, E., Elefanty, A., 2005. The primitive streak gene Mixl1 is required for efficient haematopoiesis and BMP4-induced ventral mesoderm patterning in differentiating ES cells. Development 132, 873–884.

Ng, E., Davis, R., Stanley, E., Elefanty, A., 2008. A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. Nature Protocols 3, 768–776.

Ng, E.S., Davis, R.P., Azzola, L., Stanley, E.G., Elefanty, A.G., 2005. Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. Blood 106, 1601–3.

Ohgushi, M., Matsumura, M., Eiraku, M., Murakami, K., Aramaki, T., Nishiyama, A., Muguruma, K., Nakano, T., Suga, H., Ueno, M., Ishizaki, T., Suemori, H., Narumiya, S., Niwa, H., Sasai, Y., 2010. Molecular pathway and cell state responsible for dissociation-induced apoptosis in human pluripotent stem cells. Cell Stem Cell 7, 225–239.

Olson, E., Arnold, H., Rigby, P., Wold, B., 1996. Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene MRF4. Cell 85, 1–4.

Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Deun, K., Smith, P., Berger, B., Heller, A., 2000. Concordance of the toxicity of pharmaceuticals in humans and in animals. Regulatory toxicology and pharmacology : RTP 32, 56–67.

Osafune, K., Caron, L., Borowiak, M., Martinez, R., Fitz-Gerald, C., Sato, Y., Cowan, C., Chien, K., Melton, D., 2008. Marked differences in differentiation propensity among human embryonic stem cell lines. Nature Biotechnology 26, 313–315.

Park, J., Cho, C., Parashurama, N., Li, Y., Berthiaume, F., Toner, M., Tilles, A., Yarmush, M., 2007. Microfabrication-based modulation of embryonic stem cell differentiation. Lab on a Chip 7, 1018–1028.

Pevny, L., Nicolis, S., 2010. Sox2 roles in neural stem cells. The international journal of biochemistry & cell biology 42, 421–424.

Pierce, G., Dixon, F., Verney, E., 1959. Testicular teratomas. I. Demonstration of teratogenesis by metamorphosis of multipotential Cells. Cancer 12, 573–583.

Pina, C., Fugazza, C., Tipping, A., Brown, J., Soneji, S., Teles, J., Peterson, C., Enver,

T., 2012. Inferring rules of lineage commitment in haematopoiesis. Nature Cell Biology 14, 287–294.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., Hadjantonakis, A.-K., 2008. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development 135, 3081–3091.

Rappolee, D., Basilico, C., Patel, Y., Werb, Z., 1994. Expression and function of FGF-4 in peri-implantation development in mouse embryos. Development 120, 2259–2269.

Rossant, J., Tam, P., 2009. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development 136, 701–713.
Rubin, L., 2008. Stem cells and drug discovery: the beginning of a new era? Cell 132, 549–552.

Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L., Robertis, E., 1994. Xenopus chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. Cell 79, 779–790.

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

Schöler, H., Hatzopoulos, A., Balling, R., Suzuki, N., Gruss, P., 1989. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. The EMBO journal 8, 2543–2550.

Schrode, N., Saiz, N., Talia, S., Hadjantonakis, A.-K., 2014. GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. Developmental cell 29, 454–467.

Schumann, J., 2010. Teratogen screening: state of the art. Avicenna journal of medical biotechnology 2, 115–21.

Simeone, A., Acampora, D., Arcioni, L., Andrews, P., 1990. Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. Nature 346, 763–766.

Singh, A., Reynolds, D., Cliff, T., Ohtsuka, S., Mattheyses, A., Sun, Y., Menendez, L., Kulik, M., Dalton, S., 2012. Signaling Network Crosstalk in Human Pluripotent Cells: A Smad2/3-Regulated Switch that Controls the Balance between Self-Renewal and Differentiation. Cell Stem Cell 10, 312–326.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., Rogers, D., 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature 336, 688–90.

Smith, J., Vallier, L., Lupo, G., Alexander, M., Harris, W., Pedersen, R., 2008. Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. Developmental biology 313, 107–117.

Smith, W., Harland, R., 1992. Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in Xenopus embryos. Cell 70, 829–840.

Smyth, N., Vatansever, H., Meyer, M., Frie, C., Paulsson, M., Edgar, D., 1998. The Targeted Deletion of the LAMC1 Gene. Annals of the New York Academy of Sciences 857, 283–286.

Solter, D., Knowles, B., 1978. Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). Proceedings of the National Academy of Sciences of the United States of America 75, 5565–5569.

Spielmann, H., 1997. The embryonic stem cell test (EST), an in vitro embryotoxicity test using two permanent mouse cell lines : 3T3 fibroblasts and embryonic stem cells. In vitro toxicolpgy 10, 119–127.

Stephens, T., 1988. Proposed mechanisms of action in thalidomide embryopathy. Teratology 38, 229–239.

Stephens, T., Bunde, C., Fillmore, B., 2000. Mechanism of action in thalidomide teratogenesis. Biochemical pharmacology 59, 1489–1499.

Stephenson, R., Rossant, J., Tam, P., 2012. Intercellular interactions, position, and polarity in establishing blastocyst cell lineages and embryonic axes. Cold Spring Harbor Perspectives in Biology 4, 1–17.

Stevens, L., Little, C., 1954. Spontaneous Testicular Teratomas in an Inbred Strain of Mice. Proceedings of the National Academy of Sciences of the United States of America 40, 1080–1087.

Stewart, M., Bossé, M., Chadwick, K., Menendez, P., Bendall, S., Bhatia, M., n.d. Clonal isolation of hESCs reveals heterogeneity within the pluripotent stem cell compartment. Nature Methods 3, 807–815.

Strickland, S., Mahdavi, V., 1978. The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell 15, 393–403.

Sumi, T., Tsuneyoshi, N., Nakatsuji, N., Suemori, H., 2008. Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. Development 135, 2969–2979.

Tada, S., Era, T., Furusawa, C., Sakurai, H., Nishikawa, S., Kinoshita, M., Nakao, K., Chiba, T., Nishikawa, S., 2005. Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell

differentiation culture. Development 132, 4363–4374.

Tahamtani, Y., Azarnia, M., Farrokhi, A., Sharifi-Zarchi, A., Aghdami, N., Baharvand, H., 2013. Treatment of human embryonic stem cells with different combinations of priming and inducing factors toward definitive endoderm. Stem cells and development 22, 9, 1419-

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., Reik, W., Bertone, P., Smith, A., 2014. Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. Cell 158, 1254–1269.

Tam, P., Behringer, R., 1997. Mouse gastrulation: the formation of a mammalian body plan. Mechanisms of development 68, 3–25.

Teo, A., Arnold, S., Trotter, M., Brown, S., Ang, L., Chng, Z., Robertson, E., Dunn, N., Vallier, L., 2011. Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes & Development 25, 238–250.

Tesar, P., Chenoweth, J., Brook, F., Davies, T., Evans, E., Mack, D., Gardner, R., McKay, R., 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448, 196–199.

Theunissen, T., Powell, B., Wang, H., Mitalipova, M., 2014. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell stem cell 15, 471–487.

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

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

Tonge, P., Olariu, V., Coca, D., Kadirkamanathan, V., Burrell, K., Billings, S., Andrews, P., 2010. Prepatterning in the stem cell compartment. PLoS ONE 5, e10901.

Tonge, P.D., Shigeta, M., Schroeder, T., Andrews, P.W., 2011. Functionally defined substates within the human embryonic stem cell compartment. Stem cell research 7, 145–53.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., Niwa, H., 2008. Identification and characterization of subpopulations in undifferentiated ES cell culture. Development 135, 909–918.

Ungrin, M., Joshi, C., Nica, A., Bauwens, C., Zandstra, P., 2008. Reproducible, ultra high-throughput formation of multicellular organization from single cell suspension-derived human embryonic stem cell aggregates. PloS one 3, e1565.

Valamehr, B., Robinson, M., Abujarour, R., Rezner, B., Vranceanu, F., Le, T., Medcalf, A., Lee, T., Fitch, M., Robbins, D., Flynn, P., 2014. Platform for induction and maintenance of transgene-free hiPSCs resembling ground state pluripotent stem cells. Stem Cell Reports 2, 366–381.

Vallier, L., Alexander, M., Pedersen, R.A., 2005. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. Journal of cell science 118, 4495–509.

Waddington, C., 1957. The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. 262.

Wang, P., Rodriguez, R., Wang, J., Ghodasara, A., Kim, S., 2011. Targeting SOX17 in human embryonic stem cells creates unique strategies for isolating and analyzing developing endoderm. Cell Stem Cell 8, 335–346.

Wang, Z., Oron, E., Nelson, B., Razis, S., Ivanova, N., 2012. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell 10, 440–454.

Ware, C., Nelson, A., Mecham, B., Hesson, J., Zhou, W., Jonlin, E., Jimenez-Caliani, A., Deng, X., Cavanaugh, C., Cook, S., Tesar, P., Okada, J., Margaretha, L., Sperber, H., Choi, M., Blau, C., Treuting, P., Hawkins, R., Cirulli, V., Ruohola-Baker, H., 2014. Derivation of naive human embryonic stem cells. Proceedings of the National Academy of Sciences 111, 4484–4489.

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J., Nishikawa, S., Nishikawa, S., Muguruma, K., Sasai, Y., 2007. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nature Biotechnology 25, 681–686.

Wells, J., Melton, D., 2000. Vertebrate endoderm development. Cell and Developmental Biology 15, 393–410.

Wu, H., Xu, J., Pang, Z., Ge, W., Kim, K., Blanchi, B., Chen, C., Südhof, T., Sun, Y., 2007. Integrative genomic and functional analyses reveal neuronal subtype differentiation bias in human embryonic stem cell lines. Proceedings of the National Academy of Sciences 104, 13821–13826.

Xu, R.-H., Chen, X., Li, D., Li, R., Addicks, G., Glennon, C., Zwaka, T., Thomson, J., 2002. BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nature Biotechnology 20, 1261–1264.

Xu, R.-H., Sampsell-Barron, T., Gu, F., Root, S., Peck, R., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R., Thomson, J., 2008. NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. Cell Stem Cell 3, 196–206.

Yamaguchi, T.P., Harpal, K., Henkemeyer, M., Rossant, J., 1994. fgfr-1 is required

for embryonic growth and mesodermal patterning during mouse gastrulation. Genes & development 8, 3032–44.

Yamanaka, Y., Lanner, F., Rossant, J., 2010. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development 137, 715–724.

Yang, D., Smith, E., Roland, I., Sheng, Z., He, J., 2002. Disabled-2 is essential for endodermal cell positioning and structure formation during mouse embryogenesis. Developmental Biology 251, 27–44.

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

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., Smith, A., 2008. The ground state of embryonic stem cell self-renewal. Nature 453, 519–523.

Yu, P., Pan, G., Yu, J., Thomson, J., 2011. FGF2 Sustains NANOG and Switches the Outcome of BMP4-Induced Human Embryonic Stem Cell Differentiation. Cell Stem Cell 8, 326–334.

Zhang, P., Behre, G., Pan, J., Iwama, A., Wara-aswapati, N., Radomska, H., Auron, P., Tenen, D., Sun, Z., 1999. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proceedings of the National Academy of Sciences of the United States of America 96, 8705–8710.

Zhang, S., Wernig, M., Duncan, I., Brüstle, O., Thomson, J., 2001. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. Nature Biotechnology 19, 1129–1133.

Zhao, R., Watt, A., Li, J., Luebke-Wheeler, J., Morrisey, E., Duncan, S., 2005. GATA6

is essential for embryonic development of the liver but dispensable for early heart formation. Molecular and Cellular Biology 25, 2622–2631.