Heterogeneity Within the Stem Cell Compartment: Impact on Fate Determination of Human Pluripotent Stem Cells

By:

Jonathon Michael Carr

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

February 2018

The University of Sheffield
Centre for Stem Cell Biology
Department of Biomedical Science

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank members of the Andrews lab and CSCB both past and present for their technical knowledge and willingness for team work, even at the expense of their own work and schedules. Within the group I would like to give special thanks to Dylan Stavish, of whom I worked closely with on this project, for his creative ideas and technical support in the lab. Additionally, Chris Price, Dr. Paul Gokhale and Dr. Joanne Lacey for their knowledge and expertise of flow cytometry, the inCell Analyser and Western Blotting respectively. I would of course also like to thank Prof. Peter Andrews for giving me the opportunity to work on this project and for being a supportive manager throughout my PhD.

Outside of the Andrews lab, I would also like to give thanks to Dr. Konstantinos Anastassiadis who put time aside to support my stay in Dresden and teach me laboratory techniques in his lab. Additionally, I would like to thank members of the Enver lab, especially Dr. Charlotta Böiers, for their collaboration within this project.

Away from Research, I would like to acknowledge my parents for their support and willingness to go beyond the call of duty to help when needed. I would also like to acknowledge my friend Ben Tayler-Barrett for his support and source of motivation in his own unique way. Finally, I would like to personally acknowledge my partner Ella Blanquet for her unending support and clear logical thinking, serving as a source of motivation and a shoulder to lean on.

Abbreviations

BCL- (a mixture of 0.48µM LPA with the addition of 0.1% BSA, 2µM Cholesterol and 10nM β-merceptoethanol in E8 medium)

bFGF- Basic Fibroblast Growth Factor

CHIRON- Chemical Wnt agonist Chir99021

DMEM- Dulbecco's Modified Eagle's Medium

DMSO- Dimethyl Sulfoxide

E8- Essential 8 media

EB- Embryoid Body

EC cell- Embryonal Carcinoma Cell

EMT- Epithelial- Mesenchyme Transition

EpiSC- Epiblast Derived Stem Cell

ES cell- Embryonic Stem Cell

FACS- Fluorescence-Activated Cell Sorting

FBS- Fetal Bovine Serum

hPSC- Human Pluripotent Stem Cell

ICM- Inner Cell Mass

KOSR- Knock-Out Serum Replacement

LPA- Lysophosphatidic Acid

MEF- Mouse Embryonic Fibroblast

mESC- Mouse Embryonic Stem Cell

NMP- Neuromesodermal Progenitor

PBS- Phosphate Buffered Saline

PFA- Paraformaldehyde

Priming Media- (a mixture of 0.48µM LPA with the addition of 0.1% BSA, 2µM Cholesterol, 10nM β-merceptoethanol, 3µM CHIRON and 1µM IWP2)

QPCR- Quantitative Polymerase Chain Reaction

RNA- Ribonucleic Acid

SSEA3- Stage Specific Embryonic Antigen 3

T- Brachyury gene

Abstract

Human pluripotent stem cells (hPSC) have the capacity to differentiate to all functional somatic cell types, which offers enormous potential for regenerative medicine and further understanding into the molecular mechanisms that drive early development. These cells however exhibit heterogeneity when grown in culture, with respect to cell surface antigen expression, gene expression patterns and functional properties such as clonogenicity. The central hypothesis of this project is that the heterogeneity exhibited in culture is the result of hPSC existing in different states within the undifferentiated stem cell compartment. The aim was then to determine whether cells occupying particular states exhibit a bias in their propensity to differentiate along alternative lineages. This project focused on cells that may be biased towards the mesoderm lineage.

One strategy to isolate cells occupying different sub-states was to use cell surface antigens, specifically the well-defined antigens CD9 and SSEA3 and a novel monoclonal antibody-defined antigen, BF4. Subsets of hPSC were isolated based on levels of expression that showed different propensity for differentiation, despite being able to form colonies and self-renew. In single cell transcriptome studies, the SSEA3/CD9 subsets exhibited heterogeneity in respect to genes associated with mesoderm/endoderm. Functionally, the SSEA3 High/CD9+ subset exhibited increased propensity to differentiate towards mesoderm derivatives in embryoid body conditions. In contrast, cells defined as being SSEA3+/CD9- exhibited increased propensity to differentiate towards ectoderm derivatives in these conditions.

An alternative strategy to identify and isolate a mesoderm biased population was to use a combination of SSEA3 and a *BRACHYURY* (T) gene reporter, involving a histone 2B- Venus fluorescent protein knock-in to the *BRACHYURY* locus. In KOSR/MEF self-renewal conditions, the SSEA3+/*BRACHYURY*+ subset demonstrated increased expression of genes associated with mesoderm/endoderm differentiation, as well as exhibiting increased propensity to mesoderm differentiation in Neutral EB conditions. This SSEA3+/*BRACHYURY*+ subset was recreated in a defined system, that while it

could self-renew and clone similarly to the KOSR/MEF subset, it differed at the transcriptome level and through propensity for differentiation.

Taken together, the results obtained in this thesis provide evidence that SSEA3+ cells are heterogeneous containing many potential states with different propensities for differentiation. It also provides evidence for lineage-biased sub-states of hPSC that can be maintained and analysed, providing a tool for further study into early fate determination of hPSC.

Chapter 1 : Main Introduction

1.1 Foreword

Human pluripotent stem cells (hPSC) contain characteristics that distinguish them from their differentiated counterparts. One crucial characteristic is their capability to self-renew indefinitely under appropriate growing conditions, while maintaining the capacity to differentiate to derivatives of the three germ layers. This capacity to differentiate to any somatic cell type in the body is what makes the stem cell such a desirable tool in regenerative medicine and studies in early development. Such applications can range from furthering understanding into early developmental processes, through to translational applications in making specific cell types for clinical applications. However, to achieve these desired applications, it is necessary to efficiently and reproducibly differentiate these cells. While there are many examples in the literature of directed differentiation of pluripotent stem cells, the underlying mechanisms of early fate determination from the pluripotent stem cell compartment is poorly understood. Consequently, one of the biggest technical difficulties in working with these cells is achieving reproducible and efficient results in generating specific cell types. This can become a very problematic and expensive concern when applied to translational applications, such as industrial scale up for generating large quantities of differentiated derivatives. These applications demand protocols that are cost-effective and achieve high purity and yield of a desired cell type. This consequently points to a need to understand the cellular and environmental ques that are driving heterogeneous cell fate decisions.

There is growing evidence in the literature that heterogeneity is present within stem cell culture. This is demonstrated with differences in cell surface antigen expression and gene expression patterns that can directly correlate to differences in cellular behaviour. Although there is evidence in these studies of cells expressing genes associated with lineage specification, there is very little evidence on the functional consequences of these expression patterns and how it relates to stem cell fate decisions.

This work presented is based on exploring heterogeneity within the stem cell compartment through identification of subsets of cells using a combination of various techniques that can allow analysis of the transcriptome, as well as functionally. The objective is to identify subsets of cells with bias to differentiation to mesoderm. The purpose is to understand the molecular mechanisms involved in creating this bias in culture, so it can be controlled and maintained in culture. This will allow greater understanding of the molecular processes involved in fate determination, as well as serving as a tool that could be used in regenerative medicine to improve yield and purity of differentiation to a desired cell type.

1.2 Origin of Pluripotent Stem Cells From *in vivo* to *in vitro*

The human body is a complex network of specialised cell types that serve specific roles depending on their location. The intricate interactions between the different cells in this network result in the development of the human body from a single cell through to a complex structure of tissue and organs. Indeed, this single cell, the zygote, is formed from the fertilisation of a maternal egg with a paternal sperm, resulting in a single cell that contains the genetic information required to form a fully functioning human body.

This process starts once fertilisation occurs, where the zygote undergoes a series of cleavage events until it reaches the 16-cell stage which results in a structure called the morula. After a few days post fertilisation, the totipotent morula forms a cavity from the movement of sodium ions and water from trophoblast cells, resulting in a 32-cell structure called the blastocyst. The blastocyst stage is the first evidence of cell lineage specification, where the outer cells of the blastocyst develop into the first epithelium known as the trophectoderm. During this, the cells that are within the interior of the blastocyst form what is known as the inner cell mass (ICM), a mass of cells that are pluripotent [1]. This capacity for being pluripotent means that these cells can form all functional somatic cell types in the adult. But crucially, they are the source of cells that when isolated and grown *in vitro*, gave rise to the human embryonic stem cell lines that are used the world over.

1.3 Embryology: From the ICM to Formation of the Three Primary Germ Layers

Human pluripotent stem cells have the capacity to differentiate into any derivative of the three primary germ layers. To understand the processes that form these cells, information can be gathered from the embryo in development.

Following formation of the blastocyst and ICM, in week 2 of implantation, the trophoblast layer that was formed with the blastocyst penetrates further into the endometrium. The ICM proceeds to form a bilaminar embryonic disk comprising of the epiblast and embryonic endoderm. The epiblast is the structure that will give rise to the three primary germ layers of the mesoderm, endoderm and ectoderm through the procedure of gastrulation [2] .

Gastrulation marks the creation of a trilaminar embryo through the conversion of the two-layered epiblast into the three primary germ layers through a series of morphological changes, cell proliferation and differentiation [2,3]. This gastrulation process features recruitment of cells from the embryonic ectoderm of the epiblast forming a transient structure called the primitive streak. This primitive streak is a grooved structure that extends from the midline of the bilaminar disk [2]. To form this, cells from the midline ridge of the epiblast undergo an epithelial to mesenchyme transition (EMT) that arrange along this midline creating the left-right and cranial-caudal embryonic axis [2]. Additionally, these epiblast cells proceed to ingress through the primitive streak, firstly forming the mesoderm and definitive endoderm [2]. As the primitive streak reaches its mid-stage, it marks the beginning of formation of the head mesoderm, somites, intermediate, paraxial and lateral plate mesoderm [2]. As well at this mid-stage, it marks the formation of the notochord, a rod-like structure laying on the anteroposterior axis [2] that induces the formation of the neural plate that forms the basis of formation of the nervous system [4].

Clonal analysis has shown that epiblast cells do not exhibit lineage commitment prior to ingression through the primitive streak [5]. Indeed, mesendodermal cells demonstrate a lineage choice between the mesodermal and endoderm. Fate mapping analysis of the primitive streak has shown that cells of the axial mesoderm and anterior definitive endoderm ingress through the most anterior

region, where as the paraxial mesoderm, lateral plate mesoderm and extra-embryonic mesoderm emerge from the anterior third region, the middle region and the posterior region respectively [6]. Effectively, the time and site of ingression influences the allocation of cells to specific tissue types and their destination along the anterior-posterior-axis [7,8].

As described, the process of ingression is essential for cells to undergo the EMT and ingress through the primitive streak to form lineage specific cells. Cell signalling processes (which will be described in more detail within this introduction) play a key role in cell movement through the primitive streak. In respect to FGF signalling, FGFR-1 mutant embryos result in epiblast cells that are unable to undergo the EMT and consequently fail to ingress through the primitive streak [9]. These FGFR-1 mutant embryos have also shown to have reduced expression of TBX6 and T (*BRACHYURY*) which also affects movement of the mesoderm. In these embryos, high E-cadherin activity results in a reduction of nuclear localisation of β-Catenin. This results in a down-regulation of *BRACHYURY* expression, a direct target of WNT/ β-Catenin signalling [9]. In the chick embryo, mutations in FGF8 results in accumulation of mesenchyme in the primitive streak, highlighting its role in supporting ingression through the primitive streak [10].

Movement of both mesoderm and endoderm cells through the primitive streak can be affected by different genes. Indeed, it was shown that mutations in embryos for the gene *MESP1*, resulted in negatively affecting mesoderm ingression, but not endoderm [11]. In respect to endoderm formation, mutant embryos for the mesendoderm gene *MIXL1*, resulted in fewer definitive endoderm cells being recruited to the endodermal layer [12]. Additionally, mouse MIXL1-null embryos were shown to have complex defects in both the axial mesoderm and formation of gut endoderm [13]. In MIXL1-null embryos, BRACHYURY expression was also affected, where its expression localised in a poorly organised structure that resembled the notochordial plate [13]. BRACHYURY's role in formation of the notochord was also shown in BRACHYURY null embryos, where the result was defective primitive streak formation and notochord morphogenesis [14].

Formation of different mesoderm and endoderm cells depends on different levels of Nodal and Wnt signalling within the primitive streak. It has been demonstrated that loss of Nodal signalling results in a failure to form the axial mesoderm, paraxial mesoderm and definitive endoderm [15–17]. Conversely, increased Nodal signalling results in an over expression of mesoderm associated genes [18–20]. Formation of definitive endoderm relies on activation of MIXL1 to down-regulate Nodal signalling [13]. Other signalling pathways such as BMP signalling have shown to have a role in mesoderm formation. BMP -/- mutant embryos were shown to be deficient in posterior derivatives of the primitive streak such as the haematopoietic and extra mesodermal lineages [21]. In respect to the paraxial mesoderm, it has also been shown to be affected by Wnt signalling. It was shown that in WNT3A null mutant mice, neuroectoderm formed in place of somites [22]. It therefore highlights that formation of the middle-primitive streak is a result of co-operation between the Nodal, FGF and Wnt signalling pathways [23], whereas the posterior-primitive streak depends on BMP signalling activity.

Evidence of early lineage bias to the mesoderm and endoderm can be seen both *in vivo* at this stage of development and comparatively in Epiblast Stem Cells (EpiSC) and ES cells *in vitro*. Within embryo analysis *in vivo*, BRACHYURY and OCT4 were observed to be expressed in the pre-streak and early-streak embryos, where in the latter, it marked the posterior epiblast prior to formation of a nascent layer of mesoderm [24]. Single cell analysis of early post-implantation embryos revealed that a subset of cells identified as being from a day 6.5 embryo, co expressed both *OCT4* and the endoderm associated gene *GATA6*. Clusters of these cells expressed genes associated to both the mesoderm and definitive endoderm, suggesting the majority of these cells represented those of the mesendoderm [25]. During development, a key process during formation of the mesoderm and endoderm, is the co-expression of the mesendoderm gene *MIXL1* with the pluripotency associated gene, *OCT4* [26]. It was shown in differentiating ES cells, that there existed a sub population of cells that co-expressed *MIXL1* and *OCT4*, along with other primitive streak associated genes. It was proposed that these cells were the *in vitro* equivalent of primitive streak cells *in vivo* [26]. EpiSC have been shown to be undifferentiated cells, yet they heterogeneously exhibit co-expression of pluripotency associated genes such as *OCT4* with lineage associated genes such as *BRACHYURY* [27,28]. It was shown by (Tsakiridis

et al., 2014) [29] that primitive-streak like EpiSC exhibit a lineage bias to mesoderm and endodermal fates, while maintaining pluripotency. The primitive streak like characteristics of EpiSC were mediated by endogenous Wnt Signalling, while elevated Wnt signalling restricted the EpiSC to primitive streak like cells with mesendodermal and neuromesodermal characteristics. In another study, (Song et al., 2016) [28] showed that *BRACHYURY* Positive EpiSC co-expressed *OCT4* while being maintained as undifferentiated. The *BRACHYURY* Positive EpiSC exhibited a lineage bias to differentiate to the mesendoderm, while *BRACHYURY* Negative EpiSC exhibited a bias to differentiate to the neural ectoderm.

During development, the process of formation of cells of the brain and spinal cord arrive from the procedure of neurulation that is divided into two sections, primary and secondary neurulation [30]. At the primary neurulation stage, the derived ectoderm is divided into two neural plate cells, the epidermis and the neural crest. In the mouse, the neural crest is formed along almost the entire axis at the neural plate border. Here, neuroepithelial cells from the neural tube proceed with their own EMT, resulting in delamination and migration to the periphery of the neural tube. After this migration step, it marks the formation of the four neural crest populations of the cranial, cardiac, vagal and trunk [2]. These four progenitor populations of neural crest will continue to further differentiate into various derivatives, including neural, cartilage and bone for the cranial neural crest [31,32].

In summary, the above describes the basic procedures of the transition of the ICM through to gastrulation and forming of the neural plate. However, as briefly described, the process within the embryo involves a complex network of signalling cascades and gradients to give rise to specific derivatives. These signalling cascades are recapitulated in stem cell fate determination in culture, both for maintaining the pluripotent state as well driving differentiation. Understanding the mechanisms that control these signalling networks is integral to understanding fate determination mechanisms of pluripotent cells. However, a lot of our early knowledge of pluripotency came from other sources, such as the malignant counterpart of the hPSC, the embryonal carcinoma (EC) cell. Furthermore, prior to the derivation of hPSC lines, mouse pluripotent stem cell lines were derived and were used for understanding early developmental processes. A research field which continues to this day. As well as providing knowledge for characterising pluripotency, these different cell types provide tools

for comparison with hPSC and are necessary contextual background for this study.

1.4 Prior Research to hPSC: Embryonal Carcinoma (EC) cells

1.4.1 Mouse EC cells

Prior to the use of human embryos for cell line derivation in 1998 [33], a lot of the knowledge of the mechanisms and hallmarks that define pluripotency and differentiation were learnt from embryonal carcinoma (EC) cells. The origin of these pluripotent cells came from a type of malignant tumour, known as the teratocarcinoma.

The teratocarcinoma and its non-malignant counterpart, the teratoma, have been of research interest for many decades as these tumours have biological comparisons to the early embryo. Within these tumours, cell types can be observed that are derived from the three germ layers in embryo development. What was intriguing with the teratocarcinoma however, was that it also contained a clonogenic, pluripotent cell type that became known as the EC cell. One of the first reported studies of stable growth and maintenance of these cells *in vitro* was that of Finch and Ephrussi [34]. Here it was demonstrated that the use of a sub-cell line derived from a teratocarcinoma of mouse testis origin, could be maintained on a layer of irradiated, senescent mouse embryonic fibroblasts (MEF) with Dulbecco's Modified Eagle's Medium (DMEM) and fetal calf serum. This use of a layer of MEF was proposed to be important for the maintenance of these cells as undifferentiated, presumably through providing the necessary factors for self-renewal. Following this study, many more EC lines were derived that all contained the key characteristics of being able to self-renew and being pluripotent.

It was considered whether EC cells were equivalent to cells of the ICM, as evidenced by the generation of chimeric mice that were formed through injections into mouse blastocysts [35]. Another example of the similarity of these cells to those of the ICM was shown in embryoid body (EB) differentiation, three-dimensional aggregates that resemble the inner cell mass of embryos *in vitro.* A study by (Martin and Evans) [36] showed that EB formed by EC cells resulted in the

generation of a variety of differentiated cell types of different lineages. Their capability to differentiate to specific lineages was demonstrated through the addition of retinoic acid to the F9 EC line, providing evidence that these cells could respond to exogeneous signals that would influence their fate decision [37]. This F9 line would become prevalent in a further study that showed that these cells expressed a cell surface antigen, known as the F9 antigen. Here it was shown that this antigen was only expressed on cleavage stage embryos and not on differentiated derivatives [38]. Following the derivation of monoclonal antibodies, an antibody known as MC-480 or SSEA1, was shown to have a similar expression pattern to that of F9 [39]. This observation led to the hypothesis that there were other cell surface antigens whose expression was lost following differentiation. Nevertheless, this evidence that the mouse EC cells resembled that of the mouse ICM was a major driving force for deriving the human equivalent.

1.4.2 Human EC cells

Following the identification and characterisation of mouse EC cells, growing interest in the human equivalent resulted in derivation of human EC cell lines that were successfully maintained and expanded *in vitro* [40,41]. Intriguingly, characterisation of the human EC cells highlighted phenotypic differences between them and the mouse equivalent. Through the use of monoclonal antibodies, it again was apparent that there were differences in the expression patterns of these cell surface antigens. Human EC cells were shown to express the surface antigens SSEA3, SSEA4, TRA160, GCTM2, THY1 and MHC. Though conversely, mouse EC cells were shown to not express these markers, but instead expressed SSEA1, which was not expressed on human EC cells [42]. While the reason for this difference in expression is unknown, hypotheses have considered whether it is a result of species differences or the cells corresponding to different embryonic cell stages. However, through discovery of transcription factors associated with the maintenance of pluripotency, namely *OCT4*, *SOX2* and *NANOG*, these were shown to be expressed in EC cells of both species [43–46].

Aside from the evidence of these EC cells corresponding to those of the ICM, studies using human EC cells also contributed to understanding mechanisms in early development. Indeed, it was shown using the NT2/D1 line that adding precise concentrations of the morphogen Retinoic Acid (RA), induced HOX gene expression that resembled HOX gene patterning in the embryo. This was shown with the generation of cells corresponding to those of the hind brain or spinal cord at lower or higher concentrations of RA respectively [47]. In addition to studying early development, EC cell lines were also used in early attempts at regenerative medicine. Such an example is the use of neurons that were derived from the NT2/D1 line being transplanted in patients that had suffered a stroke [48].However, one of the limitations of EC cells observed is that some of the lines were variant, with the cells being nullipotent [42], leading to questions as to whether these cells truly behaved like those of the ICM. This concern would have direct implications to studies involving early development, especially involving those in early fate determination or in the mechanisms behind self-renewal and maintenance of pluripotency. Additionally, being of tumour origin, the cells may have inherent malignant properties which render their potential use in transplantation procedures risky. Consequently, future studies began focusing on cells that were shown to be derived directly from the ICM, firstly in the mouse.

1.4.3 Prior Research to hPSC: Mouse Embryonic Stem cells

Following the characterisation of EC cells, attention turned towards isolating cells that were derived directly from the ICM. However, the procedure for isolating the cells and maintaining them *in vitro* was shown to be technically difficult. A study by (Evans and Kaufman,1981) [49] followed a hypothesis that isolating enough cells at the necessary stage of the developing embryo and growing them in the appropriate culture conditions would allow expansion and maintenance of these cells *in vitro*. By isolating day 2.5 mouse blastocysts and culturing them, they gave rise to trophoblast cells and egg-like cylinder structures, that were postulated to have been derived from the isolated cells from the ICM. To verify this, these cells were cultured on a senescent fibroblast layer, in a similar way to the EC cell lines previously and were found to adopt a morphology that resembled

the previously derived EC lines. In another study by (Martin, 1981) [50] these ICM derived cells were maintained in conditioned media that were used in maintaining teratocarcinomas. This study provided the term of “embryonic stem cell” due to the cells direct derivation from the blastocyst. These landmark studies paved the way for a new field of research to study and characterise early developmental mechanisms and processes *in vitro*.

New approaches utilised the advancing technologies in genetic engineering from insertional mutagenesis using viral vectors [51], through to precise gene targeting using homologous recombination [52,53] to characterise these newly derived cells. Further studies involving deciphering of the regulatory network governing pluripotency, identified that similarly to EC cells, the transcription factors OCT4 [43], SOX2 [44] and NANOG [46] were important for the maintenance of the pluripotent state in these embryonic stem (ES) cells. Furthermore, study into the signalling networks that are involved in the maintenance of pluripotency revealed a role for Leukemia Inhibitory Factor (LIF), a member of the interleukin 6 family [54]. Following binding of LIF to its target receptor, a signalling cascade is triggered that results in activation of JAKs (Receptor associated Janus Kinases) and phosphorylation of STAT3 (Signal Transducers and Activators of Transcription 3). This in turn results in STAT3 translocation into the nucleus to activate transcription. One such gene target for STAT3 was GABP (GA-binding protein) which has a direct role in regulating expression of the pluripotency associated transcription factor, OCT4 [55].

The mouse is an experimental model that has provided much insight into early development, both in terms of understanding the mechanisms in embryogenesis as well as early developmental processes *in vitro* using ES cells. Together with EC cells, these cells have provided much of the knowledge that was gained on the characterisation of pluripotency prior to the derivation of the human ES cell equivalent. However, while mouse ES cells have and continue to provide a model for studying early development, the question remained on how relevant species differences were in respect to the mechanisms involved in pluripotency and early development. For understanding the mechanisms of early development in humans, it was hence necessary to attempt to derive and utilise the human equivalent of the mouse ES cell.

1.5 Human Embryonic Stem Cells

Following the success of derivation of cells isolated directly from the mouse ICM, attempts at achieving the same using human embryos were underway. The first successful derivation was achieved in the study by (Thomson et al. 1998) [33] where in a similar vain to previous derivations using mouse and human EC cells, a mouse fibroblast layer was used as the matrix. Though this was not the only similarity, as once derived, these human ICM derived cells were shown to have the same self-renewal and differentiation capabilities of their mouse cell counterparts. Further characterisation of these newly derived human ES lines revealed a similar trend to their EC cell counterparts in that they expressed the cell surface antigens SSEA3, SSEA4 and TRA-160, but again similarly to EC cells, they lacked expression of SSEA1 [33]. This observation provided the first insights into the potential species differences between mouse and human ES cells and inspired future studies into characterising the mechanisms involved in human ES cell pluripotency.

Studies into the signalling networks involved in regulating pluripotency in these cells, revealed that unlike mouse ES cells, human ES cells did not LIF for maintaining pluripotency [33]. Furthermore, studies into the role of BMP signalling, which features prominently in maintaining pluripotency in mouse ES cells, were shown to induce differentiation in human ES cells [56]. These studies provided justification for further study in to understanding the differences in early human development and provided platform to understand the signalling networks involved in maintaining pluripotency and driving differentiation.

1.6 Signalling Pathways Involved in Self-Renewal and Differentiation of Human Embryonic Stem Cells

In order to understand early fate determination process of human pluripotent cells in culture, it's necessary to understand the signalling cascades involved in maintenance of pluripotency as well as drivers of differentiation. By activating or repressing precise signalling pathways, transcription factors can be expressed that drive expression of genes involved in either maintaining the stem cell state or leaving it to a differentiated derivative. Important pathways involved in early stem cell fate determination include the BMP, TGF-β, FGF, WNT and HIPPO pathways, where co-ordination of these pathways can directly influence stem cell fate decisions.

1.6.1 TGF-β Signalling Role in Maintaining Pluripotency

The TGF-β signalling pathway is one that features prominently in both the developing embryo and the adult. It is also involved in a variety of cellular processes ranging from stem cell maintenance and differentiation through to cell growth and apoptosis. TGF-β is a cytokine and a member of the transforming growth factor (TGF) superfamily. The superfamily can be divided into two functional sub-group of ligands. The first set consists of Activin A, Nodal and TGF-β where the signalling cascade resembles that of the JAKs pathway, where a type 2 receptor kinase phosphorylates its target type 1 receptor kinase equivalent, resulting in activation of a signalling cascade [57]. Here, the cytoplasmic signalling molecules SMAD2 and SMAD3 are activated following signalling through the ALK4/5/7 receptors (known as ACVR1B, TGFBR1 and ACVR1C respectively) [58,59] where they become phosphorylated at the C-terminal serine residue [60]. Following this process, this phosphorylated SMAD 2/3 complex forms a trimeric complex with SMAD4, which permits translocation into the nucleus where it can serve as an activator for target genes. The second set involves the BMP receptor family where signalling proceeds through the type 1 receptors ALK1,2,3 and 6 (known as ACVERL1, ACVER1, BMPR1A and BMPR1B respectively) resulting in phosphorylation of SMAD1 and SMAD5 which again

forms a trimeric complex with SMAD4, where it can translocate to the nucleus and serve in activating target genes. [59,61,62].

Specific to human ES cells, the TGF/Activin/Nodal activation of the SMAD2/3 signalling cascade has been shown to have a role in the maintenance of the pluripotent state. Such examples include the pluripotency associated gene NANOG, which has been shown to be regulated by TGF/Activin/Nodal signalling [63]. Conversely, studies involving the inhibition of this signalling pathway using the ALK5 chemical inhibitor SB431542 resulted in down-regulation of NANOG and consequent differentiation [63]. Furthermore, studies into the TGF- β/Activin/ Nodal pathway through activation of the SMAD 2/3 signalling cascade using Activin A as a protein ligand, showed that it served a role in maintaining the pluripotent state [59]. However, while these studies highlight the role of TGF-β super family in maintaining pluripotency, there is also evidence in the literature for this signalling network serving a role in driving differentiation.

1.6.2 TGF-β Signalling Role in Driving Differentiation

As well as having a prominent role in maintaining pluripotency, the TGF-β signalling pathway has also show to feature within mesoderm formation and patterning. One of the earliest recorded examples evidencing this was performed in a Xenopus embryo [64]. In this study a truncated form of the Activin type 2 receptor was cloned and integrated into the embryo. This resulted in defects at the gastrula stage in development which resulted in a loss of mesoderm specification and an increase in ectoderm differentiation. This receptor has also been shown to feature within the BMP signalling network, where mutations resulted in inhibition of the BMP signalling cascade [65]. Activin A had also been shown to have a role in endoderm formation, where in low serum conditions, addition of Activin A at 100ng/ml resulted in up to 80% SOX17+ cells [66].

The SMAD 2/3 signalling cascade which, as previously described, has a role in maintaining the pluripotent state, has also been shown to have a role in differentiation. This has been shown through the anterior/posterior patterning of the epiblast in the mouse embryo through regulating formation of the primitive

streak [67,68]. In addition, it had also been shown that loss of SMAD 2/3 results in loss of mesoderm formation due to a failure for the embryo to proceed into gastrulation [69]. In addition to the SMAD2/3 cascade, TGF-β signalling was shown to have a prominent role in mesendoderm specification. It was shown that expression of the TFG-β antagonists Lefty and Cerberus or inhibition of the ALK receptors using SB431542 resulted in increased neural ectoderm differentiation [70].

These studies highlight the prominent role TGF-β signalling has in both promoting maintenance of pluripotency, as well as promoting mesendoderm differentiation both *in vivo* and *in vitro*. The contrast, highlights the complexity underlying cell fate decisions in early development. This is further highlighted when studying the co-operation of this signalling network with other signalling pathways. Such an example can be seen with the PI3K/Akt pathway, where cell fate choices are altered depending on the level of activity of the signalling network. At higher levels it results in promotion of the Activin A/ SMAD2/3 pro-self-renewal network, whereas at lower levels, it results in promotion of a Wnt-induced SMAD2/3 pro-differentiation network [71]. This complexity of cell fate decisions promoted by these signalling cascades can be seen in another member of the TGF-β signalling family, BMP signalling.

1.6.3 BMP Signalling Role in Driving Differentiation

The BMP signalling cascade, a member of the TGF-β family, has similarities to the TGF-β signalling cascade, but differs through its mechanism. In this pathway, binding of the BMP ligands to both type 1 and type 2 receptors, results in the phosphorylation of SMAD1/5/8 that forms a trimeric complex with SMAD4 before translocation to the nucleus. Such an example of this cascade at work can be seen in formation of trophoblast cells, where supplementation of recombinant BMP4 ligand in culture results in differentiation of human ES cells to this lineage [72]. BMP4 has also been shown to have a role in formation in other cellular derivatives such as the mesendoderm. Here, through activation of the TGF-β/Activin/Nodal pathway, this ligand results in down-regulation of the pluripotency gene associated gene and early ectoderm gene *SOX2*, to drive a endoderm cell

fate [73]. This drive towards the mesendoderm can also be seen through its role in activating gene expression of genes associated with the Wnt signalling pathway, such as WNT3, resulting in early mesoderm/mesendoderm specification [74]. Co-ordination between the BMP pathway and the Wnt pathway is associated with formation of either neural crest or early mesendoderm [75]. Here, a combination of the level of activity of the Wnt signalling cascade with inhibition of the BMP pathway, resulted in differing cell fate decisions to either the neural crest or mesendoderm respectively. This again is another example of the complexity in the co-ordination of these signalling cascades to influence cell fate decisions. This complexity can also be seen in comparison to other species.

As mentioned previously, the role of the BMP signalling cascade in the mouse has a contrasting effect where it promotes maintenance of the pluripotent state [76]. While not the same in human ES cells, it was however shown that in conjunction with FGF signalling, the BMP inhibitor Noggin could block the pro-differentiation effects of BMP signalling [77].However, independently, Noggin was shown to be essential for driving neural induction in the Xenopus [78] [79] as well as driving neural ectoderm differentiation in hPSC [80]. As an alternative to Noggin, chemical inhibitors such Dorsomorphin were discovered that could work in a mechanistically similar way, as well as improving the efficiency of neural differentiation in hPSC [81]. However, by adding inhibitors of the TGF-β/Activin/Nodal pathway such as SB431542 which inhibits ALK5 [70], neural differentiation was further improved. This co-ordination of the two signalling pathways to improve the efficiency of differentiation was further demonstrated by Chambers [82] where using a combination of Noggin and SB431542, the percentage of PAX6+ cells generated was increased relative to the percentage generated from using either inhibitor independently. Mechanistically, this co-ordination could be described by three processes. Firstly, by inhibiting the TGF-β/Activin/Nodal pathway, the NANOG regulated pluripotency network becomes destabilised, allowing generation of ectodermal cell types [63,83]. In the second stage, inhibition of the BMP cascade would prevent prolonged BMP signalling, resulting in trophectoderm differentiation [56]. Finally, the result of inhibition of both the TGF-β/Activin/Nodal pathway and BMP pathway would mean both mesoderm and endodermal cell fates would be unfavoured. This is due to both signalling pathways being shown to feature within early mesendoderm differentiation [66,82].

Further adding to the complexity of these signalling networks is the co-ordination with another signalling pathway, the FGF pathway.

1.6.4 FGF Signalling Role in Maintaining Pluripotency

The FGF family of ligands are growth factors that feature in a variety of cellular functions including angiogenesis and early embryogenesis [84]. Specifically, to human pluripotent cells, FGF2 (otherwise known as bFGF) features prominently as a media component as it serves a role in maintenance of the pluripotent state when in combination with agonists of the TGF-β pro-self-renew signalling network [85]. FGF2 has also been shown to have a role in promoting stem cell survival post-dissociation and in the presence of oxidative stress, though the exact mechanism is not elucidated [86].

Within human pluripotent cells, FGF signalling proceeds through two pathways, the MEK/ERK pathway and the PI3K pathway. The role of ERK signalling independent is subject to debate within the field as contrasting reports reveal it can serve in promoting both self-renewal [87,88] and differentiation [89]. In combination with the PI3K signalling cascade, these two pathways have been proposed to regulate WNT signalling. It was demonstrated that when the PI3K cascade is inhibited and the ERK cascade is activated, the result is a loss of GSK3β and the triggering of the release of β-catenin, resulting in target gene expression [90]. In contrast, when both PI3K and ERK are inhibited, GSK3β is not lost, resulting in a failure of β-catenin to accumulate and drive mesoderm differentiation [71]. The PI3K pathway has been shown to prominently have a role in maintaining the pluripotent state when in combination with the TGF-β/Activin/Nodal pathway. It has been shown that inhibition of the PI3K signalling cascade through the chemical compound LY294002 results in pro-differentiation gene expression [91].

1.6.5 FGF Signalling Role in Driving Differentiation

As described through the co-operation of the PI3K and ERK pathways, FGF signalling can serve a role in both promoting self-renewal or promoting differentiation. In the developing embryo, FGF signalling has been shown to have a role in gastrulation. Using the Xenopus as a model, it has been shown that expression of a mutant version of the FGF receptor results in abnormalities and a failure to successfully complete gastrulation [92]. In mice it was also shown that knockout of the FGFR-1 gene that encodes the Fibroblast Growth Factor Receptor 1, results in abnormal patterning of the mesoderm in the developing mesoderm resulting in a loss of formation of somites [93]. Though in human cells, it has also been shown that FGF features prominently in driving lineage specification in a dose-dependent manner. Depending on the level of FGF signalling, human ES cells would differentiate towards different endoderm fates with higher levels driving the cells towards hepatocytes and lower levels towards pancreatic fates respectively [94].

Further elaborating on the complexity of these signalling pathways in tandem, it has been shown that FGF signalling through the MEK/ERK pathway can guide BMP induced differentiation towards mesendoderm fates through regulation of NANOG expression [95]. This again highlights the co-ordination of these pathways and the complexity on their influence in cell fate decisions. Another pathway which has been described in passing in this section which features prominently in pro-differentiation of hPSC is the Wnt pathway.

1.6.6 Wnt Signalling Role in Driving Differentiation

Wnt signalling features in a variety of cellular functions and consists of three distinct cellular pathways, the canonical β-catenin pathway and the non-canonical, cell polarity and Wnt/calcium pathways. Within hPSC, the canonical β-catenin pathway is the one that has been most intensely studied and has been shown to feature prominently in differentiation of hPSC. In this pathway, the mechanism of activation involves regulation of β-catenin. When Wnt signalling is

not active, β-catenin becomes phosphorylated by the protein kinase GSK3β, resulting in β-catenin being targeted for ubiquitination and degradation by the proteasome. When Wnt signalling is active through the binding of Wnt ligands to the Frizzled and LRP5/6 receptors, it results in GSK3β being dissociated from its complex with Axin, resulting in an inability to phosphorylate β-catenin. Consequently, the stabilised β-catenin can accumulate in the cytosol and translocate into the nucleus where it can bind to its transcription transactivation partner, TCF, and activate target gene expression. As well as interacting with β-catenin, TCF has also been shown to interact with SMAD4, suggesting a potential role in TGF-β signalling [96].

Though while mechanistically, the canonical-Wnt is well defined, species specific differences can be seen once this pathway is active. Culturing mouse ES cells in serum free conditions has shown to be stable with the addition of two inhibitors, a MEK1/2 inhibitor and a GSK3β inhibitor. Within human ES cells, the inhibition of GSK3β results in the activation of the canonical Wnt pathway resulting in differentiation [97,98]. This observation is supported by evidence that active Wnt signalling can lead to increased gene expression of early mesendoderm associated genes in hPSC [99]. Conversely, the use of the endogenous Wnt inhibitor IWP2, supports long term maintenance of hPSC in culture, as well as causing more uniform morphology [74]. However, contrastingly, there is some evidence to support the use of GSK3-β inhibitors to support self-renewal of hPSC, [100]. Though in this example, the explanation could be attributed to the culture conditions used, as one method to derive human naïve ES cells is to inhibit GSK3-β and activate Wnt signalling [101–103]. Together these examples elaborate the complexity of the regulation of these pathways and their influence on cell behaviour.

1.6.7 HIPPO Signalling Role in Maintaining Pluripotency

The HIPPO signalling pathway is involved in the regulation of organ size during development and has shown to be highly conserved across the species [104–106]. Considered a kinase cascade, the pathway proceeds from the activation of HIPPO/MST1/MST2 resulting in phosphorylation of YAP and TAZ where it can translocate to the nucleus and activate target gene expression [107]. One of the key functions of YAP/TAZ is as transcriptional activators as evidenced for its role in activating gene expression for the TEAD family of transcription factors [105,106] which feature within mammalian development including regulation of proliferation and apoptosis [108].

In contrast to the other described signalling pathways, this pathway does not appear to require the use of extracellular ligands and receptors to initiate activation. Rather, it utilises a network of upstream components that are involved in the regulation of cell adhesion and cell polarity [106,109,110]. However, there is evidence suggesting that there is co-ordination with other established signalling pathways. Therefore, the HIPPO pathway is better described as a complex signalling network [106]. Within human ES cells, the HIPPO pathway has shown to interact with the TGF-β pathway. This has been observed with YAP/TAZ with the SMAD2/3 complex which features in TGF-β signalling. Impairment of TAZ translocation has shown to result in sequestration of TAZ and SMAD2/3 in the cytosol [111]. Additionally, nuclear YAP/TAZ has shown to form a complex with SMAD2/3 in binding to TEAD transcription factors [112], as well the pluripotency associated transcription factor, OCT4 [113]. In respect to BMP signalling, YAP has also been shown to interact with SMAD1 following activation of BMP [114].

YAP/TAZ have also been shown to feature in regulation of Wnt signalling. As described, Wnt signalling features prominently in driving differentiation of human ES cells. It has been shown that YAP/TAZ activation results in sequestration of β-catenin in the cytoplasm [115]. YAP/TAZ has also been shown to interact with the dishevelled/DVL, resulting in inhibition of target Wnt-related gene expression [116,117]. The tyrosine phosphatase SHP2 features within β-catenin mediated gene expression dephosphorylates CDC73 and in turn binds to β-catenin to form a CDC73/β-catenin complex to drive Wnt target gene expression [118,119]. SHP2 has

shown to move between the cytoplasm and nucleus, however it is excluded from nuclear localisation in the presence of phosphorylated YAP/TAZ in the nucleus. This suggests that there is a dynamic where cytoplasmic YAP/TAZ inhibits Wnt signalling, whereas nuclear YAP/TAZ promotes it. To further add to this complexity, it has also been observed that YAP is a β-catenin target gene [120], meaning that an increase in Wnt-signalling can result in an increase in YAP expression and hence increased transcriptional activity that would feedback into regulating Wnt- signalling. In addition to other signalling pathways described, this is a clear example on how the co-operation of signalling pathways can directly influence the fate choices on human ES cells.

1.6.8 Using the Knowledge of the Signalling Pathways to Guide Cell Fate

Through characterising these numerous signalling pathways, our understanding into the processes that drive cell-fate decisions become apparent. Differentiation protocols consist of the addition of compounds and ligands that can activate or inhibit these pathways at specific times during the protocol to guide cell fate choices in a specific direction. However, when assessing differentiation efficiency in respect to both purity and yield, it becomes apparent that the decisions that these cells make in culture is not uniform across the population. There is increasingly growing evidence to suggest that hPSC do not reside in culture as a homogeneous population, but in fact exist as a heterogeneous population that respond differently to the extrinsic cues provided. How the combination of cell signalling, gene expression and metabolic processes all coincide to influence this heterogeneity is a sub-field of research that is currently being explored.

1.7 Heterogeneity and States Within the Pluripotent Stem Cell Compartment

When considering pluripotent stem cell fate choices, their fate determination decision can be separated into four choices. These are self-renewal, differentiation, apoptosis and quiescence [121]. However, these fate choices can be further complicated when lineage specification is applied, as a differentiating stem cell can become a derivative of any of the three primary germ layers. What does connect these different cell fates is that they are regarded to be largely governed by specific gene expression patterns. This is evidenced during development when a stem cell existing within a pluripotent state, transitions into a differentiated cell that is now residing within a restricted state determined by its lineage [121].

One of the earliest depictions of cell states was the famous image of the "ball rolling down the landscape" [122]. In this image, the ball (which represents the cell) is rolling down a sloped hill containing several branching routes with these routes representing final stable states. This model was expanded on by (Andrews, 2002) [123] where rather than the ball rolling down the hill at a smooth trajectory, the ball divots. This represents the cells entering separate meta-stable states down the landscape, that in terms of biology, represent the cells entering different intermediate cell types, such as progenitor cells (Figure 1). Within the landscape, these intermediate states are impressions along the landscape, which in a real-world context, can be modelled mathematically as attractors. These attractors represent dynamic states in an evolving system over time. Within a biological context, an attractor would represent the most stable state that the cell could reside in at that moment, as shown by their stable gene expression network. Therefore, within development, these different cell states with their own distinct stable gene regulatory network, represent points of equilibrium within a progressive dynamic system that continues to evolve [121].

Within the context of the stem cell compartment however, there is evidence to suggest that cells can exist within separate states, or sub-states that extend along the spectrum of cell fate choices. In culture, the hypothesis is these sub-states are interconvertible meaning a cell could transition between one state or another

and that they represent the probability of a cell making a fate choice, such as lineage specific differentiation and this probability alters, depending on the sub-state the cell is residing in [121]. Within the attractor model, a cell could conceptually exist within a state that is at the bottom of the attractor and would therefore represent the most stable state at that given time (Figure 1). Conversely, the cell could exist within a state that is at the top of the attractor and would therefore represent an unstable state within the stem cell compartment. In a biological context, this would be reflected by the gene expression profile that the cell was exhibiting. Within the attractor model, the cell could transition within the states as depicted by the ball circulating around the attractor. The cause of this could be attributed to the stochastic network of signals within the system resulting in heterogeneity, as shown by the gene expression levels in that network [124]. This concept has been shown to be presented in the adult stem cell system, in both the erythroid and myeloid lineages prior to terminal differentiation to either the erythroid or granulocytic lineages respectively [125].

Within the pluripotent stem cell field, one of the most documented examples of cells existing with separate states within the stem cell compartment is the naïve and primed states. Within the mouse biological system, the naïve cells are thought to be akin to the cells of the ICM with distinct functional characteristics such as increased cloning efficiency compared to the primed cells, that were thought to be akin to the late epiblast [97,126].

Within the human biological system, naïve and primed states have also been proposed [101–103,127]. Although the naïve and primed states are different to the proposed sub-state hypothesis, as interconversion is not seamless and requires a reprogramming process to convert the primed cells to naïve cells [128]. There is however growing evidence in human ES cells, that within the so called “primed” state, there are heterogeneous sub-states containing cells with differential cell surface antigen and gene expression patterns. This has been evidenced in studies in which subsets of hPSC with different functional properties can be separated based on Positive or Negative antigen expression using SSEA3 and TRA160 [124,129], or by levels of antigen expression in the case of CD9, EPCAM and GCTM2 [130,131]. However, what was similar with these studies was the expression of pluripotency associated genes in the sorted populations, suggesting that despite the differential antigen profiles, the cells sorted contained

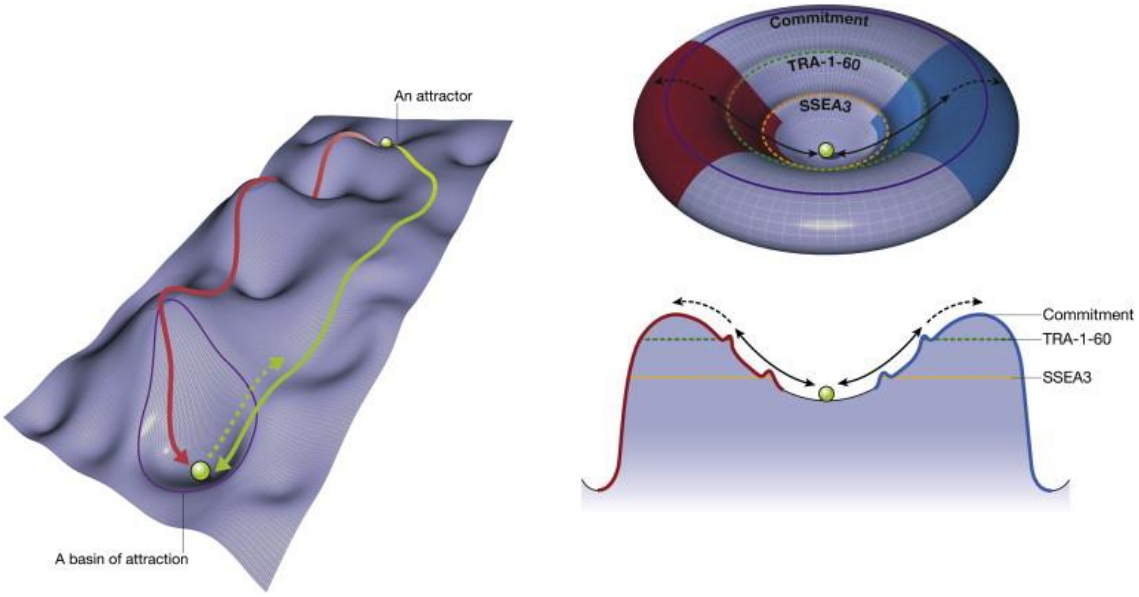
undifferentiated cells. Furthermore, there is evidence that within these antigen sorted subsets, there is differential expression of lineage specific genes, suggesting a readiness to differentiate in that direction [130–132]. Within the attractor model, the hypothesis would be that the cells expressing the highest levels of pluripotency associated genes and the lowest level of lineage specific genes would be at the bottom of the attractor. Conversely, those cells with the lower levels of pluripotency associated genes and higher levels of lineage specific genes would be at the top of the attractor near the commitment barrier. This was demonstrated by (Tonge et al. 2011) [129] who found the SSEA3 “High” cells were more clonogenic and had the least amount of spontaneous neural differentiation compared to the SSEA3 Low and SSEA3 Negative (-ve) cells. Heterogeneous pluripotentstates have also been shown *in vivo*.

Within the mouse system, it was shown that while NANOG has an important role in maintaining the pluripotent state of ES cells [46]. It was shown to be heterogeneously expressed within the mouse embryo in day 3.5 ICM cells [133]. It was also shown that NANOG Negative expressing cells could interconvert to NANOG Positive cells, although these were more prone to differentiation [134]. Although not conclusive, it could otherwise suggest that these interconverting NANOG Negative cells were still pluripotent [121].

Together this evidence suggests not only the existence of heterogeneity, but that heterogeneity has a role in fate determination of ES cells in culture, resulting in non-homogeneous differentiation and differing cell numbers post-differentiation. The central hypothesis of this project is that by identifying and characterising this heterogeneity, it would be possible to understand the mechanisms involved in early fate determination of hPSC. This knowledge could be applied to a multiple of applications from understanding early development processes, as well as serving as a useful tool for differentiation applications in culture.

Figure 1: Heterogeneity and states within the stem cell compartment

The central hypothesis of this study is that pluripotent cells exist within separate states within the stem cell compartment. This is depicted as an energy landscape (bottom left) which contains different stable states (attractor) that a cell could exist within. Within an attractor (bottom right) the cell (ball) can exist within any position within the attractor, but which relates to the cell either self-renewing at the bottom of the attractor or committing to differentiation out of the attractor. These probabilities are linked to specific marker expression, such as the antigen expression of SSEA3 and TRA-160. In this example, loss of SSEA3 and expression of TRA-160 is a state closer to differentiation.



Stem Cell States, Fates, and the Rules of Attraction [121]

1.8 Aims

The central aim of the research described in this report was to assess whether subsets of cells can be identified, using antigens or gene reporters, that exhibit a lineage bias to mesoderm. Once identified, the next aim was to create a defined system using cross antagonism of signalling pathways that allows recreation and maintenance of this lineage biased state of cells in culture. By achieving this, a platform will be created that will allow the study of the mechanisms that underlie the formation of the biased state as well as a tool for differentiation to mesodermal derivatives. This will aid our understanding into early developmental processes involved in lineage fate determination of hPSC.

Chapter 2 : Methods and Materials

2.1 Cell culture

2.1.1 Human pluripotent stem cell culture

Human pluripotent stem cells (hPSC) were cultured in both feeder and feeder-free conditions, dependent on specific experimental requirements as described below. In both culture systems, hPSC were incubated at 37^{o}C in a 5% CO_2 humidified incubator.

Table 1: Cell lines

Name	Type of cell line	Origin/ Construction	Notes
2102ep	Embryonic Carcinoma	Human Testicular Teratocarcinoma	[40]
H9	hPSC	Fresh embryo, Isolated ICM plated	[135]
HES3	hPSC	Frozen embryo	[136]
NCRM1	iPSC	CD34+ Cord Blood Reprogrammed with Episomal Plasmid	NIH CRM Lonza Contract
H9TVD3	H9 *BRACHYURY* reporter	Roger A. Pedersen University of Cambridge	Venus-H2B fusion to *BRACHYURY* locus [137]
SHEF4-GATA6-GFP	SHEF4 GATA6 GFP reporter	Andrew Smith University of Edinburgh	Zinc Finger nuclease targeted locus of GATA6
NCRM1-MSGN1-Venus	NCRM1 MSGN1 Venus reporter	Konstantinos Anastassiadis Biotechnology Center of the TU Dresden (BIOTEC)	BAC Recombineered Venus-targeted MSGN1 locus

Media was made under sterile conditions and filtered using a 500ml Stericup (Millipore) before being stored at either 4^{o}C and used within 1 week or frozen as aliquots at -20^{o}C and used within 2 weeks.

Table 2: Composition of Knock-Out Serum (KOSR) media for MEF culture

Component	**Final volume (ml)**	**Supplier**	**Cat #**
Knockout DMEM	400	Life Technologies	10829018
Knockout Serum Replacement	100	Life Technologies	10828010
1% NEAA	5	Life Technologies	11140035
1mM L-Glutamine	5	Life Technologies	25030081
0.1mM β-mercaptoethanol	1	Life Technologies	31350010
4ng/ml Human bFGF	0.5	RnD Systems	233FB01M

Media was made in-house using the following reagents based on the protocol from [138]. Media was made under sterile conditions and filtered using a 500ml Stericup (Millipore) before being stored at either 4°C and used within 1 week or frozen as aliquots at -20°C and used within 2 weeks.

Table 3: Composition of Essential 8 (E8) media for MEF-free culture

Component	Concentration per Litre	Supplier	Cat #
DMEM/F12		Sigma	D6421
L-ascorbic acid	64mg/L	Sigma	A8960
Sodium selenium	14ug/L	Sigma	S5261
Insulin	19.4mg/L	Sigma	91077C
NAHCO3	543mg/L	Sigma	S5761
Transferrin	10.7mg/L	Sigma	T0665
Glutamax	10ml/L	Thermofisher	35050038
FGF2	100µg/L	Peprotech	10018B
TGFB1	2µg/L	Peprotech	100-21

2.1.2 Mouse embryonic fibroblast (MEF) preparation

Human pluripotent cells were grown on mitomycin inactivated mouse fibroblasts (feeders) from laboratory stocks prepared from MF1 outbred mice. Fibroblasts were expanded through approximately 4 population doublings in DMEM (Sigma cat # D6046) and 10% FBS (Hyclone, GE Life Science Cat #SV30160.03). The cells were then inactivated with Mitomycin C (Sigma, cat # M-4287). The mitomycin C was diluted in DMEM/ 10% FBS at a concentration of 1µg/ml and incubated at 37°C for 2 hrs. After incubation, the cells were washed with 1x PBS (w/o Ca^{2+}, Mg^{2+}) (diluted from 10x Sigma, cat # D1408)

three times and incubated with 0.25% Trypsin/ 0.2% EDTA for 2 minutes at room temperature to achieve a single cell suspension, followed by inactivation of the enzyme with DMEM/10% FBS at a ratio of 1:10 media. The cell suspension was then transferred to a falcon tube and centrifuged at 200 RCF for 4 minutes to pellet and the supernatant aspirated. The pellet was then mixed with a freezing solution of 10% DMSO (Sigma cat # D2650) and 90% FBS, prior to approximately $3x10^6$ cells being aliquoted as a 0.5m solution into cryovials and frozen at -80ºC for 24 hours. The following day, the frozen cryovials were stored in liquid nitrogen for long term storage.

2.1.3 Vitronectin preparation

Human pluripotent stem cells were cultured on Vitronectin (Stem cell technologies) for feeder-free condition experiments. The vitronectin was thawed at room temperature and diluted 1:50 with 1xPBS (w/o Ca^{2+}, Mg^{2+}). The solution was coated to tissue-treated plastic-ware and incubated at room-temp for 1hr prior to plating.

2.1.4 Growing hPSC in KOSR/MEF culture

To prepare culture-ware for MEF plating, T-25 flasks were coated with 0.1% Gelatin (Sigma cat #G2500) and left to incubate at room temperature for 1hr. Following incubation, the gelatin was aspirated and MEF were thawed and seeded at a density of 14000 cells/cm^2 in DMEM/10% FBS and left to attachovernight prior to use at 37ºC and 10% CO_2. On day of passage, the DMEM/FBS was aspirated and replaced with KOSR media to equilibrate. To passage, KOSR media was aspirated and the cells were incubated with 1ml/12.5cm^2 of 1mg/ml collagenase IV (Sigma) for 7 min at 37ºC. After incubation, the collagenase was aspirated and replaced with 3ml KOSR media. The cells were then scraped with a Pasteur pipette and split typically at a ratio of 1:3. Based on the protocol by [139].

2.1.5 Growing hPSC in MEF-free culture

For MEF-free cultures, hPSC were cultured in E8 media and grown on Vitronectin. Cells were passaged by firstly washing with 1x PBS (w/o Ca^{2+}, Mg^{2+}) and incubated briefly with 0.25ml/cm^2 ReLeSr (Stem Cell Technology cat # 05873). After approximately 10 seconds, the ReLeSr was aspirated and the cells were left to incubate at room temperature for 6 minutes. Following incubation, E8 was added to the cells to create a cell suspension consisting of small clumps. The clumps were then transferred to freshly prepared, vitronectin coated plastic-ware following aspiration of the vitronectin solution and replaced with E8 media. The hPSC colony clumps were typically passaged at a ratio of 1:3.

2.1.6 hPSC freezing

hPSC grown on KOSR/MEF conditions were frozen and banked for long-term storage. The cells were separated into clumps following the same protocol as stated above. Once separated into clumps, the cells were centrifuged at 200 RCF for 4 min to pellet. Once pelleted, the supernatant was aspirated and the cells were resuspended gently as clumps in freezing media. The cells were aliquoted into cryovials and placed in Isopropanol freezing containers and placed into -80^{o}C overnight. The following day, cells were transferred to liquid nitrogen for long term storage.

2.1.7 hPSC thawing

The hPSC lines were thawed into KOSR/MEF growing conditions. Once the pellet was thawed in KOSR media, the cell solution was pelleted at 200 RCF for 4 min. Once pelleted, the supernatant was aspirated and replaced with fresh KOSR media. The cell suspension was then aliquoted into a MEF culture vessel (typically a single well of a 6 well plate) with the addition of 10µM of the Rho Kinase inhibitor Y-27632 (Tocris) for improved viability [140].

2.1.8 Single Cell dissociation

To create single cell suspensions when required, hPSC were incubated with 0.25% Trypsin/ 0.2% EDTA for 1 minute at room temperature prior to enzyme inactivation with DMEM/ 10% FBS at a ratio of 1:10 of media. The single cell suspension was then transferred to a falcon tube and centrifuged at 200 RCF for 4 min to pellet. Once pelleted, the cells were resuspended in the appropriate media for the required downstream application and used accordingly.

2.2 Antibody Assays

2.2.1 Flow Cytometry analysis

For flow cytometry, hPSC were dissociated into a single cell suspension as stated above. The cells were then counted with a haemocytometer and centrifuged at 200 RCF for 4 min. Once pelleted, the supernatant was aspirated and resuspended in FACS buffer (DMEM/ 10% FBS) at a density of $1x10^7$/ml. For analysis, 100µl of the sample were transferred to a 5ml FACS tube (BD Falcon) and incubated with the required primary antibody (table 4) on ice for 20 min with gentle rocking. Once incubated, the cell solution was washed twice with FACS buffer and centrifuged at 200 RCF for 4 min to pellet. Once pelleted, the cells were resuspended in 200 µl of FACS buffer and incubated with the required secondary antibody (table 4) on ice for 20 min with gentle rocking. After incubation the cell solution was then washed twice with FACS buffer and centrifuged at 200 RCF for 4 min to pellet. Once pelleted, the cell solution was resuspended in 500µl FACS buffer and analysed via flow cytometry on a BD FACS JAZZ.

Prior to every analysis, the lasers on the flow cytometer were calibrated and aligned using Rainbow Calibration Particles (Biolegend). For setting the fluorescence baseline for each required channel, either hPSC incubated with the negative control antibody P3X63AG8 or the unlabelled parental line for the respective reporter cell line was used. As a Positive control for the secondary

antibody, the pan human antibody TRA-185 was used. The flow cytometry analysis software FlowJo was used for all plots and statistical analysis.

Antibodies derived from hybridomas were concentrated in-house to a 10x concentrate and stored -20^{o}C. These antibodies were dilution titre on the 2102ep embryonic carcinoma line and the H9 or HES3 ES cell line to determine optimal concentration. Commercial antibodies were used at a range of concentrations within the suggested range from the manufacturer to determine optimal titre.

Table 4: Antibodies used in Flow Cytometry and in situ immunofluorescence

Antibody/Antigen	**Species**	**Company/Ref**	**Cat #**
P3X 63/Ag8	Mouse	[141]	Hybridoma
TRA-185 (Oka)	Mouse	[142]	Hybridoma
AG10	Mouse	[143]	Hybridoma
BF4	Mouse	[143]	Hybridoma
CH8 (CD9)	Mouse	[143]	Hybridoma
MC 631 (SSEA3)	Rat	[144]	Hybridoma
MC 813-70 (SSEA4)	Mouse	[145]	Hybridoma
TRA-160R	Mouse	[146]	Hybridoma
TRA-160S	Mouse	[146]	Hybridoma
TRA-181	Mouse	[146]	Hybridoma
OCT4A	Rabbit Mouse	Cell Signalling Santa Cruz	2840 C-10 sc-5279
NANOG	Rabbit	Cell Signalling	D73G4
SOX2	Goat/ Mouse	RnD Systems	AF2018 MAB2018

SOX17	Mouse	RnD Systems	MAB1924
FOXA2	Goat	Santa Cruz	C-20 sc-6553
TBX6	Goat	RnD Systems	AF4744
CDX2	Rabbit	Abcam	Ab76541
PAX6	Rabbit	Abcam	Ab195045
HOXC9	Mouse	Abcam	Ab50839
AlexaFluor Goat anti-mouse 647 IgG+ IgM (H+L)	Goat	Jackson Immuno Research	115-605-044
DyLight 405-AffiniPure Goat Anti Mouse IgG, Fcγ Fragment Specific	Goat	Jackson Immuno Research	115-475-008
Alexa Fluor 488-AffiniPure Goat Anti-Mouse IgG, Fcγ Fragment Specific	Goat	Jackson Immuno Research	115-545-008
Alexa Fluor 647-AffiniPure Goat Anti-Mouse IgM, μ Chain Specific	Goat	Jackson Immuno Research	115-605-020
AlexaFluor Donkey anti-rabbit IgG (H+L) 647	Donkey	ThermoFisher	A-31573
AlexaFluor Donkey anti-Goat IgG (H+L) 594	Donkey	ThermoFisher	A-11058
Goat anti-Rabbit IgG (H+L) Rhodamine Red	Goat	ThermoFisher	R-6394

2.2.2 Flow Cytometry Cell sorting

To prepare cells for sorting, the cells were immunostained in the same way as described above. Following laser alignment on the sorter, the drop delay was determined using Accudrops (BD Biosciences cat#345249). For each sort, the cells were sorted into 3ml of the appropriate media for the downstream application. For live/dead discrimination, cells for sorting were incubated with 1µg/ml DAPI (Thermofisher cat#D1306). Following sorting, the cells were assessed for purity via reanalysis of the sorted population, with acceptable purity determined at greater than 90%. For downstream applications that required plating, the cells were counted as a 1:2 mix with Trypan Blue (Sigma cat# T8154) for live/dead discrimination, with only the none blue cells counted as alive. For each plating experiment, the seeding cell media had 10µM Y-27632 (Tocris cat#1254) to aid plating and survival, and 50µg/ml Gentamycin (Sigma cat#G1397) to prevent contamination from sorting.

2.2.3 Neuraminidase Test

hPSC were dissociated to a single cell suspension as previously described. Cells were then either treated with 0.1 units/ml Neuraminidase (Sigma cat# 11080725001) in 1x PBS (w/o Ca^{2+}, Mg^{2+}) for 30 min at 37^{o}C or kept on ice as a control. After incubation, the neuraminidase was inactivated with DMEM + 10% FBS and centrifuged at 200 RCF for 4 min. The media was then aspirated and both treated and control samples were immunostained and analysed via FACS as previously described.

2.2.4 Co-Immunoprecipitation Assay

HES3 hPSC grown at 90% confluency in T-25cm^2 were washed with 1x PBS and lysed with 200µl of lysis buffer (Pierce RIPA buffer (Thermo Scientific cat# 89900) and 1x HALT protease inhibitor cocktail (Thermo Scientific cat #78430). Following lysis, the cell solution was sonicated on ice through 3x 5 second bursts at 10 micron amplitude. The cell solution was then centrifuged at 1400rpm for 5 min at 4^{o}C to pellet. Once pelleted, 40µl of supernatant was transferred to a separate Eppendorf tube to be used as a total lysate sample. To extract the antigen of interest, protein G coated Dynabeads (ThermoFisher cat #10003D) were chosen. The beads were rotated within the vial to resuspend, followed by 50µl of the bead solution being transferred to a new Eppendorf tube and placed on a DynaMag-2 Magnet (ThermoFisher cat#12321D) to separate the beads from the supernatant. The supernatant was then removed and replaced with 10µg of Goat anti-mouse IgG+ IgM (H+L) secondary antibody (Jackson ImmunoResearch, cat #115-605-044) in 200µL of binding buffer (ThermoFisher, cat #10007D) and incubated with rotation at 4^{o}C for 30min. After incubation, the tube was placed back on the magnet and supernatant removed. The beads were then washed 3x with 200µL washing buffer (ThermoFisher, cat #10007D).

After washing the beads, approximately 10µg of primary antibody was diluted in 200µL of binding buffer and added to the beads, followed by incubation with rotation at 4^{o}C for 1 hour. The beads were then washed as described above. After washing, 50µL of lysate supernatant was added to the beads and incubated with

rotation at 4ºC for 1 hour. After incubation, the tube was placed back on the magnet and the supernatant transferred to a separate tube to be used as an unbound lysate control. The beads were then washed 3x with washing buffer and transferred to a tube for antigen elution. The tube was placed back on the magnet and the supernatant removed. 20µL of elution buffer with the addition of 10µL 4x SDS-PAGE loading buffer (240 mM Tris/HCl pH 6.8, 8% SDS, 0.04% bromophenol blue, 5% beta-mercaptoethanol) was then added to the beads. 20µL of elution buffer/SDS PAGE loading buffer was also added to the total lysate and unbound lysate samples and all samples were heated at 95ºC for 10 min to denature the protein. The tube was then put back on the magnet and the supernatant removed, to be run on a SDS PAGE gel for Western blot.

2.2.5 Western Blot

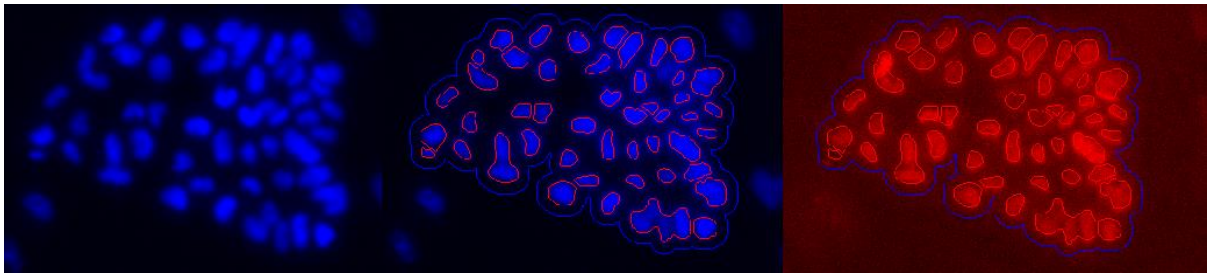
Following elution of the antigen from the Dynabeads as stated above, the elution was run alongside a protein ladder (ThermoFisher, cat# 26619) on a set 4% SDS-PAGE gel at 150v in 1x SDS PAGE Running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS in 2L ddH_2O) for 2 hours. Afterwards, the gel was transferred to a cassette containing two sponges, four pieces of cut to size Whatman paper (Sigma) and a methanol equilibrated PVDF membrane (ThermoFisher). Once complete, the cassette was placed in Electrophoretic transfer apparatus (Biorad) and run at 200mA for 2 hours on ice. Following the transfer, the membrane was briefly washed with Ponceau S solution (Sigma) and washed with 1x TBS (50 mM Tris-Cl, pH 7.6, 150 mM NaCl in 1L ddH_2O). After staining, the membrane was washed with TBS+ 0.1% Tween (Sigma). After washing the membrane was blocked in 5% milk (milk powder, Tesco) diluted in TBS+ 0.1% Tween) for 1 hour, room temperature with gentle agitation. Following blocking, the membrane was the washed 3x with TBS+0.1% Tween. The TRA160-R antibody was then diluted in 5ml of 5% milk/TBS/Tween solution and incubated overnight at 4ºC with gentle agitation. The following day, the membrane was washed 3x with TBS+ 0.1% Tween. The secondary antibody (anti-mouse HRP conjugate) was diluted to a 1:2000 dilution in 5% milk/TBS/Tween solution and incubated at room temp for 1 hour with gentle agitation. Following incubation, the membrane was then washed

3x with TBS+ 0.1% Tween. Finally, the membrane was incubated with Amersham ECL Western Blotting Detection reagent (RPN2106) for 1 min and imaged using a CCD camera.

2.2.6 Immunostaining

All intracellular immunostaining was performing using the same fixation and permeabilization protocol as previously described. The plates were then blocked with species specific serum to which the secondary antibody was raised (Biosera) in 1x PBS (w/o Ca^{2+}, Mg^{2+}) at 4°C for 1 hr. The primary antibody was resuspended in 1X PBS+ 10% serum (species specific) at the appropriate concentration (table 4) and incubated for 1 hour at 4°C. After incubation, the antibody was gently pipetted away, and the plates were washed twice with 1x PBS (w/o Ca^{2+}, Mg^{2+}). The appropriate secondary antibody was then applied with the addition of 10µg/ml of Hoechst 33342 and incubated for 1hr at 4°C. The plates were then washed as the same manner as the primary antibody. The plates were then stored in 1x PBS (w/o Ca^{2+}, Mg^{2+}) at 4°C for long term storage. All immunostained plates were scanned and analysed via the inCell analyser 2000 and inCell Investigator Developer Toolbox software. All immunostaining plates were quantified by setting a fluorescence threshold using a secondary only control or unlabelled hPSC line where appropriate. Positive cells were determined as nuclei (Hoechst 33342 immunostaining) having higher fluorescence values than the set baseline. An example of nuclei masking for immunofluorescence analysis can be seen in the image below.

Hoechst immunostain Hoechst immunostain with mask CY5 channel with mask



2.3 Transcriptome Analysis

2.3.1 RNA Extraction

All RNA extractions were performed on TRIzol (ThermoFisher, cat# 155960260) through lysis of live cells. All samples (either sorted or monolayer) were pelleted at 200 RCF for 4 min and dissolved in 600µl TRIzol. For RNA extraction, the samples were left to thaw at room temperature and quickly incubated with chloroform (Sigma) at a ratio of 200µl/ml of TRIzol for three minutes at vortexing. The samples were then centrifuged at 10000rpm for 10 min at 2-8^{O}C. Following centrifugation, the top layer of the chloroform-phenol mix was transferred to a new Eppendorf tube and mixed with equal volume 70% ethanol via vortexing. The solutions were then added to spin columns from the "Total RNA extraction kit "(Norgen cat# P4-0015) and centrifuged at 14000rpm at room temperature for 1 min. The samples were then washed three times with Wash Solution A provided in the kit. Following the third wash, the spin columns were spun again at 1400rpm for two minutes to thoroughly dry the resin in the columns. Following this, 35µl of elution buffer provided by the kit was added to each spin column and were then centrifuged at 2000rpm for 1min, followed by 14000rpm for 1 min. Each RNA solution was then quantified for concentration using a NanoPhotometer (Implen).

2.3.2 Reverse Transcription

All reverse transcription reactions were performed using the High Capacity cDNA Reverse Transcription Kit (ThermoFisher cat# 4368814). For reverse transcription to create cDNA to be using for QPCR analysis, 10µl of RNA solution was added to 10µl of mastermix (2µl 10x RT buffer, 0.8µl 25x dNTP mix, 2µl 10x RT Random Primers, 1µl Reverse transcriptase. 4.2µl dH_2O). The thermal cycler was set to 25^{o}C for 10min followed by 37^{o}C for 2 hours and finally 85^{o}C for 10 min before cooling to 4^{o}C.

2.3.3 QPCR

All population level QPCR assays were performed using 5ng/µl dilutions of cDNA, diluted with dH_2O. Gene analysis was performed with either individual primers designed using the Roche Universal Probe library (table 5) or using Human Stem Cell pluripotency panel Taqman arrays (ThermoFisher cat# 4385344). For individual gene analysis, all preparations consisted of a mix containing 5µl Taqman Fast Universal Master Mix (Life Technologies), 2µl of 0.2µM primer mix (consisting of sense and anti-sense primers), 1µl of probe and 2.7µl dH_2O per well of a 384 well PCR plate. 8µl of mix was then added to each respective well, followed by addition of 2µl of 5ng/µl cDNA. Once aliquoted, the plates were sealed with adhesive PCR plate seals and centrifuged for a few seconds up to speed at 200 RCF. The plates were then analysed using a QuantStudio 12K Flex Real Time PCR System (Life Technologies) using the following parameters (95^oC- 10 mins, 95^oC- 15 seconds, 60^oC-1min for a total of 40 cycles). Once complete, the data was exported and cycling threshold (Ct) values were converted into 1/ΔCt in Excel. All data gathered from these experiments were normalised to *ACTB* and *GAPDH* in their respective assays.

For TAQMAN arrays, a mastermix consisting of 225µl Taqman Universal mastermix (ThermoFisher cat# 4324018), 90µl of cDNA and 135µl dH2O was made and 100µl was aliquoted into each well of the Taqman array. The array was then centrifuged at 200 RCF for 1 min (x2) and sealed. The array was then run using the same parameters as described above as with the data processing and analysed either via Graphpad or Genesis software (Thallinger Lab)

2.3.4 Single Cell QPCR

Single cell QPCR was performed on single hPSC seeded in each well of a 96 well plates. For each plate a lysis master mix was made consisting of 17µl 10%NP-40 and IPEGAL (sigma), 2.8µl 10mM dNTP, 10µl 0.1M DTT, 5.3µl RNaseOUT (Invitrogen) and 390µl dH_2O. 4µl of this mix was then added to each well. Single cells were then sorted into the plates through the FACS JAZZ sorter using the same sorting procedure as described above. The plates were then

sealed and stored at -80ºC until use. Once ready to use, a master mix consisting of 6.25ul 2x reaction mix, 1ul SSIII/Taq and 1.5ul primer mix was added to each well. The plates were then spun down and run using the following procedure: 50ºC 60min, 95ºC 2min, 95ºC 15sec and 25ºC 10 sec, for a total of 25 cycles. Primers were designed using the Fluidigm/Biorad PrimePCR tool with the top recommended primer set chosen. The primers used in the assay were designed by TAQMAN and fit the criteria of being 100% efficient (10%+/-) meaning the gene expression levels can fairly be compared to one another. The raw data was then processed using the SCEXV online tool into Bean plots.

2.3.5 Gene List:

ACTB, BMP2, BMP4, BMPR1A, CER1, CITED2, CLDN6, COL1A1, CXCR4, DKK1, DNMT3B, EOMES, FN1, FOXA2, FOXD3, FRZB, FST, GAL, GATA4, GATA6, GSC, HAS2, HHEX, ITGA5, KRT19, LEFTY1, LEFTY2, LGALS1, LHX1, LIN28A, MIXL1, MMP2, MYL7, NANOG, NODAL, PAF1, POU5F1, RPS18, SMAD2, SNAI1, SOX17, T, TAGLN2, TDGF1, WNT3

Primers were designed using the Primer3 and Assay Design Centre software (Roche) and were designed to span intron junctions where possible. The table below contains the gene name, primer sequences and corresponding probe number.

Table 5:Primer and Probe combinations used for QPCR to detect mRNA level

Gene	Sense	Anti-Sense	Probe #
ACTB	AGAGCTACGAGCTGCCTGAC	CGTGGATGCCACAGGACT	9
GAPDH	AGCCACATCGCTCAGACAC	GCCCAATACGACCAAATCC	60
OCT4	AGCAAAACCCGGAGGAGT	CCACATCGGCCTGTGTATATC	35
NANOG	AGATGCCTCACACGGAGACT	TTTGCGACACTCTTCTCTGC	31
SOX2	TTGCTGCCTCTTTAAGACTAGGA	TAAGCCTGGGGCTCAAACT	35
REX1	TCTGAGTACATGACAGGCAAGAA	TCTGATAGGTCAATGCCAGGT	65
DKK1	TTCTCCCTCTTGAGTCCTTCTG	CTACCATCGCGACAAAGACC	21
CER1	CACTGAACTTTCCTCCGTGAT	CTCCGTCTTCACCTTGCACT	9
MIXL1	GACACAGATGAGGGGCAGTT	CCCGTTTTCAGCTACCATTC	6
T	GCTGTGACAGGTACCCAACC	GGAGAATTGTTCCGATGAGC	23
GATA2	AAGGCTCGTTCCTGTTCAGA	GGCATTGCACAGGTAGTGG	76
GATA4	GGAAGCCCAAGAACCTGAAT	GCTGGAGTTGCTGGAAGC	69
GATA6	AATACTTCCCCCACAACACAA	CTCTCCCGCACCAGTCAT	90
SOX17	GCTTTGAATGTGTCCCAAAAC	CACACCCAGGACAACATTTCT	69
FOXA2	CGCCCTACTCGTACATCTCG	AGCGTCAGCATCTTGTTGG	9
PAX6	AGGGCAACCTACGCAAGA	CGTTGGAACTGATGGAGTTG	12
SOX1	ACCAGGCCATGGATGAAG	CTTAATTGCTGGGGAATTGG	37
TUBB3	GCGGATCAGCGTCTACTACA	GGCCTGAAGAGATGTCCAAA	85
PAX3	CAGCCCACATCTATTCCACA	CGTGCTTTGGTGTACAGTGC	69
SOX10	GACCAGTACCCGCACCTG	GCGCTTGTCACTTTCGTTC	75
CD34	GCGCTTTGCTTGCTGAGT	GGGTAGCAGTACCGTTGTTGT	8
HAND1	AACTCAAGAAGGCGGATGG	GGAGGAAAACCTTCGTGCT	24
MESP1	CTGTTGGAGACCTGGATGC	CGTCAGTTGTCCCTTGTCAG	27
TBX6	GAACGGCAGAAACTGTAAGAGG	GTGTGTCTCCGCTCCCATAG	5

MSGN1	AGCTCAGGATGAGGACCTTG	CTGGCCTCTCTGGCTGTAGA	87
MEOX1	AAATCATCCAGGCGGAGAA	AAGGCCGTCCTCTCCTTG	2
HOXC9	GCAGCAAGCACAAAGAGGA	CGTCTGGTACTTGGTGTAGGG	85
NKX1.2	GTCGAAGCGGGGAAAGAT	GATCCTCCGCATCCTCCT	78
CDX1	GGTCTACACCGACCACCAA	CGGTTTTGGAACCAGATCTTC	1
CDX2	ATCACCATCCGGAGGAAAG	TGCGGTTCTGAAACCAGATT	34

2.4 Clonogenic and Differentiation assays

2.4.1 Monolayer Differentiation

hPSC were disassociated into a single cell suspension as stated above and pelleted in a centrifuge at 200 RCF for 4 min. During centrifugation, freshly prepared vitronectin coated plastic-ware was aspirated to remove the PBS/vitronectin solution and replaced with mesoderm differentiation media (5 days E8 +3μM Chiron), endoderm differentiation media (Day 1 RPMI (Sigma) 1% B27 (ThermoFisher) 3μM CHIR99021 and 100ng/μL Activin A. Day 2 RPMI + 100ng/μl Activin A for additional 2 days) and ectoderm differentiation media (5 days E6 + 1μM DMH-1 and 10μM SB431542) and 10μM Y-27632 (Tocris) to aid plating and survival. The cells were then counted using the method as stated above and seeded at a density of 5000 cells/cm^2. For the duration of each differentiation, the respective differentiation media was replaced with media without 10μM Y-27632. After completion of the protocol, the cells were either immunostained, harvested for RNA or harvested for flow cytometry (mesoderm only) using previously stated methods. Mesoderm protocol based on Lippmann's protocol [147] but CHIR99021 concentration reduced to 3μM, endoderm protocol based on Kunisada's [148], ectoderm protocol based on Neely's [149].

2.4.2 Embryoid Body Differentiation

The assays were set up in 96 well U Bottom plates (Grenier) with the outside wells containing 100µl 1xPBS to prevent evaporation. 50ul of each condition were then plated into the desired number of wells (minimum 10) and mixed with 50ul of 3000 sorted hPSC using the standard procedure. 10µM Y-27632 (Tocris) to aid survival, and 50µg/ml Gentamycin (Sigma) to prevent contamination from sorting were also added. All embryoid body (EB) assays were formed using four defined conditions: Final concentration, Mesoderm (20ng/ml Activin A and 20ng/ml BMP4), Endoderm (100ng/ml Activin A and 1ng/ml BMP4), Ectoderm (100ng/ml bFGF, 1µM DMH-1 and 10µM SB431542) and Neutral (APEL media). The Neutral media is defined as a basal media that does not contain any growth factors or chemical compounds that may direct differentiation. All EB were left to grow for 10 days in a 37^{o}C incubator with 5% CO_2. On the 10th day, the EB were pooled together in their respective condition and harvested for RNA by being dissolved in TRIzol. Samples were stored at -80^{o}C until use. EB size was determined using Fiji, ImageJ software [150]. Length of EB was determined by measuring the length of the EB in their respective images using a scale from a length of known value. All EB images were taken from x20 magnification using an inCell analyser.

2.4.3 Clonogenic Assays to determine lineage bias through expression of lineage associated genes

Clonogenic assays were performed on H9TVD3 hPSC grown on MEF and MEF-free conditions with all assays being completed within KOSR/MEF conditions. The hPSC were dissociated into single cells as described above and immunostained with the appropriate antibodies as described (table 4). The cells were then sorted into the desired sub-populations using the procedure as described above. After sorting, the cells were centrifuged at 200 RCF for 4 min as described above and countedusing the standard procedure. For the assay, cells were seeded at a density of 500 cells/cm^2 into 24 well plates (ThermoFisher) in KOSR/MEF media with 10µM Y-27632 (Tocris) was added to aid plating and

survival, and 50µg/ml Gentamycin (Sigma) was added to prevent contamination from sorting. The plate was then placed in a 37ºC 5% CO_2 incubator and left for 2 days. On day 3, the media was changed to fresh KOSR media. On day 5, the plates were washed with 1xPBS (w/o Ca^{2+}, Mg^{2+}) fixed with 4% PFA (Sigma) for 15min and permeabilised with 0.1% TritonX (Sigma) for 1hr. The plates were then stored in 1x PBS at 4ºC until use.

2.4.4 Single Cell Deposition

One day prior to sorting, MEF cells were seeded at a density of 10000 cells/well of a flat bottom 96 well plate (Grenier) and left at 37ºC 10% CO_2 as previously described. hPSC grown on KOSR/MEF were dissociated into a single cell suspension and immunostained to be ready for cell sorting. To calibrate the machine for cell sorting, green and red fluorescent beads (Bangs Laboratory and BD biosciences respectively) were mixed together and sorted singularly into single wells of a 96 well plate with a row of red and a row of green beads respectively.The plates were then viewed on a inCell analyser 2000 to determine specificity of the sorting, ensuring only the specified bead was sorted into the well.Once calibrated, the cells were then sorted via the BD FACS JAZZ into KOSR/MEF plates with one cell being depositedin each well. To the KOSR media, 10µM Y-27632 (Tocris) was added to aid plating and survival, and 50µg/ml Gentamycin (Sigma) as added to prevent contamination from sorting. The plates were then spun down for a few seconds at 200 RCF and put in 37ºC 5% CO_2. After 3 days, the KOSR media was changed in each well. After 7 days, colonies were identified in the respective wells and were then passaged into a 48 well plates containing KOSR/MEF and 10µM Y-27632. The colonies were then left to grow to 80% confluency before being passaged into 24 well plates. This process continued with growing the colonies up to 80% confluence and being passaged into a smaller plate, up to a 6 well plate. At the 24 well stage, 12 colonies were chosen to be passaged into a 12 well plate, with the remaining 12 being washed with 1xPBS (w/o Ca^{2+}, Mg^{2+}) fixed with 4% PFA (Sigma) for 15 min and permeabilised with 0.1% TritonX (Sigma) for 1 hr, to be immunostaiined for NANOG. A total of 6 colonies were taken forward in each respective sort, with the

remaining colonies being immunostained for NANOG. These 6 colonies (now clonal lines) were analysed via FACS on KOSR/MEF and E8/Vitronectin after 5 passages. The clonal lines were also differentiated via single cell dissociation as previously described and seeding at a density of 500cells/$cm^{2.}$ The cells were differentiated to mesoderm (5 days E8 +3µM Chiron), endoderm (Day 1 RPMI (Sigma) 1% B27 (ThermoFisher) 3µM CHIR99021 and 100ng/µL Activin A. Day 2 RPMI + 100ng/µl Activin A) and ectoderm (5 days E6 + 1µM DMH-1 and 10µM SB431542) from Vitronectin. The cells were then immunostained for TBX6 (mesoderm), SOX17 and FOXA2 (endoderm) and PAX6 (ectoderm) at the appropriate concentration (table 4).

Chapter 3 : Use of Antigens to Identify and Assess Subsets of hPSC for Lineage Bias

3.1 Introduction

3.1.1 The differences between cell fate choices *in vivo* and *in vitro*

The process of differentiation *in vivo* from pluripotent cells in the ICM to terminally differentiated, functional somatic cells is a complex, systematic process. This process can be divided into several steps where biological events occur that allow the cell to make key fate choices that determine a final stable cell type. These events involve the cell receiving and transmitting signals that promote specific gene expression and repression processes that trigger fate choices during differentiation from the ICM to germ layer specification, through to final somatic cell type. These signalling network processes are still being understood natively *in vivo*. Trying to study, analyse and recreate these processes involved in early differentiation decisions *in vitro*, provides a key challenge for understanding fate determination within hPSC in culture.

3.1.2 Connecting heterogeneity within the stem cell compartment and lineage-bias

The evidence that hPSC can exist with different states in culture poses the question of the relevancy of these states to the fate determination of these cells. As will be described in this section, there are numerous reports that have showing heterogeneity in the stem cell compartment. This can be seen in differences on the transcriptome as well as cell surface antigen expression. Additionally, heterogeneity has also been shown in the signalling networks involved in both maintenance of the pluripotent and promoting differentiation.

As previously described, an early example of this heterogeneity was demonstrated by (Chambers et al. 2007) [134] where it was shown that mouse

NANOG negative cells were able to convert into a NANOG positive cells. This provided evidence that at least a subset of these NANOG Negative cells were undifferentiated. Another example of heterogeneity in the mouse was shown with the pluripotency associated genes REX1 and OCT4. Here, similarly to NANOG, there was evidence for the existence of two separate pluripotent states. It was postulated that cells that co-expressed both REX1 and OCT4 were more akin to cells from the ICM, whereas cells that were REX1-/OCT4+ were more akin to the early primitive ectoderm [151]. However, crucially, both REX1/OCT4 states were able to interconvert to the heterogeneous starting population that contained both states. This suggested that at least a subset of cells from both states were still pluripotent. Though while this phenomenon existed within mouse ES cells, this was not evidenced in human ES cells. It was shown that only REX1 Positive cells were able to generate REX1 Negative cells, highlighting the differences between the two species regarding states in the stem cell compartment [152].

With cell surface antigen heterogeneity, species differences can also be seen between the human and mouse. In the mouse, the antigens SSEA1 and PECAM-1 are expressed on pluripotent cells [153]. Loss of either SSEA1 and PECAM-1 was shown to result in up-regulation of lineage associated genes and an increase in propensity for differentiation, although the cells retained the capability of self-renewal [154]. In respect to human ES cells, heterogeneity has been observed with several cell surface antigens included SSEA3 [129,155], CD9 and GCTM2 [130,131]. In regard to SSEA3, SSEA3+ and SSEA3- cells were shown to be able to self-renew [129]. Further studies on the single cell level revealed considerable heterogeneity within SSEA3+ and SSEA3- cells regarding expression of lineage associated markers [155]. Among these cells was evidence for subsets that co-expressed pluripotency associated genes such as *NANOG* and *OCT4* in addition to expressing lineage associated genes such as *GATA6*. This observation provided evidence of the existence of subsets of pluripotent cells that were existing within a lineage-biased state.

Further studies into the existence of SSEA3 Positive subsets of pluripotent cells with a lineage bias was performed in the Andrews lab using a human ES cell line containing a GFP gene reporter construct for the endoderm associated gene *GATA6* (Andrews lab, Unpublished). Here it was shown that by growing the reporter cell line in KOSR/MEF self-renewal conditions, a subset of SSEA3+ cells

also co-expressed *GATA6*. The hypothesis was that while both the SSEA3+/*GATA6*+ and SSEA3+/*GATA6*- subsets were pluripotent, the SSEA3+/*GATA6*+ cells had a lineage bias to differentiate towards the endoderm. Following a series of functional tests on the bulk and single cell level of these subsets, there was evidence to suggest that indeed, the SSEA3+/*GATA6*+ subset did exhibit increased differentiation propensity to the endoderm.

The objective of this study was to identify a subset of human pluripotent cells that exhibited a lineage bias to differentiate to the mesoderm. As described, there are different methods to achieve this, such as using a gene reporter construct containing a mesoderm associated gene, or by using a combination of cell surface antigens. Both methods have shown to be successful in identifying heterogenous subsets of pluripotent cells with different functional properties. In a complimentary study, a gene reporter containing the mesendoderm marker *MIXL1* was used to identify subsets for cells with a mesoderm bias (Andrews lab, unpublished). This was performed initially using a similar work-flow to what had been previously been done using the *GATA6* reporter (Andrews lab, unpublished). The results showed that similarly to *GATA6,* there was evidence of a subset of SSEA3 Positive cells from KOSR/MEF self-renewal conditions that co-expressed *MIXL1*. Following transcriptome analysis on both the bulk and single cell level, as well as functional analysis comparing the SSEA3+/*MIXL1*- and SSEA3+/*MIXL1*+ subsets, the results revealed evidence that the SSEA3+/*MIXL1*+ showed increased expression of mesoderm/endoderm associated genes.

An alternative approach to using gene reporters is to follow a similar work-flow to what had been performed using CD9 and GCTM2. Initially using just these two markers [130] and then additionally with EPCAM [131], it was shown that functionally different subsets of cells could be isolated based on levels of antigen expression. Here, cells that were isolated based on High/Mid/Low/Negative expression of these respective markers were shown to have differences in respect to expression of lineage associated genes, clonogenicity and propensity for differentiation. Cells that were sorted based on “High” and “Mid” expression showed increased propensity for mesoderm/endoderm gene expression and differentiation potential. In contrast, cells that were “Low” and “Negative” expressing of these antigens, showed increased expression of ectoderm

associated markers and differentiation propensity to cells with a neural phenotype. However, crucially, these sorted subsets demonstrated that they all contained pluripotent cells due to the ability of the cells to form single cells colonies and expression of pluripotency associated genes.

The evidence from (Hough et al. 2009 and 2014) [130] showed that subsets of pluripotent cells can be identified based off levels of cell surface antigen expression. Hypothetically, this could be applied to different antigen combinations including SSEA3. Additionally, other candidate antibodies exist that are expressed on pluripotent cells such as the well-defined TRA family and novel antibodies from the Andrews lab [156]. SSEA3 however has shown to be a useful marker for isolating subsets of cells as evidenced by (Tonge et al 2011) [129] and (Andrews lab, unpublished). Together with CD9, sorting cells with SSEA3 provides a novel approach to identifying subsets of cells with lineage bias. This would provide another means of exploring the heterogeneity within SSEA3 Positive cells. Achieving this without the need for a gene reporter would have practical benefit such as for translational applications in regenerative medicine. However, gene reporters can be used to identify candidate subsets of cells for further analysis.

3.1.3 Experimental Objectives

Determining the correlation between levels of antigen expression and gene reporter expression in self-renewal conditions, would provide information of potential subsets to sort independently with the antibodies. For this study we obtained a selection of reporters attributed to early mesoderm/endoderm expression. These reporters were the endoderm associated marker *GATA6*, the paraxial mesoderm marker *MSGN1,* the mesendoderm marker *MIXL1* and the early mesoderm marker *BRACHYURY*. Utilising these reporters with the use of antibodies or sorting non-reporter cells independently with just antibodies, the aim is to identify subsets of cells that could be analysed for mesoderm lineage bias.

3.2 Results

3.2.1 Comparison of SSEA3 and CD9 expression against gene reporter expression.

The *BRACHYURY*, *GATA6*, *MIXL1* and *MSGN1* gene reporter lines were grown on KOSR/MEF conditions and analysed in conjunction with SSEA3 and CD9 independently using flow cytometry. In each flow cytometry analysis presented in this thesis, negative gating is set using the P3X antibody for antigen expression and a wildtype, unlabelled cell line for the gene reporter expression. This is detailed in Chapter 2 Materials and Methods.

The analysis was designed to see if there was a difference in the correlation between antigen and reporter expression (Figure 2 A-D). Within this experimental set-up, the fluorescence intensity range was 10^0-10^4 with 10^0-10^1 being defined as “Negative”, 10^1-10^2 as “Low” and 10^2-10^4 as “High/Mid”. The early mesoderm and endoderm markers *BRACHYURY* (Figure 2 A) and *GATA6* (Figure 2 C) were shown to be differentially expressed in KOSR/MEF conditions. Approximately ~4-5% of the cells were*GATA6* High (>10^2) in fluorescence intensity. *BRACHYURY* in contrast was clustered around 10^2 in fluorescence intensity and over multiple biological repeats, the amount of either CD9+/*BRACHYURY*+ or SSEA3+/*BRACHYURY*+ cells varied between20->80% with the majority of *BRACHYURY* Positive cells also co-expressing the respective antigen. However, in a typical culture, more than 50% of the cells were *BRACHYURY* Positive. The *MIXL1* (Figure 2 D) reporter line showed variable expression in this condition, typically ~1-3% but had been shown to increase up to 7% in some samples. The majority of *MIXL1* expressing cells also co-expressed SSEA3 and CD9 respectively. The expression level of *MIXL1* was also variable between Low (10^1) and High (10^3) in each replicate. The reporter line for the paraxial mesoderm marker *MSGN1* (Figure 2 B) was expressed at <1% in this condition. This suggests that this was not a gene that was strongly expressed in self-renewing hPSC and therefore not a suitable candidate for further analysis.

With the *GATA6* reporter, the majority of the *GATA6* Positive cells were clustered predominantly in the CD9 High/Mid region (10^2-10^4 fluorescence intensity). This was also observed with the *BRACHYURY* reporter and together suggests that

CD9 High/Mid cells were a candidate population for isolating a subset of hPSC readying for mesoderm/endoderm differentiation. In respect to SSEA3, the distribution of expression was different to CD9, with a broad expression distribution ranging from the High to Low range (10^1-10^4 fluorescence intensity) in respect to co-expressing *GATA6* High cells. This was similar with the *BRACHYURY* reporter, however what was notably different, was a large cluster of co expressing SSEA3 and *BRACHYURY* cells residing within the SSEA3 High region (10^3-10^4 fluorescence intensity). In *MIXL1*+ cells, expression of both CD9 and SSEA3 was distributed across the Low to High range.

3.2.2 SSEA3 and CD9 expression patterns to isolate subsets of cells

The results from Figure 2 showed there was a subset of SSEA3 Positive and CD9 Positive cells that co-expressed *GATA6*, *BRACHYURY* and *MIXL1*, through these genes were differentially expressed. To identify a mesoderm biased subset of cells, Positive CD9 expression was chosen as a suitable candidate. SSEA3 is a marker that is primarily expressed on undifferentiated cells, but as shown in Figure 2, there is evidence of a subset of those cells expressing *BRACHYURY*, notably in the SSEA3 High expressing region. This also suggested SSEA3 as a candidate for identifying a mesoderm biased subset of cells.

To test whether SSEA3 and CD9 could be used together to identify a mesoderm biased subset, cells grown and sorted from KOSR/MEF conditions that had High/Mid CD9 expression were chosen and sorted with SSEA3 with either High or Low expression. This was to determine if separating SSEA3 expression resulted in identifying a subset of cells with greater propensity to mesoderm differentiation. For comparison, the SSEA3 Positive (+)/CD9 Negative subset was sorted to determine if this subset resulted in more ectoderm differentiation. Therefore, the three populations sorted and analysed were SSEA3 High/CD9+, SSEA3 Low/CD9+ and SSEA3+/CD9. A representative figure describing the sort set-up can be seen in Figure 3. The experiments were performed three pluripotent lines for comparison. These consisted of two ES cell lines (HES3 and H9) and one iPSC (NCRM1).

3.2.3 Functional assessment of SSEA3/CD9 sorted subsets: Cloning Efficiency

Before investigating differentiation potential of these SSEA3/CD9 subsets, they were assessed to determine if there was evidence these subsets contained undifferentiated hPSC. Cells were plated directly from sorting into KOSR/MEF self-renewal conditions in the presence of Rho Kinase inhibitor and left to form colonies for 5 days. After 5 days, the cells were immunostained for the pluripotency associated markers OCT4 and SOX2 (Figure 4 A). The plating efficiency for these subsets was different across 3 independent biological repeats, with SSEA3 High/CD9+ subset having ~15% cloning efficiency, the SSEA3 Low/CD9- subset having ~8~10% efficiency and the SSEA3+/CD9- subset having ~ 3-5% plating efficiency in the presence of Rho Kinase inhibitor. However, all three sorted subsets could form colonies with > 80% of the cells in the colonies co expressing OCT4 and SOX2 (Figure 4 A). The results showed that the SSEA3/CD9 sorted subsets contained undifferentiated hPSC.

3.2.4 Functional assessment of SSEA3/CD9 sorted subsets: EB Differentiation

To investigate whether these three SSEA3/CD9 sorted subsets exhibited a lineage bias to differentiation, embryoid body (EB) experiments were performed based on directed and undirected, “Neutral” differentiation. The directed conditions were to assess whether the three SSEA3/CD9 subsets could differentiate to derivatives of the three germ layers, evidencing that these subsets contain undifferentiated cells. Additionally, efficiencies of differentiation can be compared to assess whether any of the subsets have lowered efficiency of differentiation to a specific lineage. The Neutral condition is designed to assess lineage bias by allowing the cells to differentiate to what they have a propensity to differentiate to. This is without addition of components that could influence fate decisions.

The SSEA3/CD9 subsets were cell sorted and seeded into EB growing conditions in 3 independent biological replicates and put under four differentiation methods

mesoderm, endoderm, ectoderm (Figure 4 B-D) and neutral (Figure 5 A-C and Figure 6) as described in Chapter 2 Material and Methods; Mesoderm (20ng/ml Activin A and 20ng/ml BMP4), Endoderm (100ng/ml Activin A and 1ng/ml BMP4), Ectoderm (100ng/ml bFGF, 1µM DMH-1 and 10µM SB431542) and Neutral (APEL media)[157]. These conditions were assessed via QPCR for expression of the extra embryonic mesoderm genes *HAND1* and *GATA2* for the mesoderm differentiation, *GATA4*, *GATA6* and *SOX17* for endoderm differentiation and *PAX6*, *TUBB3* for the ectoderm differentiation. For the Neutral differentiation, the EB were formed under the APEL basal media without the addition of growth factors or other known chemical inducers of differentiation in hPSC.

The results of these EB experiments demonstrated that while all three SSEA3/CD9 subsets could differentiate in all three induced conditions, their propensity for differentiation was different (Figure 4). In the mesoderm conditions, all three SSEA3/CD9 subsets had higher expression levels of *HAND1* and *GATA2* compared to the undifferentiated control (Figure 4 B). This trend was also shown in the endoderm (Figure 4 C) and ectoderm conditions (Figure 4 D) with higher expression levels of *GATA4*, *GATA6* and *SOX17* in the endoderm conditions and *PAX6* and *TUBB3* expression in the ectoderm conditions (Student T test $p<0.05$). However, the SSEA3+/CD9- subset also exhibited increased gene expression for endoderm associated genes *GATA4*, *GATA6* and *SOX17* in the endoderm condition compared to the other two SSEA3/CD9 subsets.

While all three subsets could differentiate to all three derivatives of all three germ layers, there was differences in the size of the EB. A total of 8-10 EB across three biological replicates were analysed and averaged. Both the SSEA3 High/CD9+ and SSEA3 Low/CD9+ subsets formed EB that were larger than the SSEA3+/CD9- subset in the mesoderm and endoderm conditions (Figure 5 A-B) ($p<0.05$ Tukey Multiple Comparison Test). In this condition there was also more observable cell death and debris with the SSEA3+/CD9- subset. In the ectoderm conditions (Figure 5 C), there was not a significant difference in the size of EBs. EBs analysed using Fiji, ImageJ software [150].

In respect to the Neutral conditions, the three sorted subsets also exhibited different propensities to differentiation (Figure 6 A-C). The SSEA3+/CD9- subset followed the prediction and exhibited significant increase in expression of genes

associated with neural ectoderm (*PAX6* and *SOX1*) and neural crest *(PAX3* and *SOX10*) which are two ectoderm derived lineages. The SSEA3 High/CD9+ subset exhibited significant increased expression of mesoderm associated genes (*CD34* and *HAND1*). The SSEA3 Low/CD9+ subset did not follow the trend of the other two sorted subsets, though did appear to have gene expression profiles closer to the SSEA3+CD9- subset, with elevated expression of ectoderm associated genes compared to the SSEA3 High/CD9+ subset. The results confirmed that by segregating the hPSC sorted population into differential levels of expression for SSEA3 with CD9, subsets of cells could be isolated that had different propensities for differentiation.

3.2.5 Single cell transcriptome analysis of SSEA3/CD9 sorted subsets

When considering the strategy for the transcriptome analysis, it was necessary to consider the depth of information that can be gathered from population transcriptome analysis compared to single cell analysis. As the focus on the study is based on heterogeneity and differences between single cells that may have functional implications, such observations can be lost in the average gene expression levels derived from analysis of the whole population. Consequently, single cell QPCR was chosen to analyse the transcriptome of these sorted populations. The gene list featured 48 genes chosen based on the expression profiles of SSEA3/*MIXL1* sorted subsets with the top ranked differentially expressed genes based on fpkm values from previous bulk RNA seq analysis (Andrews lab, Unpublished). The gene list features predominately early mesoderm/mesendoderm genes that were chosen based on the highest variation of expression between the SSEA3+/*MIXL1*+ and SSEA3+/*MIXL1*- subsets. This list also contained the pluripotency associated genes *OCT4* and *NANOG*.

For the experiment, the SSEA3/CD9 subsets were analysed and sorted via FACS from KOSR/MEF conditions, but with a single cell being sorted and deposited in lysis buffer, in a single well of a 96 well plate. A QPCR assay was then performed on the plate. Ct values obtained from the QPCR were converted into Z-scores (standard deviation scores) for each gene in each cell and analysed in GENESIS

[158]. The results revealed that within each sorted subset, there was heterogeneity, as shown by existence of cells that either expressed or did not express the gene of interest. Additionally, these genes were differentially expressed at varying levels. This was shown with some cells having either high, lower or non-expression of the respective gene in comparison to the other cells in the assay (Figure 7). Due to this variation when attempting to apply hierarchical clustering of these populations, the resulting heatmap (Figure 7) showed there were different sub-groups of cells from the different SSEA3/CD9 subsets clustering together, as opposed to cells from the same subset clustering together only.

When applying gene clustering in the analysis, the pluripotency associated genes *OCT4* and *NANOG* clustered together, while early mesendoderm/endoderm associated genes including *MIXL1*, *FOXA2*, *CXCR4* and *SOX17* also clustered close together. When analysing the range of gene expression in each subset (Figure 8, Figure 9, Figure 10), the SSEA3 High/CD9+ subset was shown to have more cells that expressed early endoderm/mesendoderm associated genes such as *HHEX*, *CXCR4* and *SOX17*. In contrast, the SSEA3+/CD9- cells had more cells that did not express these markers. The SSEA3 Low/CD9+ subset was the population that had the most variation in expression of these genes, including the pluripotency associated genes. The PCA analysis shown in (Figure 11) provides a representation for the heterogeneity within the SSEA3/CD9 subsets. Similarly, to the heatmap, the PCA plots show a scattering of cells within all the subsets, as opposed to distinct groups.

However, when comparing expression levels of the genes in the panel, all three SSEA3/CD9 subsets had either lower expression levels or fewer cells expressing these genes, compared to the SSEA3+/*MIXL1*+ and SSEA3-/*MIXL1*+ subsets which were analysed in parallel (Figure 12). Additionally, *BRACHYURY* was not expressed in any cell from any of the SSEA3/CD9 subsets. Unpublished results had shown that these SSEA3/*MIXL1* subsets did contain pluripotent cells and these results suggest, that the *MIXL1*+ subsets were closer to mesoderm/endoderm differentiation compared to the SSEA3/CD9 subsets. Consequently, the SSEA3/CD9 subsets did not appear as a suitable candidate for identifying a mesoderm biased subset, as differences between these subsets were only clearly observable once they had differentiated. In the case of SSEA3

High/CD9+, the results indicate that that this subset is heterogenous and contains some mesoderm biased cells and some not. This observation is supported by the two-colour fluorescence plots using the gene reporters (Figure 2) where not every SSEA3 High expressing cell is co-expressing the mesoderm reporter gene.

From these results it was decided to consider other candidate antibodies to identify a mesoderm biased subset of undifferentiated cells.

3.2.6 BF4 as a tool to isolate subsets of cells

As demonstrated using SSEA3 and CD9, subsets of hPSC with different functional properties in respect to differentiation and clonogenicity can be isolated. Considering other candidate antibodies that could be assessed from the literature, the antibody BF4 was intriguing due to its broad expression profile range from High through to Negative. This expression profile was first observed by (Wright et al 2011) [143]. To determine whether it was reproducible, pluripotent cells were analysed for BF4 expression via flow cytometry. Furthermore, to explore whether this expression profile related to different pluripotent sub-states, BF4 expression was segregated into “High”, “Low” and “Negative (-ve)” and sorted from KOSR/MEF conditions based on these gated populations. A representative figure of this procedure can be observed in (Figure 13).

3.2.7 Analysis of BF4 sorted subsets reveal they contain hPSC and exhibit differential expression of Lineage Associated Genes

To determine whether BF4 was a candidate for identifying a mesoderm biased subset of cells, transcriptome analysis was performed on the BF4 subsets from cells growing in KOSR/MEF self-renewal conditions. The results showed that differential expression of genes associated with either self-renewal or differentiation was observed in the BF4 sorted subset (Figure 14 A). The BF4 High cells had higher expression of genes associated with self-renewal and the

pluripotent state compared to BF4 Negative cells, whereas the BF4 Low and BF4 Negative cells had higher levels of expression of genes associated with mesendoderm differentiation (*CER1* and *DKK1*) endoderm (*GATA6* and *SOX17*) and neural ectoderm (*PAX6*) (Tukey Multiple Comparison Test $p<0.05$). The results showed that all three sorted subsets were shown to have relatively higher levels of expression of the pluripotency associated genes, however increased expression of differentiation associated genes was observed in the BF4 Low and BF4 Negative subsets. The three BF4 subsets were also seeded in KOSR/MEF self-renewal conditions to assess if they contained undifferentiated cells that could form colonies (Figure 14 B). All three BF4 subsets contained cells that could form colonies, albeit at different cloning efficiencies: BF4 High at ~7-10%, BF4 Low at ~3-5% and BF4 Negative (-ve) at ~1-3% in the presence of Rho Kinase Inhibitor. However, these colonies all predominantly contained cells that co-expressed the pluripotency associated markers SOX2 and OCT4. Together with the transcriptome results, it indicates that the BF4 sorted subsets contain undifferentiated hPSC.

3.2.8 Comparison of BF4 and gene reporter expression

The transcriptome results from the BF4 High/Low/Negative (-ve) sorted cells revealed that BF4 Low and BF4 Negative cells had higher expression of lineage associated genes. With analysis of the gene reporters shown in (Figure 2) both SSEA3 and CD9 had different antigen/gene expression profiles despite both having comparable percentages of co-expressing antigen/gene for *BRACHYURY* and *GATA6*. To determine if this was the same with BF4, the reporter cells were analysed with BF4 from self-renewal conditions in the same manner (Figure 15 A-D). Similarly, to SSEA3 and CD9, there was less than 1% of the culture composed of BF4/*MSGN1* co expressing cells (Figure 15 B). The *BRACHYURY* expressing cells (Figure 15 A) were expressed across the BF4 High/Low range similarly to SSEA3 and CD9. However, in contrast, the BF4 Negative subset had significantly more *BRACHYURY* expressing cells compared to the SSEA3 Negative and CD9 Negative subsets. This trend was repeated with *GATA6* (Figure 15 C), where the *GATA6* higher expressing cells were clustering

predominantly in the BF4 Negative region. For *MIXL1* expression (Figure 15 D), the majority of *MIXL1* expressing cells were also clustered primarily in the BF4 Negative region. Similarly, to the transcriptome analysis, the results suggest that a loss of BF4 expression coincides with upregulation of lineage specific genes.

3.2.9 Relationship between BF4 and other glycoprotein antigens

The BF4 antibody did appear to be a useful candidate for identifying subsets of cells with differential expression of lineage associated genes. However, as a novel antibody, there was little known about its expression profile on pluripotent cells. Analysis of the BF4 antigen [143] revealed it was a ~250 kDa, glycoprotein, similarly to TRA160R/S and TRA181 and that it was also neuraminidase sensitive, similarly to TRA160S. Repeating the Co-IP experiment, with TRA160R (Figure 16 A), the result showed that the TRA160R antibody did recognise the protein pulled down with the BF4 antibody, at a molecular weight of ~250kDa.

Following the Co-IP, a mesoderm differentiation assay was performed to assess both BF4 and TRA-1-60S expression patterns (Figure 16 B). The results showed that BF4 expression did not coincide with TRA-1-60S expression, with substantial loss of expression of BF4 at <30% compared to TRA-1-60S which maintained expression greater than 90%. This indicates that BF4 was not recognising the TRA-1-60S epitope.

To test whether the BF4 epitope contained sialic acid, hPSC were treated with neuraminidase and immunostained for the respective antigens (Figure 16 C-D). The results confirmed that while BF4 and TRA160S did show reduced expression when treated with neuraminidase compared to the control, BF4 was more sensitive to treatment. Both TRA160R and TRA181 did not lose expression when treated and demonstrated clustered expression towards the “high” region (10^3-10^4). This result with TRA160R and TRA181 coincided with a previous observation seen with TRA181 in Badcock [159].

Together, the results suggest that while the BF4 antibody does not recognise the epitopes from TRA160R and TRA181, the epitopes are, however, related. This

has been shown with the TRA-1-60R epitope being on the same glycoprotein antigen as BF4 and TRA-1-81.

3.2.10 Mesoderm differentiation analysis to study antigen expression dynamics

Analysis from the gene reporters (Figure 2 and Figure 15) showed that SSEA3, CD9 and BF4 have different expression patterns in conjunction with the reporter genes. *GATA6* and *MIXL1* expression was shown to cluster primarily in the BF4 Negative region, while there was also a large cluster of *BRACHYURY* expressing cells in the BF4 Negative region (Figure 15). To further assess the differences in antigen expression and whether loss of BF4 expression coincided with differentiation, a 3-day time course, mesoderm differentiation assay was performed to observe the dynamics of antigen expression. The results demonstrated that all three antigens had differential expression profiles across the three days (Figure 17).

On day 1, all three antigens were expressed on over 90% of the cells. BF4 exhibited a broad expression profile from High to Low, where in contrast both CD9 and SSEA3 exhibited more expression in the High region. This correlation resembled the two-colour assays performed in self-renewing conditions (Figure 2). On day 2, all three antigens continued to exhibit differential expression patterns. A loss of BF4 expression was observed as shown by ~20% of cells being BF4 Negative. In contrast, CD9 and SSEA3 were still expressed at over 90% of the cells, with CD9 expression clustering more to the Mid region and SSEA3 expression was broadly distributed across the High to Low region. On Day 3, both BF4 and SSEA3 expression was substantially lost. BF4 expression was observed at ~10-20% of the cells, while SSEA3 expression was more varied at 10-50% of the cells. CD9 expression in contrast, was observed in over 90% of the cells with CD9/*BRACHYURY* co-expression being grouped in the “Mid” CD9 expressing region.

Altogether, in combination with the results from the two-colour assays in self-renewal conditions (Figure 16), the results demonstrated that BF4 is a marker sensitive to differentiation and that both CD9 and SSEA3 differ in their expression

profiles in these conditions. In both self-renewal and day1 mesoderm inducing conditions, both CD9 and SSEA3 were expressed in the High region, while also co expressing *BRACHYURY*. This result correlated with the evidence from the functional assays that a subset of CD9 and SSEA3 expressing cells were differentiating towards the mesoderm.

Figure 2: Two colour fluorescence analysis with SSEA3 or CD9 vs gene reporters in self-renewal conditions reveal differential expression of lineage associated genes

Flow cytometry plots showing SSEA3 or CD9 expression against *BRACHYURY* (A), *MSGN1* (B), *GATA6* (C) or *MIXL1* (D) in KOSR/MEF self-renewal conditions. A subset of CD9 High/Mid cells co-expressed *BRACHYURY* and *GATA6* respectively. A large subset of SSEA3 High expressing cells co-expressed *BRACHYURY*, but expression was spread across the SSEA3 High/Low/ Negative (-ve) as with *GATA6*. Less than 1% MSGN1 expression was observed in these conditions. *MIXL1* expression was broadly distributed across SSEA3 and CD9 High to Low expression respectively. n=3

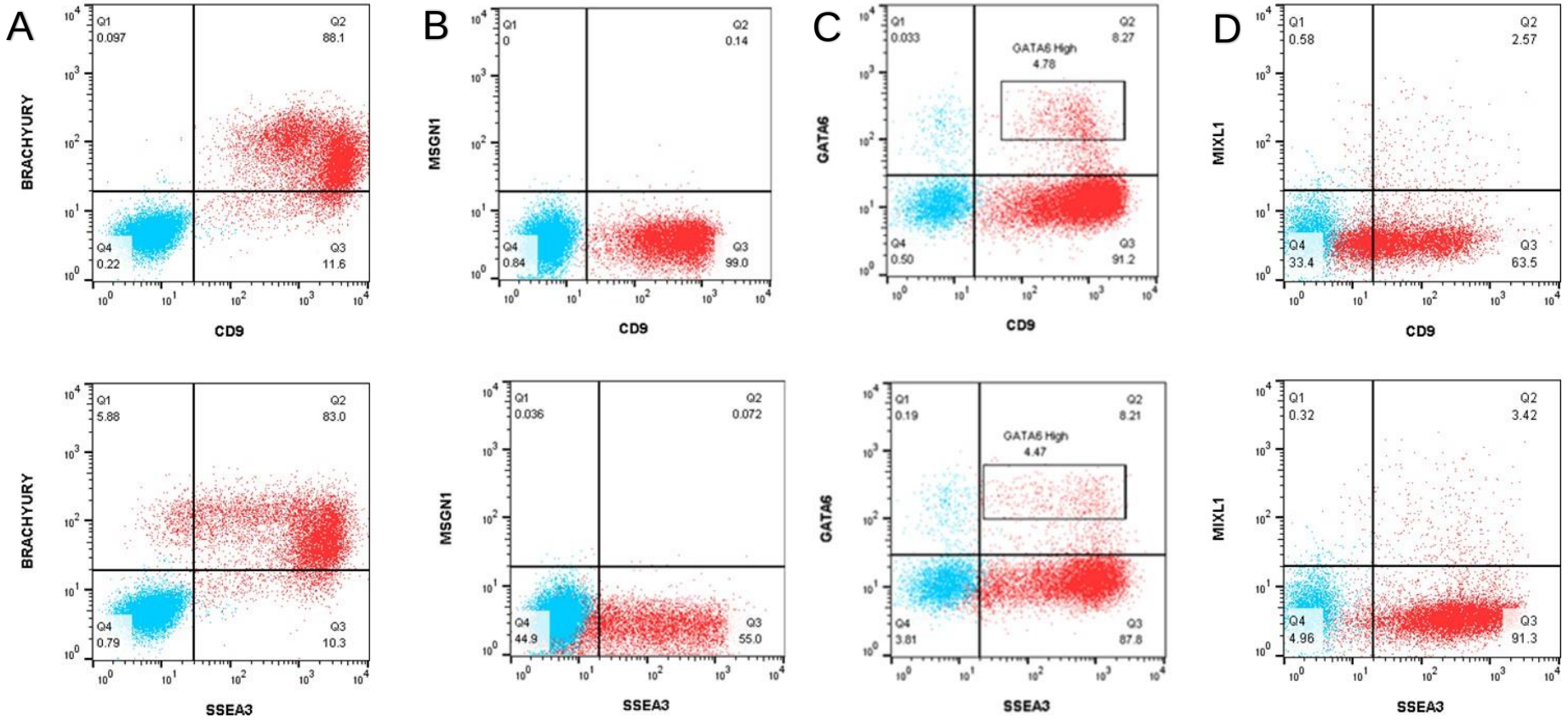
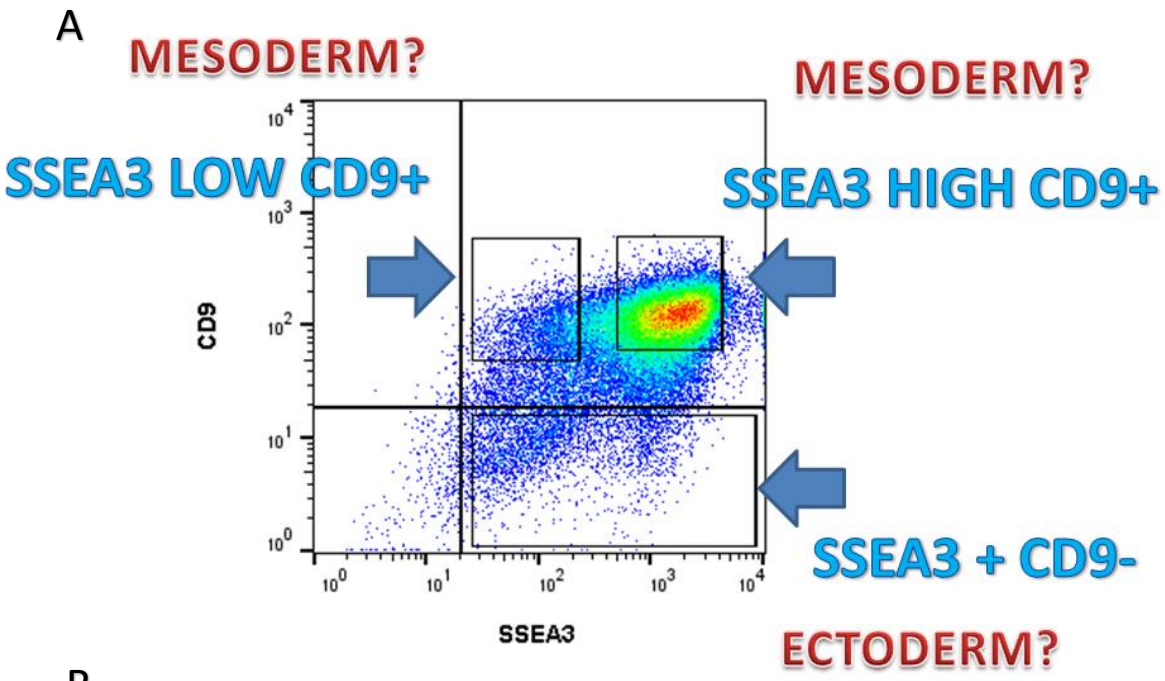


Figure 3: Representative figure of SSEA3 and CD9 co expression to highlight potential lineage bias of sorted subsets of cells

(A) Representative SSEA3/CD9 flow cytometry histogram plot showing how antigen expression was separated into SSEA3 High/CD9+, SSEA3 Low/CD9+, SSEA3+/CD9- for cell sorting. SSEA3/CD9 subsets tested functionally and on the transcriptome to assess lineage bias to mesoderm or other lineages. (B) Example of reanalysis of cell post sort, showing the purity of the sort. Indirect immunofluorescence was used with IgM and IgG specific secondary antibodies. The fluorescence plots below show there was no cross reactivity with the respective antibodies.



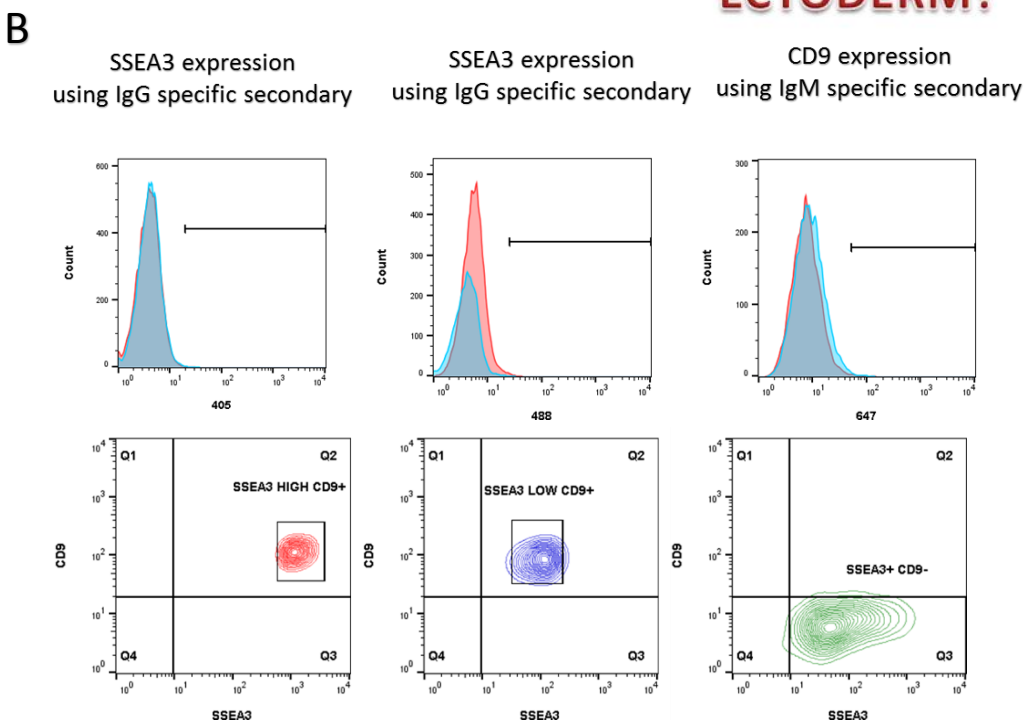


Figure 4: SSEA3/CD9 sorted subsets demonstrates the bulk population contains pluripotent cells but with different propensities for differentiation

EB differentiation assays. (A) All three subsets were able to form colonies that co express the pluripotency associated markers SOX2 and OCT4 in self-renewal conditions direct from sorting, demonstrating that the three sorted subsets contain pluripotent cells 1/ΔCT for mesoderm differentiation (*HAND1* and *GATA2*) (B), endoderm differentiation (*GATA6*, *SOX17* and *GATA4*) (C) and ectoderm differentiation *(PAX6* and *TUBB3*) (D) , comparing the following subsets: SSEA3 High/CD9+, SSEA3 Low/CD9+, SSEA3+/CD9- and E8 undifferentiated control. The results revealed that all three subsets could differentiate to the respective derivatives. Scale bar set at 50µm. Statistical significance shown by asterisks (*) p<0.05 n=3

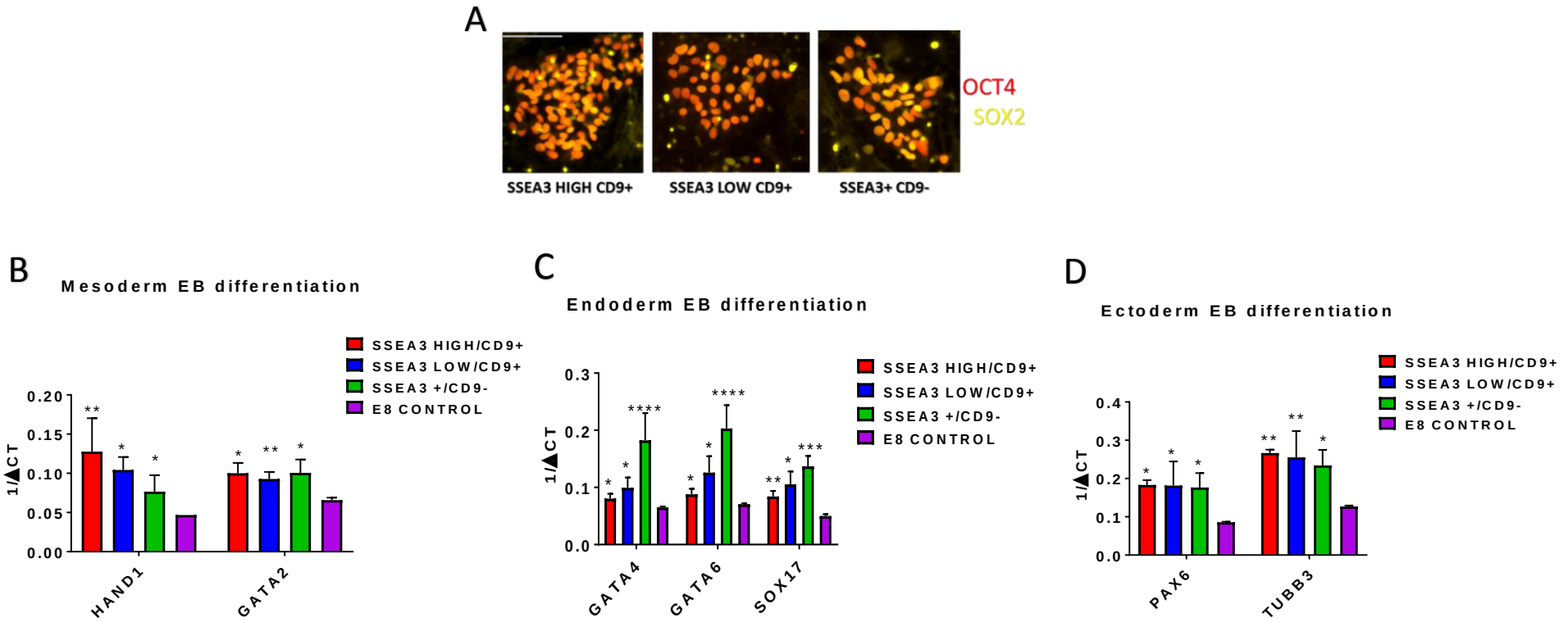
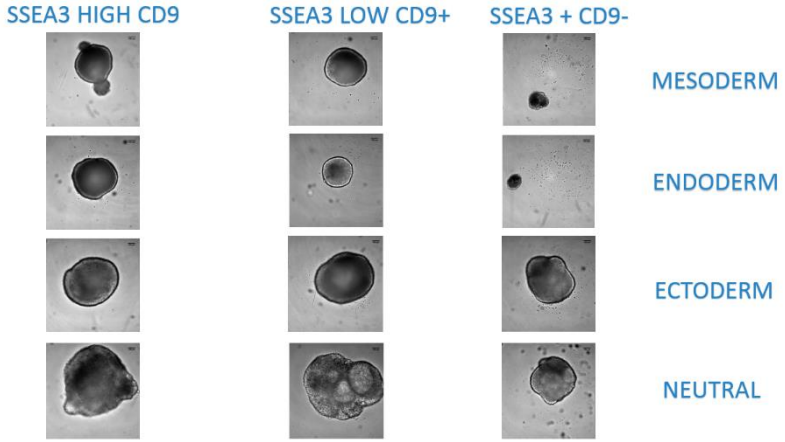


Figure 5: A comparison of size of EB in induced differentiation conditions reveals different propensities for differentiation between the SSEA3 and CD9 subsets

Analysis of size of EB grown in mesoderm (A), endoderm (B) and ectoderm (C) differentiation conditions. Length of EB given in µm. SSEA3+/CD9- subset grew smaller EB in mesoderm and endoderm conditions. Average from 8-10 EBs grown from HES3, H9 and NCRM1 lines. EB imaged at x20 magnification. EB's analysed using Fiji, ImageJ software. Statistical significance shown by asterisks (*) $p<0.05$.



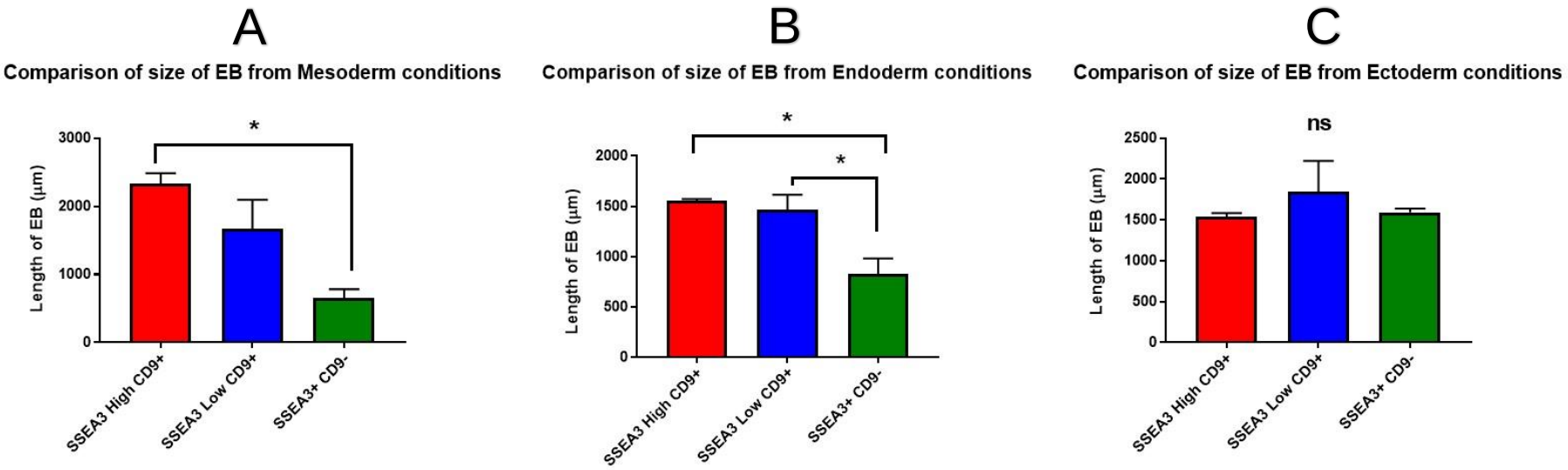


Figure 6: Embryoid body Neutral differentiation of SSEA3/CD9 sorted subsets demonstrates different propensities for differentiation

Comparison of the gene expression of *CD34, HAND1, GATA4, GATA6, SOX17, TBX6, PAX6, SOX1, PAX3* and *SOX10* from Neutral EB differentiation conditions comparing the three subsets: SSEA3 High CD9+, SSEA3 Low CD9+ and SSEA3+ CD9-. Gene expression depicted as 1/∆Ct for three different cell lines: NCRM1 (A), H9 (B) and HES3 (C). SSEA3 High CD9+ subset exhibited increased expression of *CD34* and *HAND1* suggesting a propensity for mesoderm differentiation. SSEA3+ CD9- subset exhibited increased expression of *PAX6, PAX3* and *SOX10* suggesting a propensity for neural ectoderm and neural crest differentiation. Statistical significance shown by asterisks (*) p<0.05

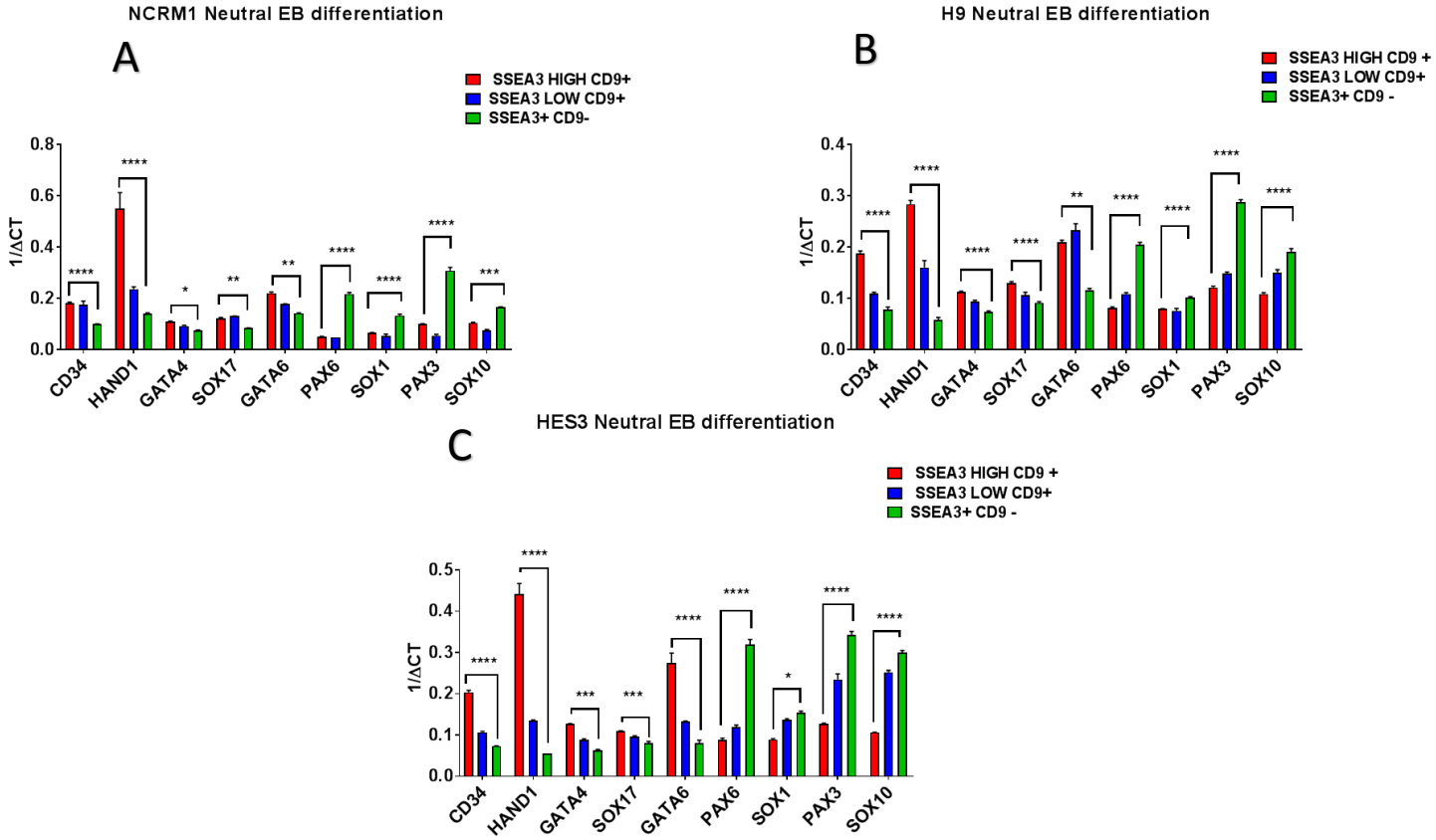


Figure 7: Single cell QPCR heatmap showing hierarchical clustering of SSEA3/CD9 subsets

Single Cell QPCR analysis of 48 genes on SSEA3/CD9 subsets. Hierarchical clustering applied to cluster single cells together. The heatmap exhibits the clustering of individual cells that have similar expression patterns for the respective genes. The map highlights the considerable heterogeneity within the respective subsets as cells from the same sorted subset are clustered with cells from other subsets. Hierarchical gene clustering performed using GENESIS software.

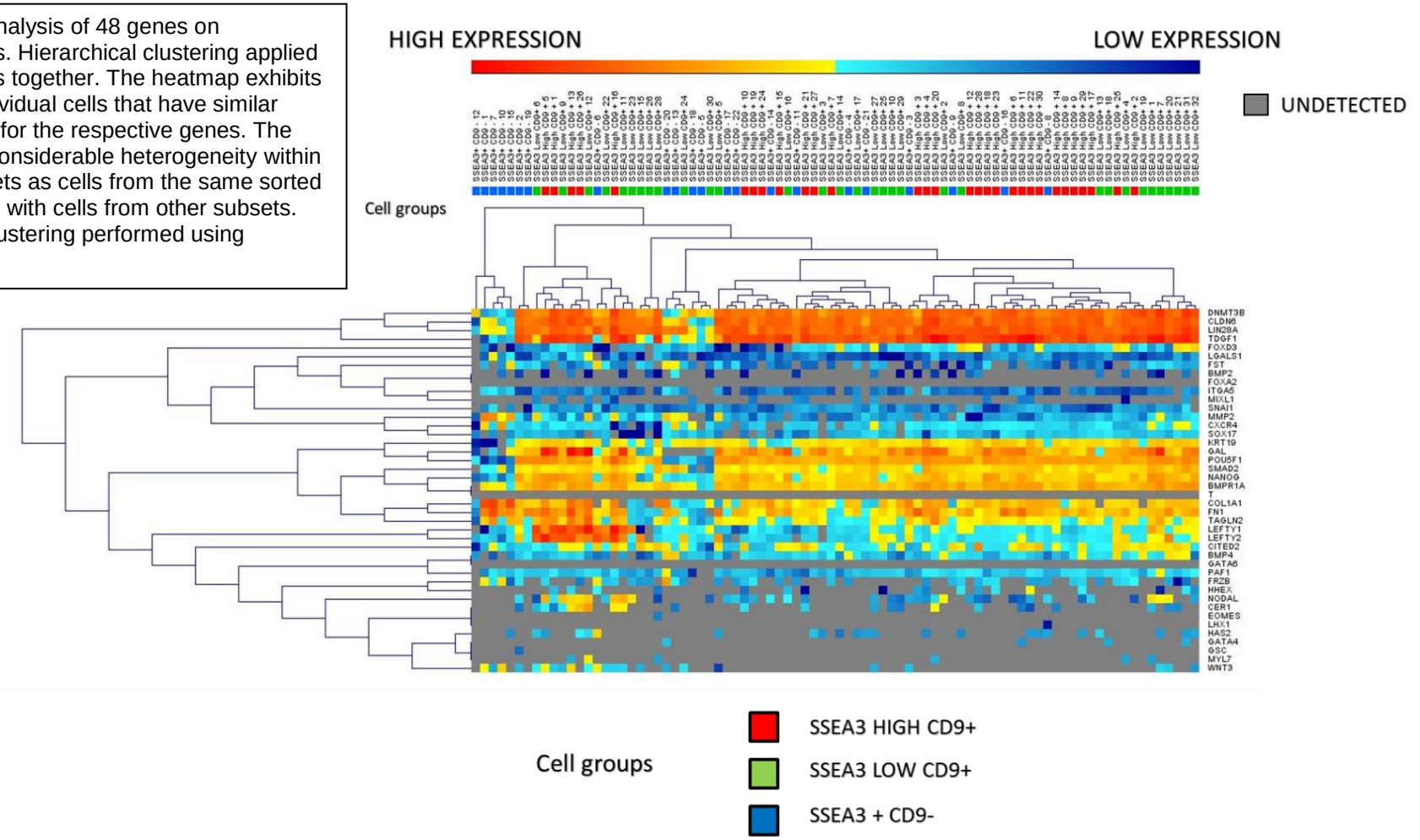


Figure 8: Single cell gene expression analysis of SSEA3/CD9 sorted subsets part 1

Bean-plots comparing expression levels for the respective genes chosen for QPCR within the sorted subsets: SSEA3 High CD9+, SSEA3 Low CD9+ and SSEA3+ CD9-. The plots demonstrate the heterogeneity present within the SSEA3/CD9 sorted subsets with high and low expression of the genes in this panel seen within each subset. Plots are depicted with high expression at the top and low expression at the bottom.

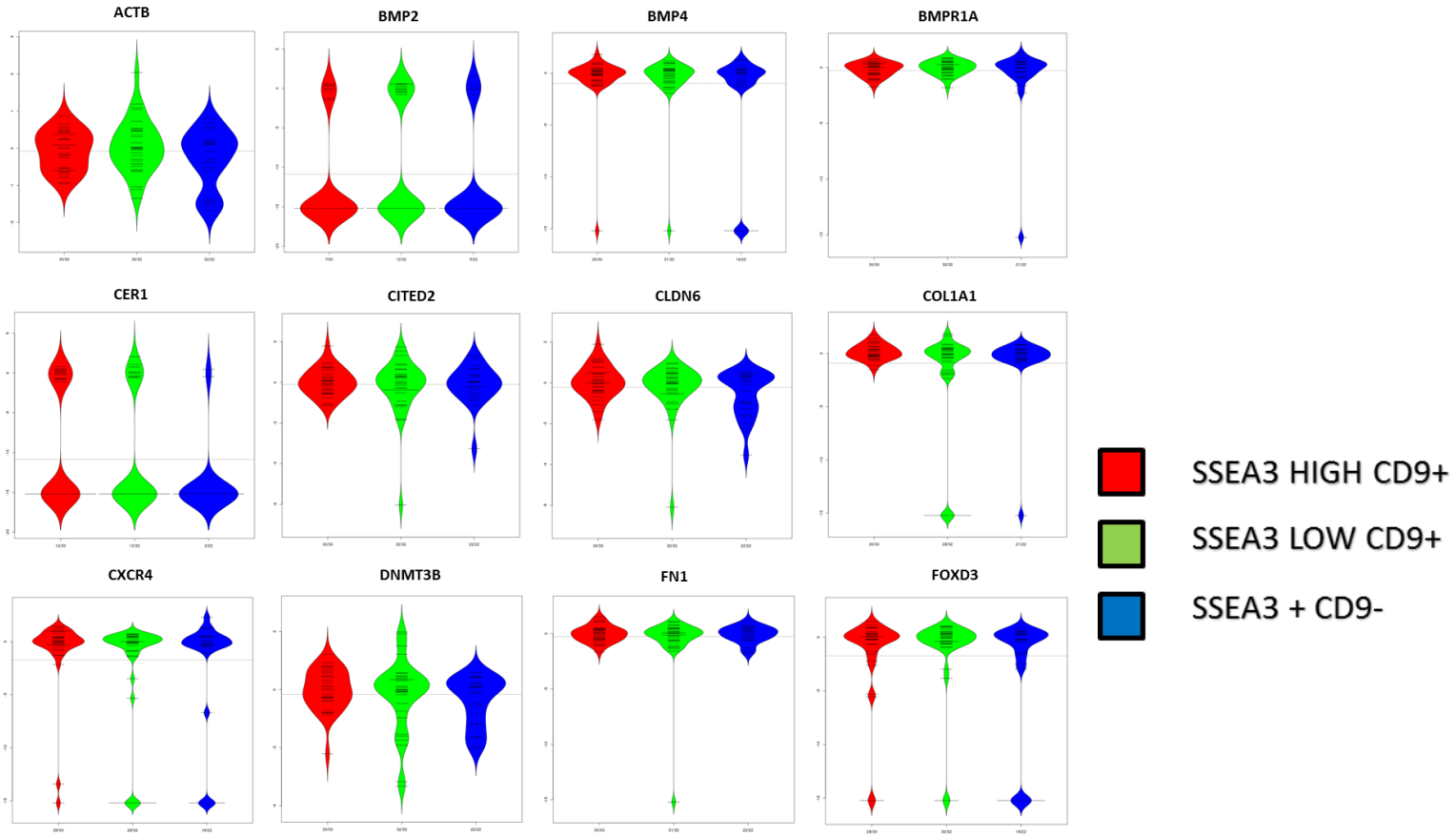


Figure 9: Single cell gene expression analysis of SSEA3/CD9 sorted subsets part 2

Bean-plots comparing expression levels for the respective genes chosen for QPCR within the sorted subsets: SSEA3 High CD9+, SSEA3 Low CD9+ and SSEA3+ CD9-. The plots demonstrate the heterogeneity present within the SSEA3/CD9 sorted subsets with high and low expression of the genes in this panel seen within each subset. Plots are depicted with high expression at the top and low expression at the bottom.

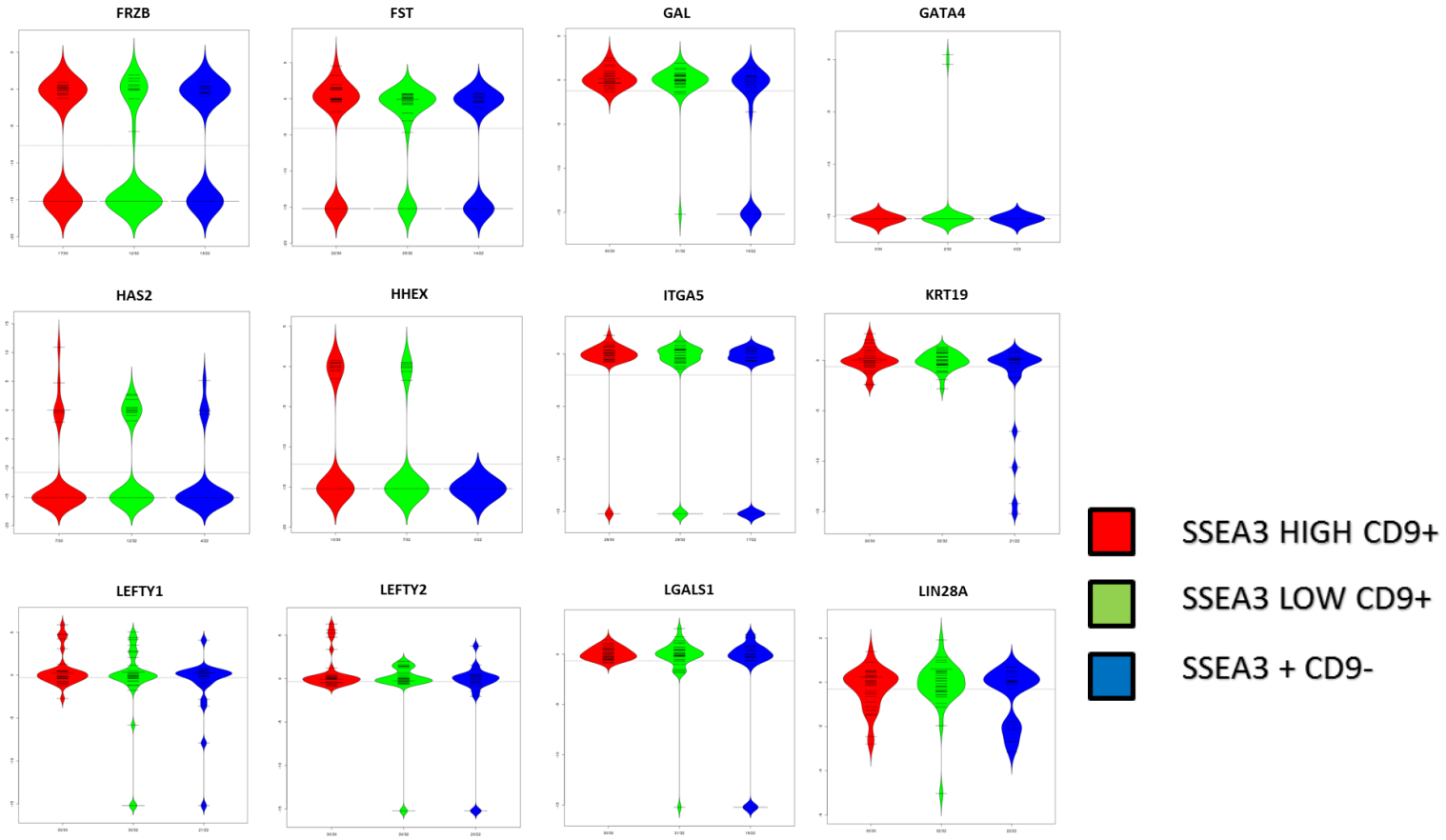


Figure 10: Single cell gene expression analysis of SSEA3/CD9 sorted subsets part 3

Bean-plots comparing expression levels for the respective genes chosen for QPCR within the sorted subsets: SSEA3 High CD9+, SSEA3 Low CD9+ and SSEA3+ CD9-. The plots demonstrate the heterogeneity present within the SSEA3/CD9 sorted subsets with high and low expression of the genes in this panel seen within each subset. Plots are depicted with high expression at the top and low expression at the bottom.

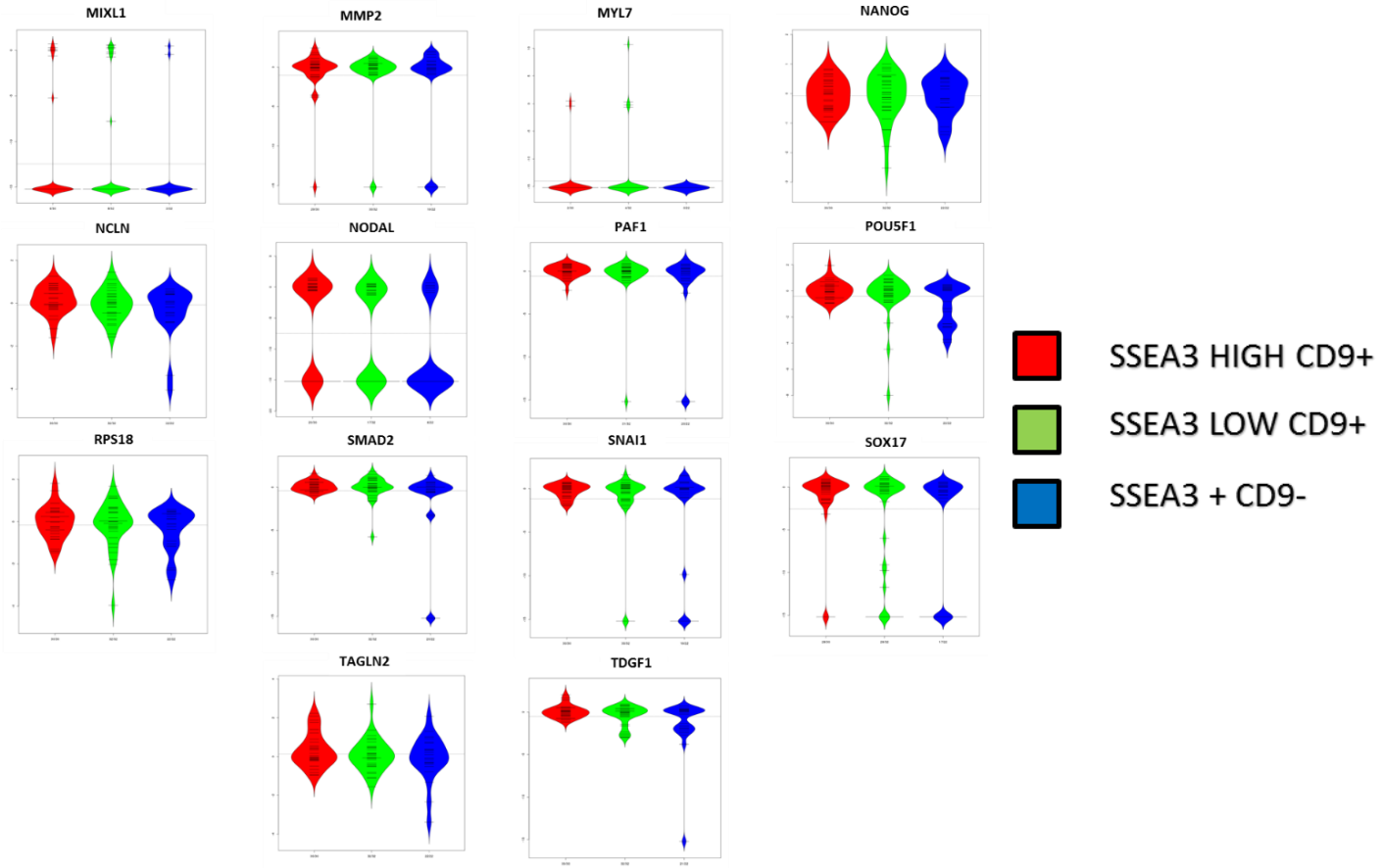


Figure 11: Single cell QPCR PCA analysis demonstrating the relationship between SSEA3/CD9 sorted subsets.

PCA analysis based on complete gene clustering of the SSEA3/CD9 subsets. The SSEA3/CD9 subsets exhibit considerable heterogeneity with scattering in each of the three PCA plots, however the SSEA3 High/CD9+ subset exhibited closer clustering to other cells from that cluster and secondly to the SSEA3 Low/CD9+ subset.

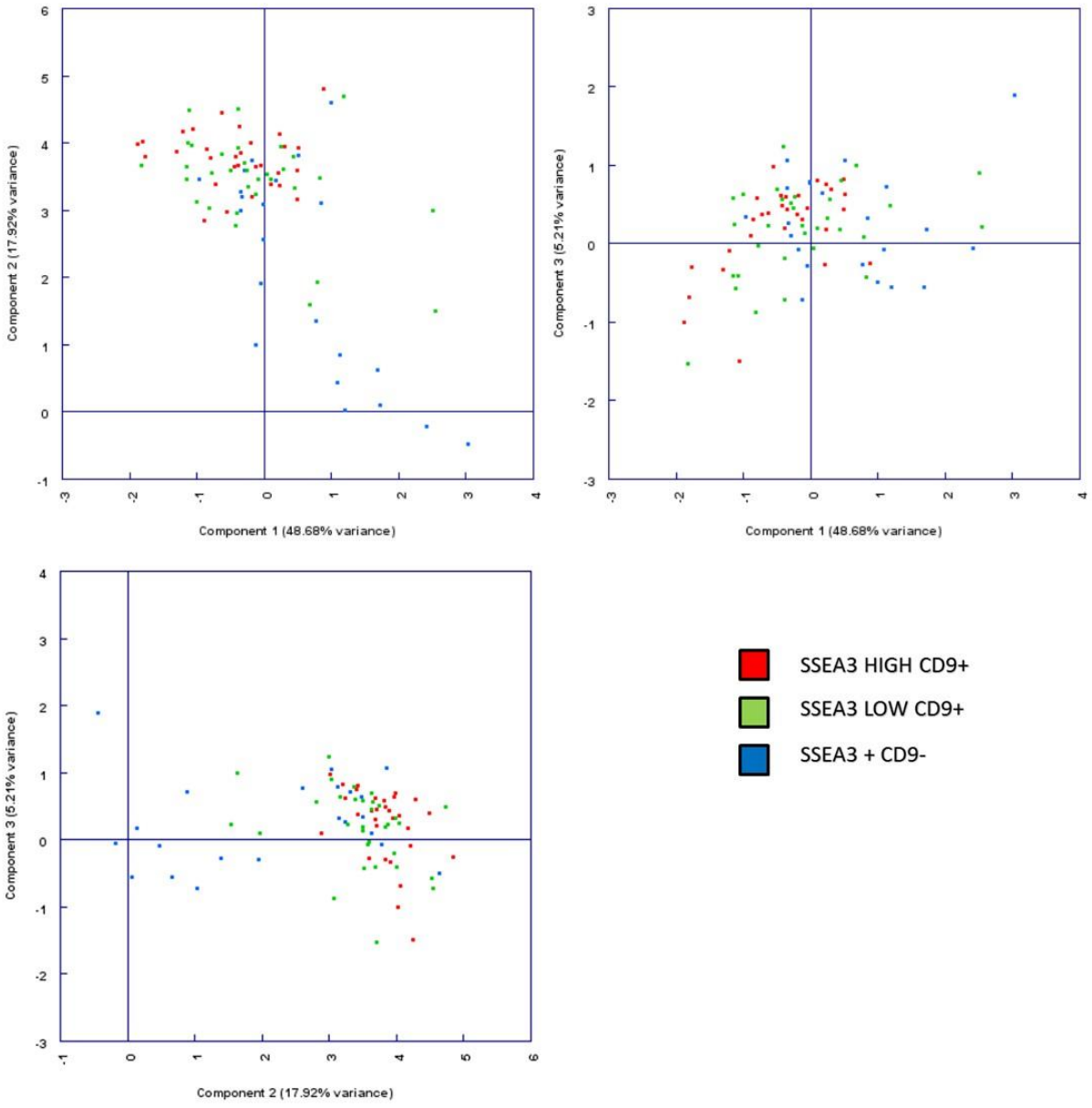


Figure 12: Comparison of Single cell QPCR Analysis of SSEA3/CD9 and SSEA3/MIXL1 subsets reveal the *MIXL1*+ subsets have higher expression of mesoderm/endoderm genes compared to the SSEA3/CD9 subsets

Comparison table of averaged Ct values for selected mesoderm and endoderm genes from SSEA3/CD9 and SSEA3/*MIXL1* single cell QPCR. The *MIXL1*+ subsets have increased expression of mesoderm and endoderm genes compared to the other subsets.

	SSEA3 HIGH CD9+	**SSEA3 LOW CD9+**	**SSEA3+ CD9-**	**SSEA3+ *MIXL1*-**	**SSEA3+ *MIXL1*+**	**SSEA3- *MIXL1*+**
NANOG	11.06	11.59	12.96	11.91	10.12	13.86
POU5F1	10.25	10.59	12.70	10.89	9.49	14.09
BMP2	35.25	31.99	34.88	38.19	25.72	11.68
BMP4	14.95	15.40	17.96	17.58	19.47	17.11
CER1	29.72	32.06	37.88	34.69	12.20	12.53
EOMES	40.00	39.26	40.00	38.85	20.64	16.91
FOXA2	40.00	40.00	40.00	39.37	21.91	15.32
GATA4	39.20	38.45	38.85	37.52	18.65	8.89
GATA6	40.00	40.00	40.00	39.46	22.76	12.18
GSC	40.00	40.00	38.95	40.00	35.61	31.45
HHEX	31.73	34.96	38.78	35.06	31.54	32.89
MIXL1	35.36	34.14	37.83	35.84	18.45	23.65
T	40.00	40.00	40.00	38.74	24.11	33.53
WNT3	36.42	30.49	31.67	40.00	22.74	25.74

Figure 13: Strategy for isolating subsets of hPSC through separation of antigen expression into HIGH/LOW/Negative (-ve)

Use of BF4 antibody to identify and assess lineage bias in hPSC. A) Representative BF4 flow cytometry histogram plot showing how antigen expression was separated into HIGH/MID/LOW/Negative (-ve) for cell sorting. B) Example of reanalysis of cell post sort, showing the purity of the sort. C) Examples of BF4 expression on hPSC in self-renewal conditions: H9 (Left), HES3 (centre), NCRM1 (right). Positive expression= +ve

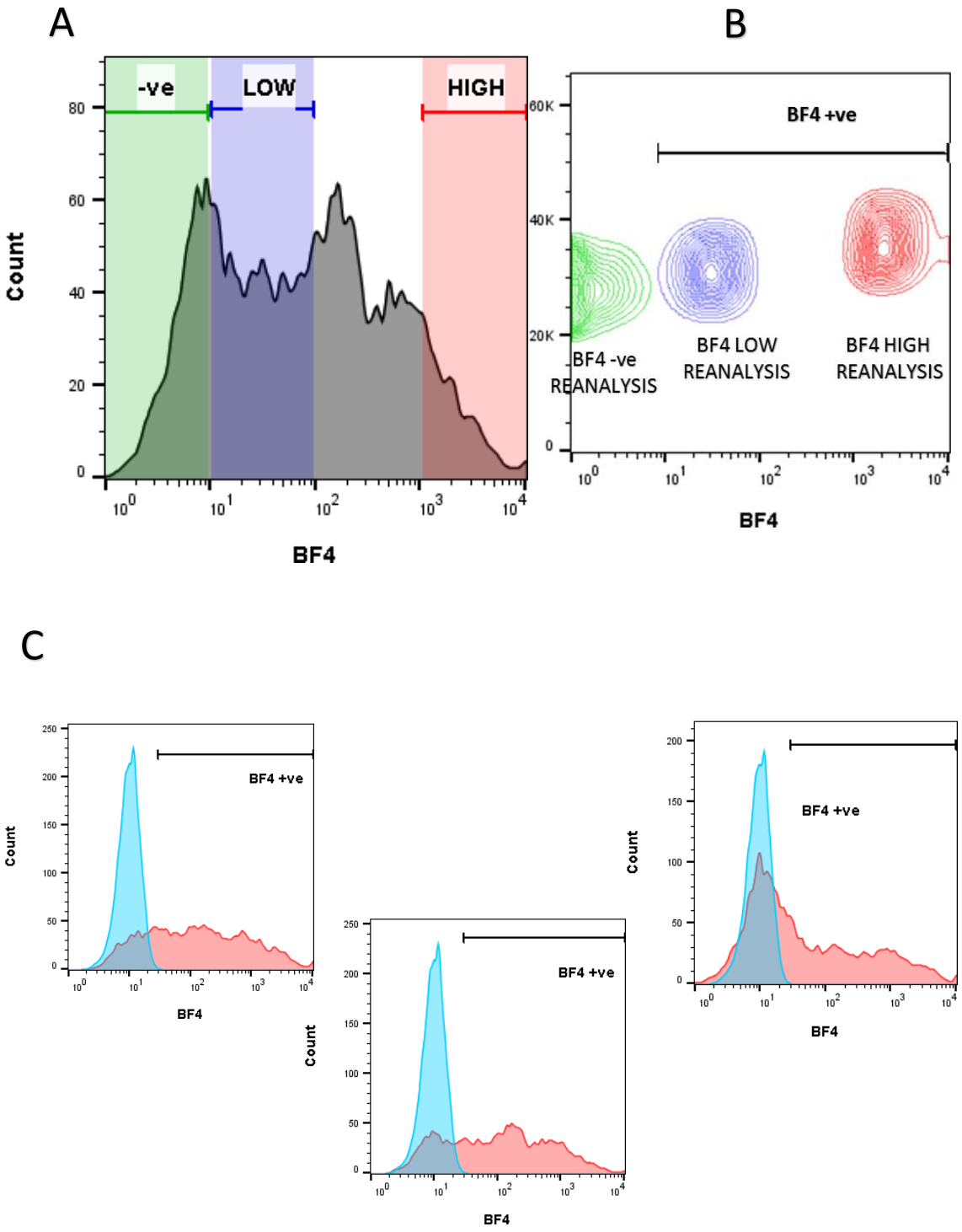


Figure 14: Comparison of BF4 sorted fractions reveal they contain hPSC and exhibit differential expression of differentiation associated genes

Transcriptome analysis to assess whether there was a difference in the BF4 sorted subsets. A) Comparison of the gene expression of *OCT4, NANOG, ZFP42, GATA6, SOX17, MIXL1, BRACHYURY, CER1, DKK1, PAX6, SOX1, PAX3 and SOX10* from H9, HES3, and NCRM1 cell lines comparing the three BF4 subsets: BF4 High, BF4 Low and BF4 Negative (-ve). Gene expression depicted as 1/ΔCt on 3 separate cell lines. BF4 Low and BF4-ve subsets exhibited increased gene expression of mesoderm, endoderm and ectoderm genes compared to BF4 High. B) Example of BF4 sorted subset cells seeded and forming colonies in KOSR/MEF self-renewal conditions. Colonies contain cells co-expressing pluripotency associated markers OCT4 and SOX2. Scale bar set at 50µm. Statistical significance shown by asterisks (*) p<0.05 ns= not significant

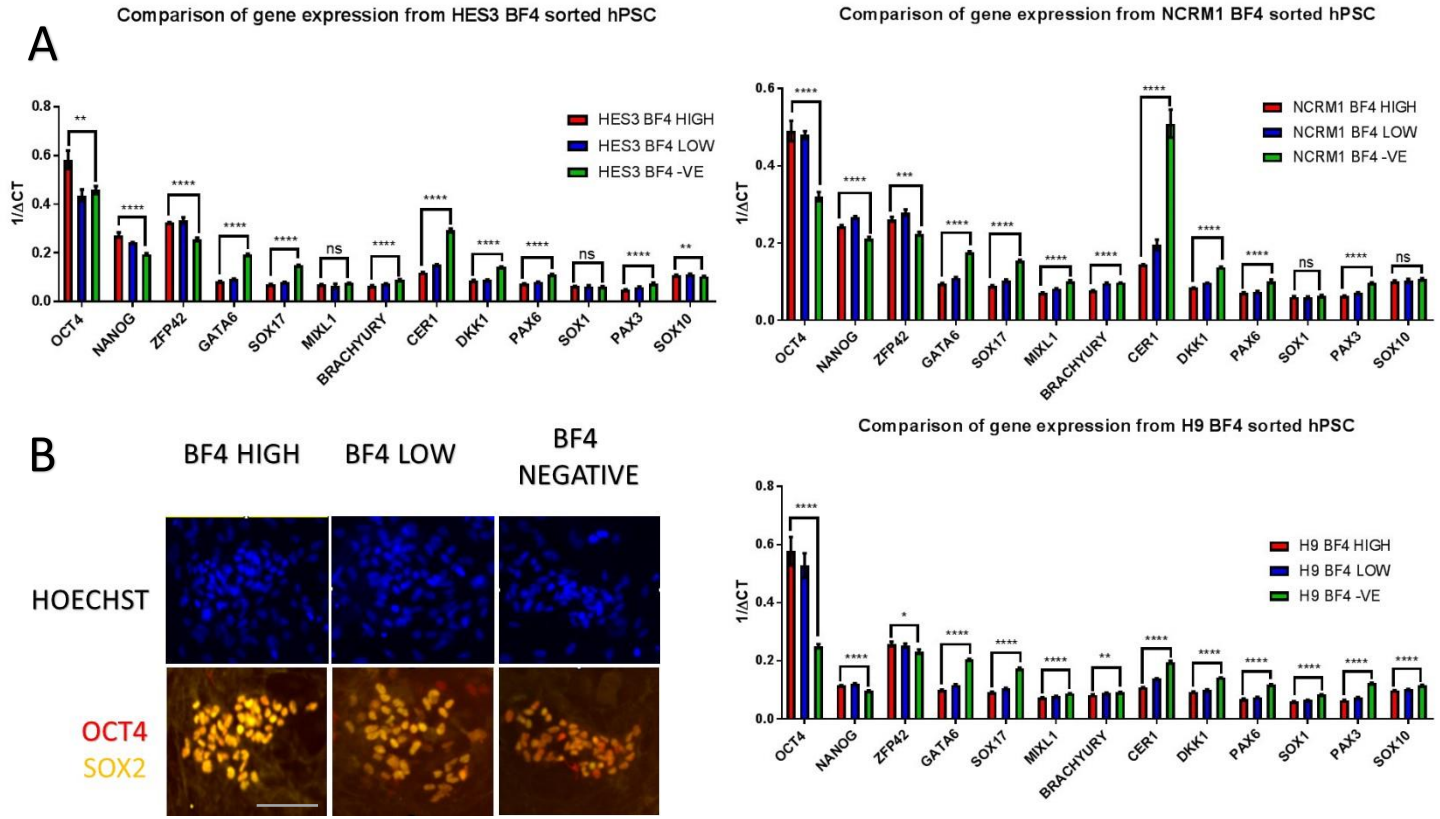


Figure 15: Two colour fluorescence analysis with BF4 and the gene reporters reveal differential expression compared to CD9 and SSEA3

Flow cytometry plots showing BF4 expression against *BRACHYURY* (A), *MSGN1* (B), *GATA6* (C) or *MIXL1* (D) in KOSR/MEF self-renewal conditions. In contrast to SSEA3 and CD9 respectively, *BRACHYURY* expression was spread across BF4 High/Low/Negative with a large cluster in BF4-ve range. Less than 1% MSGN1 expression was also observed in these conditions. Both *GATA6* and *MIXL1* expressing cells clustered predominantly in BF4-ve range evidencing relationship between antigen expression and lineage associated gene expression is different to CD9 and SSEA3. n=3

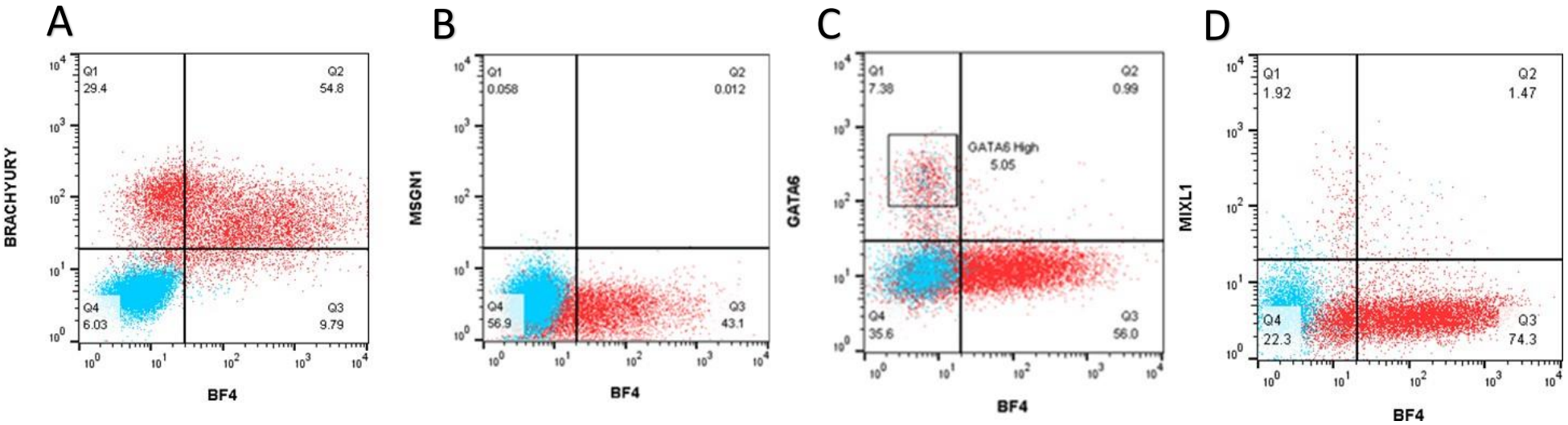


Figure 16: Co-IP of BF4 reveals the epitope is present on the same glycoprotein as TRA160R and TRA181 but is not the TRA160S epitope

Co-IP/Western blot, neuraminidase treatment and mesoderm differentiation to compare BF4, TRA160R, TRA160S and TRA-181. (A) Western blot immunostained with TRA160R showing Positive reactivity for TRA181 and BF4. (B) Mesoderm differentiation comparing BF4 with TRA-160S. E8 control in red compared to the differentiation in orange and the negative control in blue (C) Flow cytometry histogram for BF4, TRA160S and *BRACHYURY* expression in different conditions: for the E8 control in red compared to negative control in blue D) After neuraminidase treatment in red compared to negative control in blue.

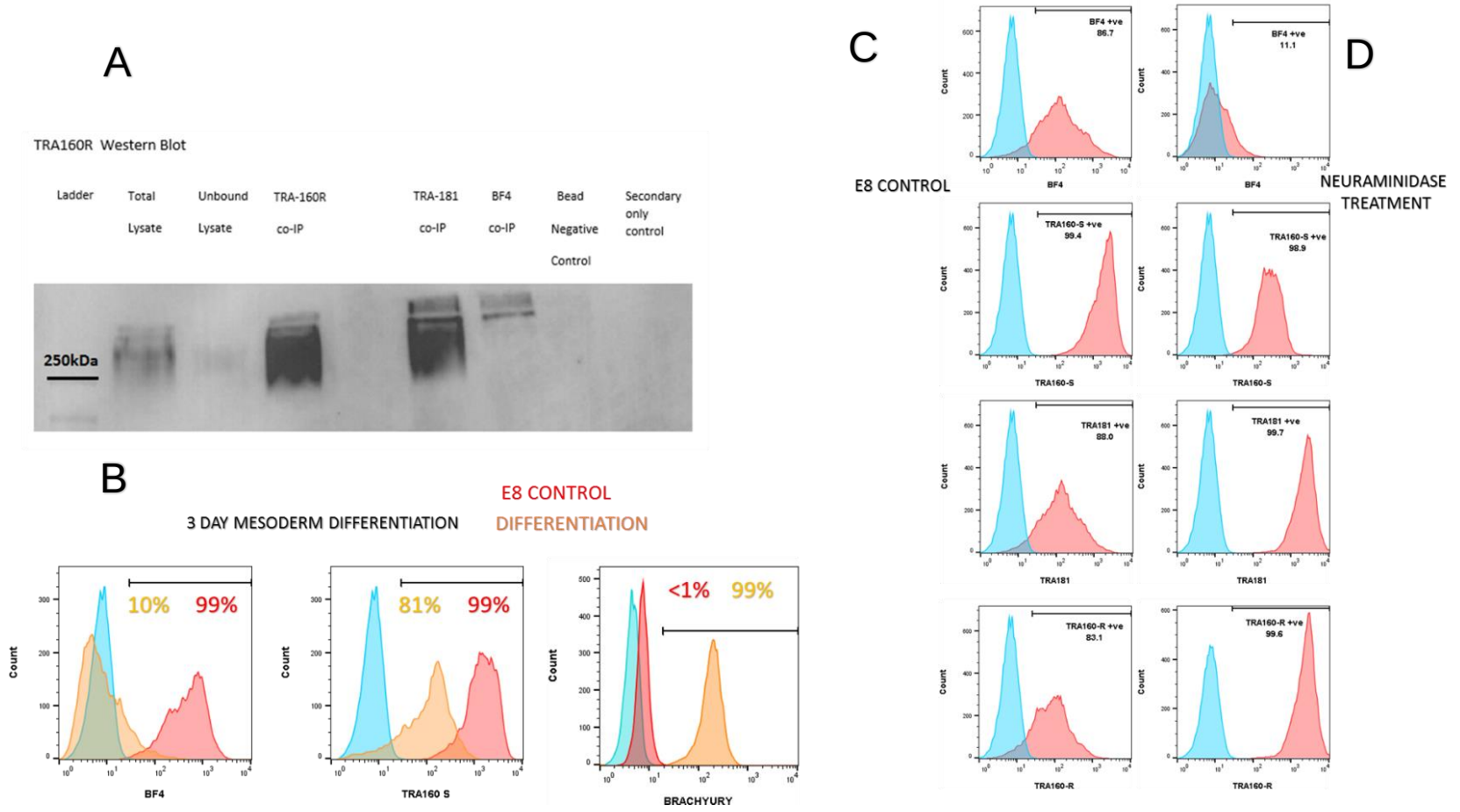
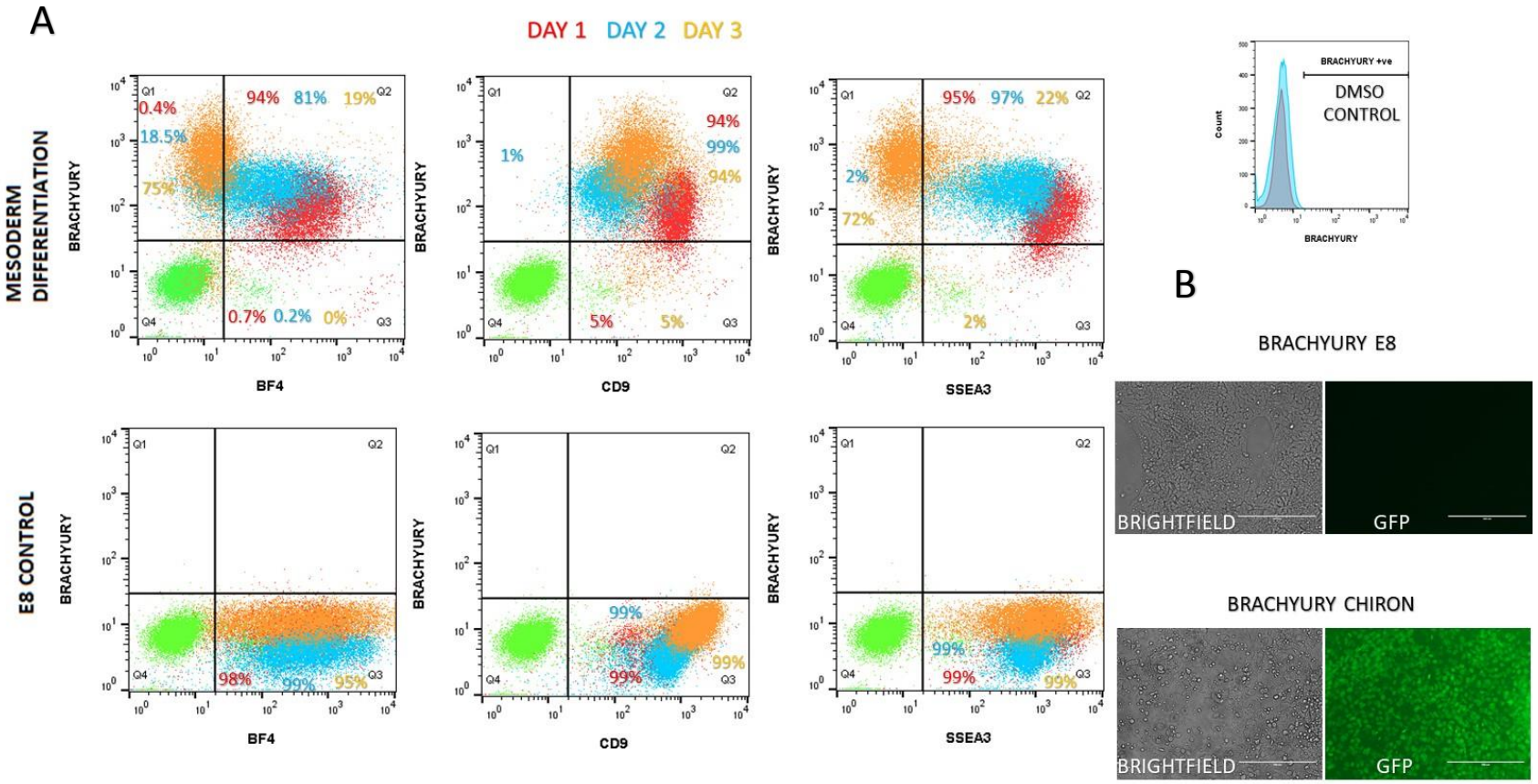


Figure 17: 3-day mesoderm induction assay to assess cell surface antigen expression dynamics during differentiation

3-day mesoderm time course assays to assess the dynamics of expression of BF4, CD9 and SSEA3. The results showed different expression profiles for the respective antigens. (A) Representative flow cytometry plots for BF4, CD9 and SSEA3 against *BRACHYURY* for the Mesoderm differentiation compared to the E8 control. Each day after introduction was compared: day 1 in red, day 2 in blue and day 3 in yellow. (B) Bright field and GFP images verifying *BRACHYURY* expression in mesoderm condition. CD9 Mid expression and loss of BF4 and SSEA3 coincides with strong *BRACHYURY* expression. Scale bar set at 200µm. n=3



3.3 Discussion

The purpose of this set of experiments was to determine whether cell surface antigens could be used to identify a subset of cells with lineage bias to mesoderm. The dynamics of SSEA3 and CD9 in self-renewal conditions in combination with the *BRACHYURY* and *GATA6* reporters (Figure 2) provided evidence that both SSEA3 and CD9 could be used together to identify a lineage-biased subset. Both antigens differed to BF4 (Figure 15) in that they had large clusters of cells co-expressing both antigens and reporter gene in the “High” expressing antigen region (> 10^2 expressing region). However, in respect to other self-renewal media conditions, CD9 was also observed to be strongly expressed in E8/Vitronectin at >10^2 fluorescence intensity in this experimental set-up.

In the case of CD9, strong CD9 expression could be simply attributed to fully differentiated cells just existing within the culture, as CD9 is a marker expressed in a variety of differentiated cells types across different lineages [160–162] as well as undifferentiated hPSC [130,163]. While CD9 has been shown to be expressed on multiple different lineages, this has not been shown with SSEA3. However there are examples of evidence to show it is present in some differentiated cell types such as human skin cells [164] and red blood cells [165]. Though unlike CD9, SSEA3 is lost upon differentiation of hPSC to immediate derivatives as shown by the mesoderm differentiation in Figure 17 and also neural ectoderm [129].

In the context of identifying a mesoderm biased subset of hPSC, the results showed that while there existed a large cluster of SSEA3 “High” cells that co-expressed *BRACHYURY*, SSEA3 also down-regulated significantly in mesoderm inducing conditions (Figure 17). Therefore, considering the SSEA3 High/*BRACHYURY*+ cells were identified in self-renewal conditions, it would not be expected that these would already be differentiated cells. The hypothesis from these observations was that this SSEA3 “High” population contained a subset of hPSC that contained a mesoderm bias. The CD9 antigen was added due to the previous evidence in the literature showing that High/Mid expression levels of this antigen isolated more cells that exhibited propensity to mesoderm/endoderm differentiation [130,131]. A potential caveat associated with KOSR/MEF self-renewal conditions is the density and quality of the feeder layer. Through preliminary work performed with the *MIXL1* reporter, it was determined that a feeder layer of

~500,000 cells per 25cm^2 vessel resulted in a very dense feeder layer. This feeder layer caused the stem cell colonies to become very tightly packed in these conditions, which after approximately 3 days, resulted in a yield of ~3-7% *MIXL1*+ cells in a culture. On low density feeders, the percentage of *MIXL1*+ cells were typically <0-2% in a culture. Considering this observation, all of the reporter lines in this study were analysed on high density feeder layers. However, considering the effect the feeder layer can have on the expression of the gene reporters, the variable expression of *BRACHYURY* in particular could be attributed to batch variation of the feeder cells.

To assess whether SSEA3 “High” cells could be used for isolating a mesoderm biased population, CD9 expression was kept consistent with only a change in expression for SSEA3. The results the EB differentiation showed that the SSEA3 High/CD9+ subsets had greater gene expression of genes associated with lineage specific differentiation to mesoderm. This was observed with expression of *HAND1* and *CD34*, two markers associated with mesoderm differentiation. *HAND1* has been shown to be expressed in lateral and extraembryonic mesoderm [166–168] while *CD34* expression has been associated with a variety of mesoderm derivatives including haematopoietic [169,170] and endothelial [171,172].

While in the literature, CD9 Negative cells had been attributed to more neural ectoderm differentiation [130,131] it had not been shown whether CD9 Negative cells attributed to other ectodermal derivatives, such as neural crest. With the Neutral EB differentiation (Figure 6) both neural ectoderm and neural crest associated genes were upregulated with *PAX6* [173,174] and *TUBB3* [175,176], and *PAX3* and *SOX10* [177–179] respectively in the SSEA3+/CD9- subset. The SSEA3+/CD9- cells were able to form EB in all three inducing conditions, but they grew considerably smaller EB in the mesoderm and endoderm inducing conditions (Figure 5 A-B).

One possibility is that this subset already contained differentiated cells that were put into growth conditions not compatible for their maintenance and consequently died, leading to smaller EB. Another possibility is that these cells were strongly biased to differentiate to an ectoderm cell fate but died when pushed to go to a mesoderm or endoderm fate. However, when these small EBs were analysed for gene expression (Figure 4 B-D), they showed significantly greater expression for genes associated with endoderm, namely *GATA4*, *GATA6* and *SOX17* [180–182]. An

explanation for this result, is that the smaller EBs contained more cells that had specific differentiation to endoderm. The larger EBs formed from the other two subsets could have had more nonspecific differentiation which explained the lower expression of the endoderm genes.

Another explanation is that the SSEA3+/CD9 subset contained different subsets within it, with bias to different lineages (such as endoderm) and that these are the ones that survive in this inducing condition. Evidencing this is difficult due to the nature of population transcriptome studies, in that few cells greater expressing a gene of interest can significantly skew results representing the entire subpopulation. This consideration was a primary driving force in performing single cell QPR on these subsets, to see whether these populations were heterogenous in containing some cells that greater expressed lineage specific genes.

The single cell QPCR results (Figure 8-10) showed that the SSEA3 High/CD9+ subset had more cells expressing mesoderm/endoderm associated genes, while the SSEA3+/CD9- subsets had more cells that did not express these genes. From this result and the Neutral EB differentiation, the evidence suggests that the SSEA3+/CD9-ve subset had more cells going towards other lineages rather than mesoderm/endoderm.

The genes chosen to analyse these subsets were the top genes that showed the most significant differential expression between the SSEA3+/*MIXL1*+ and SSEA3+/*MIXL1*- subsets. With *MIXL1* being a mesendoderm associated gene [183], the lineage specific genes chosen were primarily early mesoderm/endoderm associated. The results reflected this fact with the *MIXL1*+ single cell subsets showing significant upregulation of the early mesoderm/endoderm genes in the panel (Figure 12). Interestingly, in respect to the SSEA3/CD9 subsets, no *BRACHYURY* expression was detected in any cell. This is unexpected given that BRACHYURY-*Venus* expression was detected in self-renewal conditions (Figure 2), it would not have been expected that there would be no detectable *BRACHYURY* expression. As the same reagents and experimental set-up was performed on the SSEA3/*MIXL1* subsets which did contain *BRACHYURY* Positive cells, the explanation likely lies with the expression of the gene itself. A caveat with using the *BRACHYURY* reporter for isolating subsets of cells is the

half-life of the Venus fluorescent protein. The Venus protein is a variant of YFP, which had shown to have an *in vivo* half-life of 12.8 hr in some experimental systems [184]. However, half-life of fluorescent proteins can vary in different experimental set-ups, with some reporting upwards of 24 hr [185]. Therefore, it is possible that cells that were being identified as BRACHYURY Positive, were actually BRACHYURY Negative at the time of analysis. Another explanation is that the expression level of the BRACHYURY gene was too low to be amplified and accurately detected during the PCR reaction. For analysing the single cell QPCR data, Z-scores were used and applied to an auto-scale that standardised the data in a way that the expression levels of genes could be compared in a meaningful manner. When analysing Ct values, a range of expression from High to Low cannot be seen due to the log scale those values are obtained from. By using Z scores, gene expression is presented in a greater range from High to Low expression. This results in a heatmap that visually represents the subtleties in gene expression that is occurring in the cell, such as an increase in *CER1* and a decrease in a pluripotency gene which implies a lineage biased cell to mesoderm/endoderm. This observation would otherwise be missed using Ct values on the log scale, resulting in the genes being coloured the same on the heatmap.

The PCA analysis shown in (Figure 11) provided a representation for the heterogeneity within the SSEA3/CD9 subsets. The PCA plots show a scattering of cells within all of the subsets, as opposed to distinct groups. While this could be attributed to the heterogeneity within the sorted subsets, that some cells are closer in similarity to those from other subsets, this again is subject to the genes used for the correlation analysis. While speculative, having a gene list containing genes associated with other lineages, such as ectoderm, may have given more distinct groups from the PCA analysis. The relevance for this would be for further study into the gene networks that govern these states of cells.

Due to the considerable heterogeneity within these sorted subsets, there is not enough evidence that the gene list chosen contain candidate genes that are distinct to that subset. That is, that the sorted SSEA3/CD9 subsets cannot be separated based on expression of early mesoderm/ mesendoderm associated genes. This becomes problematic when the objective is to identify key candidate gene networks that govern early transition away from pluripotency to lineage

specification. While the evidence shows that the SSEA3 High/CD9+ subset formed more mesoderm derivatives when allowed to differentiate, this propensity for mesoderm differentiation cannot be seen directly from sorting for this population in self-renewal conditions, as shown by the single cell QPCR and PCA analysis. When comparing expression levels to the SSEA3/*MIXL1* subsets, the expression levels of mesoderm/endoderm associated genes were much lower in the SSEA3/CD9 subsets. However, both SSEA3/CD9 and SSEA3/MIXL1 subsets had shown to contain pluripotent cells. Therefore, in a future direction to gain better insight into heterogeneity with hPSC cultures, single cell RNA seq is required. However, within this study, the single cell QPCR result contributed to the decision to look at other cell surface markers, namely BF4.

The QPCR results of cells from self-renewal conditions, demonstrated that in all three cell lines, BF4 Low and Negative subsets had upregulation of genes associated to differentiation to different lineages such as *CER1* and *DKK1* for mesendoderm/endoderm and *PAX6* for ectoderm (Figure 14). This suggested that the BF4 Low contained a population of cells further differentiated compared to the BF4 High subset, while the BF4 Negative subset likely contains more fully differentiated cells. The evidence of this can be seen in the differences in cloning efficiency between the sorted subsets with BF4 Negative cells have reduced cloning efficiency and also the increased gene expression of differentiation associated genes in the BF4 Low and Negative subsets (Figure 14). Additional evidence is seen in the 3-day mesoderm time course (Figure 17) and the clustering of *GATA6* High cells and *MIXL1* Positive cells in self-renewal conditions (Figure 15). Though it should be noted that expression of pluripotency associated genes (*OCT4*, *NANOG* and *ZFP42*) can be seen in BF4 High, Low and Negative populations, albeit at varying levels. The BF4 Negative subset was shown to have lower expression of these genes. This suggests that these BF4 subsets contain within them a population of hPSC. However, as the objective of this study is to identify subsets of hPSC that are biased to lineage specific differentiation, BF4 did not appear a useful marker in this context.

3.4 Conclusion

The results presented in this chapter highlight how previously undefined antigens together can be used to demonstrate heterogeneity within hPSC through isolation and analysis via the transcriptome and functional assays to induce differentiation. The results from the functional analysis of the SSEA3/CD9 subsets correlated with previous results using *GATA6* and *MIXL1* (Andrews lab, unpublished), that SSEA3+ sorted cells contain heterogenous subsets with different propensities for differentiation. It had previously been shown that SSEA3+/*GATA6*+ and SSEA3+/*MIXL1*+ cells exhibited increased expression of genes associated with the endoderm and mesendoderm respectively. It was therefore considered that the SSEA3+/*BRACHYURY*+ cells that were observed in self-renewal conditions, may exhibit a more homogeneous propensity to mesoderm differentiation. Similarly, to the SSEA3+/*MIXL1*+ subset, the SSEA3+/*BRACHYURY*+ subset could provide another subset of cells that have propensity to mesoderm differentiation that could be isolated and analysed to expand on our knowledge of early fate determination of hPSC.

Chapter 4 : Use of the H9 *BRACHYURY* Reporter to Assess Whether SSEA3+/*BRACHYURY*+ cells Exhibit a Mesoderm Bias

4.1 Introduction

4.1.1 Overview

The results in chapter 3 concluded that being able to identify undifferentiated hPSC that exhibit a lineage bias to differentiation is technically difficult and was unsuccessful with using antigens alone. A previously successful approach to identify lineage biased cells, was to use a combination of antigen and gene reporter to identify and functionally test, putative subsets of cells for lineage bias. The mesoderm-associated marker *BRACHYURY* had shown to be expressed in a subset of cells in self-renewal conditions (Figure 2 and Figure 15), therefore was a candidate for identifying a subset of pluripotent cells with a lineage bias to mesoderm.

4.1.2 BRACHYURY expression in development and its role in lineage specification both *in vivo* and *in vitro*

BRACHYURY is a member of the highly conserved T-Box group of nuclear transcription factors [186] where the mechanism involves binding to the palindromic T site activating transcription [187]. It is a transcription factor that has a long-recorded history in developmental studies, where the first reported evidence of BRACHYURY was described as a mutation in the T (*BRACHYURY*) locus [188]. The name BRACHYURY itself means “short tail” and is derived from the observation of shortened tails in heterozygous mice that had mutations of the T (*BRACHYURY*) locus [188]. Following its discovery as a transcription factor discovery of conserved genes with similar sequence homology to T (*BRACHYURY*) were found in both mice and drosophila, resulting in the coined-term “T-box genes” to describe this group [189].

Within development, BRACHYURY has been shown to be expressed during early-mesoderm differentiation and plays an important role in early development

in respect to formation of the primitive streak and notochord. Mice that are homozygous for *BRACHYURY* mutations die mid-gestation [190]. Through mouse mutations in *BRACHYURY*, it has also been shown that BRACHYURY serves a role in movement of mesodermal cells during formation of the primitive streak. Mice with these mutations result in primitive streaks with impaired elongation of the body axis posterior to the forelimbs [188,190]. It has also been shown that mice with homozygous mutations for *BRACHYURY* result in severe axial-skeletal effects due to paraxial mesoderm apoptosis [191–193].

In *BRACHYURY* homozygous embryos, while the notochord process is initiated, a trunk notochord fails to form. This results in phenotypic observations of abnormalities in the spinal cord and somites [194,195] as well as a defective allantois and a disrupted primitive streak and node [196]. Additionally, in mouse chimeras, ES cells with homozygous mutations for BRACHYURY were introduced by blastocyst injection into wildtype host embryos. The presence of these cells in the allantois resulted in a failure to form the correct placental connection, resulting in arrest later in development. In chimeras where these cells were predominantly in the tail, it resulted in severe abnormalities including foreshortening, branching and haemorrhagic cavities[197].

The role of BRACHYURY in the notochord can also be observed in the formation of chordomas, a rare type of cancer that originates from the notochord. BRACHYURY has been shown to be highly expressed within the majority of cells within chordoma tumours [198]. Furthermore, to BRACHYURY's role with the primitive streak and notochord, mice with homozygous and heterozygous mutation with *BRACHYURY* have shown to have other phenotypic defects such as left-right patterning defects and abnormalities within heart development [195,199].

BRACHYURY has also shown to have a role in primordial germ cell (PGC) formation. This is the result of a precursor population in the epiblast that gives rise to the mature spermatozoa and ova of adults [200]. Indeed, it was shown that BRACHYURY was essential for *BLIMP1* and *PRDM14*, two germline determinants. *BRACHYURY* binds to distinct regulatory elements of these genes, directly upregulating them [201]. This upregulation has shown to be perturbed by a loss of BMP4 signalling. This relationship between BRACHYURY and BMP4 signalling is also evident in axolotl embryos, where induction of BRACHYURY

and BMP4 signalling is observed in formation of PGC [202]. BRACHYURY has also been shown to be expressed in germ cell tumours, including testicular seminomas [203].

For lineage specification *in vitro*, BRACHYURY has served to play an important role in early human and mouse ES cell differentiation and is observed to be both a mesoderm and mesendoderm marker [204,205]. The haemangioblast, a common progenitor that generates haematopoietic and vascular cells, is a mesoderm subset that co-expresses BRACHYURY and FLK-1 [206]. BRACHYURY has been shown to physically interact with the mesendodermal marker MIXL1, resulting in overlapping expression during ES cell differentiation and regulation of transactivation of the promoter of the primitive streak marker *GOOSECOID* (GSC) [207]. The relationship between *MIXL1* and *BRACHYURY* is further highlighted in the xenopus embryo, where the *BRACHYURY* homolog *XBRA* represses *MIXL1* expression and *MIXL1* represses *XBRA* expression [208]. Additionally, RNAi mediated knock-out of *MIXL1* was shown to enhance *BRACHYURY* expression, where-as over expression of *MIXL1* resulted in suppressed *BRACHYURY* expression [209]. Expression of MIXL1 was observed through the emerging mesoderm in mid to late-streak embryos, though becoming restricted to the posterior primitive streak at the head-fold stage [13]. The role of BRACHYURY in regulating endodermal differentiation is complex, as it has been demonstrated that both knockout and upregulation of BRACHYURY can result in upregulation of both *GSC* and *SOX17* in human ES cells [210]. Furthermore, BRACHYURY has also shown to mark mesoderm and endoderm formation. BRACHYURY Positive cells have shown to have the capacity to differentiate to both mesoderm and endodermal lineages as shown by formation of cardiomyocytes and hepatocytes respectively [211–213]. However, it was shown that in partnership with eomesodermin (EOMES) and activin-SMAD2/3 signalling, that BRACHYURY is not necessary to activate endodermal gene expression in human ES cells [214].

BRACHYURY has also been shown to have a role in specifying mesodermal fates within the lineage. Within neuromesodermal progenitors, a multi-potent cell type that can give rise to derivatives from both the mesoderm and ectodermal lineages [215], BRACHYURY was shown to specify mesodermal fates through co-operation with Wnt signalling [215]. In respect to Wnt/β-Catenin signalling which

induction promotes mesoderm differentiation [98] two TCF binding sites were located within a BRACHYURY promotor fragment that bind to the LEF-1/β-catenin complex. This was shown to specifically respond to β-catenin dependent transactivation [216]. It was also shown in mouse ES cells *in vitro* that had differentiated to a primitive-streak like state, that both the epithelial to mesenchymal transition (EMT) and transient BRACHYURY expression was observed during this differentiation. Similarly to *in vivo* development, this differentiation was regulated by BRACHYURY expression that was dependent on the EMT and β-catenin activity [217].

Aside from Wnt signalling, BRACHYURY has also been shown to interact with downstream effectors of Activin and BMP4 signalling networks. Specifically, this has been shown in BMP induced SMAD1 in mesoderm progenitors and Activin induced SMAD2/3 in endoderm progenitors [214]. Furthermore, co-operation between BMP4-SMAD1 signalling and BRACHYURY was shown to be a requirement for mesodermal differentiation while simultaneously suppressing endodermal differentiation [214]. Another role in BMP signalling was demonstrated by BMP co-operating with FGF2 to promote mesoderm differentiation and resulting in cells that had high levels of BRACHYURY and CDX2 [218], the latter a marker that is required for somitic differentiation [219].

In respect to undifferentiated cells, it was shown in the pluripotent Embryonal Carcinoma line NTERA2, that BRACHYURY protein was detectable in the undifferentiated cells [220]. In mouse EpiSC, BRACHYURY was shown to co-express with OCT4 in some cell lines and negatively correlated with these cells ability to form ectoderm. These BRACHYURY expressing EpiSC could also not be converted to a mESC state, suggesting these cells could be in a lineage primed state [221].Additionally, it was shown by (Tsakiridis et al., 2014) that primitive-streak like EpiSC exhibit a lineage bias to mesoderm and endodermal fates, while maintaining pluripotency. Here, heterogeneity was observed in respect to *BRACHYURY* expression, with some EpiSC co-expressing *BRACHYURY* with pluripotency associated genes, *OCT4*, *NANOG* and *SOX2* [29]. Altogether, the evidence from the literature demonstrates the essential role of BRACHYURY in mesoderm/mesendodermal differentiation both *in vivo* and *in vitro*. Together it supports the suitably of BRACHYURY to be used as a marker for identifying a mesoderm/mesendodermal biased population of hPSC.

4.1.3 Experimental Objectives: SSEA3 and the *BRACHYURY* reporter as a tool to identify subsets of cells

SSEA3 has been used previously to isolate pluripotent subsets of cells with lineage bias, as shown by cells being SSEA3+/*GATA6*+ and SSEA3+/*MIXL1*+ having increased propensity to endoderm and mesendoderm respectively (Andrews lab, unpublished). SSEA3 and *BRACHYURY* have not be analysed together in this manner, but the evidence in chapter 3 (Figure 2) is that there are cells in self-renewal conditions that are SSEA3+/*BRACHYURY*+ and SSEA3+/*BRACHYURY*-. Using a combination of transcriptome analysis and functional tests, it can be determined whether both subsets contain undifferentiated cells and whether the SSEA3+/*BRACHYURY*+ subset has a lineage bias to mesoderm differentiation.

4.2 Results

4.2.1 Analysis of SSEA3 and *BRACHYURY* in self-renewal conditions

H9 *BRACHYURY* cells were grown in KOSR/MEF self-renewal conditions and analysed via flow cytometry. (Figure 18) is an image of a SSEA3/*BRACHYURY* expression profile that would be separated into the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ sorted subsets. The percentage of SSEA3+/*BRACHYURY*+ cells varied as described in chapter 3 discussion, but was typically in the range of 20->80% (An example of the majority of the population expressing *BRACHYURY* is shown in Figure 2) which was significantly more than what was observed for SSEA3+/*GATA6*+ cells (~8%) and SSEA3+/*MIXL1*+ cells (typically ~0-7%) (Figure 2 and Figure 15). However, while there were more cells expressing *BRACHYURY*, the levels of expression were limited to ~10^2 fluorescence intensity when analysed via Flow cytometry which was lower compared to the *GATA6* (10^3) and *MIXL1* (up to 10^4) in this experimental set-up.

4.2.2 Gene expression analysis of the SSEA3+*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets to assess lineage bias

To determine whether the SSEA3+/*BRACHYURY*+ subset exhibits increased propensity to mesoderm differentiation, SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ cells were FACS sorted from KOSR/MEF conditions and their transcriptomes were analysed on by QPCR(Figure 19-20). The results showed that on average, genes associated with early mesoderm or endoderm differentiation such as *MESP1*, *SOX17*, *BRACHYURY*, *GATA4*, *FOXA2* and *EOMES* were upregulated in the SSEA3+/*BRACHYURY*+ subset compared to the SSEA3+/*BRACHYURY*- subset (Figure 19 and Figure 20). The genes associated with pluripotency, such as *SOX2*, *NANOG*, *POU5F1* and *ZFP42* were shown to be expressed at comparable levels between the two subsets. This indicates that both subsets contained undifferentiated cells, but that the SSEA3+/*BRACHYURY*+ subset had upregulation of early mesoderm/endoderm associated genes, suggesting a lineage biased state.

4.2.3 Single cell cloning of the SSEA3+/*BRACHYURY*+ subset to determine if it contains undifferentiated cells

The results from (Figure 19 and Figure 20) showed that both subsets had cells expressing pluripotency associated genes, but that the SSEA3+/*BRACHYURY*+ subset had upregulation of mesoderm/endoderm genes, suggesting a lineage biased state. However, it is not possible to conclude whether the subset contained a mix of pluripotent and differentiated cells from that analysis. It was therefore necessary to determine whether that subset contained functional, pluripotent cells. To answer this, single cells from the SSEA3+/*BRACHYURY*+ subset from KOSR/MEF conditions were FACS sorted and seeded as single cells in a 96 well plate and left to grow as individual colonies. These colonies were then passaged and assessed functionally to determine if the starting cell was pluripotent. A schematic of this procedure can be seen in (Figure 21).

A total of 24 colonies were obtained and assessed morphologically for evidence of stem cells before being passaged twice into a 24 well plate. Due to the technical limitations of growing and passaging 24 colonies, 12 randomly selected colonies were chosen the following passage and finally reduced to 6 colonies. The colonies that were not selected were immunostained for NANOG and all contained NANOG Positive cells. Four of the colonies that were selected for further passage were from indexed cells from the initial sort and were shown to be within the SSEA3+/*BRACHYURY*+ sorting gate (Figure 222). The other 2 colonies were selected at random.

The following passage, the 6 colonies (now clonal lines) were divided into two growing conditions, KOSR/MEFS and E8/Vitronectin. This was to determine whether the clonal lines grown from a single SSEA3+/*BRACHYURY*+ cell contained cells that were now SSEA3+/*BRACHYURY*-. This was an observation found using the SSEA3+/*MIXL1*+ subset, that the population now contained SSEA3+/*MIXL1*- cells (Andrews lab, unpublished)). The significance of this is the evidence of interconversion of the stem cell population into a different state. Interconversion was observed on both media/matrix combinations the clonal lines were passaged onto upon analysis after 5 passages respectively. The 6 clonal lines on KOSR/MEFS had returned to a heterogeneous population of cells containing both SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ cells

(Figure 23). The starting population of cells that had been originally sorted from had ~40% SSEA3+/*BRACHYURY*+ cells, all 6 clones had less than 20% SSEA3+/*BRACHYURY*+ expression at the time of analysis. There was more variation observed with the SSEA3+/*BRACHYURY*- subset where two of the clones had ~50% of the cells that were SSEA3+/*BRACHYURY*-, while three of the colonies had more than 75% of the cells that were SSEA3+/*BRACHYURY*-.

On E8/Vitronectin, the clonal lines had interconverted predominantly to a SSEA3+/*BRACHYURY*- state as demonstrated by less than 1% *BRACHYURY* expression (irrespective of SSEA3+ expression) and over 95% of the cells being SSEA3+/*BRACHYURY*- (Figure 24). The interconversion of these colonies to a SSEA3+/*BRACHYURY*- state provided evidence that the starting cell was indeed pluripotent. However, to validate the starting cell that created these clonal lines was a pluripotent cell, the clonal lines were immunostained for the pluripotency associated marker NANOG and differentiated into mesoderm, endoderm and ectoderm. All clonal lines exhibited the capability of differentiating to cell derivatives of the three germ layers, as demonstrated by expression of *TBX6* and *MSGN1* (Figure 22 C-D) for paraxial mesoderm, *FOXA2*, *GATA6* and *SOX17* for endoderm (Figure 22 C-D) and *PAX6*, *SOX1* and *TUBB3* for ectoderm (Figure 22 C-D). This was shown in both increased gene expression compared to the undifferentiated control and via immunostaining for TBX6 in mesoderm conditions, SOX17 in endoderm conditions and PAX6 in ectoderm conditions (Figure 22 C). The clones were also shown to express NANOG in self-renewal conditions (Figure 22 C). The results demonstrated that a pluripotent subset of cells existed within the SSEA3+/*BRACHYURY*+ subset.

4.2.4 Assessment of lineage bias of the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ populations through high-content clonogenic assay

The results from (Figure 19 and Figure 20) showed that both subsets had cells expressing pluripotency associated genes, but that the SSEA3+/*BRACHYURY*+ subset had upregulation of mesoderm/endoderm genes, suggesting a lineage biased state. To assess this functionally, single cell derived colonies formed by the SSEA3+/*BRACHYURY*+ and SSEA3+/*BRACHYURY*- subsets were immunostained for lineage associated markers. This was to determine whether the SSEA3+/*BRACHYURY*+ subset generated more mesoderm cells within the colonies formed.

Single cells from both SSEA3+/*BRACHYURY* subsets were seeded and left to form colonies for 5 days. Colonies chosen for analysis were selected based on the size of the colony being no bigger than 64 cells in size, due to the doubling time of a stem cell being ~18 hours[222]. Colonies were scored based on containing at least one OCT4 Positive cell and were analysed based on co expression of lineage associated markers within the colony. A schematic of this procedure can be seen in (Figure 24).

The results showed that both SSEA3+/*BRACHYURY* subsets exhibited heterogeneity in respect to expression of lineage markers (Figure 26 and Figure 27). Both subsets generated colonies that had a variable number of cells expressing the lineage markers of interest. By way of comparison, cells that were SSEA3-/*BRACHYURY*+ were also sorted and seeded at the same density, but which failed to form colonies.

Both SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets generated colonies that predominantly contained NANOG and OCT4 expression within them, providing evidence that the starting cell in forming the colony was an undifferentiated stem cell. In respect to lineage specific markers, both subsets generated colonies that contained cells expressing markers associated with early endoderm (SOX17+/OCT4+) (FOXA2+/*BRACHYURY*-) and early mesoderm (FOXA2+/*BRACHYURY*+) or (TBX6+). The SSEA3+/*BRACHYURY*+ subset however generated more colonies that had greater than 60% of the total number

of cells expressing either (SOX17+/OCT4+) or (FOXA2+/BRACHYURY+) where in contrast the SSEA3+/*BRACHYURY*- subset predominantly generated colonies where less than 50% of the total number of cells expressed those markers. In respect to FOXA2+/*BRACHYURY*- expression, the SSEA3+/*BRACHYURY*+ subset generated more colonies expressing this marker combination, though they were predominantly under 50% of the total cell number. In respect to TBX6 expression, both SSEA3+/*BRACHYURY* subsets generated colonies with a similar number of TBX6+ cells, with the majority of colonies having less than 40% of the cells expressing TBX6.

For the ectoderm associated marker PAX6, both SSEA3+/*BRACHYURY* subsets generated colonies expressing this marker, providing evidence that both subsets could form ectoderm cells. However, both subsets generated significantly fewer colonies expressing PAX6. For the SSEA3+/*BRACHYURY*- subset the total number of cells expressing PAX6 within the colony was fewer than 40%. For the SSEA3+/*BRACHYURY*+ subset there was more colonies generated within this percentile range, but with an addition colony having more than 60% of the cells expressing of PAX6. However, while there were examples where the SSEA3+/*BRACHYURY*+ subset generated colonies that had more cells expressing the mesoderm/endoderm markers as described, the results showed that across all colonies across three biological repeats the mean percentage for both subsets were similar.

For comparison of the SSEA3+/ *BRACHYURY*- vs SSEA3+/ *BRACHYURY*+ subsets, the mean percentages of expression of the markers in cells within colonies were SOX17+/OCT4+ (10.56% vs 9.89%), TBX6 (9.03% vs 11.48%), FOXA2+/*BRACHYURY* (4.44% vs 4.45%), FOXA2+/*BRACHYURY*+ (9.71% vs 8.5%) and PAX6 (0.8% vs 2%). Therefore, there was no evidence that either SSEA3+/*BRACHYURY* subset had greater propensity for more mesoderm/endoderm or mesendoderm differentiation within the colony (unpaired student T Test $p>0.05$).

The results also showed that across three biological replicates there was not a statistically significant difference in respect to total number of colonies containing at least one cell that expressed a lineage specific marker (Figure 29) (Student T

test unpaired p>0.05). Suggesting that both subsets could comparably generate colonies that contained differentiated cells to the respective lineages. This similarity was also observed with the distribution of these lineage markers in either OCT4 Positive or OCT4 Negative- colonies (Figure 28). Except for the FOXA2/*BRACHYURY* analysis, most of the colonies were OCT4 Positive with the lineage markers not expressed at all in the colony for both subsets. Overall the results showed that in this experiment both SSEA3+/*BRACHYURY* subsets were similar in their propensity for differentiation.

4.2.5 Assessment of Lineage Bias of the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ populations through embryoid body differentiation

An alternative approach to functionally assessing whether the SSEA3+/*BRACHYURY*+ subset had a lineage bias to mesoderm was to perform EB differentiation. As evidenced in Chapter 3 (Figure 4-6) this experiment showed that subsets of cells could be sorted and seeded in the differentiation conditions, resulting in different propensities for differentiation. The results showed from the SSEA3+/*BRACHYURY* subsets that while there was not a significant difference between the efficiency of differentiation in all three directed conditions, both subsets exhibited higher levels of expression of *HAND1* and *GATA2* for mesoderm (Figure 30 A), *GATA4*, *GATA6* and *SOX17* for endoderm (Figure 30 B) and *PAX6* and *TUBB3* for ectoderm (Figure 30 C) compared to the undifferentiated control (student T-test p<0.05). Despite both subsets being able to differentiate to derivatives of all three germ layers, there was a significant difference observed in the size of the EBs formed in these conditions (Figure 31 A-C) (student T-test p<0.05). The *SSEA3+/BRACHYURY*+ subset grew larger EBs in the mesoderm and endoderm conditions (Figure 31 A-B), but there was not a significant difference observed in the ectoderm conditions (Figure 31 C).

In Neutral conditions (Figure 30 D), there was variation exhibited between the SSEA3+/*BRACHYURY* subsets. Both subsets showed comparably low expression levels for *TBX6*, *MSGN1* and *MEOX1* compared to the other genes in the panel, suggesting there was not a propensity for paraxial mesoderm

differentiation. For neural ectoderm and neural crest differentiation, the SSEA3+/*BRACHYURY*- subset showed to have similar levels of expression across the three replicates for *PAX6* and *SOX1* for neural ectoderm and *PAX3* and *SOX10* for neural crest. In contrast, the SSEA3+/*BRACHYURY*+ subset exhibited higher variation of expression for these genes across replicates, being higher or lower comparable to the other genes in the panel. This resulted in high standard deviation across the three replicates, with expression either lower or comparable to the SSEA3+/*BRACHYURY*- subset depending on the biological replicate. This high variability was also observable for the endoderm associated genes. The SSEA3+/*BRACHYURY*- subset exhibited more consistent levels of expression for *GATA6* and *SOX17*, while being overall lower in expression compared to the SSEA3+/*BRACHYURY*+ subset. In contrast, the levels of expression of these genes varied significantly in the SSEA3+/*BRACHYURY*+ subset. This pattern was also exhibited for *CD34* and *HAND1*, where together both genes were upregulated significantly higher in the SSEA3+/*BRACHYURY*+ subset, despite the high variation of expression across the replicates (Tukey Multiple Comparison Test $p<0.05$ $n=3$). Despite the high variability in the expression of the mesoderm associated genes, the results showed that the SSEA3+/*BRACHYURY*+ subset had a greater propensity to mesoderm differentiation compared to the SSEA3+*BRACHYURY*- subset.

Figure 18: Use of the H9 *BRACHYURY* reporter line with SSEA3 to identify subsets of hPSC with mesoderm lineage bias

Representative Flow cytometry plot showing the distribution of SSEA3 and *BRACHYURY* expression in KOSR/MEF self-renewal conditions (A). Two arrows show the SSEA3+/*BRACHYURY* + subset and the SSEA3+/*BRACHYURY* – subset from KOSR/MEF self-renewal conditions which were assessed to determine a mesoderm lineage bias. (B) The two plots below the SSEA3/*BRACHYURY* plot are examples of purity reanalysis of cells post sort.

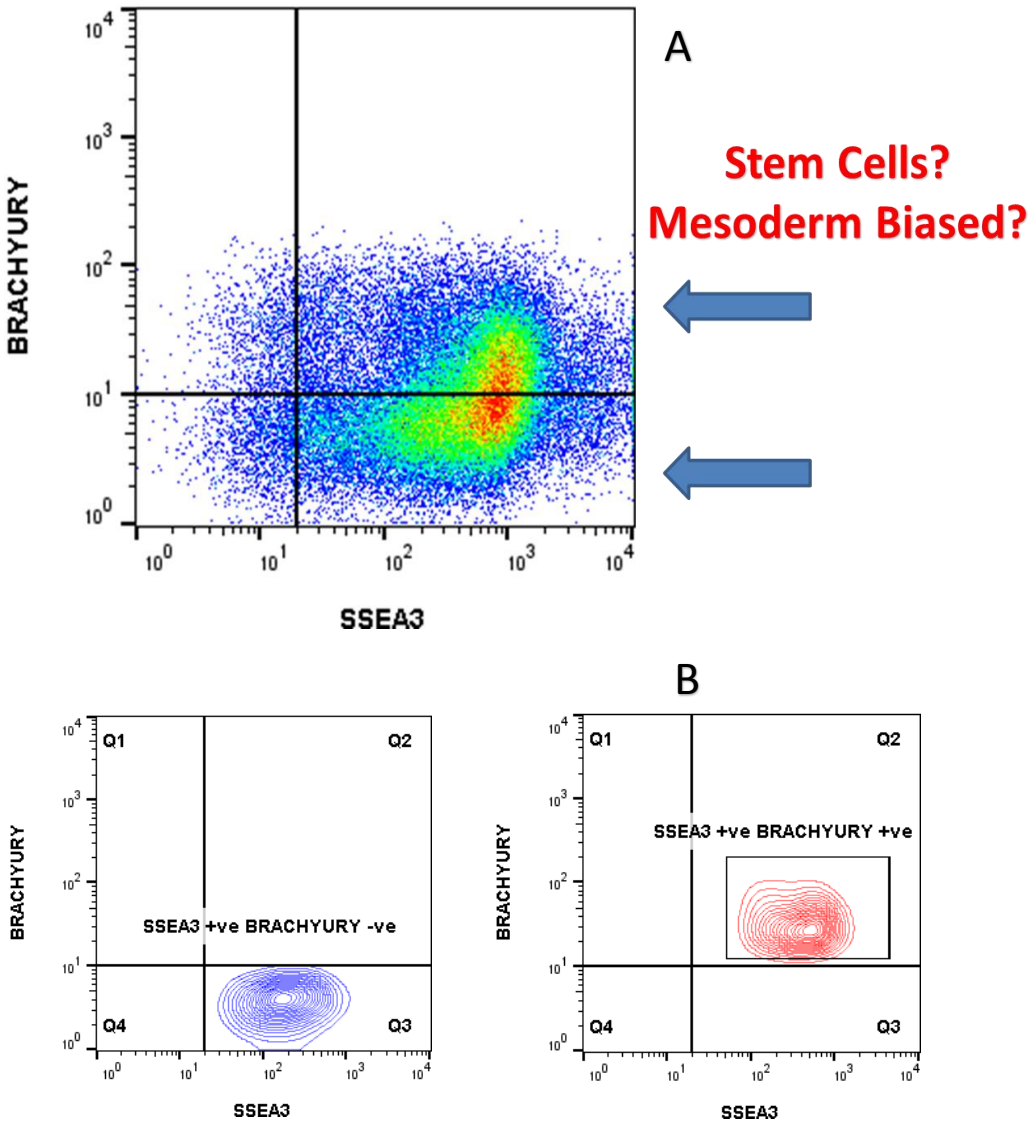


Figure 19: Population transcriptome analysis comparing SSEA3+/*BRACHYURY*- against the SSEA3+/*BRACHYURY*+ sorted subsets reveal differential expression of genes associated with differentiation and self-renewal

Comparison of the gene expression between the SSEA3+/*BRACHYURY* (T) + subset and the SSEA3+/*BRACHYURY* (T) – subset from KOSR/MEF self-renewal conditions. The results demonstrated that the SSEA3+/*BRACHYURY*+ subset had upregulation of genes associated with the mesendoderm such as T(*BRACHYURY*), *SOX17, GATA4, FOXA2 and EOMES*. Results are shown as ∆Ct with increased gene expression relative to values closer to 0. Selected genes compared in the following figure. Average of 3 biological repeats.

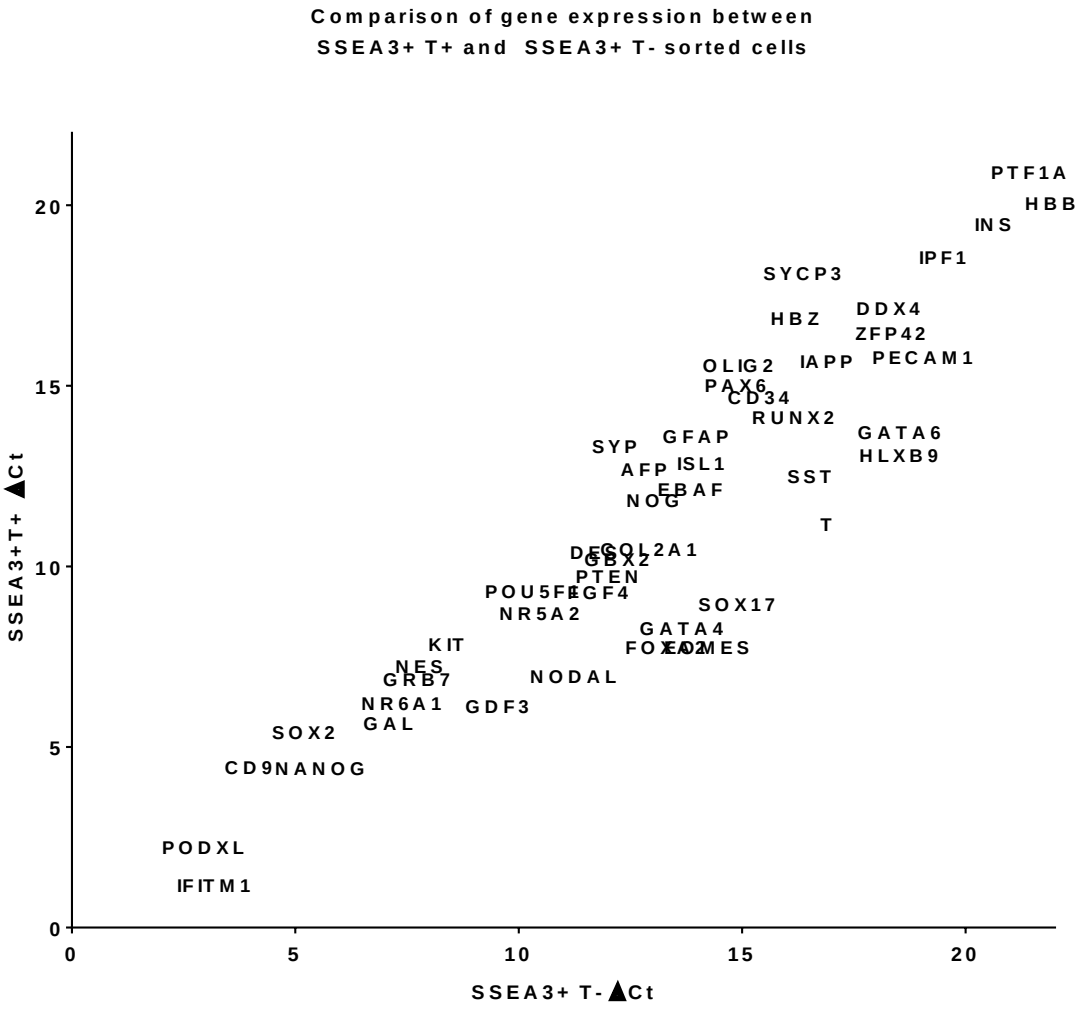
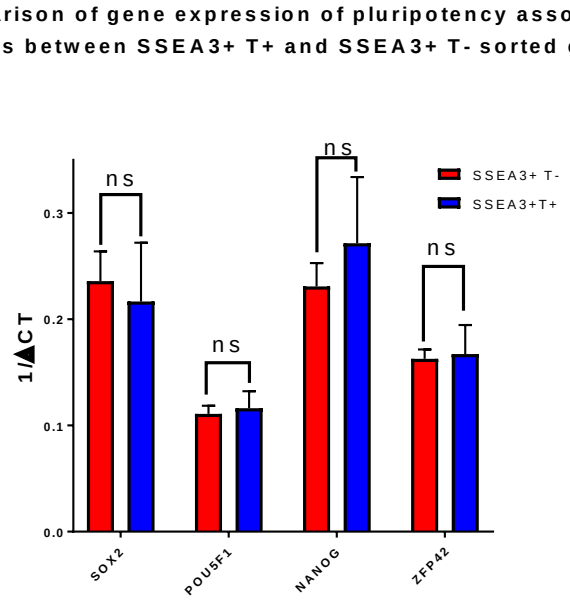


Figure 20: Population transcriptome analysis comparing SSEA3+/*BRACHYURY*- against the SSEA3+/*BRACHYURY*+ sorted subsets reveal differential expression of genes associated with differentiation and self-renewal

Comparison of the expression of selected genes including from Figure 20*: SOX2, POU5F1, NANOG, ZFP42* (A)*, EOMES, FOXA2, GATA4, GATA6, MIXL1, BRACHYURY* and *MESP1* (B) between the SSEA3+/*BRACHYURY* (T) + subset and the SSEA3+/*BRACHYURY* (T) – subset from self-renewal conditions. Results show upregulation of mesoderm and endoderm genes from the SSEA3+/*BRACHYURY*+ subset in these conditions. Statistical significance shown by asterisks (*) p<0.05. Gene expression depicted as 1/ΔCt as an average of 3 biological repeats. ns= not statistically significant p>0.05

A



B

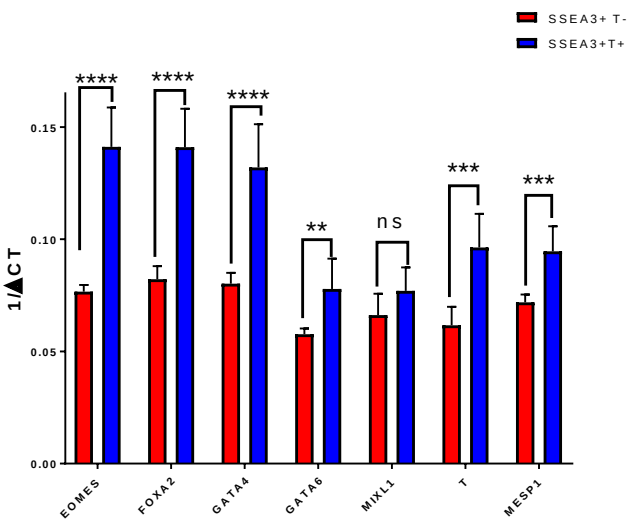


Figure 21: Schematic of the single cell plating experiment to demonstrate the SSEA3+/*BRACHYURY*+ subset contains hPSC

Schematic showing work-flow of single cell cloning experiment. To calibrate the FACS sorter for high stringency, accurate sorting the red and green fluorescing mixed beads were sorted into individual wells to show the machine could accurately sort individual populations. Single SSEA3+/*BRACHYURY*+ cells were sorted as single cells and left to grow into colonies. A selection of those were then taken forward and further analysed for functional capabilities to differentiation and self-renewal.

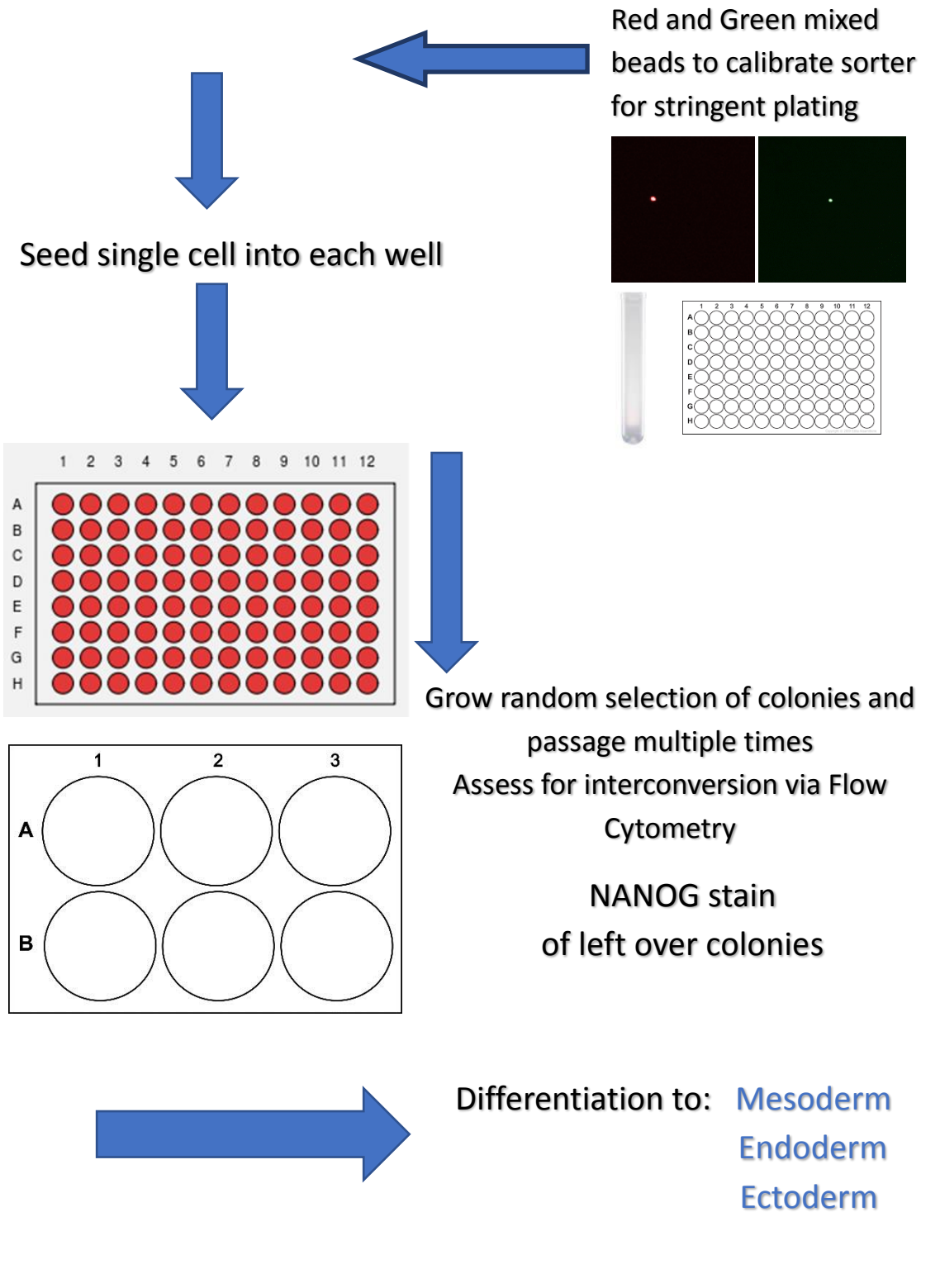


Figure 22: Single cell cloning of SSEA3+ *BRACHYURY*+ sorted subset exhibits that starting single cells were pluripotent

To verify whether the SSEA3+/*BRACHYURY* (T)+ subset contained functional hPSC, SSEA3+/*BRACHYURY*+ cells were seeded as single cells and grown as colonies before assessment. (A) Flow cytometry plot showing SSEA3+/*BRACHYURY* (T)+ sorted subset from KOSR/MEFS. (B) Position of index sorted SSEA3+ *BRACHYURY* (T)+ cells. (C) Representative picture of immunofluorescence for NANOG in self-renewal conditions, PAX6, TBX6 and SOX17 after differentiation. (D) Transcriptome analysis. Average 1/ΔCT of the clones for mesoderm differentiation (TBX6 and MSGN1), endoderm differentiation (GATA6, SOX17 and FOXA2) and ectoderm differentiation (PAX6, SOX1 and TUBB3). All 6 clones express NANOG and differentiated to derivatives of the three germ layers, demonstrating the starting cell of each clone was pluripotent. Scale bar set at 50µm. Statistical significance shown by asterisks (*) p<0.05

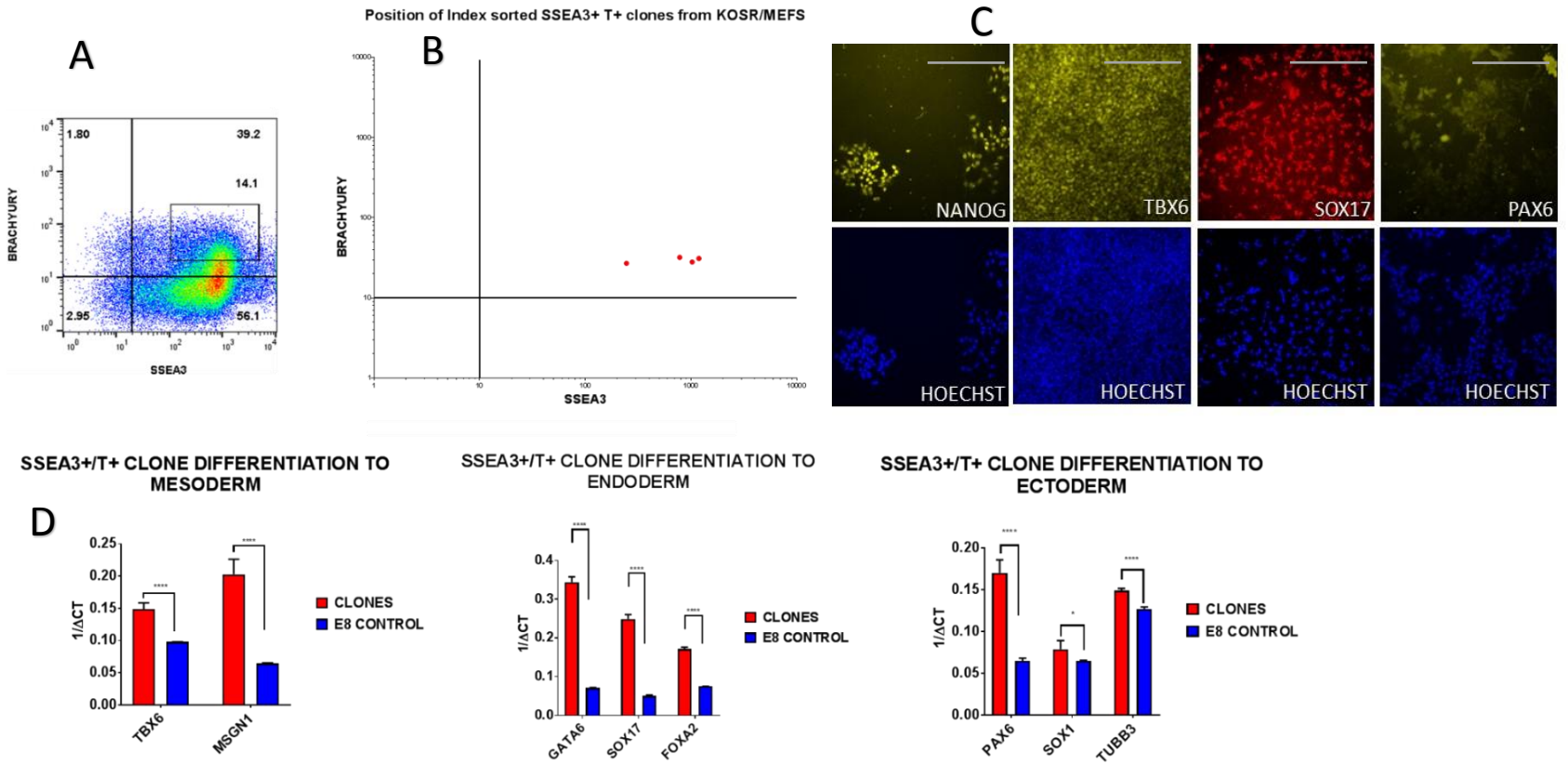


Figure 23: Analysis of SSEA3+/*BRACHYURY*+ clones grown on KOSR/MEFS after 5 passages demonstrate interconversion through emergence of SSEA3+/*BRACHYURY*- subset

Flow cytometry plot showing SSEA3 expression from the single cell clones in KOSR/MEF after 5 passages. (A) After 5 passages, the 6 clonal lines had interconverted from a SSEA3+/*BRACHYURY*+ subset into a heterogeneous population including SSEA3+/*BRACHYURY*- cells. After 5 passages, the 6 clonal lines had returned to a similar distribution to the starting population that the sort was performed on, as well as a distribution to what is natively found in the KOSR/MEF conditions. (B) Reference of the FACS plot from the initial cell sort.

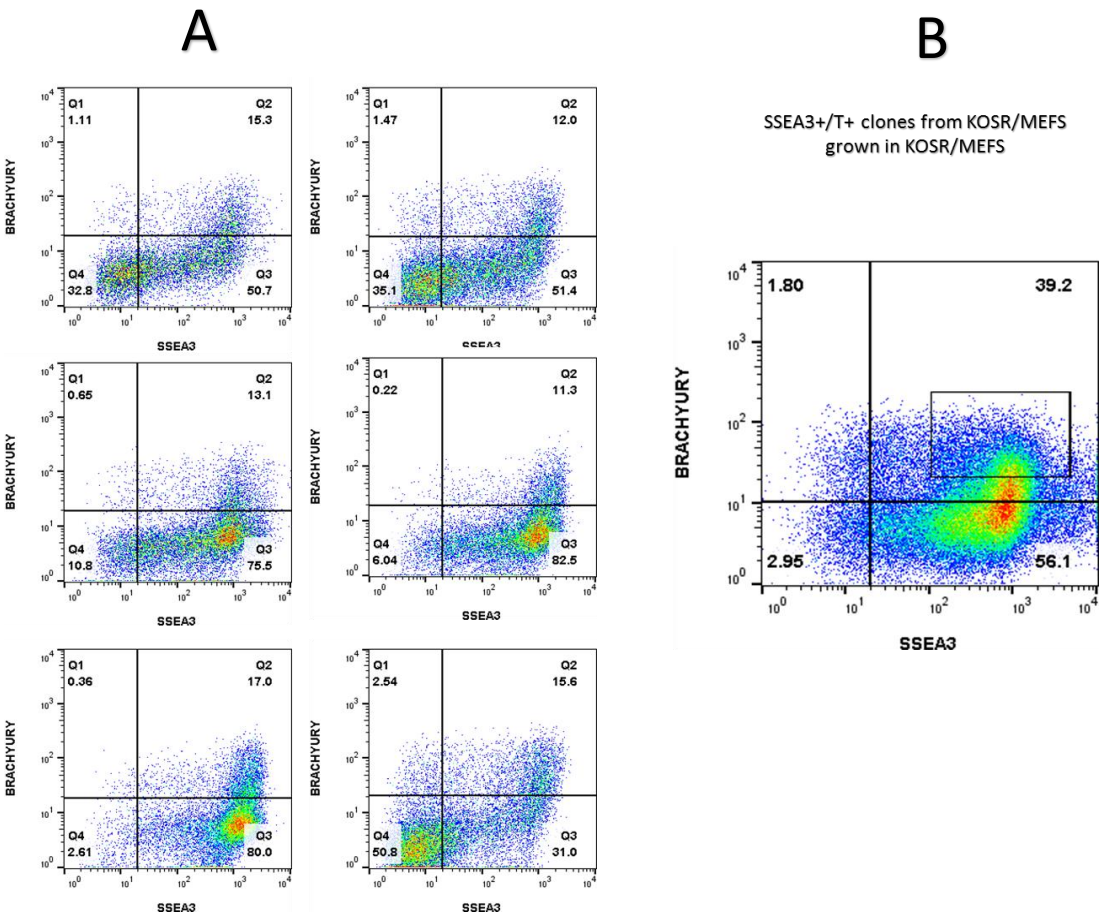


Figure 24: Analysis of SSEA3+/*BRACHYURY*+ clones grown on E8/Vitronectin after 5 passages reveal interconversion through loss of *BRACHYURY* expression

Flow cytometry plot showing SSEA3 expression from the single cell clones in E8/Vitronectin after 5 passages. (A) After 5 passages, there was interconversion of the 6 clonal lines from the SSEA3+/*BRACHYURY*+ subset they were originally sorted from into a SSEA3+/*BRACHYURY*- subset with over 95% of each colony being SSEA3+/*BRACHYURY*-. (B) Reference of the FACS plot from the initial cell sort.

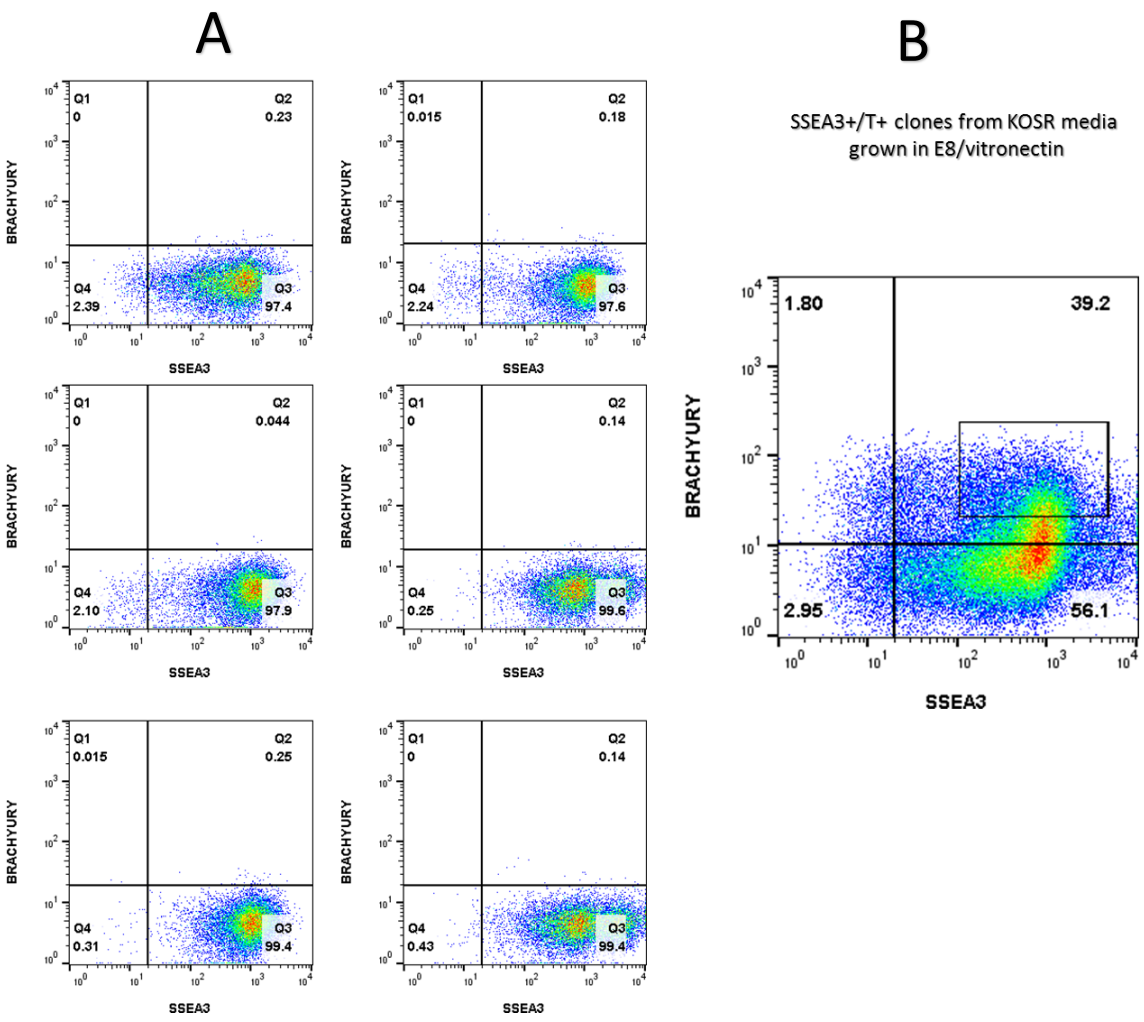


Figure 25: Schematic of the high-content clonogenic assay workflow to assess lineage bias of the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets

A schematic detailing the experimental procedure and analysis for the high-content clonogenic assays. An example of an immunostained, analysed colony can be seen at the bottom. A mask is drawn around a colony to analyse individual nuclei within the colony. Once individual nuclei have been isolated (middle) each individual cell nuclei is then analysed for marker expression (right) based on fluorescence higher than threshold set from Negative control. A detail of this immunofluorescence analysis can be found in Chapter 2 Methods.

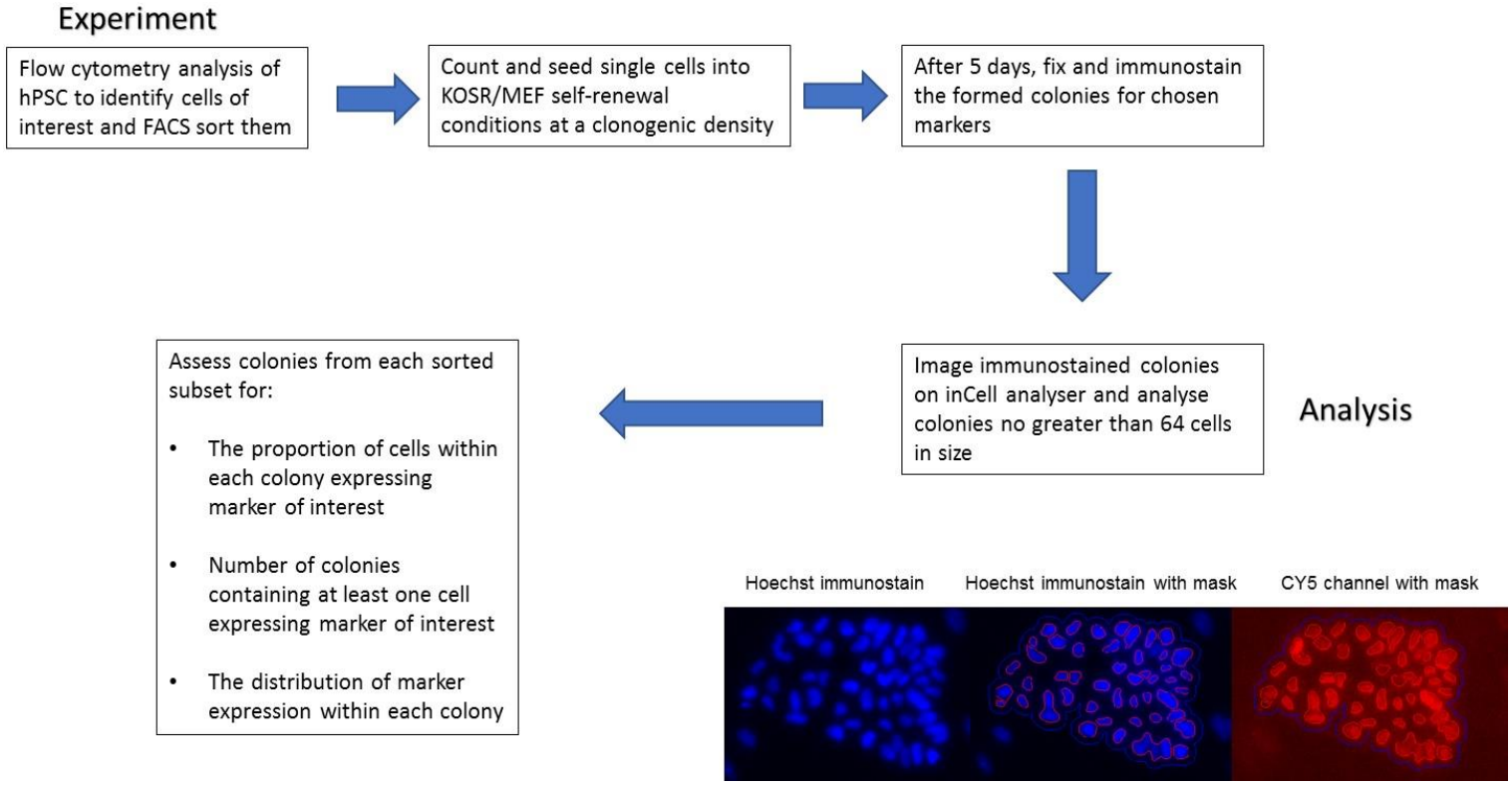
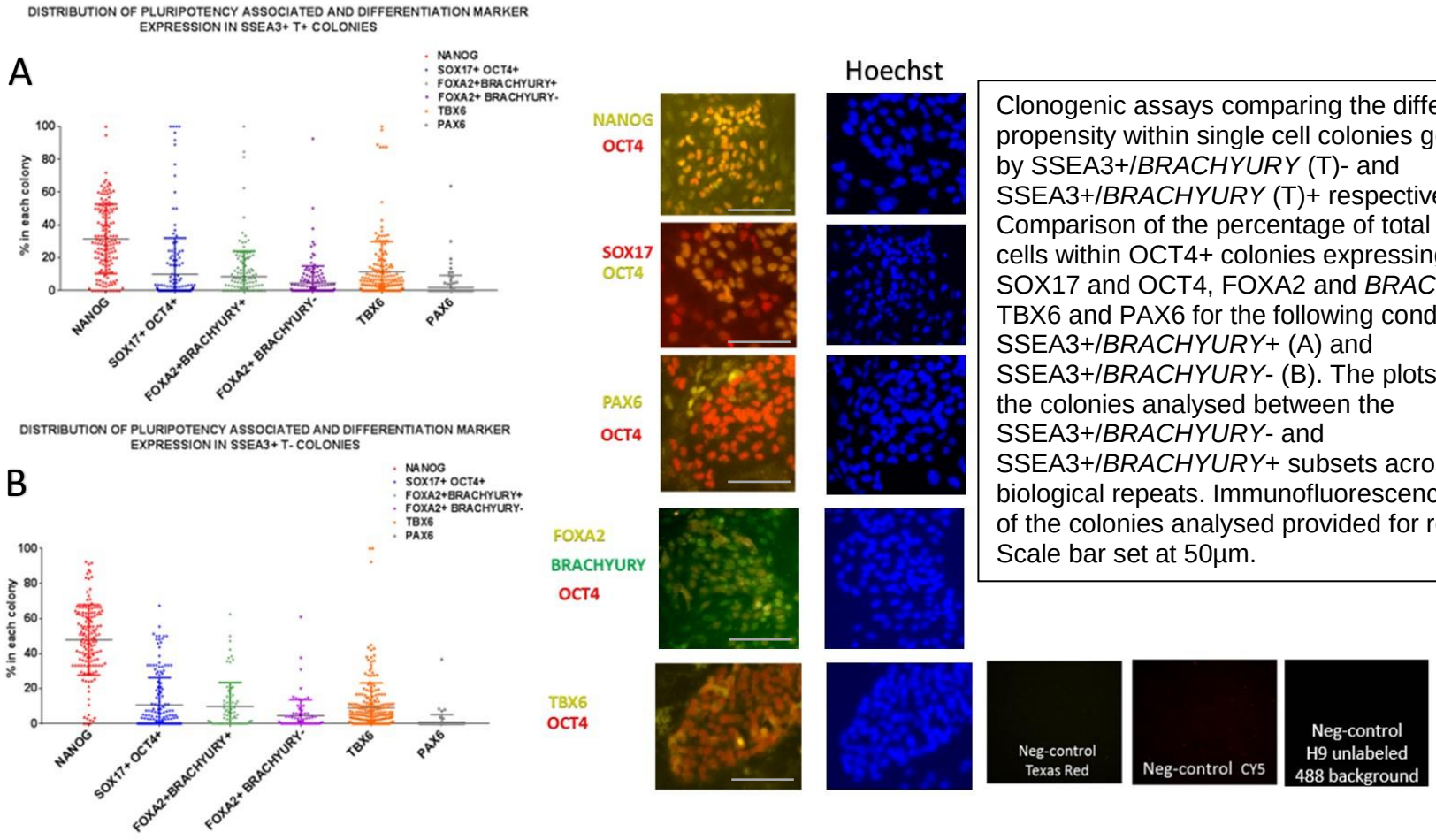


Figure 26: High content clonogenic assays comparing the SSEA3+/*BRACHYURY*- subset against the SSEA3+/*BRACHYURY*+ subset



Clonogenic assays comparing the differentiation propensity within single cell colonies generated by SSEA3+/*BRACHYURY* (T)- and SSEA3+/*BRACHYURY* (T)+ respectively. Comparison of the percentage of total number of cells within OCT4+ colonies expressing NANOG, SOX17 and OCT4, FOXA2 and *BRACHYURY*, TBX6 and PAX6 for the following conditions: SSEA3+/*BRACHYURY*+ (A) and SSEA3+/*BRACHYURY*- (B). The plots contain all the colonies analysed between the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets across three biological repeats. Immunofluorescence images of the colonies analysed provided for reference. Scale bar set at 50µm.

Figure 27: High content clonogenic assays comparing the SSEA3+/*BRACHYURY*- subset against the SSEA3+/*BRACHYURY*+ subset

Clonogenic assay comparing the differentiation within single cell colonies generated by SSEA3+/*BRACHYURY* (T)- and SSEA3+/*BRACHYURY* (T)+. Comparison of the percentage of total number of cells within OCT4+ colonies expressing NANOG (A), SOX17(B), TBX6 (C), FOXA2 and *BRACHYURY*- (D), FOXA2 and *BRACHYURY* (T)+ (E), and PAX6 (F) within OCT4 Positive colonies for the following conditions:

SSEA3+/*BRACHYURY* (T)- in red and SSEA3+/*BRACHYURY* (T)+ in blue.

The plots below exhibit the frequency of the percentage of the total numbers of cells in each colony expressing the respective lineage marker. The plots contain all the colonies analysed between the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets across three biological repeats. ns= not statistically significant p>0.05

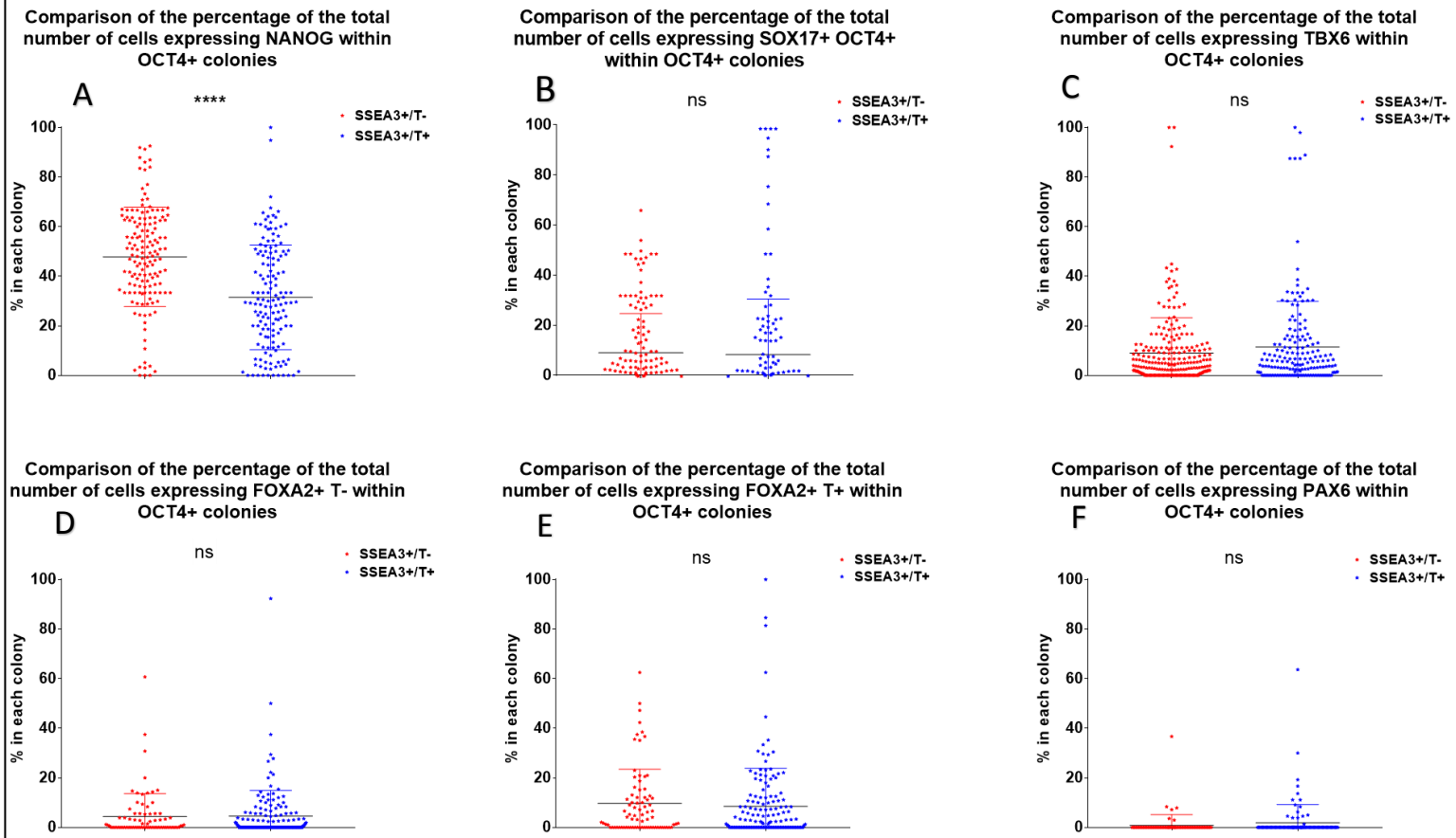


Figure 28: SSEA3+/*BRACHYURY* subsets exhibit heterogeneity in colonies in respect to expression of markers associated with self-renewal and differentiation

Comparison of the distribution of NANOG (A), SOX17 (B), TBX6 (C), PAX6 (D) and FOXA2 and *BRACHYURY* (E) within the OCT4+/OCT4- colonies in the following conditions:

SSEA3 (3) +/*BRACHYURY* (T)- and

SSEA3 (3) +/*BRACHYURY* (T)+.

Results reveal heterogeneous expression of these markers but a similar trend between the SSEA3/*BRACHYURY* subsets. Average of 3 biological repeats.

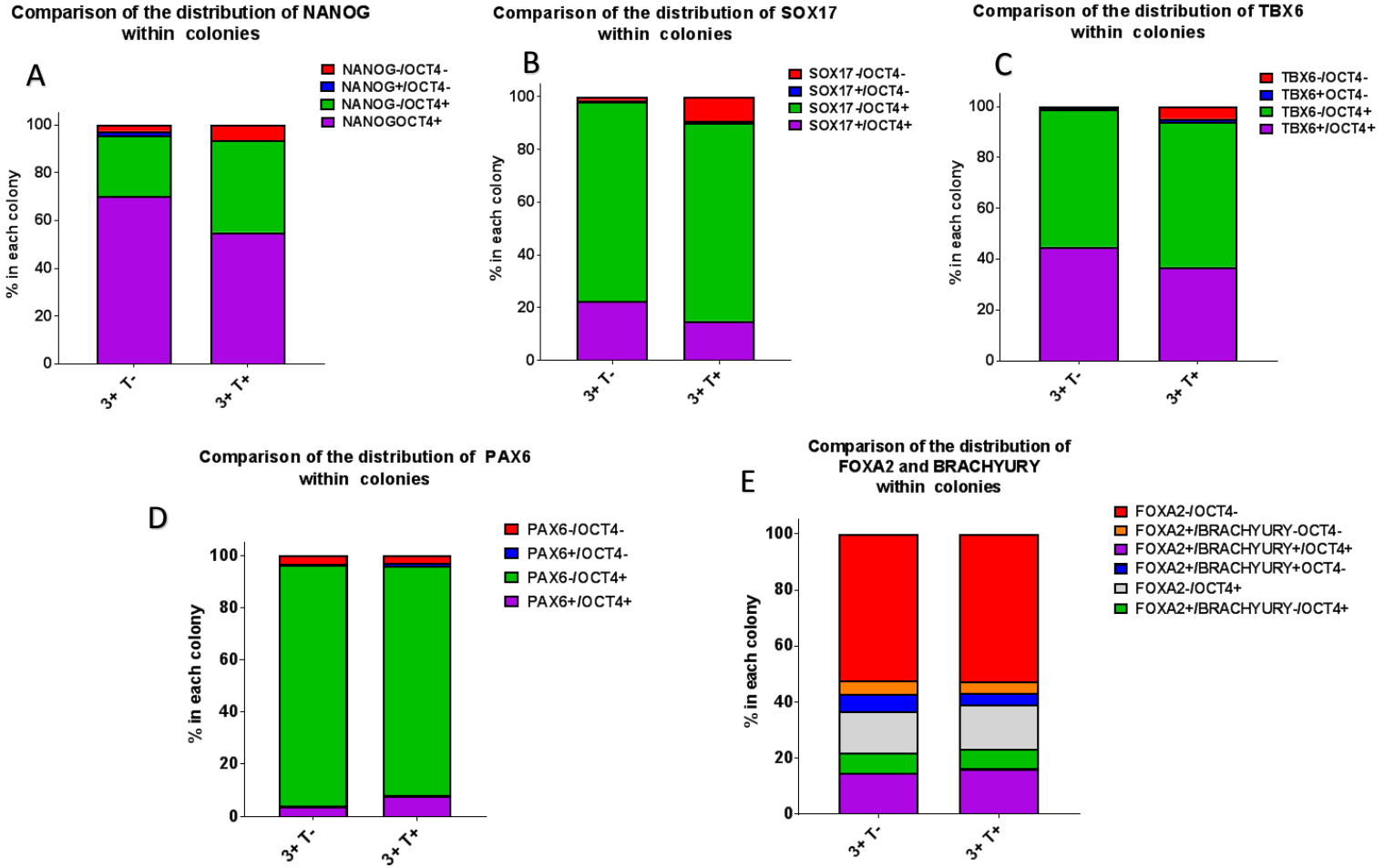


Figure 29: Analysis on the distribution of pluripotency associated and lineage specific marker expression in OCT4+ colonies

Comparison of the expression percent of NANOG (A), SOX17 (B) FOXA2 and *BRACHYURY* (C), TBX6 (D) AND PAX6 (E) within the OCT4+ Positive colonies in the following conditions:

SSEA3+/*BRACHYURY* (T)-, SSEA3+/*BRACHYURY* (T)+, SSEA3+/*BRACHYURY* (T)- to determine if any subset generate more OCT4+ colonies that contained at least 1 cell with a respective marker. The results reveal heterogeneous expression of these markers but a similar trend between the SSEA3/Brachyury subsets. Average of 3 biological repeats. ns= not statistically significant p>0.05

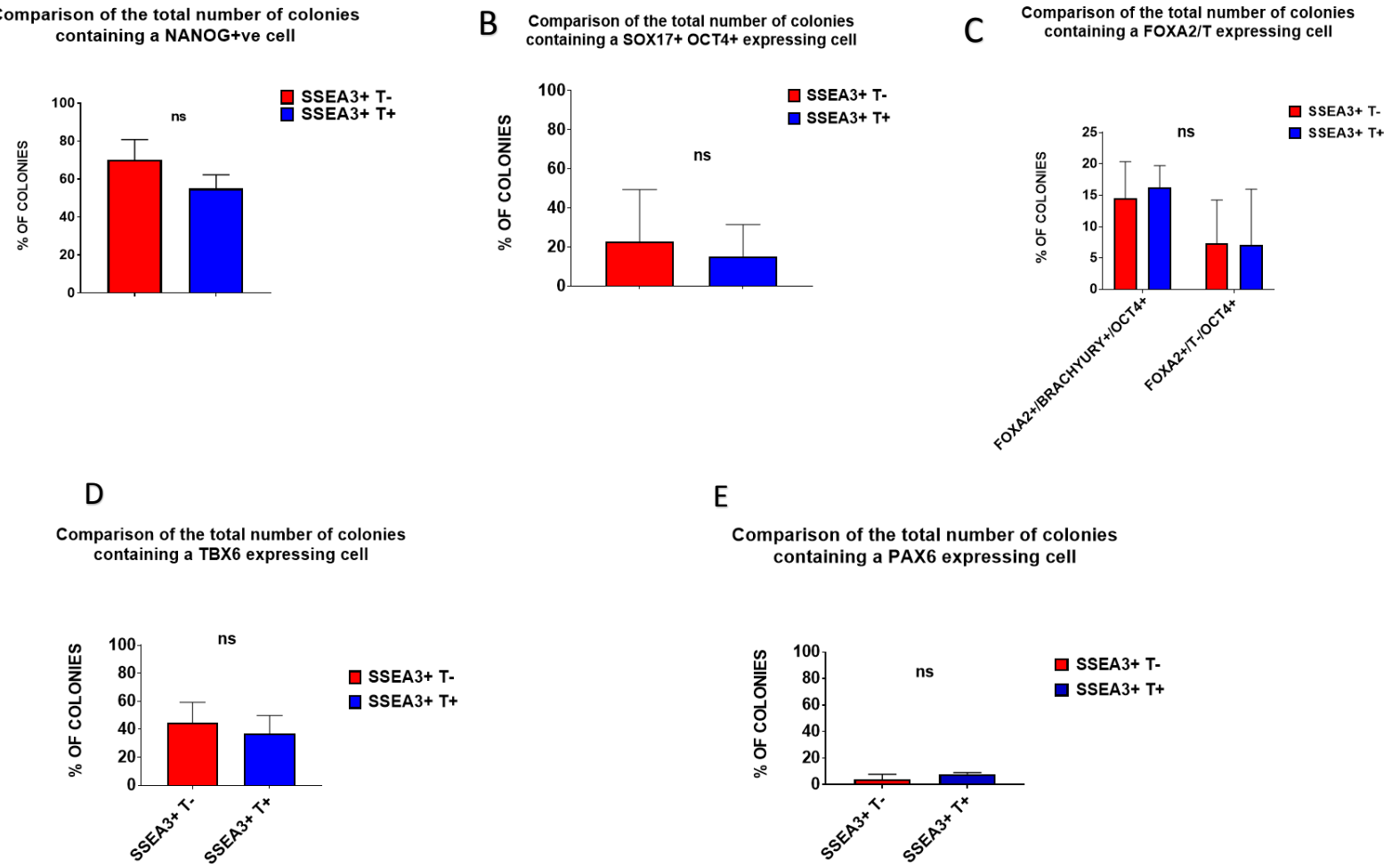


Figure 30: Lineage specific and neutral differentiation of SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets demonstrate both contain hPSC and have different propensities for mesoderm differentiation in Neutral conditions

Comparison of the gene expression comparing the SSEA3 (3) +/*BRACHYURY* (T)+ subset and the SSEA3 (3)+/*BRACHYURY* (T)- subset and mesoderm differentiation (HAND1 and GATA2) (A), endoderm differentiation (GATA6, SOX17 and GATA4) (B) and ectoderm differentiation (PAX6 and TUBB3) (C) with the E8 undifferentiated control. *CD34, HAND1, GATA4, GATA6, SOX17, PAX6, SOX1, PAX3*, *SOX10, TBX6, MSGN1* and *MEOX1* from Neutral EB differentiation condition (D). Gene expression depicted as 1/ΔCt as an average of 3 biological repeats. EBs imaged at x20 magnification

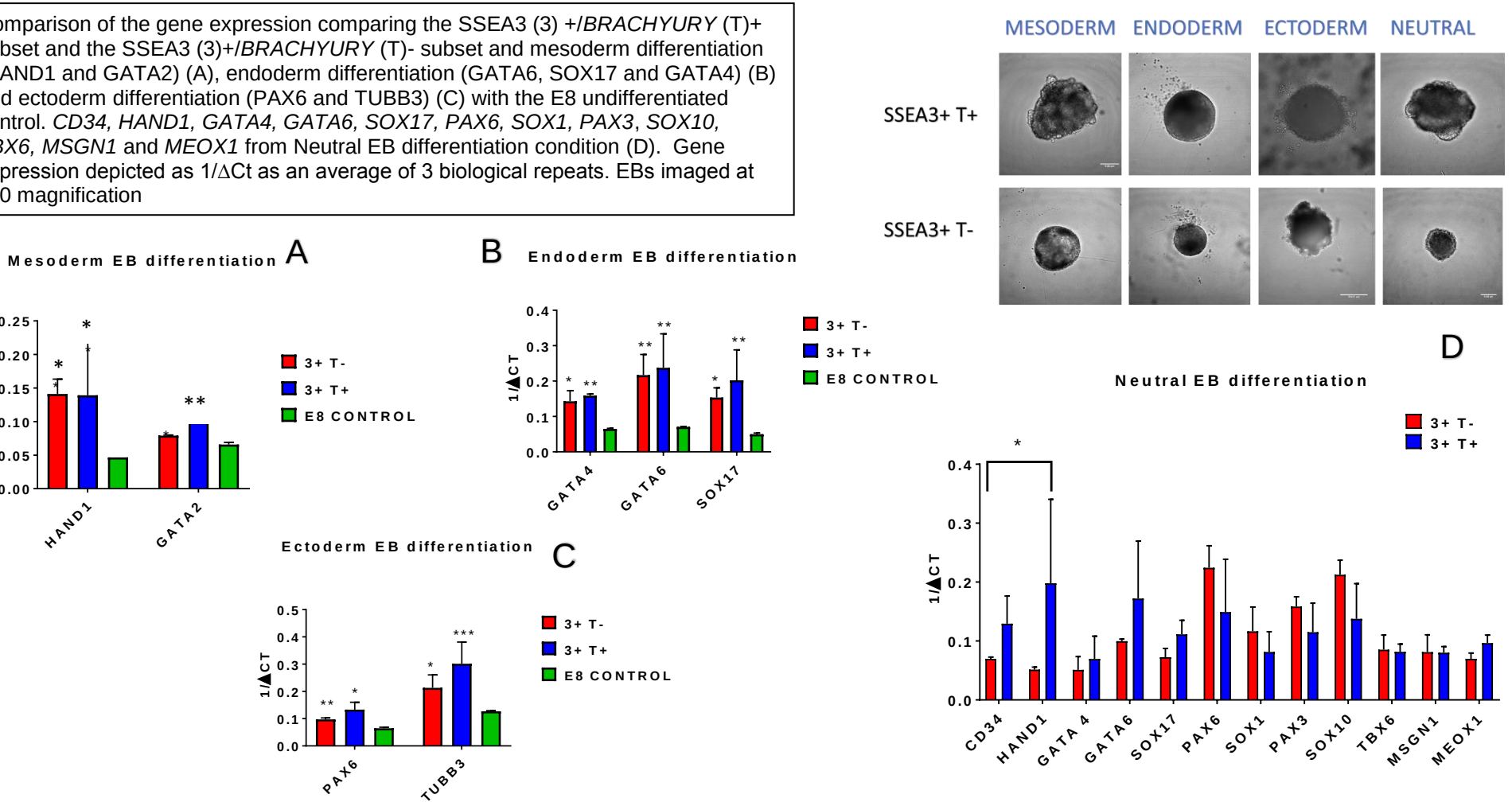
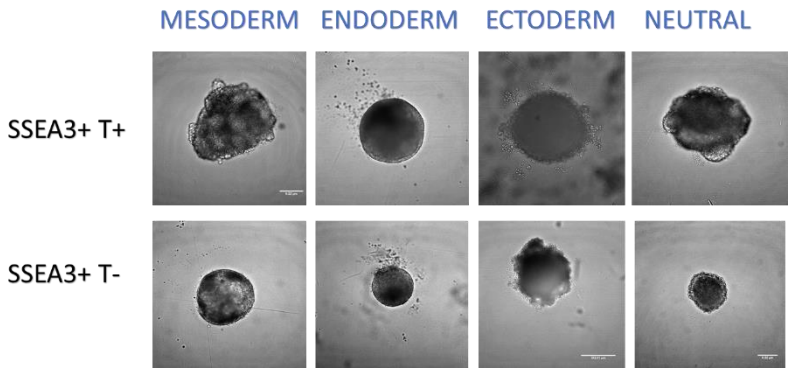
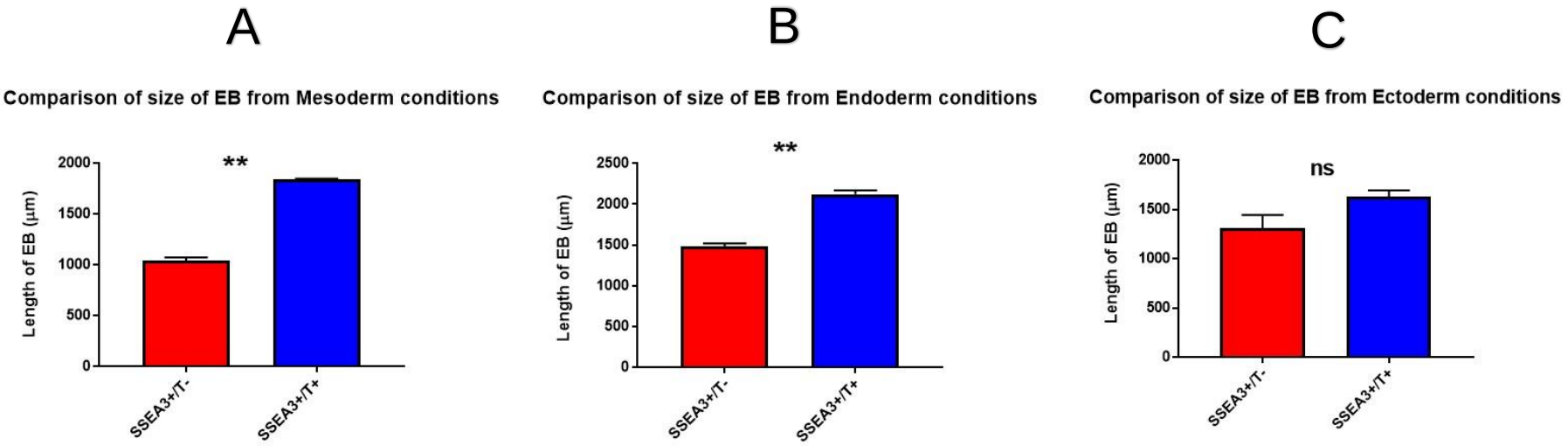


Figure 31: A comparison of size of EBs in induced differentiation conditions reveals the SSEA3+/*BRACHYURY* (T)+ subset forms larger EBs in mesoderm and endoderm conditions

Analysis of size of EBs grown in mesoderm (A), endoderm (B) and ectoderm (C) differentiation conditions. Length of EB given in µm. SSEA3+/T+ subset grew larger EB in mesoderm and endoderm conditions. Average of 8-10 EBs from each subset. n=3 Statistical significance shown by asterisks (*) $p<0.05$





4.3 Discussion

The gene analysis of the SSEA3+/*BRACHYURY* subsets from self-renewal conditions demonstrated that the SSEA3+/*BRACHYURY*+ subset had upregulation of early mesoderm and endoderm associated genes compared to the SSEA3+/*BRACHYURY*- subset. However, the ΔCt was lower than that of the pluripotency associated genes, to which both SSEA3+/*BRACHYURY* subsets were comparable (Figure 19 and Figure 20). This suggests that both subsets contain undifferentiated cells. Differential expression was observed with the lineage associated genes such *GATA6* and *AFP*, two endoderm associated genes [223,224]. Both *GATA*6 and *AFP* were much lower expressed to that of *GATA4* in both subsets, while *SOX17* and *FOXA2* and were similar in expression together in both subsets, but higher expressed in the SSEA3+*BRACHYURY*+ subset. While *GATA6* has been shown to be important for successful endoderm differentiation in vitro [182] both *GATA6* and *AFP* are expressed in the extra embryonic endoderm [225]. In contrast, while *GATA4* and *SOX17* are also considered pan-endodermal [226] the comparably high expression of *FOXA2* which is also considered an endodermal marker when not co-expressed with BRACHYURY [227] suggesting the SSEA3+/*BRACHYURY*+ subset may have an endoderm biased subset within it. However, *BRACHYURY* was also upregulated in the SSEA3+/*BRACHYURY*+ subset as expected, though not as highly as *FOXA2*. In respect to mesoderm differentiation, *MESP1* was higher expressed in the SSEA3+/*BRACHYURY*+ subset. *MESP1* has shown to have a role in patterning mesoderm to cardiac, haematopoietic or skeletal myogenic derivatives [228]. Together, this result indicates that the SSEA3+/*BRACHYURY*+ subset is heterogeneous in respect to early endoderm or mesoderm differentiation.

While it could also be argued that there are simply small numbers of cells that have already differentiated which explains the transcriptome result from (Figure 19 and Figure 20), the single cell cloning experiment demonstrated that single cells from the SSEA3+/*BRACHYURY*+ subset could be seeded in self-renewal conditions, grown as colonies, be passaged multiple times and then induced to differentiate to derivatives of the three germ layers. While it was not feasibly possible to perform this analysis on every colony, 6 were chosen (four of which from the index sort) and all 6 demonstrated similar behaviour in respect to

interconversion back to a heterogeneous population consisting of SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ cells on KOSR/MEF conditions and predominantly, SSEA3+/*BRACHYURY*- cells on E8/Vitronectin conditions. While the total number of cells that were SSEA3+/*BRACHYURY*- in each subset, varied within each clone, this could be explained by spontaneous differentiation resulting variable amounts of SSEA3 Negative cells on analysis.

However, when analysed for NANOG expression, all 6 clonal lines generated colonies that expressed NANOG, suggesting the colonies contained pluripotent cells. When selecting markers for analysis post differentiation, markers were chosen based on their strong correlation with differentiation to specific lineages. For mesoderm differentiation, TBX6 was chosen. This had previously been shown to be a paraxial mesoderm marker [229]. For endoderm, SOX17 and FOXA2 were chosen. These two markers independently can be attributed to other lineages that endoderm, but co expression together are considered endoderm associated markers [230,231]. For ectoderm, the well-defined neural ectoderm marker PAX6 was chosen.

All SSEA3+/*BRACHYURY*+ clones exhibited comparable efficiencies in differentiating to these lineages. An explanation for this is that the clonal lines were induced to differentiate after multiple passages and once on E8/Vitronectin conditions. The reason being that these differentiation protocols are optimised for feeder-free conditions and this could result in non-specific differentiation that could skew the conclusions being made. The observation that hPSC grown on E8/Vitronectin lose expression of the early mesoderm/endoderm reporter genes was observed in this experiment. This could explain the comparable differentiation results, that due to the cells interconverting back to a “unbiased” state, all clones were starting from a relative unbiased starting point and hence differentiated comparably to all lineages. Though while this experiment demonstrates that the starting cell was a hPSC, it does not demonstrate whether that starting cell contained a lineage bias.

These results again provide evidence for the ability of pluripotent states of cells to interconvert. The inducers of these states in a non-defined system are unknown. Additionally, without a means of maintaining this state, there is the potential of cells interconverting prior to their use in a specific application. This

could have significant consequences when using these cells translationally and justifies the need to create a system that can induce and maintain the biased state.

To determine whether the SSEA3+/*BRACHYURY*+ subset contained a lineage bias functionally, the clonogenic assay experiment was designed to determine whether the SSEA3+/*BRACHYURY*+ subset could form colonies in self-renewing conditions and if differentiation occurred, what lineage was it associated to (Figure 26-29). The results showed that both SSEA3+/*BRACHYURY* subsets could form colonies that expressed NANOG, while differing in the number of cells expressing the early endoderm and early mesoderm markers SOX17+/OCT4+, FOXA2+/*BRACHYURY*+, FOXA2+/*BRACHYURY*- within their respective colonies. The markers were chosen based on their lineage specific association with either endoderm or mesoderm. SOX17 is a marker than while is largely attributed to endoderm it has also been reported to be express in a variety of other lineage specific derivatives such as primordial germ cells [232] and cardiac mesoderm [233]. There is however reported evidence that cells co-expressing both SOX17 and OCT4 together are an indication of early endoderm differentiation [234,235].

In respect to FOXA2 and BRACHYURY co-expression, there is reported evidence that cells expressing both FOXA2 and BRACHYURY co-expression are associated with early primitive streak differentiation [236] or axial mesoderm [237], whereas cells which are FOXA2+/*BRACHYURY*- are more likely indictive of early endoderm differentiation due to the presence of FOXA2 expression in endoderm that is BRACHYURY Negative [66]. The results demonstrated that the SSEA3+/*BRACHYURY*+ subset could generate more colonies that had greater than 60% expression of these mesoderm and endoderm associated markers. This was not observed for the paraxial mesoderm marker TBX6. An explanation is this is attributed to the subset of cells having bias to either an entirely different mesodermal lineage or that the paraxial mesoderm was a stage later than most of the mesoderm associated differentiation occurring in the assay. It should be noted that there were a comparable number of colonies in both subsets that expressed TBX6 with total cell number less than 40%, most of which in the 20% range.

The SSEA3+/*BRACHYURY*+ subset also generated more colonies that expressed the ectoderm marker PAX6, though again the majority of colonies had under 20% of the cells expressing PAX6 per colony. This provides evidence that within both SSEA3+/*BRACHYURY* subsets there exists heterogeneity in respect to propensity for differentiation. One explanation for why the SSEA3+/*BRACHYURY*+ subset generated more ectoderm associated differentiation could be attributed to the pattern of BRACHYURY expression during differentiation, perhaps with a subset forming neuromesodermal progenitors (NMP) which can generate spinal cord neural ectoderm[238].

As described previously, it has been reported that during endodermal differentiation, BRACHYURY expression can be detected early in the differentiation and BRACHYURY Positive cells can give rise to endoderm in the mouse [211]. The results overall demonstrated in this assay, that there was no evidence of a mesoderm bias with the SSEA3+/*BRACHYURY*+ subset compared to the SSEA3+/*BRACHYURY*- subset.

One of observations from the results was how the distribution of antigen and reporter expression varied depending on whether the cells were growing on KOSR/MEF or E8/Vitronectin conditions (Figure 23 and Figure 24). One possible explanation for this phenomenon is the culture environment that the cells are growing in. It has been documented that MEF secrete various proteins including ACTIVIN A and BMP4 in culture [239] which are two protein ligands that are involved in many signalling networks involving different differentiation pathways. This is an explanation as to why both SSEA3+/*BRACHYURY* subsets generated more mesoderm/endoderm differentiation in the colonies than ectoderm. The clonogenic assay protocol was only successful in KOSR/MEF conditions where previous attempts at plating in E8/Vitronectin or mTeSR/Matrigel failed to generate any colonies. There is currently no evidence to suggest that either E8 media or vitronectin contains any chemical moiety that may be inducing differentiation in hPSC.

Considering these limitations, the embryoid body assay (Figure 30-31) was chosen to add further validity to the SSEA3+/*BRACHYURY*+ subset having greater propensity for mesoderm/mesendoderm differentiation compared to the SSEA3+/*BRACHYURY*- subset. The results demonstrated variability across

biological replicates which could be attributed to the level of *BRACHYURY* expression being too low (10^1-10^2 fluorescence intensity in this experimental set-up) to implicate readying for mesoderm differentiation in some cases. This could be explained due to *BRACHYURY* expression being no greater than the second decade (10^2) in self-renewal conditions which is lower than the *GATA6* and *MIXL1* reporters.

The results nevertheless demonstrated an increased propensity of the SSEA3+/*BRACHYURY*+ subset to mesoderm differentiation in Neutral conditions as shown by increased expression of *CD34* and *HAND1* (Figure 30 D). In respect to endoderm associated genes, *GATA6*, *SOX17* and *GATA4* were shown to be higher in the SSEA3+/*BRACHYURY*+ subset in both the endoderm and Neutral conditions. In contrast, the ectoderm associated gene *PAX6* was highly expressed in both subsets in the neutral conditions. For *TUBB3* expression, while both SSEA3+/*BRACHYURY* subsets expressed higher levels of the gene compared to the control in the ectoderm inducing condition, it was higher in the SSEA3+/*BRACHYURY*+ subset. Additionally, the EB were not comparable in size during these assays, with the SSEA3+/*BRACHYURY*+ subset generating larger EBs in the mesoderm and endoderm conditions (Figure 31). These results indicate that the SSEA3+/*BRACHYURY*+ subset found in KOSR/MEF conditions is not homogenously biased to differentiate in one direction.

As described in chapter 3, a caveat with using the *BRACHYURY* reporter for isolating subsets of cells is the half-life of the Venus fluorescent protein. The Venus protein is a variant of YFP, which had shown to have an *in vivo* half-life of 12.8 hr in some experimental systems [184]. However, half-life of fluorescent proteins can vary in different experimental set-ups, with some reporting upwards of 24 hr [185]. Consequently, a major consideration of the experimental design of using the *BRACHYURY* reporter to isolate cells, is whether cells sorted are actually still BRACHYURY+ or simple Venus+. This consideration explains the variation in the biological replicates with the clonogenic assays and neutral EB differentiation assays. Potentially, it could have been that cells were actually SSEA3+/*BRACHYURY*- when sorted and compared. On average, SSEA3+/*BRACHYURY*+ cells sorted from self-renewal conditions and analysed for *BRACHYURY* expression on the transcriptome, had higher levels of *BRACHYURY* expression compared to the SSEA3+/*BRACHYURY*- cells (Figure

20). However, considering the potentially large discrepancy between fluorescent protein half-life and gene expression, the potential of SSEA3+/B*RACHYURY*+ cells actually being *BRACHYURY* Negative is a factor to consider. This caveat provides another justify reason to attempt to induce and maintain expression *BRACHYURY* expression in a defined system.

4.4 Conclusion

The results overall demonstrate that both the SSEA3+*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets contain hPSC and there is evidence in some cases that the SSEA3+/*BRACHYURY*+ subset found in KOSR/MEF conditions can exhibit increased propensity to mesoderm and endoderm, as shown by gene analysis results and functional analysis via EB forming assays. However, the differentiation occurring across replicates was not homogeneous and included variable non-specific differentiation to other lineages. The results therefore show that this subset in KOSR/MEF conditions consists of different subsets of cells readying to differentiate to various lineages.

In order to study specifically early fate determination towards specifically mesoderm differentiation, it is a requirement to recreate this population in a defined system with mesoderm-specific signalling induction. The concept being that if this population could be recreated with a defined signalling network, it could be maintained continuously in self-renewing conditions if the balance was right between the signals promoting self-renewal and differentiation. If achieved, a system could be created that would allow study into the molecular mechanisms involving early fate determination of hPSC towards to the mesodermal lineage.

Chapter 5 : Recreation and Analysis of the SSEA3+/*BRACHYURY*+ Subset in a Defined System

5.1 Introduction

5.1.1 Overview

The results from chapters 3 and 4 demonstrate the existence of subsets of hPSC that exhibit different propensities for differentiation to different lineages. These subsets can be defined exclusively by the expression levels of cell surface antigens, or through a combination of these surface markers and a reporter cell line with a gene whose expression is associated with specific lineages following differentiation. However, there are limitations when analysing these populations that are found in KOSR/MEF conditions in respect to reproducibility. The evidence of upregulation of genes associated with different lineages in these sorted subsets prior to and after differentiation, suggests that these populations have different sub-populations within them. This presents a technical problem when trying to identify the molecular mechanisms or gene expression network that is governing the process of differentiation to a specific lineage. As these conditions are undefined, it is not possible to elucidate what signalling mechanisms are inducing and controlling these sub-states in culture. By creating a defined system to recreate and maintain a specific mesoderm biased population, these mechanisms can be further explored.

5.1.2 The balancing of signalling networks to “trap” subsets of hPSC in a lineage biased, but undifferentiated state

The process of differentiation of hPSC is a complex biological process involving co-ordination of signalling networks, resulting in functional somatic cell types. The evidence in the literature, is that the resulting cell derivative created from differentiation is the result of induction of these networks from specific chemical inducers at specific concentrations. For mesoderm differentiation, as discussed chapter 4, there are different signalling networks that can be induced that result

in generation of mesoderm derivatives. In hPSC, the Wnt pathway can be induced by the chemical Wnt agonist CHIR99021 (CHIRON). Adding CHIRON at varying concentrations has been shown to result in increased efficiency at forming either mesoderm or endoderm [240].

For forming more specific derivatives, in addition to CHIRON, a combination of compounds can be added that result in forming the desired cell type. This can be seen when forming specific differentiated progenitors/cell types such as neuromesodermal progenitors (NMP) or cardiomyocytes, where the protocols can have similarities, such as the addition of precise concentrations of CHIRON [215,241] but differ in respect to the selection and concentrations of components being added throughout the duration of the differentiation. This concept of adding in compounds at precise concentrations that can induce or inhibit signalling processes, is the principle behind the concept of inducing a lineage biased state of hPSC. However, the key biological question is whether the lineage biased state cannot only be induced but also maintained through precise balancing of the signalling networks driving self-renewal and differentiation.

Evidence in the literature for manipulating ES cell signalling networks can be seen in the example of the addition of lysophosphatidic acid (LPA), a compound found within KOSR [242], to perturb Wnt induced differentiation of hPSC. By adding in the precise concentration of 10μM LPA, which is the concentration found to be in 20% KOSR, Wnt induced mesoderm induction could be ablated [99]. Further studies had shown that LPA can feature in maintaining pluripotency by substituting for YAP in activating the HIPPO signalling pathway [243].

In the (Blauwkamp et al. 2012) study [99], the group were able to induce Wnt high and Wnt Low subsets of hPSC, but were unable to maintain these subsets as undifferentiated over multiple passages and suffered additional unwanted side-effects of the culture system, such sub-optimal growth rate of the cells. While this attempt at maintaining the hPSC under these culture conditions were unsuccessful, it nevertheless served as a conceptual model that cross-antagonism of pro-differentiation and pro-self-renewal signals can alter the signalling network within the cell culture.

5.1.3 Considerations for the culture system for maintaining a lineage biased state

As shown by (Blauwkamp et al. 2012) [99] being able to induce and maintain a state of hPSC that can be passaged indefinitely is technically very difficult. A consideration of the culture system being chemically-defined and feeder-free, is to avoid variation in results from batch variation of media components. Additionally, being chemically defined also removes unknown factors present in the system which could influence cell-behaviour.

One of the earliest examples of feeder-free culture matrix was Matrigel, a matrix that is formulated from a gelatinous protein mixture secreted by mouse Engelbreth-Holm-Swarm (EHS) sarcoma cells [244–246]. The matrix constituents contain structural proteins such as laminin, entactin and collagen which are proteins that would be found in the native environment during embryonic development [247,248]. The matrix also contains signalling ligands involved in maintaining self-renewal such as TGF-β and EGF [249], altogether providing an environment for maintaining proliferating pluripotent cells. However, one of the limitations of this matrix is that it is not chemically defined and is also prone to batch variation that can influence experimental results and ultimately affect reproducibility [250,251].

Another feeder-free cell culture matrix is vitronectin, a glycoprotein found abundantly in *in vivo* in humans, such as in the extracellular matrix and blood [252,253]. This substrate when used in a culture system for hPSC interacts with integrin αvβ5 to promote cell attachment [254,255] and it had been demonstrated that inhibition of αvβ5 and β1 perturbed cell proliferation of iPSC [255]. Vitronectin provides a viable, chemically defined alternative to Matrigel for maintaining self-renewal and pluripotency [254]. Economically, vitronectin also provides a cost-effective alternative for culturing hPSC in feeder-free conditions. However, subject to experimental application, there are requirements for using specific a cell culture matrix. Such examples include, using vitronectin as a substrate for 3D culturing of cells in microcarriers for bioreactor scale-up [256] or using Matrigel as a matrix for culturing 3D intestinal organoid structures [257]. In addition to Vitronectin, other feeder-free cell matrices also exist, including Collagen, Fibronectin and Laminin. These have shown to have different mechanisms for

cell attachment, as well as different efficiencies for maintaining undifferentiated cells [258,259]. However, in addition to the cell matrix, the media composition used for cell culturing also has a significant role in maintenance of undifferentiated cells.

Specifically, to human pluripotent cell culture, cells were traditionally grown in feeder conditions with different media components, including KOSR [139]. In feeder-free conditions, the KOSR media recipe is not chemically defined, nor suitable for maintaining undifferentiated pluripotent cells. For alternatives, while many products and protocols exist, one of the most recognised and widely used was mTeSR [260]. However, one of the drawbacks of this media is that it is only partially defined due to the BSA component. An adaptation of this media that is fully chemically defined media is Essential 8 [138]. Being chemically defined, the media is less prone to batch variations from undefined components and is more suitable for use in a culture system to study early molecular processes within the cell environment. Additionally, it was shown in Chapter 4 (Figure 24) *BRACHYURY* expression is lost when cells were grown in E8/Vitronectin. This provides evidence of a cost effective, chemically defined platform to create a culture system for recreating and maintaining a mesoderm-biased state.

5.1.4 Experimental Objectives

Utilising signalling network induction in a chemically defined system, the aim is to create a culture system through cross-antagonism of pro-differentiation and pro-self-renewal signalling. This system can induce a mesoderm biased state of hPSC, so that it can be maintained over multiple passages, resulting in lineage biased hPSC that remain pluripotent. Through this system, it can be determined if cells have greater propensity to mesoderm differentiation within this recreated population. This system could be used as a tool for further downstream applications into studying fate determination mechanisms of hPSC towards mesoderm differentiation.

5.2 Results

5.2.1 Use of Wnt signalling to induce a mesoderm-biased state

To determine whether mesoderm-induced cells could be passaged and maintained as pluripotent, the H9 *BRACHYURY* reporter cells were seeded as single cells on Vitronectin and induced to mesoderm differentiation with 3μM CHIRON for 3 days and passaged into E8 self-renewal conditions. All conditions were analysed via flow cytometry for BF4, CD9 and SSEA3 or immunostained with NANOG after 3 days. The results from (Figure 32) show that addition of CHIRON alone was not suitable for maintaining a mesoderm biased, undifferentiated state. All 3 antigens had loss of expression post passage compared to the self-renewal control (Figure 32 A-C). BF4 and SSEA3 expression decreased from ~99% expression to <10%, while CD9 expression decreased to from ~99% to <50%. NANOG expression was 50-70% in the self-renewal conditions but decreased to <1% in the post-passage differentiation conditions (Figure 32 D). Additionally, the cells in the post-passage differentiation conditions failed to form colonies and exhibited a mesenchymal morphology. The results showed that inducing mesoderm differentiation with CHIRON alone and passaging them back into self-renewal conditions was not sufficient in maintaining the cells as undifferentiated.

5.2.2 Titration of the LPA component to find the optimal level to "trap" mesoderm biased hPSC

The results from the CHIRON mesoderm induction showed that this induction pushed the balance away from pro self-renewal towards terminal differentiation. The addition of CHIRON was efficient in inducing *BRACHYURY*+ expression (Figure 32). A similar result was obtained using the *MIXL1* reporter (Andrews lab, unpublished) so it was considered whether regulating this signalling network within the pro-self-renewal conditions would be enough to maintain the balance within the system and maintain a mesoderm biased, but undifferentiated population. Previous unpublished work (Andrews lab) demonstrated that the addition of 4.8 μM LPA component (LPA with the addition of BSA, Cholesterol β-

merceptoethanol) (now referred to as BCL) to 3µM CHIRON in E8 medium (mesoderm induction media) perturbed the emergence of SSEA3+/*MIXL1*+ cells. To determine optimal levels of BCL, a 3-day titration series was devised, consisting of concentration ranging from 4.8µM through to 0.3µM with the addition of 3µM CHIRON in E8 medium (Figure 33). H9 *BRACHYURY* reporter cells were seeded as single cells with a media change to the differentiation media for a total of three days. As the concentration of BCL decreased, conversely the percentage of *BRACHYURY* Positive cells increased. However, this increased plateaued around the 0.6 µM to 0.3µM range. Alongside this titration series, the *MIXL1* reporter was also used in a titration series using the same concentration range, producing comparable results (Andrews lab, unpublished). The final concentration decided was 0.48 µM, which fell midway between the 0.3µM and 0.6µM range.

5.2.3 Adding an LPA component (BCL) to create the balance of the mesoderm-biased state

To determine the effect of adding 0.48µM of BCL in E8 medium to the expression of BF4, CD9 and SSEA3, a 3-day a mesoderm differentiation assay was designed with and without the addition ofBCL. To test the effectiveness of BCL, in addition to CHIRON induced mesoderm formation with BCL, Activin A and BMP4 were used with BCL to induce *BRACHYURY* Positive expression as this would correspond to induction of formation of different cell types. Activin A had been shown to induce endoderm formation [66], while BMP4 had been shown to induce lateral plate/ ventral posterior mesoderm formation [21].

The cells were dissociated into a single cell suspension and seeded as stated in the mesoderm differentiation protocol above. The following day, three differentiation media were made (E8 + 3µM CHIRON), (E8+ 50ng/µl Activin A (Tocris) and 40ng/µl BMP4 (Peprotech) (referred to as $AA^{50}/BMP4^{40}$), (E8+ 100ng/µl Activin A and 5ng/µl BMP4) (referred to as $AA^{100}/BMP4^{5}$) and the hPSC were left to grow in these respective media for a total of three days with daily media change. The media were prepared with or without the addition of 0.48µM BCL for comparison. All conditions were analysed via flow cytometry for BF4,

CD9 and SSEA3 after 3 days. With the addition of 0.48 µM BCL the percentage of cells that were *BRACHYURY* Positive were reduced in all three differentiation media conditions compared to the non-BCL differentiation conditions, at ~2-3% (Figure 34 A-C). Furthermore, the percentage of double Positive cells for all three antigens was reduced compared to the CHIRON only differentiation condition. However, the CHIRON differentiation condition with BCL produced the most double Positive cells for all three antigens and had the most *BRACHYURY* Positive expression. In comparing the three differentiation conditions, the CHIRON with BCL condition produced considerably more BF4/CD9/SSEA3+ *BRACHYURY*+ cells compared to the two respective Activin A/BMP4 conditions. Across the three biological replicates, in the CHIRON with BCL inducing conditions, there was 30-50% more BF4+/*BRACHYURY*+ cells, over 50% more CD9+/*BRACHYURY*+ cells and over 50% more SSEA3+/*BRACHYURY*+ cells compared to the two Activin A/BMP4 conditions. These results demonstrated that the addition of CHIRON with BCL was most optimal for maintaining a SSEA3+/*BRACHYURY*+ state compared to the addition of Activin A and BMP4.

5.2.4 Development of the media system to maintain the SSEA3+/*BRACHYURY*+ subset over multiple passages

The results from (Figure 34) show that the addition of BCL in E8 medium at various concentrations could affect the percentage of *BRACHYURY* Positive cells and the co-expressing antigen double Positive cells within the population. However, what was an unknown was whether the addition of the 0.48µM BCL could prevent the cells from differentiating post-passage which was not possible with the addition of CHIRON alone (Figure 32). In addition to adding BCL separately to E8 medium or to 3 µM CHIRON with E8 medium, the endogenous Wnt blocker IWP2 was added at various concentrations (0.25 µM to 2µM) on single cells to see how this affected the cell culture post-passage. To assess the efficiencies of these different BCL/IWP2 E8 media recipes, the cell surface antigens BF4, CD9 and SSEA3 were analysed on H9 *BRACHYURY* cells over a total of three passages. A schematic of this procedure can be seen in (Figure 35).

As demonstrated in chapter 3, these BF4, CD9 and SSEA3 have different expression profiles following mesoderm induction, with both BF4 and SSEA3 losing significant expression and CD9 moving to a “Mid” expression level (Figure 17). The results from the first passage showed that compared to the standard E8 with CHIRON control condition, the addition of IWP2 at all concentrations ranging from 0.25 to 2µM reduced the number of BF4 Negative cells in culture by ~1-5% (Figure 36 A). However, the cells in the BCL in E8 medium reduced the number of BF4 Negative cells by more than 10% compared to just the E8 medium condition.

Furthermore, in the BCL with IWP2 conditions, most of the culture was CD9 High and there was less SSEA3 Low/Negative cells (Figure 36 B-C). The amount of *BRACHYURY* Positive cells also reduced, from ~88% in the standard E8 with CHIRON control conditions to 22%-32% in the BCL with IWP2 conditions (Figure 36 D). While this reduction in *BRACHYURY* was not desired, it nevertheless served as a proof of principle that the addition of the LPA component to the culture system could maintain increased expression of the cell surface antigens that was not achievable with the addition of CHIRON alone.

The results from the first passage showed that BCL was a better basal media component to just E8 (Figure 36). In the second passage, only the BCL with CHIRON and BCL with CHIRON and IWP2 conditions were analysed (Figure 37). The results showed that the addition of IWP2 with BCL again reduced the number of BF4 Negative cells (Figure 37 A), as well as increasing SSEA3 and CD9 levels to the Mid/High range (Figure 37 B-C). The levels of *BRACHYURY* Positive cells also decreased with the addition of IWP2 to BCL compared to the BCL with CHIRON conditions (Figure 37 D). In all conditions with IWP2 added, the percentage of *BRACHYURY* Positive cells dropped to 18-26%. However, the 1 µM and 2 µM conditions of IWP2 had the higher percentage of cells that were *BRACHYURY* Positive at ~26% for the 1 µM condition and ~25% for the 2 µM condition. From both passages these two IWP2 conditions were consistently producing fewer BF4-ve cells and showed consistent percentages of *BRACHYURY* Positive cells. Following this result, both of these IWP2/BCL conditions were further assessed on the third passage.

On the third passage the cells grown in their respective conditions were analysed as before in the previous two passages. Alongside this, the cells were put into standard E8 conditions or the BCL conditions without CHIRON and with or without 1μM/2μM IWP2, to see if the cells could interconvert back to a “unbiased” standard E8 condition phenotype (Figure 38). This was to assess whether the cell cultures still contained undifferentiated cells after three passages of growth in their respective medium. The results showed that putting the cells from the BCL with 1μM of IWP2 condition back into conditions without CHIRON, resulted in fewer BF4 (Figure 38A), CD9 (Figure 38 B) and SSEA3 Negative cells (Figure 38 C) as well as increasing the number of SSEA3 “High” expressing cells and CD9 “High” expressing cells.

Keeping the cells in the same media conditions after three passages also resulted in increased numbers of BF4 and SSEA3 Positive cells. However most of the culture was BF4 and SSEA3 High/Mid expressing and consequently closer to the E8 control conditions. The percentage of *BRACHYURY* expressing cells (Figure 38 D) significantly decreased from 82% in the BCL with CHIRON and 1μM IWP2 condition to just 8% when those same cells were put into E8 control conditions. This result suggested that those cells growing in BCL with CHIRON and IWP2 were not differentiated. Further evidence for this was shown morphologically (Figure 38 E) with the BCL with CHIRON and 1μM IWP2 condition cells forming colonies in E8 conditions, which was not observed when the CHIRON-only treated cells were assessed post passage (Figure 32).

When comparing the differences between the BCL with 1 μM and 2μM IWP2 concentrations in E8, the cells from the 2μM IWP2 conditions resulted in more BF4, CD9 and SSEA3 negative cells at ~40%, ~20%, ~40% respectively compared to ~5% for all the antigens from the cells from the 1μM IWP2 conditions. The cells from the 2μM IWP2 condition also failed to form colonies in E8 conditions. It was also observed that the growth rate of cells from the 2μM IWP2 condition was less than that of the 1μM IWP2 condition with a reduction in confluency from ~ 20-30% after 3 days. Overall these results were similar for the *MIXL1* reporter line and verified that the addition of 1μM IWP2 to the BCL condition (now referred to as Priming Media) was beneficial for maintaining higher levels of expression for the cell surface antigens being analysed.

5.2.5 Verifying whether cells passaged multiple times in Priming Media conditions and interconverted back into standard self-renewing conditions were still pluripotent

To confirm whether the H9 *BRACHYURY* cells grown in Priming Media that interconverted in the E8 media condition were still pluripotent, the population of cells from Priming Media and the recreated SSEA3+/*BRACHYURY*+ subset were dissociated to single cells and either FACS sorted for the SSEA3+/*BRACHYURY*+ subset and plated followed by immunostaining for NANOG (Figure 39 A), or plated for the bulk population followed by immunostaining for SOX2, OCT4 and NANOG (Figure 39 B). The cells from the bulk population predominantly expressed SOX2 and OCT4 at greater than 70% for both across three biological repeats and that NANOG expression was in approximately 40% of the cells in the three repeats. This result provided evidence that cells "primed" towards mesoderm differentiation could be maintained and passaged multiple times, while still maintaining as undifferentiated. For cells sorted from the SSEA3+/*BRACHYURY*+ subset (Figure 39 A), adding in 1µM of IWP2 to E8 upon plating, significantly improved the maintenance of undifferentiated cells. As shown by increased percentages of cells expressing NANOG in ~80% of the culture which was similar in expression to the E8 control condition. The results overall confirmed that the addition of IWP2 to the cells growing in Priming Media conditions helped maintain them as undifferentiated either as a bulk population through multiple passages or on the SSEA3+/*BRACHYURY*+ subset post-sort. This was shown by the increased NANOG expression within the culture, as well as the evidence of the cells forming colonies when put back into standard E8 conditions (Figure 38).

5.2.6 Single cell cloning of the recreated SSEA3+/*BRACHYURY*+ subset to determine if this subset contained undifferentiated cells

The previous results provided evidence that cells growing in the Priming Media condition could be passaged multiple times while still exhibiting expression of the cell surface antigens and pluripotency associated markers. This however was not

evidence of the presence of functional hPSC in this media system. To further verify whether the SSEA3+/*BRACHYURY*+ contained functional hPSC, The SSEA3+/*BRACHYURY*+ cells from Priming Media were sorted and seeded as single cells and left to grow as colonies in a similar vein to the SSEA3+/*BRACHYURY*+ subset from KOSR/MEFS described in chapter 4. A total of over 40 colonies from the Priming Media subset were obtained and firstly assessed if they morphologically looked like stem cell colonies. After which, a total of 6 clones were further analysed functionally. These clones were randomly chosen from index sorted cells, 5 of which were from index sorted cells (Figure 40 A-B).

The remaining clones that were not further passaged were fixed and immunostained for NANOG to determine whether the starting cell was a stem cell. All remaining clones contained NANOG Positive cells (Figure 40 C). The 6 clones carried forward exhibited interconversion through generation of SSEA3+/*BRACHYURY*- cells when maintained in KOSR/MEF and E8/Vitronectin conditions. In the KOSR/MEF conditions with ~60-80% of the cells being SSEA3+/*BRACHYURY*- and 5-25% of the cells being SSEA3+/*BRACHYURY*+ (Figure 41). On E8/Vitronectin in all 6 colonies ~99% of the cells were SSEA3+/*BRACHYURY*- and ~1-2% of the cells were SSEA3+/*BRACHYURY*+ (Figure 42). This trend was similar to the SSEA3+/*BRACHYURY*+ clones from the KOSR/MEF starting population in chapter 4 (Figure 23 and Figure 24). All six clones exhibited NANOG expression in self-renewal conditions within the colonies (Figure 40 C) and the clones were able to differentiate to mesoderm, endoderm and ectoderm as shown by increased gene expression of *TBX6*, *MSGN1* in mesoderm condition , *GATA6*, *SOX17*, *FOXA2* in endoderm conditions and *PAX6*, *SOX1* and *TUBB3* in ectoderm conditions (Figure 40 D). Additionally, after differentiation, all the clones showed expression of TBX6 in mesoderm, SOX17 and FOXA2 in endoderm conditions and PAX6 in ectoderm condition following immunostaining (Figure 40 C). Together, the results provided evidence that the SSEA3+/*BRACHYURY*+ subset from Priming Media contained hPSC.

5.2.7 Comparison of expression of lineage associated markers through high-content clonogenic assays of the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets to demonstrate mesoderm lineage bias

Following recreation of the SSEA3+/*BRACHYURY*+ subset in a defined system as well as evidencing that that population contained hPSC, the next aim was to determine the differentiation potential of this recreated population through both spontaneous differentiation and directed differentiation. The first method to explore this was to use the high-content clonogenic assay method that was described in chapter 4 (Figure 25). The aim was to assess whether the differentiation that forms with the colony was similar to the results from the KOSR SSEA3+/*BRACHYURY* subsets and consequently provide detail into whether the recreated SSEA3+/*BRACHYURY*+ differed in its differentiation potential. In this experiment, single cells were seeded into KOSR/MEF self-renewal conditions and allowed to grow and form colonies for 5 days. On the 5th day, the colonies formed were immunostained for pluripotency associated and differentiated associated markers (Figure 43-46). This was to determine whether more mesoderm differentiation occurred from this recreated SSEA3+/*BRACHYURY*+ subset in the colonies formed, thereby exhibiting a mesoderm lineage bias.

The results from the comparison of the SSEA3+/*BRACHYURY* subsets from the KOSR conditions in chapter 4, revealed that there was not a significant difference in either the total cell numbers that had lineage associated differentiation or the number of colonies that had differentiation in them (Figure 26-Figure 29). In comparison with the SSEA3+/*BRACHYURY*- subset from E8/Vitronectin and the SSEA3+/*BRACHYURY*+ subset from the Priming Media, there were significant differences shown in the total number of cells that had NANOG expression and differentiation associated marker expression within the colonies (Figure 43). In respect to NANOG expression, both SSEA3+/*BRACHYURY* subsets from KOSR/MEF conditions produced colonies that's had more NANOG Positive cells within the colonies. The SSEA3+/*BRACHYURY*- subset from KOSR conditions produced a mean of 48% of cells expressing NANOG across three biological repeats, followed by 32% for the SSEA3+/*BRACHYURY*- subset from KOSR,

19% for the SSEA3+/*BRACHYURY*- subset from E8/Vitronectin and finally 13% for the SSEA3+/*BRACHYURY*+ subset from Priming Media.

For SOX17+/OCT4+ co-expression (Figure 43), the SSEA3+/*BRACHYURY*+ subset from Priming Media generated colonies containing the most expression within the colonies, with a mean of 23% of cells expressing across the biological repeats, compared to 11% for the SSEA3+/*BRACHYURY*- subset from KOSR, 10% for the SSEA3+/*BRACHYURY*+ subset from KOSR and 4% for the SSEA3+/*BRACHYURY*- subset from E8/Vitronectin. For TBX6, the mean percentages were overall lower across all the subsets, but the results were that the SSEA3+/*BRACHYURY*+ subset from KOSR generated colonies containing a mean expression of 11% of the cells compared to just 3% from the SSEA3+/*BRACHYURY*+ subset from Priming Media. The two SSEA3+/ *BRACHYURY*- subsets were closer in mean average at 9% for the KOSR subset and 7% for the E8/Vitronectin subset. For FOXA2+/*BRACHYURY*– (Figure 44) the mean averages were again similar across all of the subsets with a range of approximately 5% to 3% of cells expressing these markers within the colonies. For FOXA2+/*BRACHYURY* + (Figure 44) there was a bigger range, with the SSEA3+/ *BRACHYURY*- subset from KOSR/MEF producing a mean of 10% of the cells within colonies expressing these markers, followed by 9% from the SSEA3+/*BRACHYURY*- subset from KOSR/MEF, then 3% from the SSEA3+/*BRACHYURY*- subset from E8/Vitronectin and finally approximately 1% from the SSEA3+/*BRACHYURY*+ subset from Priming Media. Finally, for PAX6 expression, the SSEA3+/ *BRACHYURY*+ subset from KOSR/MEF produced the most at 2% of the cells in colonies expressing PAX6 compared to around 1% for the other three subsets (all analysis performed using a Tukey multiple comparison test significant results determined as $p<0.05$ n=3).

There was however no significant difference in the distribution of lineage marker expression and OCT4 expression in these colonies across the subsets (Figure 45), nor a significant difference in the number of colonies that contained at least one differentiated cell (Figure 46). The results therefore suggest that there are differences present between the SSEA3+/*BRACHYURY* subsets and specifically between the two SSEA3+/*BRACHYURY*+ subsets in respect to expression of lineage associated genes.

5.2.8 Assessment of the differentiation potential of cells growing in Priming Media

The results previously described have provided evidence of a media system that can allow induction and maintain of a *BRACHYURY* Positive state over multiple passages, as well as showing differing potential for differentiation in clonogenic assay conditions. Using this media system, the next objective was to functionally assess the differentiation potential of the bulk population of cells growing in Priming Media (Figure 47), as well as the specifically sorted, SSEA3+/*BRACHYURY*+ subset (Figure 48).

The results showed that with the population differentiation (Figure 47), there was no significant difference in the efficiency of mesoderm differentiation of cells growing in Priming Media compared to cells growing in E8 as shown by compared percentages of TBX6 Positive cells at >90% (student T test $p>0.05$ n=3) (Figure 47 A). Gene expression analysis showed that the expression levels of *TBX6* and *MSGN1* were comparable in both differentiated populations, but both significantly higher than the E8 undifferentiated control conditions (Figure 47 A) (Tukey Multiple Comparison Test $p<0.05$ n=3).

For the endoderm differentiation (Figure 47 B), while the percentages of SOX17+/FOXA2- and SOX17+/FOXA2+ were both lower than 40% which was lower than that of TBX6 in the mesoderm condition, there again was no significant difference in the efficiency of endoderm differentiation compared to the E8 condition (Tukey Multiple Comparison test $p>0.05$ n=3). However, the gene expression analysis revealed that the cells differentiated that had been grown from the E8 condition had greater expression of *GATA6*, *SOX17* and *FOXA2* compared to the Priming Media conditions when differentiated to endoderm. Though both differentiated populations had significantly higher expression of these three genes than the E8 undifferentiated control conditions. (Tukey Multiple Comparison Test $p<0.05$ n=3).

For the ectoderm differentiation (Figure 47 C) , there was a significant difference in the efficiency of differentiation as shown by the percentage of PAX6 Positive cells from the Priming Media condition at 40-65% compared to the cells from the E8 condition at >90% (student T test $p<0.05$ n=3). There were also significantly

higher levels of *PAX6* on the gene expression analysis for the E8 condition. However, both differentiated populations had significantly higher expression of *PAX6*, *SOX1* and *TUBB3* than the E8 undifferentiated control conditions as seen in (Figure 47) (Tukey Multiple Comparisons test $p<0.05$ n=3).

Analysis of the differentiation potential of the SSEA3+/ *BRACHYURY*+ Priming Media subset compared to the SSEA3+/*BRACHYURY*- E8 subset post-sort, yielded different results to the bulk differentiation in respect to endoderm and ectoderm differentiation (Figure 48). In respect to mesoderm differentiation, the number of cells after differentiation were also similar (Figure 48 A) and again, there was not a significant difference in the number of TBX6 Positive cells, both yielding around 80% (Figure 48 B) (student T test $p>0.05$ n=3).

For the endoderm condition (Figure 48 C), the efficiency of differentiation was again yielding fewer differentiated cells to that of the mesoderm condition, but there was a significant difference in the percentage of SOX17+/FOXA2+ and SOX17+/FOXA2- cells, with the SSEA3+/*BRACHYURY*- population producing more SOX17+/FOXA2+ cells and the SSEA3+/*BRACHYURY*+ population producing more SOX17+/FOXA2- cells (Tukey Multiple Comparison Test $p<0.05$ n=3) (Figure 48). There was however not a significance difference in the total number of cells post-differentiation (student T test $p>0.05$ n=3)

For the ectoderm condition (Figure 48 D), there was a significant difference in the survival of the two populations post-differentiation. The SSEA3+/ *BRACHYURY*- population failed to survive across three biological repeats, while there were pockets of cells from the SSEA3+/ *BRACHYURY*+ subset which did survive, though the numbers were significantly fewer than the mesoderm and endoderm conditions. These cells which did survive were also heterogeneous to expression of PAX6 and the HOX gene, HOXC9. This suggests that a smaller subset of cells existed within the recreated SSEA3+/*BRACHYURY*+ subset that was biased towards an NMP or posterior-related differentiated derivative.

5.2.9 Analysis of whether the SSEA3+/*BRACHYURY*+ subset contained a self-renewing NMP population in the Priming Media condition

The results from the clonogenic assays (Figure 40 and Figure 41) and the directed differentiation post-sort (Figure 48) revealed that the recreated SSEA3+/*BRACHYURY*+ subset was heterogenous and contained within it a subset of cells that potentially contained an NMP sub-population. To assess this, hPSC grown for 1 week in the Priming Media conditions with passages in between, were immunostained for markers associated with NMP differentiation (Figure 49). The results showed in respect to SOX2 and *BRACHYURY* expression, the over 50% of the SOX2 and *BRACHYURY* co-expressing cells also expressed NANOG, while a further 20% of SOX2+ cells were *BRACHYURY* Negative but also expressed NANOG. There was also approximately 20% of cells that were just NANOG Positive. One of the hallmarks of an NMP population is that of SOX2 and *BRACHYURY* co expression, but not in the presence of NANOG [238], so the results suggest that the Priming Media culture contained primarily undifferentiated hPSC and not NMP. However, over 90% of *BRACHYURY* Positive cells analysed also co-expressed CDX2, potentially suggesting the presence of a population readying to differentiate towards paraxial mesoderm.

5.2.10 Gene expression comparison between the two SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets

The results from the clonogenic assays (Figure 41) revealed that the SSEA3+/*BRACHYURY*- subsets from KOSR/MEF and E8/Vitronectin and the SSEA3+/*BRACHYURY*+ subsets from KOSR/MEF and the Priming Media were heterogeneous in respect to expression of lineage associated markers. This presence of heterogeneity was further evidenced with the subset of ectoderm differentiated cells from the SSEA3+/*BRACHYURY*+ subset that co-expressed HOXC9 and PAX6 (Figure 48). These results suggested that these two respectiveSSEA3+/*BRACHYURY*+ subsets were different in respect to their gene expression profiles, as well as their functional propensity for differentiation.

Furthermore, the SSEA3+/ *BRACHYURY*+ subset from Priming Media also showed co-expression of CDX2 and *BRACHYURY* when grown in the media for a prolonged period (Figure 49). These results together, suggest that there are subsets of cells within the SSEA3+/ *BRACHYURY*+ subsets that have different propensities for differentiation.

To further explore this, these subsets were analysed by QPCR and compared on the population level to see the differences in these gene expression patterns and provide clues as to what these subsets may be within these subsets (Figure 50). The results showed that indeed these subsets were different in respect to pluripotency associated markers and lineage specific markers attributed to early mesoderm and endoderm differentiation. The SSEA3+/*BRACHYURY*- subset from E8 showed increased levels of the pluripotency associated markers *NANOG* and *ZFP42*, while *POU5F1* was higher in the SSEA3+/*BRACHYURY*+ subset from the KOSR/MEF condition.

The SSEA3+/*BRACHYURY*+ subset from the Priming Media had the lowest levels of expression of the pluripotency associated markers compared to the other subsets. Conversely, this subset had higher levels of expression of *BRACHYURY*, *EOMES*, *CDX1*, *CDX2*, *HOXC9*, *GBX2* and *GFAP* compared to the SSEA3+/*BRACHYURY*+ subset from KOSR/MEF and the two SSEA3+/*BRACHYURY*- subsets. In contrast, the SSEA3+/ *BRACHYURY*+ subset from the KOSR/MEF condition had the highest levels of expression of *MESP1* and *TBX6*. *GATA4*, *GATA6* and *MIXL1* were comparable amongst both SSEA3+/ *BRACHYURY*+ subsets but higher than both SSEA3+/ *BRACHYURY*- subsets. Though *AFP*, *FOXA2* and *SOX17* were higher in the SSEA3+/ *BRACHYURY*+ subset from KOSR/MEF compared to the other three populations. The levels of *PAX6* expression were comparable to the KOSR/MEF SSEA3+/ *BRACHYURY*- subset, though higher than that of the SSEA3+/ *BRACHYURY*+ subset from Priming Media and SSEA3+/ *BRACHYURY*- subset from E8. Together these results demonstrate that the SSEA3+/ *BRACHYURY* subsets from the different media conditions are heterogeneous and different from one another.

5.2.11 Comparison of the differentiation potential of the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets in embryoid body conditions

In both gene expression analysis and high-content clonogenic assay analysis, comparison of the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets that are found in KOSR/MEF or E8/Vitronectin conditions respectively or the recreated SSEA3+/BRACHYURY subset from Priming Media reveal differences in upregulation of lineage associated genes. Another approach to assess the SSEA3+/*BRACHYURY* subsets functionally, was the Neutral condition, EB forming assay as previously described in chapter 3 and 4 (Figure 4 and Figure 30). The results of the Neutral EB differentiation supported the evidence from the gene expression analysis that these SSEA3+/*BRACHYURY* subsets were different in their propensities for differentiation (Figure 51 and Figure 52). As discussed in Chapter 4, there was variation amongst the biological replicates resulting in a lack of consistent reproducibility for some genes. However, for genes such as the extraembryonic mesoderm genes, *HAND1* and *GATA2*, as well as the paraxial mesoderm gene *MEOX1*, the results in this experiment showed there was a significant difference between the SSEA3+/*BRACHYURY*+ subsets.

In respect to ectoderm differentiation, this trend was also shown in the expression of *PAX6*, where the Priming Media SSEA3+/*BRACHYURY*+ resulted in less expression of *PAX6* to both SSEA3+/ *BRACHYURY* subsets from KOSR conditions in some biological replicates. However, similarly to *CD34*, the variation across the replicates resulted in a lack of a reproducible trend.

The SSEA3+/*BRACHYURY*+ subset from Priming Media did however have higher levels of *SOX1* than *PAX6* and which was comparable to the KOSR equivalent, potentially suggesting an ectoderm subset present within that population. The Priming Media SSEA3+/ *BRACHYURY*+ subset also resulted in significantly more *PAX3* and *SOX10* expression to that of the KOSR SSEA3+/ *BRACHYURY*+ suggesting increased propensity to neural crest differentiation ($p<0.05$ using Tukey Multiple Comparison Test). What was also consistent across the three biological replicates was elevated expression of *GATA4*, *GATA6* and *SOX17* in the Priming Media SSEA3+/*BRACHYURY*+ subset, however while

expression was not as high as *PAX3* and *SOX10*, *GATA4* and *SOX17* were nevertheless consistent higher expressed in the Priming Media, SSEA3+/ *BRACHYURY*+ subset compared to some of the mesoderm associated genes within that subset. This could suggest a subset of cells within this population with propensity for endoderm differentiation as reflected in the gene expression analysis (Figure 50). Overall the results support the notion that these SSEA3+/ *BRACHYURY* subsets are heterogeneous and are different in their propensities for differentiation to different lineages.

Figure 32: **Mesoderm induced hPSC cannot be maintained as undifferentiated after passage into self-renewal conditions**

H9 *BRACHYURY* single cells were induced to mesoderm differentiation and passaged into self-renewal conditions. Plots A), B) and C) are representative flow cytometry histograms showing the loss of expression of BF4, CD9 and SSEA3 post-passage. E) Control plots in Red, CHIRON post-passage in Yellow, Negative control in Blue. All 3 antigens showed loss of expression post-passage. D) Expression of the pluripotency associated marker NANOG was also lost post-passage. E) Cells post passage from CHIRON condition failed to form colonies and exhibited a mesenchymal morphology. Together with A-D, results suggest the cells have differentiated. Scale bar set at 50µm. Statistical significance marked with asterisk (*) $p<0.05$ n=3.

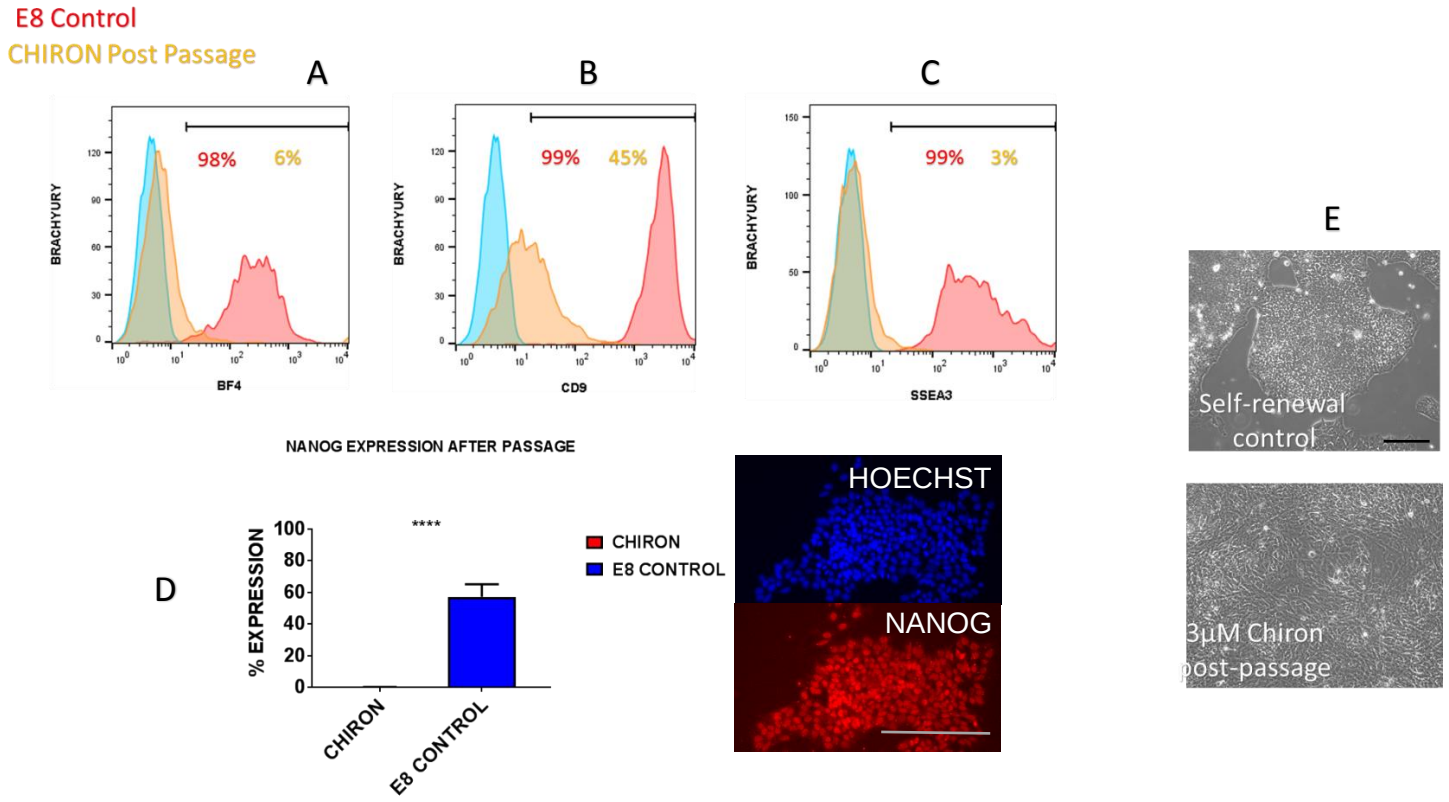


Figure 33: Titration of BCL to find optimal level to induce an SSEA3+/*BRACHYURY*+ subset to maintain it over multiple passages

Representative Flow cytometry histograms for *BRACHYURY* expression for different concentrations of BCL and the E8 control in red: 0.3 µM in blue, 0.6 µM in yellow, 1.2 µM in light green, 2.4 µM in dark green and 4.8 µM in pink. Greater than 95% *BRACHYURY* expression was observed at the 0.3 µM and 0.6 µM titres. 0.48 µM BCL was chosen for media to be comparable with the *MIXL1* reporter which obtained similar titre results to *BRACHYURY*. Experiment performed with 2 biological and 2 technical repeats, ~1-2% difference in expression between replicates.

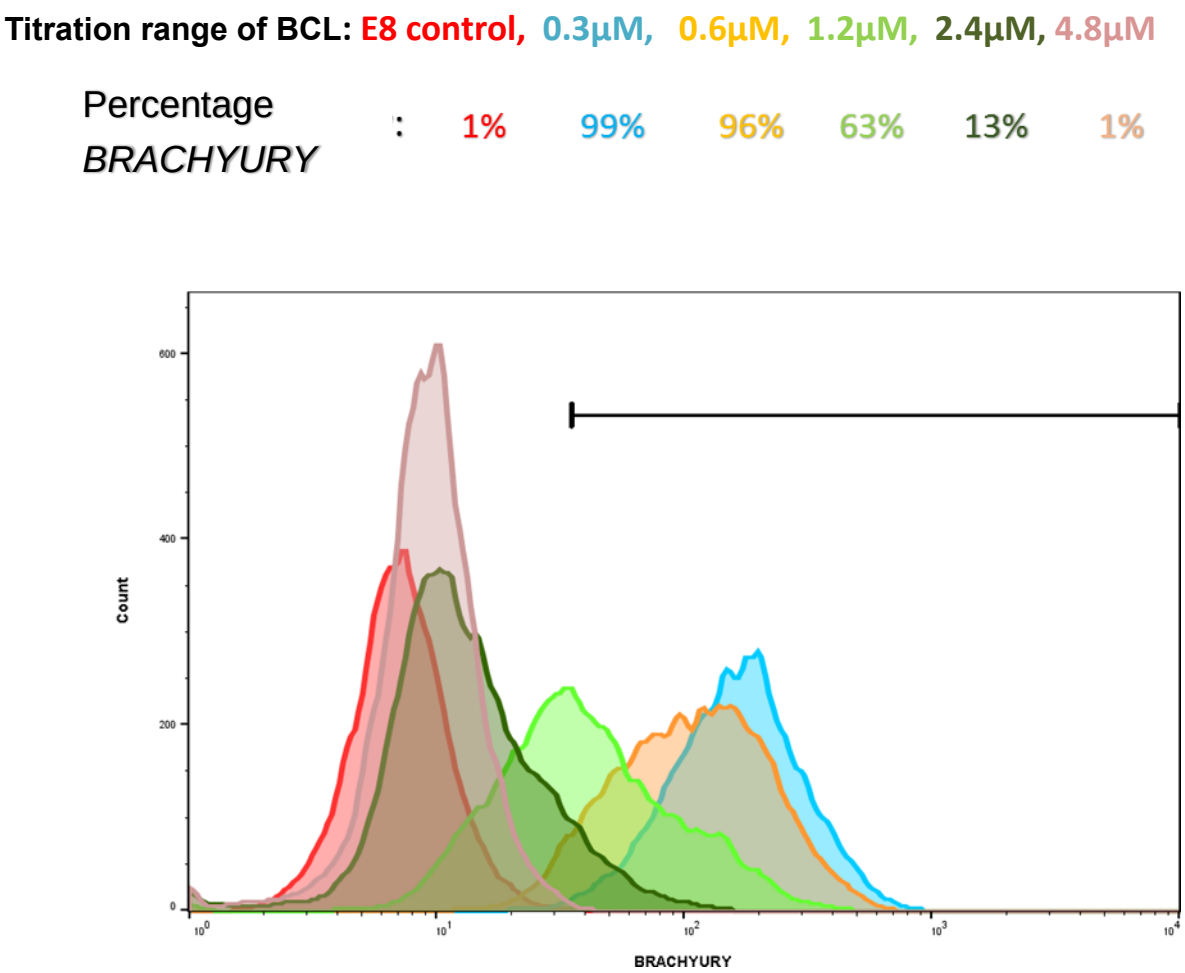


Figure 34: Addition of BCL to mesoderm differentiation protocols reduces the percentage of *BRACHYURY* Positive cells

H9 *BRACHYURY* reporter cells were seeded as single cells and induced to differentiate in different mesoderm induction medias. 3µM CHIRON, $AA^{50}/BMP4^{40}$ and AA^{100}/BMP^{5}.

Representative flow cytometry plots for (A) BF4, (B) CD9 and (C) SSEA3 against *BRACHYURY*. Addition of 0.48µM BCL reduced the amount of BF4 and SSEA3 Negative cells in all conditions. BCL+ CHIRON produced the most antigen/*BRACHYURY* double Positive cells compared to $AA^{50}/BMP4^{40}$ and AA^{100}/BMP^{5}. CHIRON+BCL chosen as ideal *BRACHYURY* inducer to trap mesoderm biased state. Control Red, mesoderm inducer Blue, Inducer+ BCL Yellow. BF4, CD9, SSEA3 on X-Axis, *BRACHYURY* on Y-axis. n=3

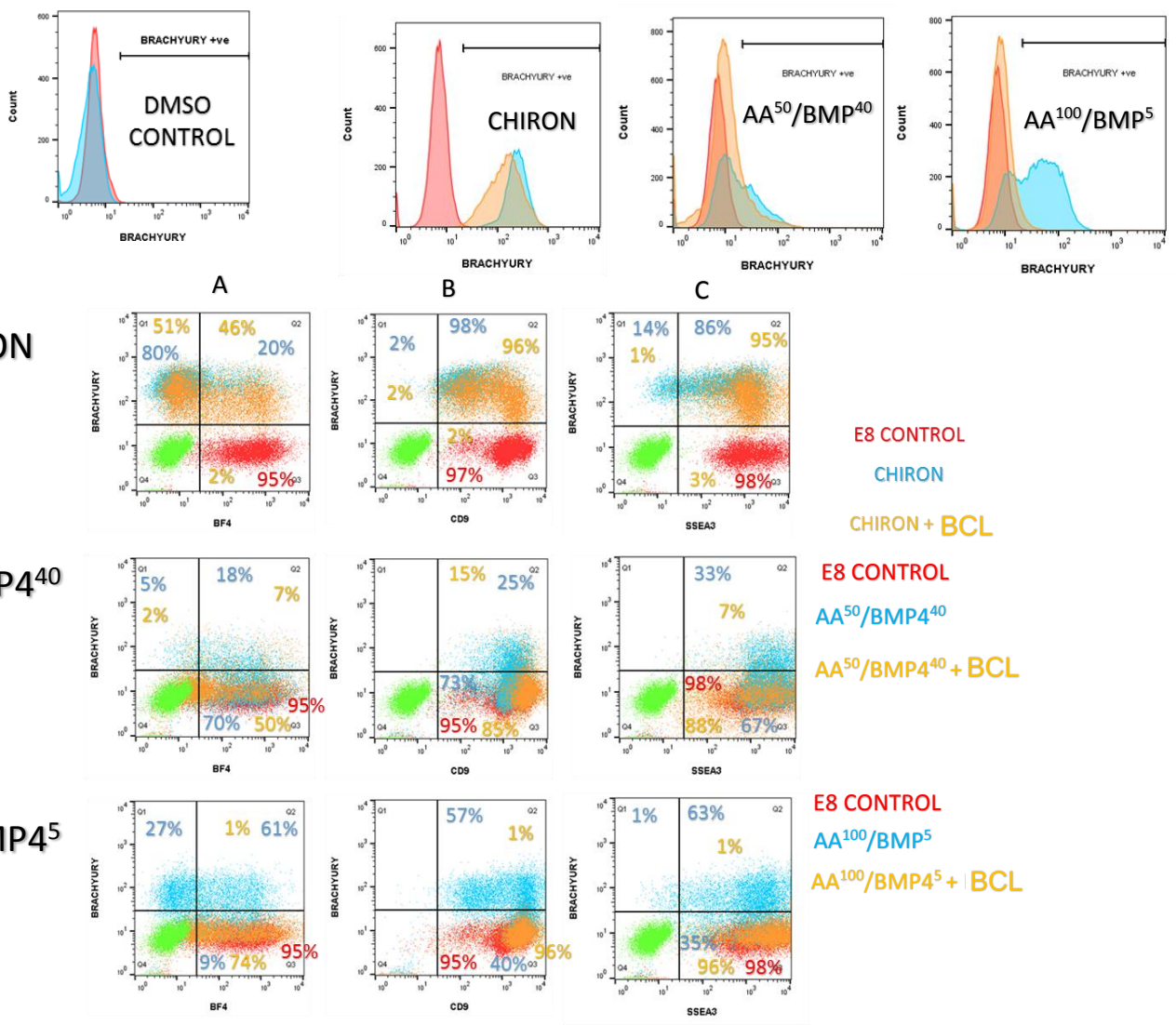


Figure 35: Schematic of workflow for determining optimal composition for Priming Media to recreate and maintain an SSEA3+/*BRACHYURY*+ state

A schematic detailing the procedure used to determine the Priming Media composition that will be used to recreate, maintain and analyse the SSEA3+/*BRACHYURY*+ subset in defined conditions. Over 3 passages, cells growing in various media compositions (detailed in results section) were analysed via flow cytometry for BF4, CD9 and SSEA3 to assess percentage of differentiation, defined by SSEA3 and BF4 Negative cells. Hypothesis for ideal composition were cells maintaining colony formation and maintaining strong levels of BF4, CD9 and SSEA3 expression over the 3 passages.

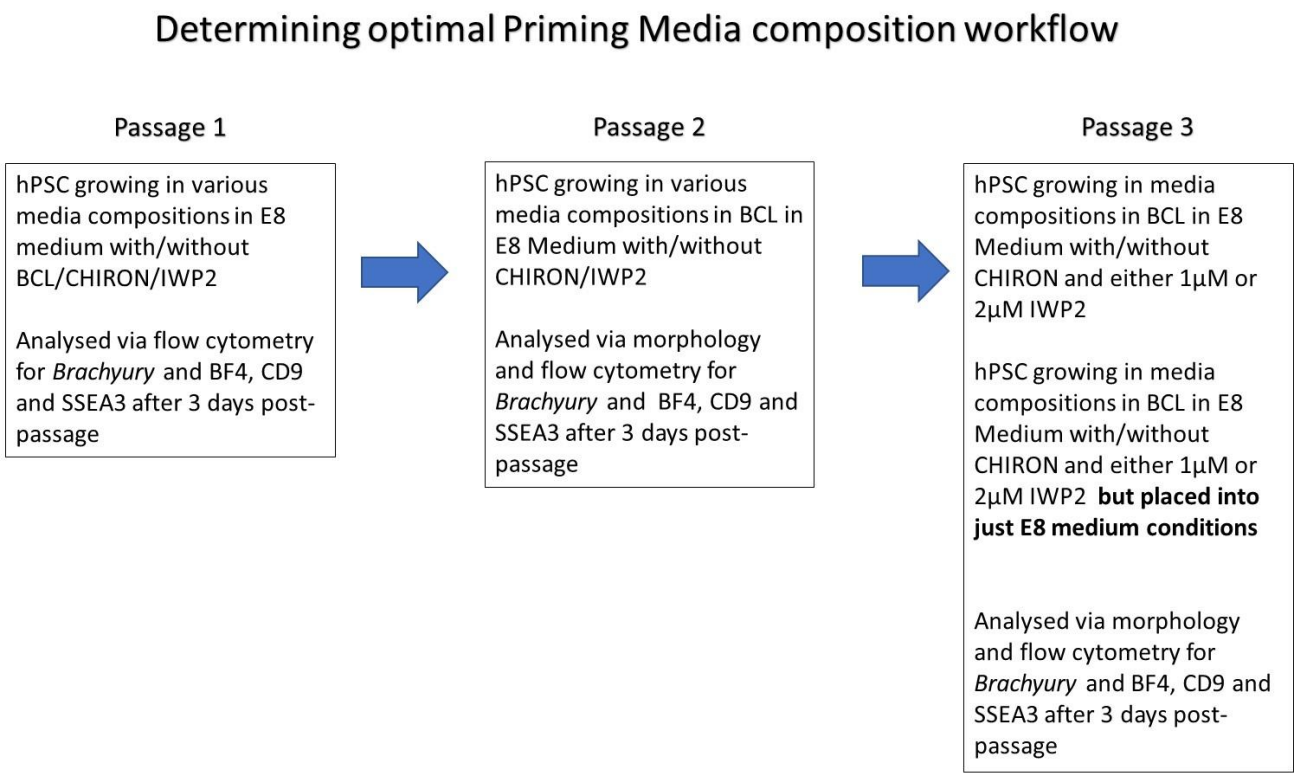


Figure 36: Comparison of antigen expression with different media compositions of LPA and the endogenous Wnt inhibitor IWP2 to find optimal conditions to maintain hPSC over multiple passages Part 1

Flow cytometry histograms exhibiting distribution of antigen expression High/Mid/Low/ Negative (-ve) post passage for A) BF4, B) CD9 and C) SSEA3 in E8 and the LPA component (now referred to as BCL) with CHIRON with/without IWP2 at the following concentrations: Control, CHIRON and CHIRON with 0.25µM through to 2 µM IWP2. D) Representative plot on how antigen expression was separated into HIGH/MID/LOW/Negative (-ve). E) Flow cytometry histogram for *BRACHYURY* expression in different conditions: E8 control in red, BCL with Chiron in yellow, BCL with Chiron 0.25 µM IWP2 in light green, BCL with Chiron 0.5 µM IWP2 in dark green, BCL with CHIRON 1 µM IWP2 in pink and BCL with Chiron 2 µM IWP2 in purple. Addition of IWP2 reduced the percentage of antigen Negative cells. Cells in BCL with CHIRON and 1 µM and 2µM IWP2 condition expressed the most *BRACHYURY* post-passage.

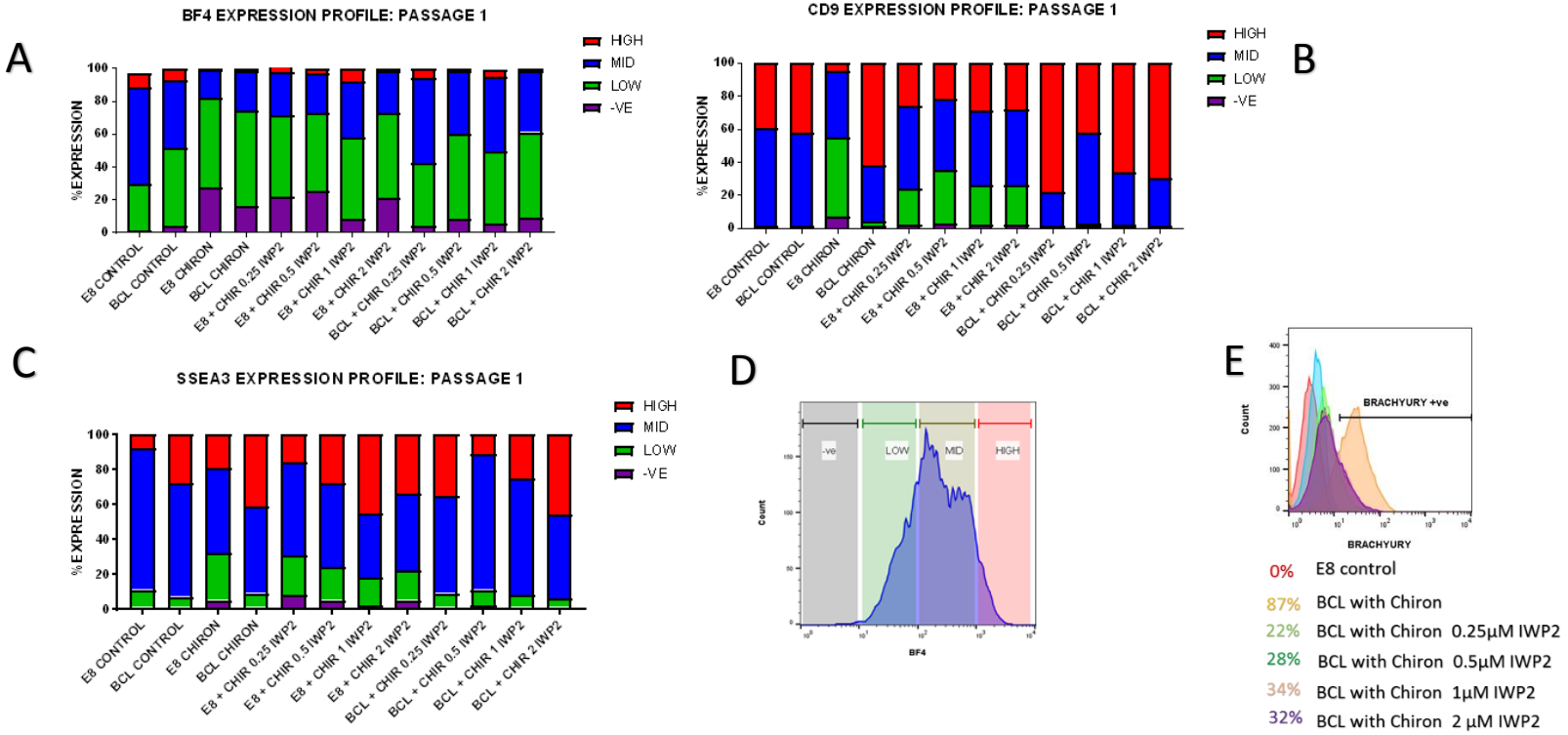


Figure 37: Comparison of antigen expression with different media compositions of LPA and the endogenous Wnt inhibitor IWP2 to find optimal conditions to maintain hPSC over multiple passages Part2

Flow cytometry histograms exhibiting distribution of antigen expression HIGH/MID/LOW/Negative (-ve) post 2nd passage for A) BF4, B) CD9 and C) SSEA3 in E8 and BCL with CHIRON with/without IWP2 at the following concentrations: Control, CHIRON and CHIRON with 0.25µM through to 2 µM IWP2. D) Flow cytometry histogram for *BRACHYURY* expression in different conditions: E8 control in red, BCL with Chiron in yellow, BCL with Chiron 0.25 µM IWP2 in light green, BCL with Chiron 0.5 µM IWP2 in dark green, BCL with CHIRON 1 µM IWP2 in pink and BCL with Chiron 2 µM IWP2 in purple. E) Brightfield images of BCL with CHIRON only and with 1 µM and 2µM IWP2. Similarly, to 1st passage, addition of IWP2 reduced the percentage of antigen Negative cells. Cells in BCL with 1 µM and 2µM IWP2 condition expressed the most *BRACHYURY* post-passage. Scale bar set at 50µm.

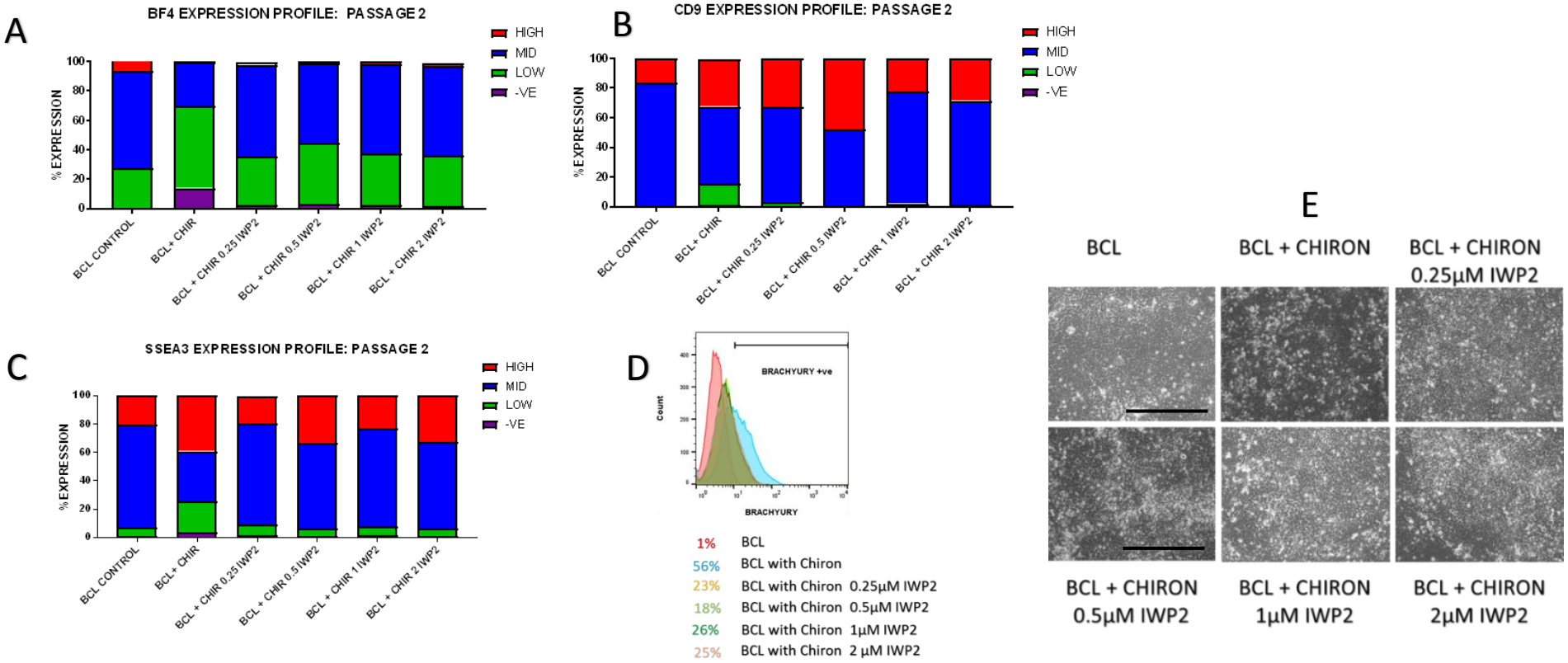


Figure 38: Comparison of antigen expression with different media compositions of LPA and the endogenous Wnt inhibitor IWP2 to find optimal conditions to maintain hPSC over multiple passages Part3

Flow cytometry histograms exhibiting distribution of antigen expression HIGH/MID/LOW/ Negative (-ve) post 3rd passage for A) BF4, B) CD9 and C) SSEA3 in following conditions: E8 control, cells from BCL into E8 media, cells from BCL1 into BCL, cells from BCL1 into BCL1, cells from BCL1 into E8, cells from BCL2 into BCL and cells from BCL2 into E8. D) Flow cytometry histogram for *BRACHYURY* expression in different conditions: E8 control RED, BCL with CHIRON in 1µM IWP2 in Yellow and BCL with CHIRON 1 µM IWP2 in E8 in Blue. E) Bright field images of cells from BCL only, BCL with CHIRON and with 1 µM and 2µM IWP2 respectively into E8 or BCL basal. BCL with CHIRON in 1µM IWP2 produced fewer antigen Negative expressing cells, loss of *BRACHYURY* expression when put into E8 conditions and formed colonies. BCL with CHIRON in 1µM IWP2 conditions was chosen as media condition to induce and maintain mesoderm biased state. Scale bar set at 50µm.

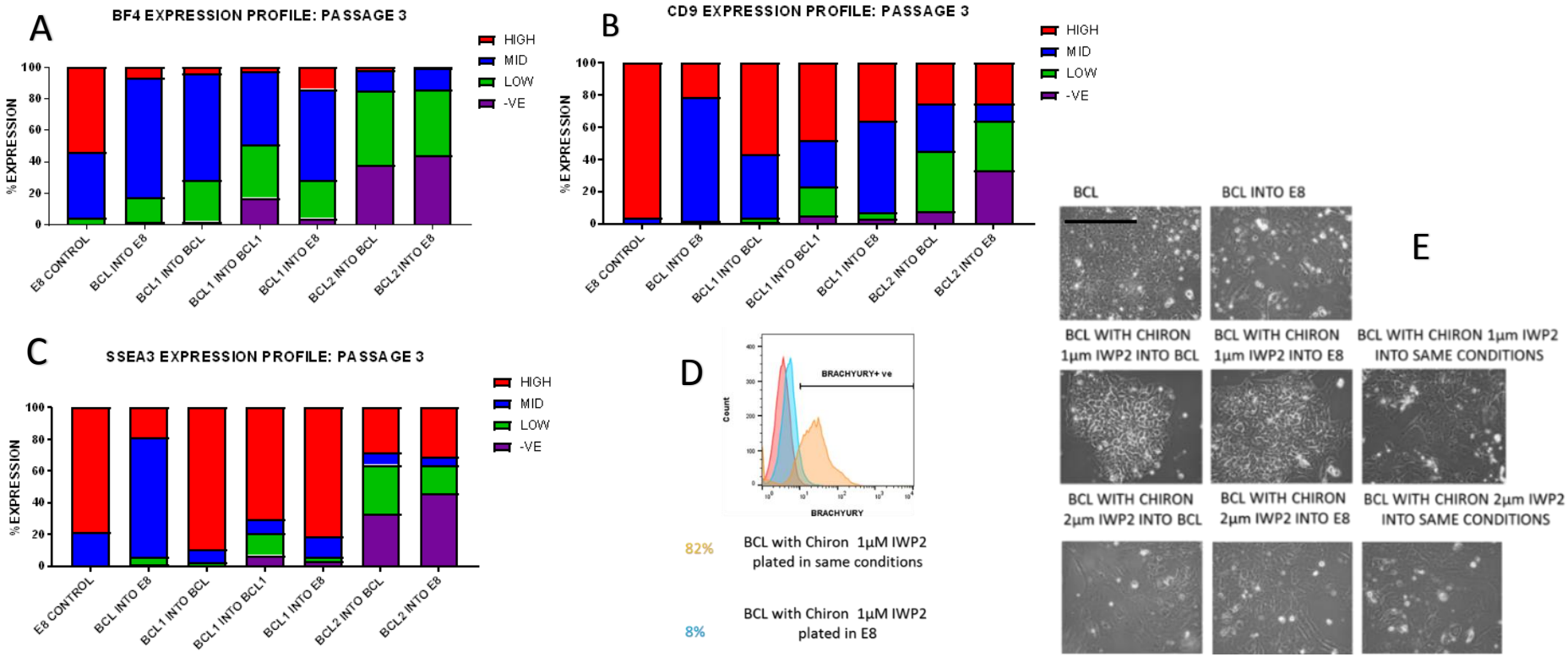
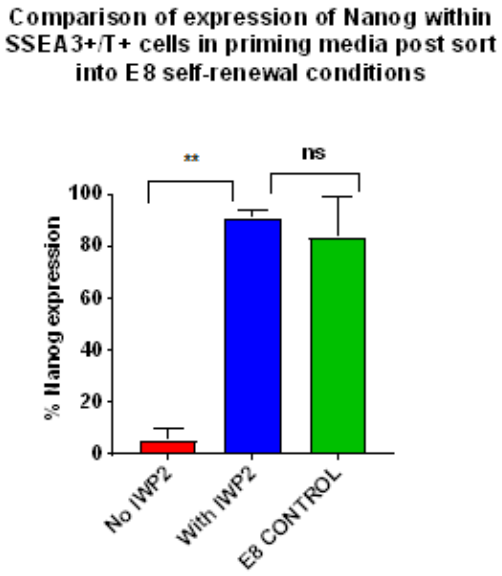


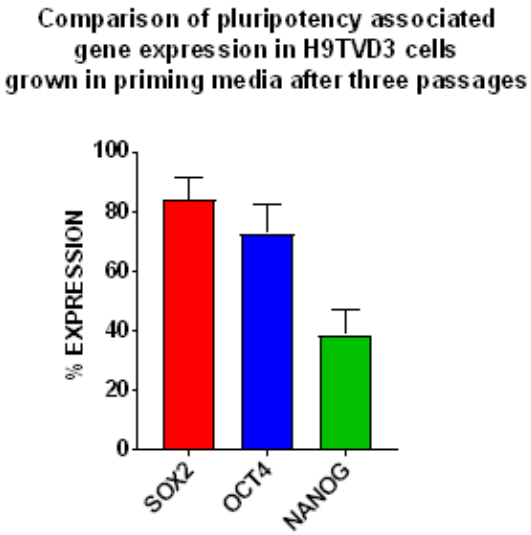
Figure 39: hPSC grown in Priming Media can be maintained over multiple passages and the addition of IWP2 helps maintain pluripotency after seeding single cells after sorting

Analysis of H9 *BRACHYURY* reporter cells (H9TVD3 cell line) grown in Priming media and the SSEA3+/*BRACHYURY* (T) + subset sorted after multiple passages. A) Percentage of expression of NANOG in SSEA3+/*BRACHYURY* (T)+ cells post sort that were sorted from Priming Media and seeded om E8 self-renewal conditions with IWP2, without IWP2 and compared to the E8 control. B) Percentage of gene expression of SOX2, OCT4 and NANOG for H9 *BRACHYURY* cells in Priming Media after 3 passages. C) Representative pictures of expression for OCT4, SOX2 and NANOG. Scale bar set at 50µm. Statistical significance marked with asterisk (*) $p<0.05$ ns=not significant n=3.

A



B



C

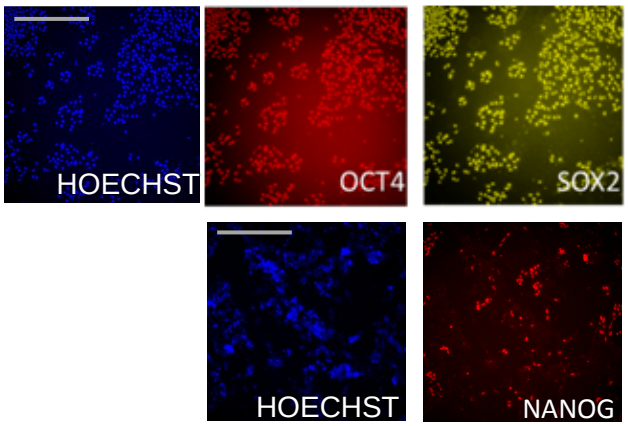


Figure 40: Summary of single cell cloning of Priming Media SSEA3+ *BRACHYURY*+ subset exhibits the starting cells were pluripotent

To further verify whether the SSEA3+/*BRACHYURY*+ subset contained functional hPSC, SSEA3+/*BRACHYURY*+ cells were seeded as single cells and left to grow as colonies before functional analysis. (A) Flow cytometry plot showing SSEA3+/*BRACHYURY*+ sorted subset. (B) Position of index sorted SSEA3+ *BRACHYURY*+ (T) cells. (C) Representative picture of immunofluorescence for NANOG in self-renewal conditions, SOX17, TBX6 and PAX6 after differentiation. (D) Transcriptome analysis. Average 1/ΔCT of the clones in Priming Media for mesoderm differentiation (TBX6 and MSGN1), endoderm differentiation (GATA6, SOX17 and FOXA2) and ectoderm differentiation (PAX6, SOX1 and TUBB3) compared to E8 undifferentiated control. All 6 clones could differentiate to derivatives of the three germ layers and express NANOG in self-renewal conditions, evidencing the starting cells must have been pluripotent stem cells. Scale bar set at 100µm. Statistical significance marked with asterisk (*) p<0.05

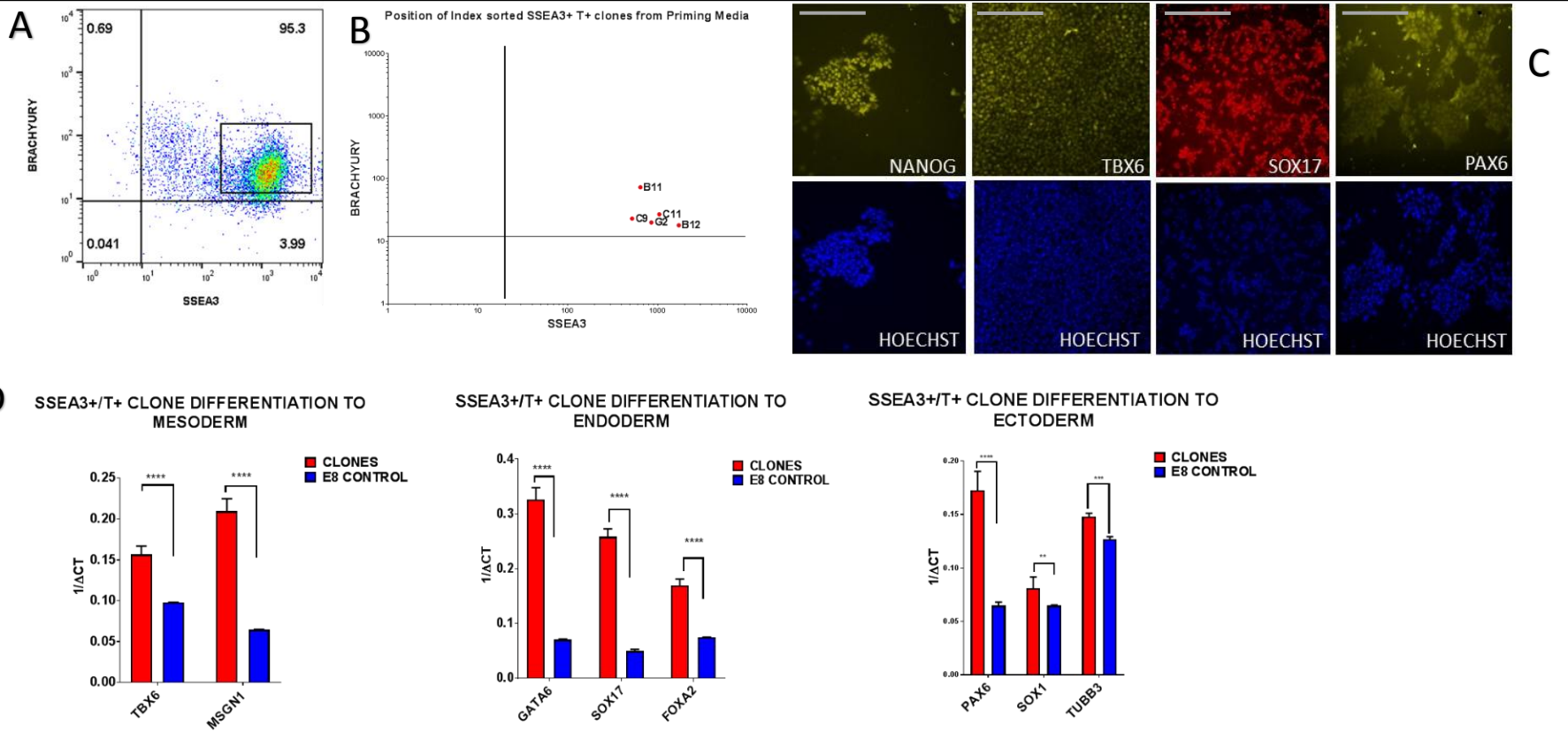
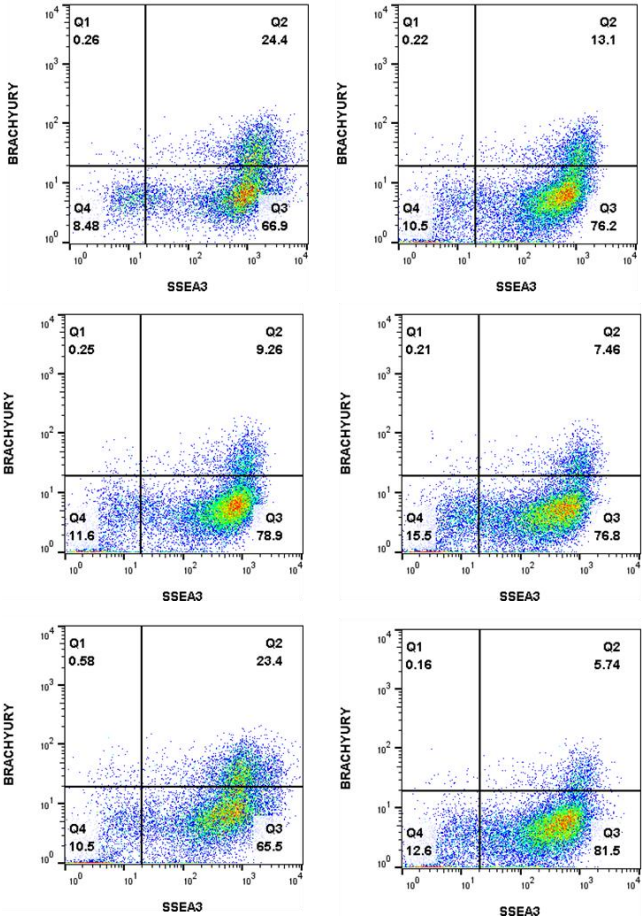


Figure 41: Analysis of Priming Media SSEA3+/*BRACHYURY*+ clones grown on KOSR/MEFS after 5 passages demonstrate interconversion through emergence of SSEA3+/*BRACHYURY*- subset

Flow cytometry plot showing SSEA3 expression from the single cell clones in KOSR/MEF after 5 passages. (A) After 5 passages, the 6 clonal lines had interconverted from a SSEA3+/*BRACHYURY* (T)+ cells to a heterogeneous population of SSEA3+/*BRACHYURY* (T)- cells and SSEA3+/*BRACHYURY*+ cells. (B) Reference of the FACS plot from the initial cell sort.

A



B

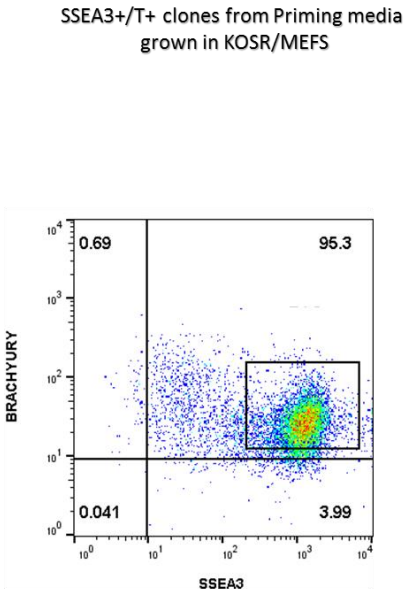


Figure 42: Analysis of Priming Media SSEA3+/*BRACHYURY*+ clones grown on E8/Vitronectin after 5 passages reveal interconversion through loss of *BRACHYURY* expression

Flow cytometry plot showing SSEA3 expression from the single cell clones in E8/Vitronectin after 5 passages. (A) After 5 passages, the 6 clonal lines had interconverted from a SSEA3+/*BRACHYURY* (T)+ cells to a predominantly homogenous population of SSEA3+/*BRACHYURY* (T)- cells. (B) Reference of the FACS plot from the initial cell sort.

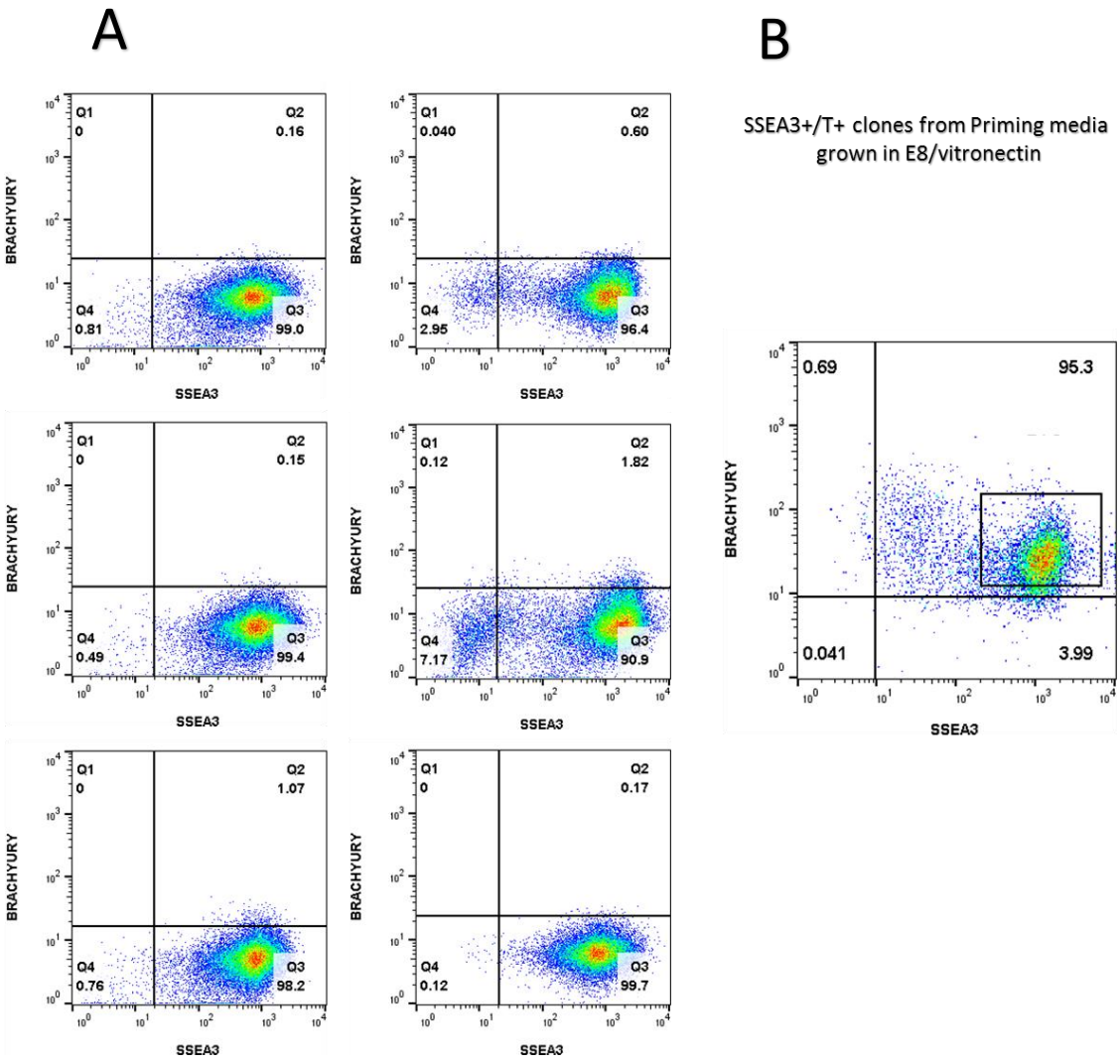


Figure 43: High content cloning assays of SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets from KOSR, Priming media and E8 reveal differential expression of markers associated with self-renewal and lineage specific differentiation

To determine whether the SSEA3+/*BRACHYURY* (T)+ subset from Priming Media exhibited a bias to differentiation, a clonogenic assay was performed in comparison with the SSEA3+/*BRACHYURY*(T)- subset from E8 and compared to the SSEA3+/*BRACHYURY* subsets from KOSR/MEF conditions. Comparison of the percentage of total number of cells expressing NANOG (A), SOX17 (B) and TBX6 (C) within OCT4+ colonies for the following conditions:

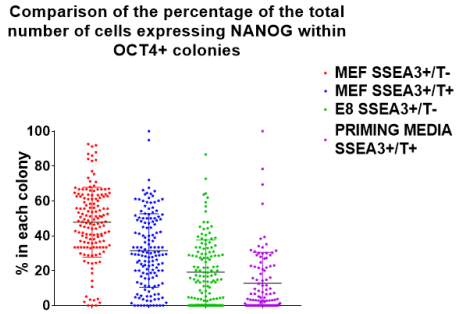
MEF SSEA3+/*BRACHYURY*-,

MEF SSEA3+/*BRACHYURY*+,

E8 SSEA3+/*BRACHYURY*-,

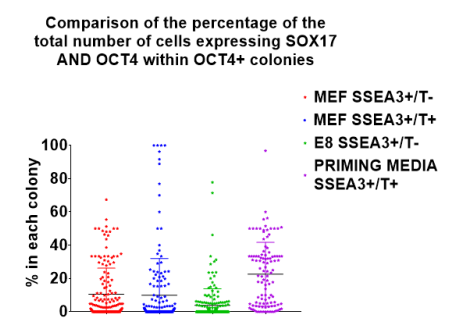
Priming Media SSEA3+/*BRACHYURY*+.

To each comparison was an associated a table registering the statistical significant of each condition between the others. KOSR/MEF data taken from Chapter 4 analysis. Average of 3 biological repeats.



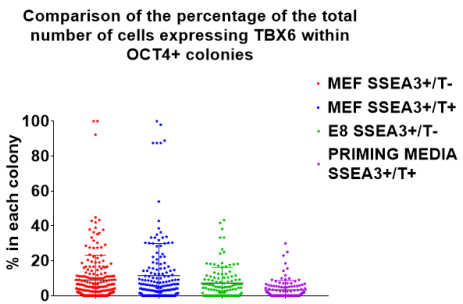
A

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Statistical Significance	Adjusted P Value
MEF SSEA3+/T- vs. MEF SSEA3+/T+	16.31	10.48 to 22.14	****	<0.0001
MEF SSEA3+/T- vs. E8 SSEA3+/T-	28.66	22.67 to 34.66	****	<0.0001
MEF SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	35.02	28.55 to 41.49	****	<0.0001
MEF SSEA3+/T+ vs. E8 SSEA3+/T-	12.35	6.348 to 18.36	****	<0.0001
MEF SSEA3+/T+ vs. PRIMING MEDIA SSEA3+/T+	18.7	12.22 to 25.18	****	<0.0001
E8 SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	6.353	-0.2732 to 12.98	ns	0.0657



B

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Statistical Significance	Adjusted P Value
MEF SSEA3+/T- vs. MEF SSEA3+/T+	0.6741	-4.092 to 5.44	ns	0.9835
MEF SSEA3+/T- vs. E8 SSEA3+/T-	6.737	2.116 to 11.36	**	0.0011
MEF SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	-12.01	-17.4 to -6.609	****	<0.0001
MEF SSEA3+/T+ vs. E8 SSEA3+/T-	6.063	1.501 to 10.62	**	0.0037
MEF SSEA3+/T+ vs. PRIMING MEDIA SSEA3+/T+	-12.68	-18.03 to -7.334	****	<0.0001
E8 SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	-18.74	-23.96 to -13.53	****	<0.0001



C

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Statistical Significance	Adjusted P Value
MEF SSEA3+/T- vs. MEF SSEA3+/T+	-2.446	-5.962 to 1.071	ns	0.2783
MEF SSEA3+/T- vs. E8 SSEA3+/T-	1.834	-2.132 to 5.801	ns	0.6326
MEF SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	5.634	1.598 to 9.669	**	0.0020
MEF SSEA3+/T+ vs. E8 SSEA3+/T-	4.28	0.1052 to 8.455	*	0.0420
MEF SSEA3+/T+ vs. PRIMING MEDIA SSEA3+/T+	8.08	3.839 to 12.32	****	<0.0001
E8 SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	3.799	-0.8213 to 8.42	ns	0.1485

Figure 44: High content cloning assays of SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets from KOSR, Priming Media and E8 reveal differential expression of markers associated with self-renewal and lineage specific differentiation

To determine whether the SSEA3+/*BRACHYURY* (T)+ subset from Priming Media exhibited a bias to differentiation, a clonogenic assay was performed in comparison with the SSEA3+/*BRACHYURY*(T)- subset from E8 and the SSEA3+/*BRACHYURY* subsets from KOSR/MEF. Comparison of the percentage of total number of cells expressing FOXA2+/BRACHYURY- (A), FOXA2+/BRACHYURY+ (B) and PAX6 (C) within OCT4+ colonies for the following conditions:

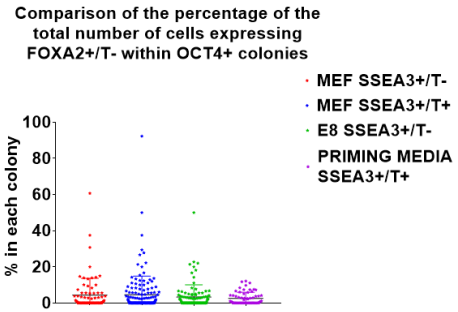
MEF SSEA3+/*BRACHYURY*-,

MEF SSEA3+/*BRACHYURY*+,

E8 SSEA3+/*BRACHYURY*-,

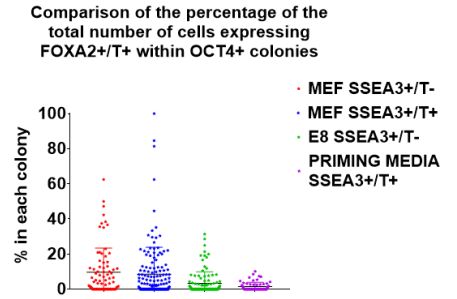
Priming Media SSEA3+/*BRACHYURY*+.

To each comparison was an associated a table registering the statistical significant of each condition between the others. KOSR/MEF data taken from Chapter 4 analysis Average of 3 biological repeats.



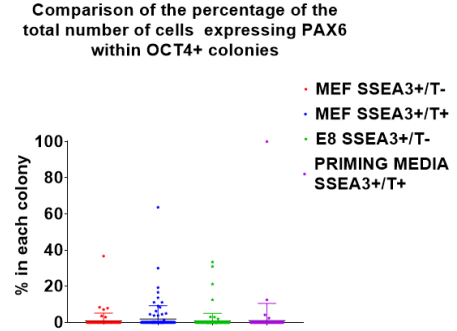
A

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Statistical Significance	Adjusted P Value
MEF SSEA3+/T- vs. MEF SSEA3+/T+	-0.1284	-3.154 to 2.897	ns	0.9995
MEF SSEA3+/T- vs. E8 SSEA3+/T-	1.079	-2.166 to 4.323	ns	0.8265
MEF SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	1.861	-1.813 to 5.534	ns	0.5591
MEF SSEA3+/T+ vs. E8 SSEA3+/T-	1.207	-1.565 to 3.979	ns	0.6753
MEF SSEA3+/T+ vs. PRIMING MEDIA SSEA3+/T+	1.989	-1.275 to 5.253	ns	0.3957
E8 SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	0.7822	-2.685 to 4.25	ns	0.9375



B

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Statistical Significance	Adjusted P Value
MEF SSEA3+/T- vs. MEF SSEA3+/T+	1.189	-3.016 to 5.393	ns	0.8853
MEF SSEA3+/T- vs. E8 SSEA3+/T-	6.346	1.838 to 10.85	**	0.0018
MEF SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	8.269	3.164 to 13.37	***	0.0002
MEF SSEA3+/T+ vs. E8 SSEA3+/T-	5.158	1.306 to 9.01	**	0.0034
MEF SSEA3+/T+ vs. PRIMING MEDIA SSEA3+/T+	7.081	2.545 to 11.62	***	0.0004
E8 SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	1.923	-2.896 to 6.742	ns	0.7321

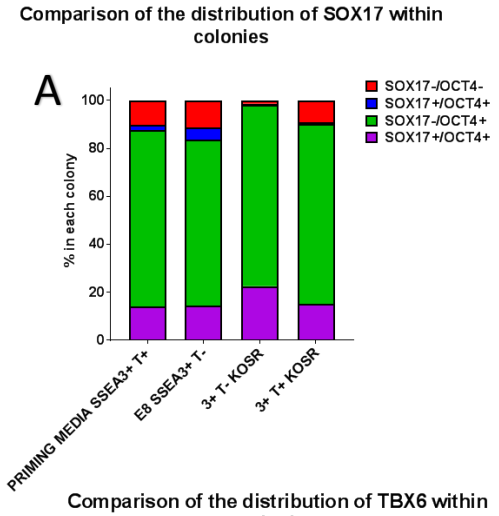


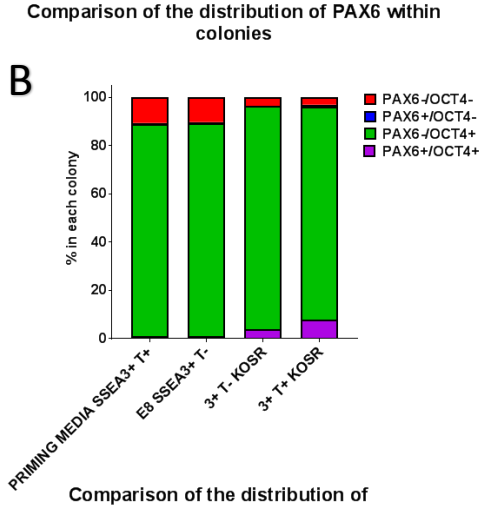
C

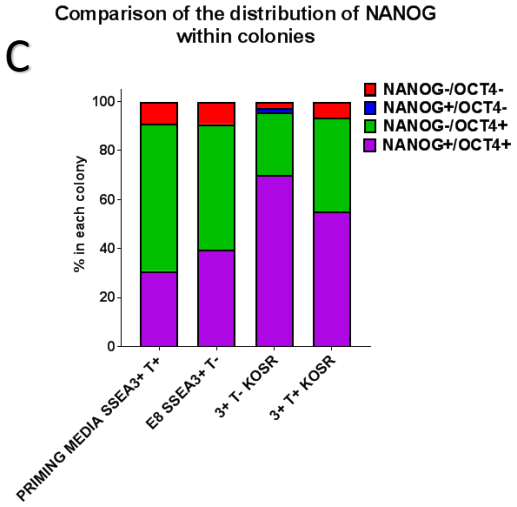
Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Statistical Significance	Adjusted P Value
MEF SSEA3+/T- vs. MEF SSEA3+/T+	-1.09	-3.616 to 1.436	ns	0.6819
MEF SSEA3+/T- vs. E8 SSEA3+/T-	0.09606	-2.299 to 2.491	ns	0.9996
MEF SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	-0.2319	-2.748 to 2.284	ns	0.9953
MEF SSEA3+/T+ vs. E8 SSEA3+/T-	1.186	-0.9907 to 3.362	ns	0.4969
MEF SSEA3+/T+ vs. PRIMING MEDIA SSEA3+/T+	0.8579	-1.452 to 3.168	ns	0.7735
E8 SSEA3+/T- vs. PRIMING MEDIA SSEA3+/T+	-0.328	-2.494 to 1.838	ns	0.9798

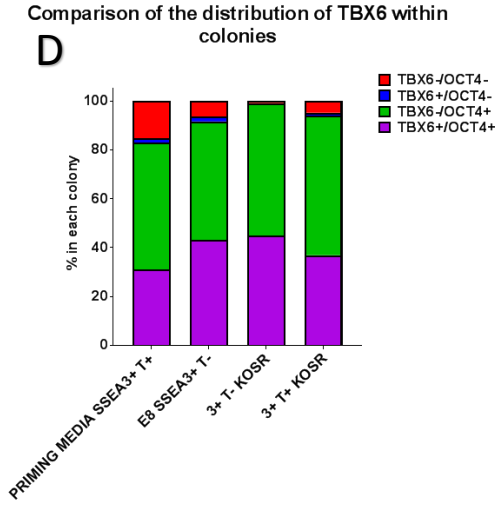
Figure 45: The various SSEA3/*BRACHYURY* subsets exhibit heterogeneity in colonies in respect to expression of markers associated with self-renewal and differentiation

Comparison of the distribution of SOX17 (A), PAX6 (B), NANOG (C), TBX6 (D) and FOXA2 and *BRACHYURY* (E) within the OCT4+/OCT4-colonies in the following conditions: Priming Media SSEA3+/*BRACHYURY*+, E8 SSEA3+/*BRACHYURY*+, SSEA3+/*BRACHYURY*- KOSR and SSEA3+/*BRACHYURY*+ KOSR. Results reveal heterogeneous expression of these markers but a similar trend between the SSEA3+/*BRACHYURY* subsets. KOSR/MEF data taken from Chapter 4 analysis. Average of 3 biological repeats.









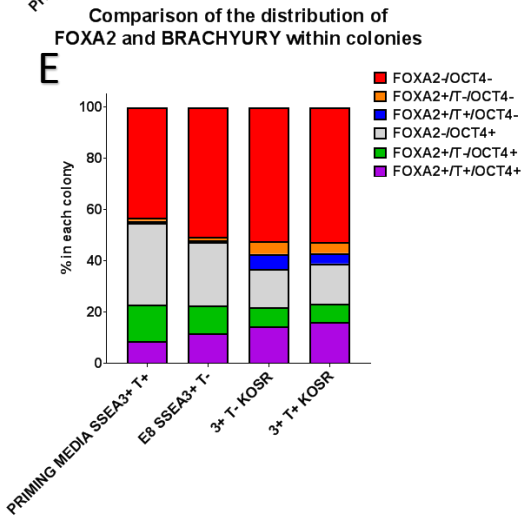


Figure 46: Analysis on the distribution of pluripotency associated marker and lineage specific marker expression in SSEA3/*BRACHYURY* sorted OCT4 Positive (+ve) colonies

Comparison of the expression percent of NANOG (A), SOX17 (B), PAX6 (C), TBX6 (D) FOXA2 (E) within the OCT4+ Positive colonies in the following conditions:

Priming Media SSEA3+/*BRACHYURY* (T) +,

E8 SSEA3+/*BRACHYURY (T)*+,

SSEA3+/*BRACHYURY*- KOSR and

SSEA3+/*BRACHYURY*+ KOSR

To determine if any subset generate more OCT4+ colonies that contained at least 1 cell with a respective marker. The results reveal heterogeneous expression of these markers but a similar trend between the SSEA3/*BRACHYURY* subsets. KOSR/MEF data taken from Chapter 4 analysis.

Average of 3 biological repeats. Statistical significance marked with asterisk (*) $p<0.05$ ns= not statistically significant $p>0.05$

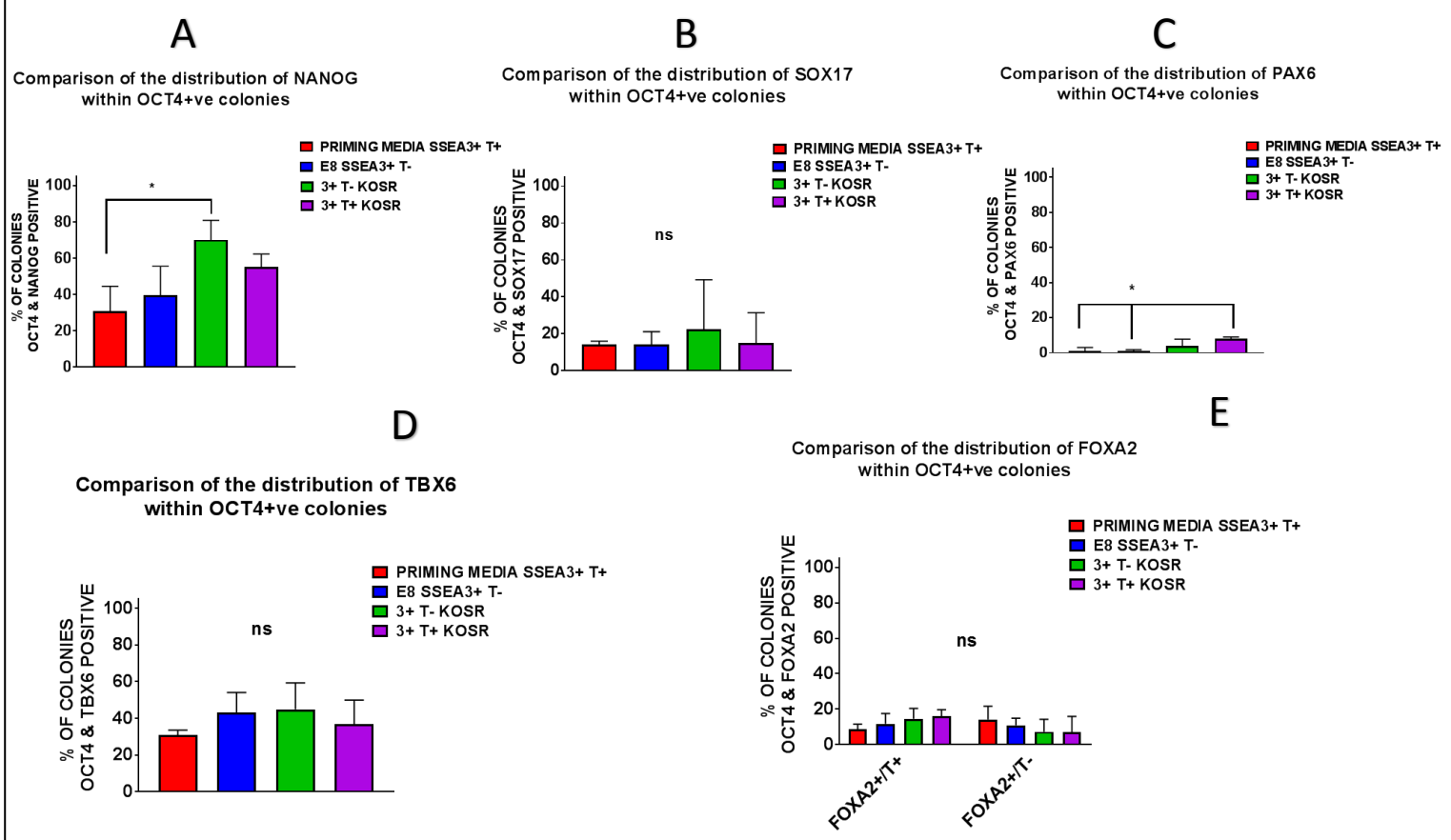


Figure 47: **Comparison of differentiation efficiency of hPSC grown in Priming Media compared to standard E8/Vitronectin conditions reveals reduced efficiency in differentiating to ectoderm**

Comparison of mesoderm (A), endoderm (B) and ectoderm (C) differentiation between cells in Priming Media and cells in E8 media. Representative pictures of HOESCHT (A.B.C), TBX6 (A), FOXA2 and SOX17 (B) and PAX6 (C). % of gene expression of PAX6 (a), SOX17+/FOXA2- and SOX17+/FOXA2+ (b) and PAX6 (c) for cells in priming media and in E8. 1/delta CT for Mesoderm differentiation (TBX6 and MSGN1 - A), Endoderm differentiation (GATA6, SOX17 and FOXA2 – B) and Ectoderm differentiation (PAX6, SOX1 and TUBB3 – C) compared to undifferentiated control. Comparable efficiency of cells from Priming Media able to form mesoderm and endoderm from immunostaining. Significant reduction in cells from Priming Media differentiating to ectoderm control. Scale bar set at 50µm. Statistical significance shown as asterisk (*) $p<0.05$ n=3

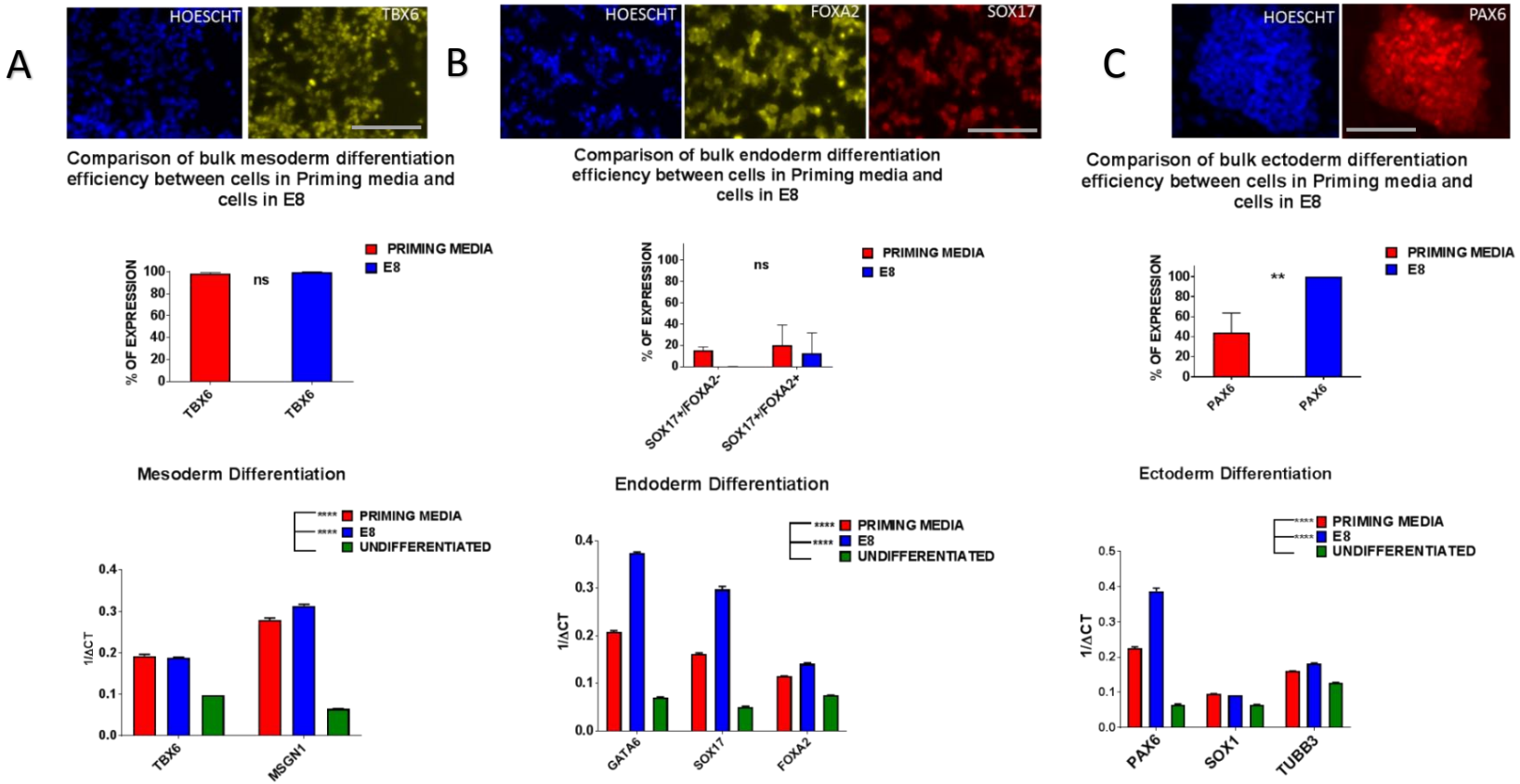
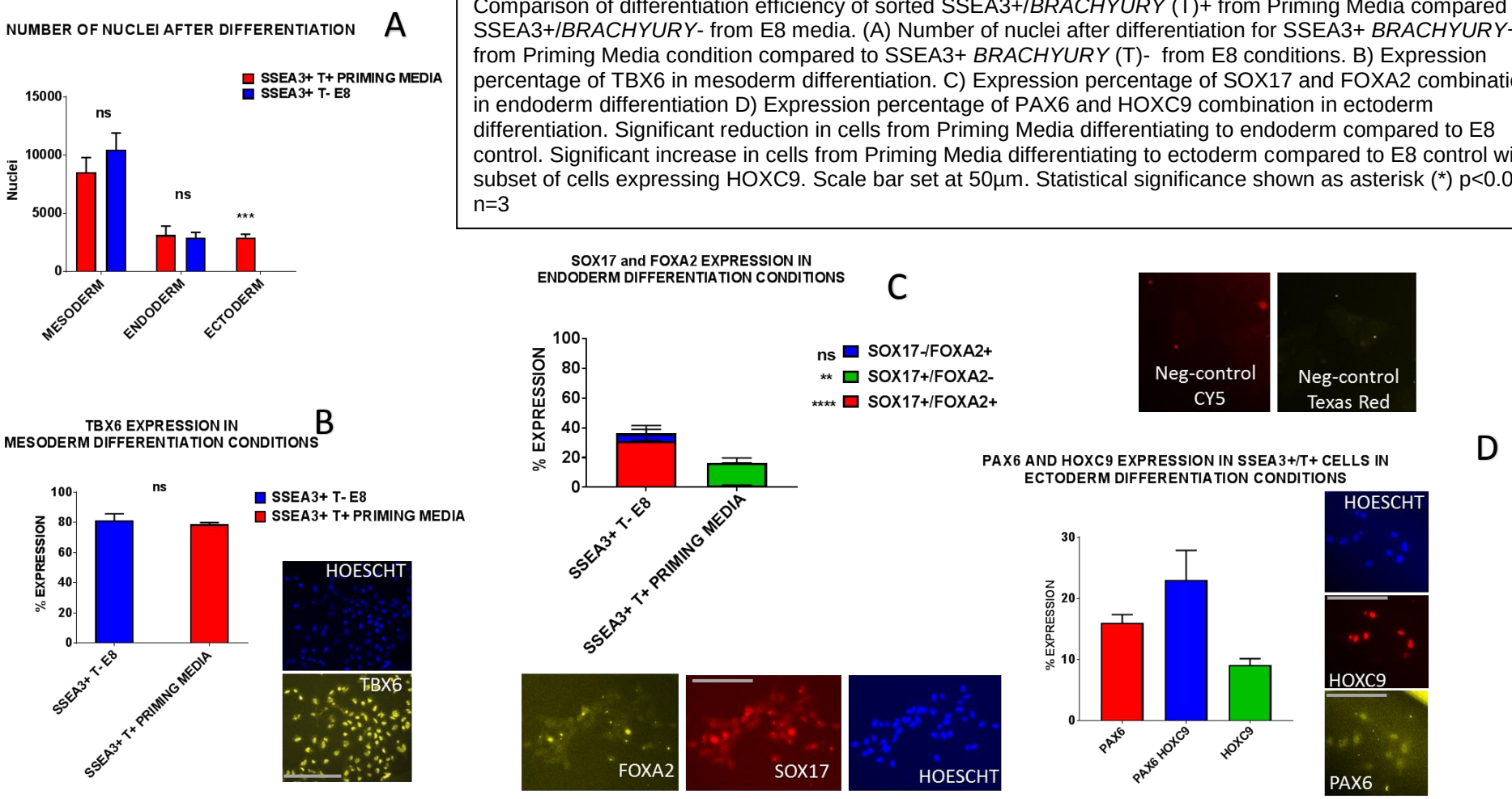


Figure 48: Comparison of differentiating the SSEA3+/*BRACHYURY*+ subset from Priming Media compared to the SSEA3+/*BRACHYURY*- subset from E8/Vitronectin demonstrates different efficiencies in forming endoderm and ectoderm



Comparison of differentiation efficiency of sorted SSEA3+/*BRACHYURY* (T)+ from Priming Media compared to SSEA3+/*BRACHYURY*- from E8 media. (A) Number of nuclei after differentiation for SSEA3+ *BRACHYURY*+ from Priming Media condition compared to SSEA3+ *BRACHYURY* (T)- from E8 conditions. B) Expression percentage of TBX6 in mesoderm differentiation. C) Expression percentage of SOX17 and FOXA2 combinations in endoderm differentiation D) Expression percentage of PAX6 and HOXC9 combination in ectoderm differentiation. Significant reduction in cells from Priming Media differentiating to endoderm compared to E8 control. Significant increase in cells from Priming Media differentiating to ectoderm compared to E8 control with a subset of cells expressing HOXC9. Scale bar set at 50µm. Statistical significance shown as asterisk (*) $p<0.05$ n=3

Figure 49: Assessment of whether the cells maintained in Priming Media contain a self-renewing NMP sub-population reveals the majority of cells co-express the pluripotency associated markers SOX2 and NANOG, as well as the posterior associated markers CDX2 and *BRACHYURY*

To determine whether the Priming Media was supporting maintenance of a self-renewing neuromesodermal progenitor (NMP) subset, cells grown in Priming Media for 7 days with passage in between, were immunostained for markers associated with NMP's and posterior mesoderm differentiation. Expression percentage of NANOG only, SOX2 only, CDX2 only, SOX2 and NANOG, SOX2 and T (*BRACHYURY*), CDX2 and T (*BRACHYURY*), SOX2 and NANOG and T (*BRACHYURY*). Representative pictures of gene fluorescence. Over 50% of the culture had cells co expressing SOX2 and NANOG, suggesting most of the culture being undifferentiated. However over 80% of the culture had cells co expressing CDX2 and *BRACHYURY*, suggesting a subset of cells heading towards a posterior associated fate. Scale bar set at 50µm. n=3

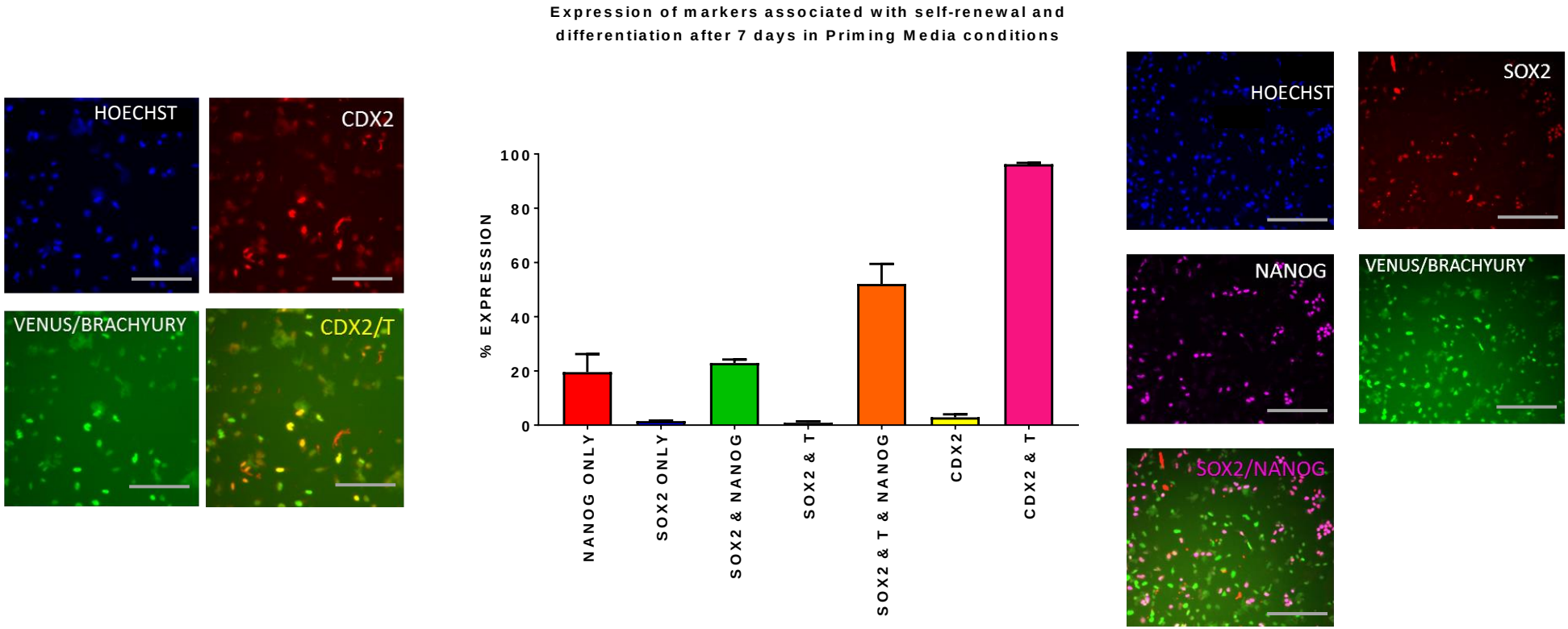
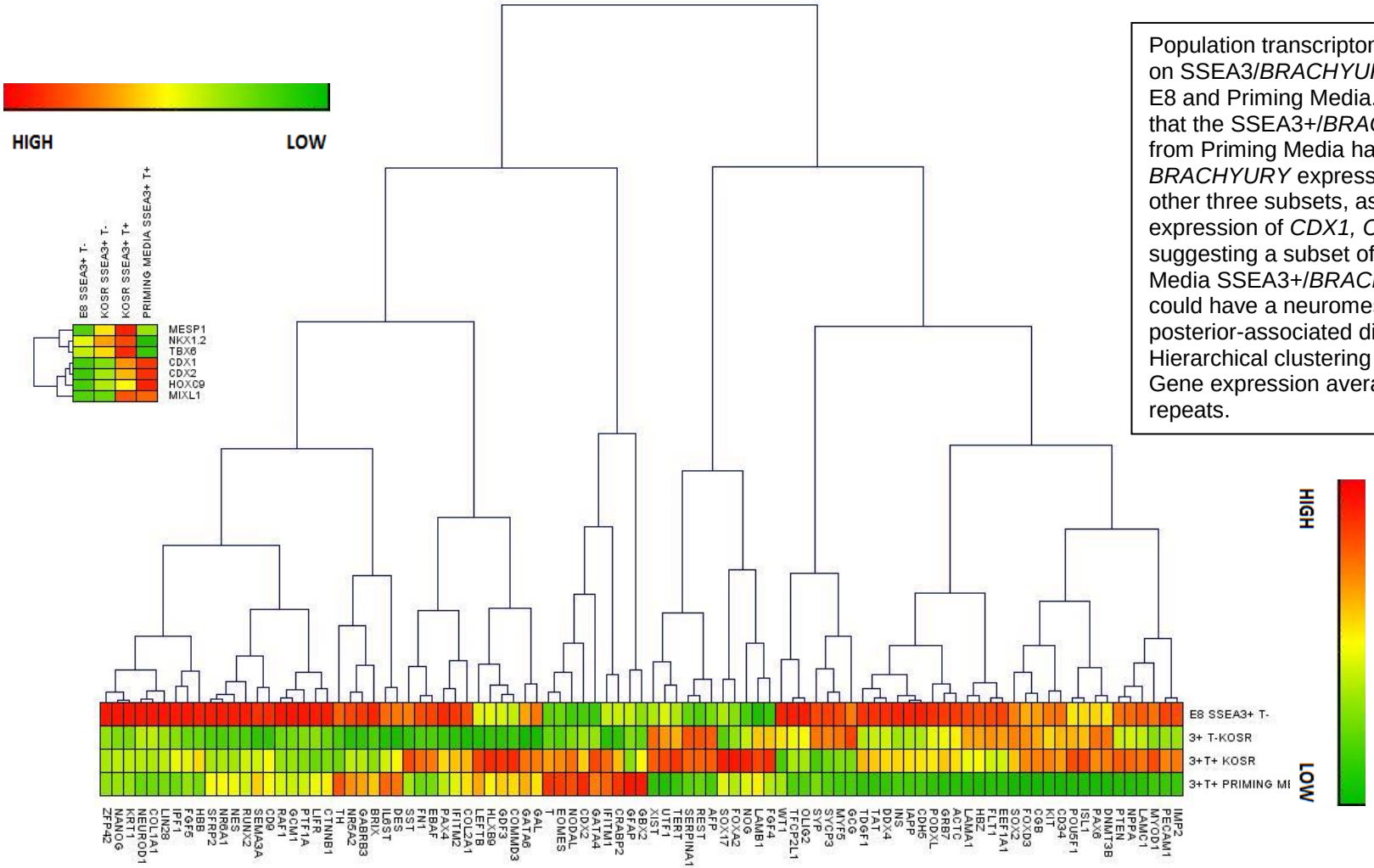


Figure 50: Gene analysis of SSEA3/*BRACHYURY* subsets reveal differential expression of genes associated with differentiation

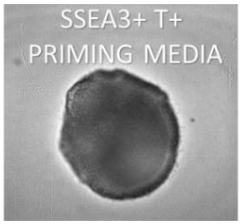


Population transcriptome analysis of 96 genes on SSEA3/*BRACHYURY* subsets from KOSR, E8 and Priming Media. The heatmap shows that the SSEA3+/*BRACHYURY* (T)+ subset from Priming Media had an increase in *BRACHYURY* expression compared to the other three subsets, as well as increased expression of *CDX1, CDX2, HOXC9* and *GBX2* suggesting a subset of cells from the Priming Media SSEA3+/*BRACHYURY*+ population could have a neuromesodermal and/or posterior-associated differentiation propensity. Hierarchical clustering applied to cluster genes. Gene expression average of 3 biological repeats.

Figure 51: Neutral EB differentiation of SSEA3+/*BRACHYURY*- from KOSR and E8 and SSEA3+/*BRACHYURY*+ from KOSR and Priming media demonstrate both the respective SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets are different in their propensities to differentiation

Comparison of the gene expression of *CD34, HAND1, TBX6, MSGN1, MEOX1, GATA2, GATA4, GATA6, SOX17, PAX6, SOX1, PAX3* and *SOX10* from Neutral EB differentiation conditions comparing the SSEA3+/*BRACHYURY* (T)+ subsets from KOSR and Priming Media conditions and the SSEA3+/*BRACHYURY* (T)- subsets from KOSR and E8 conditions.

Gene expression depicted as 1/ΔCt as an average of 3 biological repeats. EBs imaged at x20 magnification. KOSR/MEF EB data taken from chapter 4 analysis. Statistical results presented in the following figure.



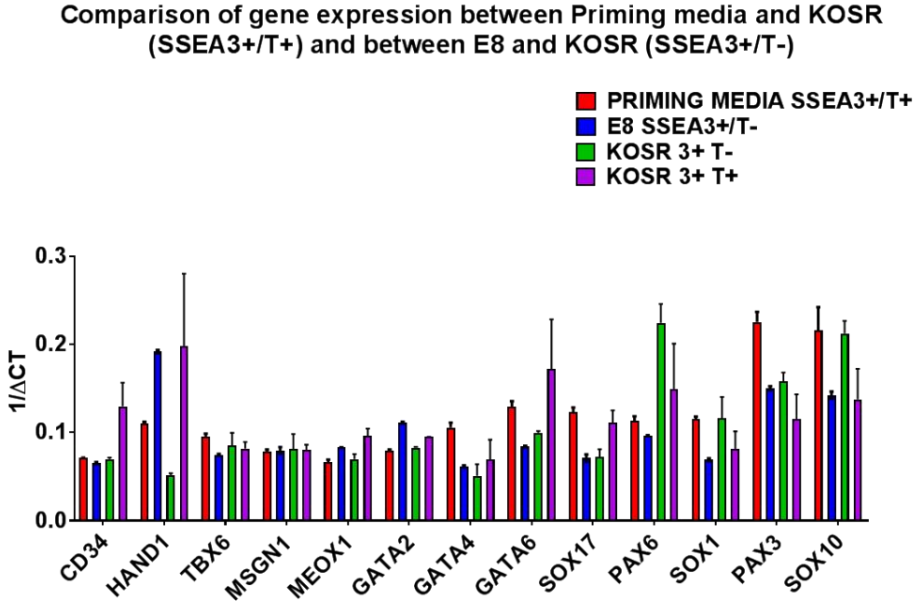


Figure 52: Neutral EB differentiation of SSEA3+/*BRACHYURY*- from KOSR and E8 and SSEA3+/*BRACHYURY*+ from KOSR and Priming Media demonstrate both the respective SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets are different in their propensities to differentiation

Statistical results of the Neutral EB differentiation of the SSEA3+/*BRACHYURY* (T)- and SSEA3+/*BRACHYURY* (T)+ subsets. Significant results were coloured in red for mesoderm associated gene, blue for endoderm, green for ectoderm and purple for neural crest. Statistical significance determined as $p<0.05$ using Tukey Multiple Comparison Test. ns= Not Significant. Statistical test performed on the average of three biological experiments.

Mesoderm Endoderm Ectoderm Neural Crest

Comparison		PRIMING MEDIA SSEA3+/T+ vs E8 SSEA3+/T-	PRIMING MEDIA SSEA3+/T+ vs. KOSR SSEA3+ T-	PRIMING MEDIA SSEA3+/T+ vs. KOSR SSEA3+ T+	E8 SSEA3+/T- vs. KOSR SSEA3+ T-	E8 SSEA3+/T- vs. KOSR SSEA3+ T+	KOSR SSEA3+ T- vs. KOSR SSEA3+ T+
CD34	Sig.	ns	ns	ns	ns	ns	ns
	P value	0.9966	0.999949	0.1612	0.9985	0.1028	0.145
HAND1	Sig.	*	Ns	*	****	ns	****
	P value	0.0216	0.1503	0.0116	0.000011	0.9965	0.000004
TBX6	Sig.	ns	ns	ns	ns	ns	ns
	P value	0.8842	0.9849	0.9638	0.9813	0.9942	0.9993
MSGN1	Sig.	ns	ns	ns	ns	ns	ns
	P value	0.999922	0.9995	0.9999	0.999944	0.999999	0.999978
MEOX1	Sig.	ns	ns	*	ns	ns	*
	P value	0.1765	0.9707	0.0128	0.3126	0.3015	0.0229
GATA2	Sig.	****	ns	***	****	***	**
	P value	0.000003	0.2087	0.0004	0.000013	0.0003	0.0029
GATA4	Sig.	ns	ns	ns	ns	ns	ns
	P value	0.3882	0.2073	0.5657	0.9818	0.9913	0.9089
GATA6	Sig.	ns	ns	ns	ns	*	ns
	P value	0.3681	0.7129	0.4153	0.9423	0.0106	0.0507
SOX17	Sig.	*	*	ns	ns	*	ns
	P value	0.0117	0.0135	0.7612	0.9995	0.0447	0.0522
PAX6	Sig.	ns	***	ns	****	ns	*
	P value	0.9307	0.0007	0.5731	0.000073	0.2383	0.0396
SOX1	Sig.	ns	ns	ns	ns	ns	ns
	P value	0.3485	0.999973	0.6144	0.3263	0.9712	0.5878
PAX3	Sig.	*	ns	***	ns	ns	ns
	P value	0.0392	0.0814	0.0008	0.9909	0.5866	0.4050
SOX10	Sig.	*	ns	*	ns	ns	*
	P value	0.0453	0.9993	0.0294	0.0618	0.9984	0.0408

5.3 Discussion

The addition of the LPA component (BCL) to the CHIRON mesoderm induction was a follow up to previously published results involving perturbing Wnt signalling, as described in the introduction to this chapter [99]. Whether this was applicable to other mesoderm inducing signalling networks was unexplored. To answer this, BMP4 and Activin A were used in protocols in comparison with the CHIRON condition. The results however showed that irrespective of the protocol used, the addition of BCL led to the desired effect of perturbing *BRACHYURY* Positive expression. To determine whether the addition of BCL to the cells was enough to maintain them as undifferentiated, the addition of the endogenous Wnt blocker IWP2 was compared alongside.

IWP2 works as a compound for blocking Wnt production and secretion by inhibiting Porcupine, a membrane-bound O-acyltransferase that results in inhibition of palmitoylation of Wnt [261]. It also causes its effects through blocking of Wnt-induced phosphorylation of the Lrp6 receptor resulting in blocking β-catenin accumulation in the nucleus [261]. There has been evidence in cancer cells that blocking porcupine can adversely affect their growth rate [262].

Within mouse ES cells, IWP2 has also been shown to have a role in preventing self-renewal and promoting conversion to an epiblast state [263]. There is as of yet no evidence to suggest IWP2 can induce adverse effects on the growth and proliferation rate of hPSC, however the results showed the concentration of 2µM IWP2 was not sustainable for this culture system. The 1µM IWP2 condition was determined as optimal in this culture set-up, a result that was supported by the *MIXL1* reporter line also analysed in parallel (Andrews lab, unpublished). The mechanism of IWP2 in blocking Wnt secretion and activity was demonstrated in the assessment of differentiated cells post-sort, where significant loss of NANOG activity was observed. This result showed that once the differentiation process was underway within the cell, it was not possible to revert it back by simply putting it back into self-renewal conditions. The results suggested that by blocking Wnt secretion and activity within the culture, the pro self-renewal circuitry was activated, allowing an interconversion process to undergo in the feeder-free

condition. However, while after three passages, the 2µM IWP2 condition resulted in more BF4 and SSEA3 Negative cells, slower cell growth and failure to form colonies was observed when put back into E8 conditions. The 1 µM IWP2 condition was the most robust over these passages for maintaining undifferentiated cells as shown by fewer BF4 and SSEA3 Negative cells, and the formation of colonies in E8 self-renewal conditions.

The dynamics of both BF4 and SSEA3 during mesoderm differentiation was a factor when considering how to analyse the data over multiple passages, as these were two markers that lost expression quickly in these conditions. This was observed in the mesoderm induction in chapter 3 (Figure 17) as well as the results post passage following CHIRON induction (Figure 32). There was no evidence presented that would imply that simply loss of both BF4 and SSEA3 implied differentiation. However, what was apparent was that to recreate a mesoderm bias in an undifferentiated population, strong expression levels of SSEA3 in the “High” region were desirable. This was shown by the SSEA3 High/CD9+ subset in Neutral EB conditions in Chapter 3. A loss of SSEA3 in this experimental set-up would imply the cells were further along terminal differentiation, which was not desirable. Therefore, the ideal result would be to maintain strong BF4 expression in respect to BF4 Positive expressing cells and also SSEA3 “High” expressing cells. With this framework in mind, it became clearer to determine optimal media conditions with the addition of BCL and IWP2.

IWP2 however was not required for the single cell cloning experiments. These experiments were performed identically to the ones in Chapter 4 and the number of colonies returned from the initial sort in this experiment produced nearly twice the number of colonies. An explanation is attributed to the media/matrix conditions being more optimal in this sort e.g. better quality MEFs in the 96 well plates. Another explanation is the media compositions being different, with the KOSR/MEF media having 10x the amount of LPA within the KOSR component. E8 media conversely does not have this component. Consequently, it is possible that a single cell was able to survive as undifferentiated and divide into daughter cells that more resembled those in the KOSR/MEF conditions. Irrespective of any differentiation that may then occur in the colony, there were enough cells that were pluripotent that could be passaged and renew into clonal lines that could differentiate to derivatives of the three germ layers.

In respect to the clonogenic assays to assess lineage bias (Figure 43-46), the results contrasted to those from chapter 4 and showed that the SSEA3+/*BRACHYURY* subsets were heterogeneous with different propensities for differentiation. This was most observable with the SOX17+/OCT4+ analysis where the Priming Media SSEA3+/*BRACHYURY*+ subset produced more expression of these markers within the colonies. One explanation for this, is the fact that the molecular pathways involved in inducing mesoderm and endoderm differentiation are closely related and that BRACHYURY Positive cells have been shown to form endodermal derivatives [211]. Therefore, the results showed there was a subset of cells within the Priming Media condition that exhibit a bias towards endoderm differentiation. Additionally, this subset produced less of the other lineage markers being analysed in this assay, notably FOXA2+/*BRACHYURY*+ and PAX6. What this could indicate is that this recreated population has less cells going towards an axial mesoderm or primitive streak like fate, as well as less neural ectoderm differentiation. However, due to the evidence of some cells from this subset differentiating towards both endoderm and mesoderm fates, it is possible that there are cells within this subsets that are more mesendoderm like.

A limitation in concluding what differentiation is occurring, is determining what the expression of these markers actually represents. SOX17 for example can indicate endodermal differentiation, but it can also indicate cardiac mesoderm differentiation [233] or primordial germ cell differentiation [232] when considered independently. This was the rationale as to why TBX6 was chosen as a robust paraxial mesoderm marker, FOXA2+/*BRACHYURY*+ was chosen as a robust mesendoderm marker and PAX6 was chosen as a robust neural ectoderm marker. This rationale is however limited to the reported evidence of expression of these markers in the literature. Nevertheless, the results indicate that this recreated SSEA3+/*BRACHYURY*+ subset was not a replication of the one found in KOSR/MEF conditions.

To further explore this differentiation potential of the recreated SSEA3+/*BRACHYURY*+ subset and also the bulk population growing in the Priming Media, differentiation assays were performed (Figure 47 and Figure 48). The evidence of the cells in the Priming Media leaning towards a mesoderm or endoderm fate can be seen in the efficiency of induction being comparable to

those from the E8 control. Only in the ectoderm inducing condition was there a significant reduction in efficiency of differentiation compared to the control. This however contrasted with the differentiation of the specific SSEA3+/*BRACHYURY*+ subset where there was increased survival and efficiency of forming ectoderm. While there was still fewer cells surviving compared to the mesoderm and endoderm conditions, a subset of these did express PAX6 and the posterior HOX gene HOXC9. HOXC9 was chosen as it is HOX gene expressed posteriorly and expressed in spinal associated regions of the developing neural tube *in vivo* [238,264]. Its expression relates to the hypothesis of the SSEA3+/*BRACHYUR*Y+ subset containing a heterogeneous population that included a subset of neuromesodermal cells readying to differentiate to a spinal/CNS derivative.

The evidence of a potential NMP subset existing with the SSEA3+/*BRACHYURY*+ subset was shown with the expression of the paraxial mesoderm marker TBX6 in some of the colonies formed from the SSEA3+/*BRACHYURY*+ subsets from both KOSR and Priming Media conditions (Figure 43-Figure 45). The evidence of HOXC9 expression following the ectoderm differentiation of the SSEA3+/*BRACHYURY*+ subset post-sort, suggests that a subset of the SSEA3+/*BRACHYURY*+ population was either biased to differentiate to an ectoderm derivative or alternatively, it already contained differentiated NMP or a pre-NMP progenitor cell type. For the mesoderm differentiation, the results were comparable to the control and with the endoderm differentiation, the control was more efficient in generating endoderm. The results nevertheless reflected those from the clonogenic assays in demonstrating heterogeneity within this recreated subset.

The results from the immunostaining from Priming Media (Figure 49) showed there was no evidence for a self-renewing NMP population, as shown by most cells in the population co-expressing the pluripotency associated markers SOX2 and NANOG. However, the co-expression of CDX2 with *BRACHYURY* within the population and its association with posterior mesoderm differentiation [219,265] could imply a link to NMP differentiation through an early bias to a NMP progenitor-like state. This hypothesis is further supported with the transcriptome analysis (Figure 50) where genes associated with NMP and posterior associated ectoderm differentiation were upregulated, namely *CDX1*, *CDX2* for NMP differentiation [238]

and *HOXC9* and *GBX2* for posterior associated ectoderm differentiation [266]. This potential bias however was not homogeneous, as there was also upregulation with primitive streak associated markers such as *EOMES* and *MIXL1* and endoderm associated markers in *GATA4*.

The Neutral EB assays (Figure 51) further supported the heterogeneity hypothesis as there was no evidence of one lineage being expressed by any of these subsets. The recreated SSEA3+/*BRACHYURY*+ subset showed significant upregulation of *PAX3* and *SOX10*, two neural crest associated markers, *GATA4* and *SOX17* for endoderm and *SOX1* for ectoderm. However, *SOX1* expression could be indictive of a NMP based neural ectoderm differentiation [238,267] further supported by the comparably low levels of *PAX6* to the KOSR subsets. The results also suggested that this recreated SSEA3+/*BRACHYURY*+ subset was leaning towards specific derivatives. This is evidenced from the expression of the mesoderm markers *CD34* and *HAND1* in the EB assays and *MESP1* in the gene expression analysis post-sort, which were comparably lower to that of the KOSR/ MEF SSEA3+/*BRACHYURY*+ subset.

Within this study, the SSEA3+/*BRACHYURY*- subset from E8 was compared against the SSEA3+/*BRACHYURY*- subset from KOSR/MEF conditions. A limitation of this decision was that a SSEA3+/*BRACHYURY*- subset from Priming Media could not be isolated. In a typical culture, the percentage of SSEA3+/*BRACHYURY*+ cells could be more than 95% of the population on analysis. Therefore, the closest subset obtainable was the SSEA3+/*BRACHYURY*- subset from E8/Vitronectin as it is obtained from the same basal media and same matrix growth conditions as the SSEA3+/*BRACHYURY*+ subset. This limitation however must be considered when directly comparing propensity for differentiation between the SSEA3+/*BRACHYURY*- subset from E8 and the SSEA3+/*BRACHYURY*+ from Priming Media. The reason is that the results could be interpreted as artefacts of the media conditions. However, what the results show is that the existence of a SSEA3+/*BRACHYURY*+ or SSEA3+/*BRACHYURY*- subset can result in different propensities for differentiation and that the two SSEA3+/*BRACHYURY*+ subsets were different.

Additionally, within this study, a caveat was that the differentiation potential assessed was only at the early stages of commitment and did not contain analysis into differentiation into later derivatives. A key question is whether cells growing in the Priming Media could create mesodermal derivatives such as cardiomyocytes, thereby demonstrating the translational application of the media. Attempts at creating cardiomyocytes in this manner were unsuccessful, as were attempts at creating NMP. In each attempt, the result was substantial cell death in all growth conditions. One explanation for this, is the large number of cells required to differentiate the cells efficiently in their respective protocols and the difficulty in gathering enough cells post-sort to perform the differentiation. The plating efficiency in feeder-free conditions has been reported to be very low (under 10%, even in the presence of Rho Kinase inhibitor) (Stem cell technologies- CloneR) meaning the number of cells required at plating is substantially greater in order to get the required density for the respective protocol. Further work would need to be performed to assess the Priming Media in this manner, which will be discussed in the following chapter.

5.4 Conclusion

The results presented in this chapter have demonstrated that cross-antagonism of signalling networks can be applied to recreating and maintaining a subset of hPSC in a defined system. By manipulating signalling pathways that are involved in inducing differentiation to specific lineages as well as driving self-renewal, it is possible to induce a state of hPSC that exhibit lineage bias both on the transcriptome and functionally, in respect to propensity for differentiation. This research serves as a proof of principle that can potentially be applied to any early differentiation away from the stem cell compartment if the appropriate signalling networks are induced. This provides potential into studying other lineage bias aside from mesoderm or endoderm, such as neural ectoderm or neural crest in a new study. However, as the results showed, despite recreating a subset of interest, there was still heterogeneity present in respect to differential propensity for differentiation. This observation provides a degree of insight into the potential subsets that may exist within a smaller sorted population of hPSC. The data presented in this chapter and in the previous two chapters overall elaborates on the enormity of the heterogeneity which exists within the stem compartment and the overall scale of possibilities for fate determination of hPSC in culture.

Chapter 6 : Final Conclusions and Future Directions

The results presented over the three chapters have demonstrated the existence of heterogeneity within the stem cell compartment *in vitro*. This heterogeneity can be observed as differences in cell surface antigen expression, gene expression patterns and ultimately in different propensities for differentiation. This is evidenced through the subsets identified through use of the BF4, CD9 and SSEA3 antibodies. The sorted subsets identified and analysed with these antibodies, served as a proof of principle, supported in the literature, that expression levels of the antigen can relate to cells with differences on the transcriptome and clonogenicity. When this concept was further expanded with the SSEA3/ CD9 subsets, it was evident that there were also functional differences in respect to propensity to differentiate, as well as differences on the transcriptome at the single cell level. One caveat to this analysis in respect to the antigen expression profiles, is the inability to compare expression levels based on gene expression. While it is possible to see the level of CD9 expression in an RNA seq analysis, SSEA3 is not a product from a single gene, so it is not possible to see the heterogeneity of this antigen on the single cell level, in this manner.

Considering the results from these chosen antibodies for identifying subsets of cells, other candidate antibodies could have been analysed. In the self-renewal experimental conditions that the cells were grown in prior to analysis, there was no observable detection of early differentiation associated markers SSEA1 or PDGFRα (Andrews, lab unpublished). Other early mesoderm associated cell surface markers include VCAM and SIRPA [268] however, these are markers that have been shown to be expressed after stem cell commitment, away from the undifferentiated state. Although, these could have been included regardless as a proof of principle, to evidence that they were not expressed on undifferentiated cells and to add more depth to the analysis performed.

One question is how relevant this hPSC heterogeneity is in an *in vivo* context. An observation is that differences in media/matrix combinations can influence stem cell functional properties, such as the disappearance of a state of hPSC in culture, as shown with *BRACHYURY* expression in E8/Vitronectin conditions. There is however evidence of heterogeneity within cells of the ICM [133], which poses the

question whether individual cells of the ICM can interconvert like the hPSC in culture, or whether they are already predetermined to go to a specific cell fate. If evidenced that they could interconvert, it could support the notion of these cells responding to extrinsic signals while maintaining the capacity for self-renewal [269]. It could also be that heterogeneity is a property of a population of plastic cell types, as heterogeneity and propensity to differentiate towards different lineages is also observed in adult stem cells and haematopoietic stem cells [269,270]. Though while this remains unanswered *in vivo*, an explanation for this is that the tightly regulated, native environment of the embryo may result in these states of pluripotent cells being transitory and otherwise difficult to observe. Being able to identify and maintain a state of undifferentiated cells was the challenge to study early fate determination of hPSC to mesoderm.

Experiments comparing the SSEA3+/*BRACHYURY*- and SSEA3+/*BRACHYURY*+ subsets were designed with the framework that having a known quantity, a mesoderm marker within the sorted subset, would allow a more homogeneous lineage biased population to be identified. There was evidence for this with the transcriptome analysis and the embryoid body experiments with the KOSR/MEF SSEA3+/*BRACHYURY*+ subset. This subset showed increased expression of mesoderm gene expression, however there was significant variation amongst the biological replicates. But what this did provide, was a logical basis to attempt to create a more refined, homogeneous population in a defined system with known signalling ques, to promote a specific mesoderm induction.

Through the knowledge gained from chapter 3 about the antigen dynamics of BF4, CD9 and SSEA3 during mesoderm differentiation, a culture system was created that allowed the expansion of a subset of hPSC that expressed *BRACHYURY*. This subset was shown to contain hPSC on both the population and single cell level through cloning experiments. In comparison with the KOSR/MEF SSEA3/*BRACHYURY* subsets, gene expression patterns and propensity for differentiation was different with the recreated SSEA3+/*BRACHYURY*+ subset. There was also evidence for different propensity to specific non-mesoderm differentiation compared to the KOSR/MEF SSEA3+/*BRACHYURY*+ subset. However, this recreated SSEA3+/*BRACHYURY*+ subset generated results that suggested the existence

of other subsets of cells that were not exhibited in the KOSR/MEF SSEA3+/*BRACHYURY*+ subset. This therefore suggests that the SSEA3+/*BRACHYURY*+ subset is a very large subset, that can contain multiple different smaller subsets of cells, that relate to different lineage bias for differentiation.

But to identify what populations may exist within the SSEA3+/ *BRACHYURY*+ subset and how these populations relate to the regulatory networks involved in fate determination of hPSC, transcriptome work would need to be performed on the single cell level. By using a combination of single cell RNA seq with statistical modelling, it could be feasible to identify what subsets of cells correlate with one another. This modelling could also be used to determine what molecular mechanisms are governing these different states within the population. This would allow the separation of the SSEA3+/*BRACHYURY*+ state into separate groups. These groups could have a transcriptome pattern that relates to differentiation to a specific derivative, such as an NMP progenitor or endoderm fate. This would ultimately provide powerful information into the undetermined gene networks that govern transition away from the stem cell compartment towards early differentiation. This data could also be compared to the SSEA3/*MIXL1* subsets.

Single cell RNA seq analysis has shown that the SSEA3+/*MIXL1*+ subset has increased gene expression of mesoderm associated genes compared to the SSEA3+/*MIXL1*- subset (Andrews lab, unpublished). Additionally, the SSEA3+/*MIXL1*+ subset also shown increased propensity to mesoderm differentiation in Neutral EB inducing conditions (Andrews lab, unpublished). Combined with the results from the SSEA3+/*BRACHYURY* subsets, a model could be made using the single cell RNA seq data that maps the progression from a non-mesoderm biased state (the SSEA3+ and *BRACHYUR*Y/*MIXL1*- subsets) to the biased state (SSEA3+ and *BRACHYURY*/*MIXL1*+ subsets). Additionally, it could help to build a bigger picture on the gene expression patterns and mechanisms involved in early mesoderm fate determination. Other studies focused on fate determination of hPSC have involved the use of antigens to identify subsets of cells [131] or to map differentiation stages throughout mesoderm commitment towards functional somatic cells [271]. Recent studies have also used single cell level transcriptome studies for determining regulators involved in

human ES cell differentiation to definitive endoderm [272] and mapping determinant stages in human ES cell neuronal differentiation [273]. These studies using antigens and transcriptomics are complimentary to the work performed on mesoderm biased states. Together they aid in our understanding of the mechanisms in fate determination of hPSC. However, the work performed using the SSEA3/*MIXL1* and SSEA3+/*BRACHYURY* subsets to recreate and maintain subsets of hPSC in a defined system has yet to be demonstrated in other model systems.

In addition to transcriptomic analysis, another immediate future direction would be to determine the mechanism of the recreated SSEA3+/*BRACHYURY*+ "trapped-state". It has been evidenced that a combination of cross-antagonism between pro-self-renewal and pro-differentiation signalling can result in a *BRACHYURY* Positive state that can be maintained and expanded over multiple passages. Though, the mechanism into how this process works is unexplored. As previously described, evidence in the literature into the mechanism of the Wnt agonist CHIRON, as well as how LPA can perturb Wnt-induced mesoderm differentiation and how IWP2 can block Wnt secretion and production provides a starting point. However, how this all comes together into a sustainable system that can maintain undifferentiated hPSC is unknown.

One candidate signalling pathway that could feature in this system is the HIPPO pathway. As described in Chapter 5, LPA can maintain pluripotency through activation of the HIPPO signalling pathway [243]. It has also shown that TAZ and YAP, two effectors for the HIPPO pathway, can regulate stem cell self-renewal in response to TGFβ signalling [111,114]. Furthermore, there is evidence of YAP and TAZ acting as effectors and modulators of Wnt signalling. Evidence from the Azzolin study [274] showed that Wnt activation can result in stabilisation of not only β-catenin, but also TAZ and that β-catenin physically interacts with TAZ, suggesting a mechanism of co-ordinated co-degradation. In respect to YAP, it has also been evidenced that Wnt/ β-catenin can activate YAP transcription [275]. Additionally, it was shown that YAP activation through inhibition of the HIPPO pathway results in activation of genes targeted by β-catenin [276]. It has also been shown that with HES3-*MIXL1* cells growing in Priming Media, there is evidence of nuclear localised YAP expression (Andrews lab, unpublished).

Together this presents a clear target for further study, with the concept being that there is co-ordination between the HIPPO pathway driving self-renewal and the Wnt-pathway driving differentiation. This results in a loop leading to a balance of a pluripotent, but lineage biased state of hPSC. Further study could also shed light on how these signalling pathways co-ordinate with other pathways, that result in cells residing within specific states that make one differentiation choice over another. Ultimately, in combination with the single cell transcriptomics, significantly more information could be achieved into the gene networks that define a state of cells. Additionally, it would provide information into the molecular mechanisms which are driving formation and maintenance of these states in culture.

The potential for this research could implicate better understanding of early developmental processes in culture when applied to different lineages. However, this potential could also be applicable to regenerative medicine by aiding in achieving greater efficiencies in generating specific cell types. An additional future direction would be to apply this media to different differentiation systems that involve generation of specific mesodermal derivatives. This would be to improve the efficiency and yield in generating these derivatives using several pluripotent lines, both ES cell and iPSC. There is evidence anecdotally and, in the literature, that behaviour of these pluripotent lines can vary in respect to their functional capability or efficiency. This has been shown in procedures such as differentiation [277,278]. This observation was the rationale for using the three pluripotent lines in the SSEA3/CD9 experiments in Chapter 3. Because of these differences, and potential effects, to create a media system that would be applicable to robustly generating these mesoderm derivatives, this media system would need to be tested and potentially optimised using different cell lines.

The lines that have been tested so far are the two ES cell lines, H9 and HES3 (with and without the reporter construct) and the iPSC line MIFF-1. The MIFF-1 line, HES3 wildtype line, the HES3 *MIXL1* line, the H9 wildtype line and the H9 *BRACHYURY* line have been passed multiple times in Priming Media and shown to maintain BRACHYURY expression in >90% of the cells (Andrews lab, unpublished *for the MIFF-1 and HES3 lines*). In addition, the H9- *BRACHYURY* and HES3 -*MIXL1* lines have been karyotyped after several passages and shown to have a normal karyotype (Andrews lab, unpublished). While these are

promising results, for robustness, more lines will need to be tested over a prolonged period to assess their performance in this media system, as well as their genetic stability. However, through combining these future directions together, a media system could be generated to serve a variety of purposes. This includes bettering our understanding into fate determination, as well as providing wider translational applications.

References

1. Srebnik, H. H. in *Concepts in Anatomy* 15–20 (Springer US, 2002). doi:10.1007/978-1-4615-0857-1_3

2. PANSKY, B. *Review of MEDICAL EMBRYOLOGY*. *McGraw-Hill,* (1982).

3. Solnica-Krezel, L. & Sepich, D. S. Gastrulation: Making and Shaping Germ Layers. *Annu. Rev. Cell Dev. Biol.* **28,** 687–717 (2012).

4. Gilbert, S. F. *Developmental biology*. (Sinauer Associates, 2010).

5. Lawson, K. A., Meneses, J. J. & Pedersen, R. A. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113,** 891–911 (1991).

6. Robb, L. & Tam, P. P. L. Gastrula organiser and embryonic patterning in the mouse. *Seminars in Cell and Developmental Biology* **15,** 543–554 (2004).

7. Tam, P. P. & Beddington, R. S. Establishment and organization of germ layers in the gastrulating mouse embryo. *Ciba Found Symp* **165,** 27–41 (1992).

8. Dumont, D. J. *et al.* Vascularization of the mouse embryo: A study of flk-1, tek, tie, and vascular endothelial growth factor expression during development. *Dev. Dyn.* **203,** 80–92 (1995).

9. Ciruna, B. & Rossant, J. FGF Signaling Regulates Mesoderm Cell Fate Specification and Morphogenetic Movement at the Primitive Streak. *Dev. Cell* **1,** 37–49 (2001).

10. Chuai, M. *et al.* Cell movement during chick primitive streak formation. *Dev. Biol.* **296,** 137–149 (2006).

11. Saga, Y. *et al.* MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development* **126,** 3437–3447 (1999).

12. Tam, P. P. L. *et al.* Sequential allocation and global pattern of movement of the definitive endoderm in the mouse embryo during gastrulation.

Development **134,** 251–260 (2007).

13. Hart, A. H. *et al.* Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo. *Development* **129,** 3597–608 (2002).

14. Wilson, V. & Beddington, R. S. P. Cell fate and morphogenetic movement in the late mouse primitive streak. *Mech. Dev.* **55,** 79–89 (1996).

15. Yamamoto, M. *et al.* Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature* **428,** 387–392 (2004).

16. Beck, S. *et al.* Extraembryonic proteases regulate Nodal signalling during gastrulation. *Nat. Cell Biol.* **4,** 981–985 (2002).

17. Episkopou, V. *et al.* Induction of the mammalian node requires Arkadia function in the extraembryonic lineages. *Nature* **410,** 825–830 (2001).

18. Perea-Gomez, A. *et al.* Nodal antaginists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. *Dev. Cell* **3,** 745–756 (2002).

19. Iratni, R. *et al.* Inhibition of excess nodal signaling during mouse gastrulation by the transcriptional corepressor DRAP1. *Science (80-.).* **298,** 1996–1999 (2002).

20. Meno, C. *et al.* Mouse lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Mol. Cell* **4,** 287–298 (1999).

21. Winnier, G., Blessing, M., Labosky, P. A. & Hogan, B. L. M. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev.* **9,** 2105–2116 (1995).

22. Yoshikawa, Y., Fujimori, T., McMahon, A. P. & Takada, S. Evidence that absence of Wnt-3a signaling promotes neuralization instead of paraxial mesoderm development in the mouse. *Dev. Biol.* **183,** 234–242 (1997).

23. Marikawa, Y. Wnt/β-catenin signaling and body plan formation in mouse embryos. *Semin. Cell Dev. Biol.* **17,** 175–184 (2006).

24. Hoffman, J. A., Wu, C.-I. & Merrill, B. J. Tcf7l1 prepares epiblast cells in the gastrulating mouse embryo for lineage specification. *Development* **140,** 1665–1675 (2013).

25. Wen, J. *et al.* Single-cell analysis reveals lineage segregation in early post-implantation mouse embryos. *J. Biol. Chem.* **292,** 9840–9854 (2017).

26. Mossman, A. K., Sourris, K., Ng, E., Stanley, E. G. & Elefanty, A. G. In Focus Mixl1 and Oct4 Proteins Are Transiently Co-Expressed in Differentiating Mouse and Human Embryonic Stem Cells. *Stem Cells Dev.* **14,** 656–663 (2005).

27. Tsakiridis, A. *et al.* Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development* **142,** 809–809 (2015).

28. Song, L., Chen, J., Peng, G., Tang, K. & Jing, N. Dynamic heterogeneity of brachyury in mouse epiblast stem cells mediates distinct response to extrinsic bone morphogenetic protein (bmp) signaling. *J. Biol. Chem.* **291,** 15212–15225 (2016).

29. Tsakiridis, A. *et al.* Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development* **141,** 1209–1221 (2014).

30. Ten Donkelaar, H. J., Bekker, M., Renier, W. O., Hori, A. & Shiota, K. in *Clinical Neuroembryology: Development and Developmental Disorders of the Human Central Nervous System* 165–217 (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-642-54687-7_4

31. Hall, B. K. The neural crest as a fourth germ layer and vertebrates as quadroblastic not triploblastic. *Evol. Dev.* **2,** 3–5 (2000).

32. Hall, B. K. in *Bones and Cartilage* (2015). doi:10.1016/B978-0-12-416678-3.00017-3

33. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science (80-.).* **282,** 1145–1147 (1998).

34. Finch, B. W. & Ephrussi, B. Retention of Multiple Developmental Potentialities by Cells of a Mouse Testicular Teratocarcinoma during Prolonged Culture In-vitro and their Extinction upon Hybridation with Cells

of Permanenet Lines. *Proc. Natl. Acad. Sci. U. S. A.* **57,** 615—621 (1967).

35. Brinster, R. L. The effect of cells transferred into the mouse blastocyst on subsequent development. *J. Exp. Med.* **140,** 1049–56 (1974).

36. Martin, G. R. & Evans, M. J. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc. Natl. Acad. Sci.* **72,** 1441–1445 (1975).

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

38. Artzt, K. *et al.* Surface antigens common to mouse cleavage embryos and primitive teratocarcinoma cells in culture. *Proc. Natl. Acad. Sci. U. S. A.* **70,** 2988–92 (1973).

39. Solter, D. & Knowles, B. B. Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). *Proc. Natl. Acad. Sci. U. S. A.* **75,** 5565–9 (1978).

40. Andrews, P. W., Bronson, D. L., Benham, F., Strickland, S. & Knowles, B. B. A comparative study of eight cell lines derived from human testicular teratocarcinoma. *Int. J. Cancer* **26,** 269–280 (1980).

41. Andrews, P. W. Human teratocarcinomas. *BBA - Reviews on Cancer* **948,** 17–36 (1988).

42. Andrews, P. W. *et al.* Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin. *Biochem. Soc. Trans.* **33,** 1526–1530 (2005).

43. Schöler, H. R., Hatzopoulos, A. K., Balling, R., Suzuki, N. & Gruss, P. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J.* **8,** 2543–50 (1989).

44. Gubbay, J. *et al.* 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 (1990).

45. Mitsui, K. *et al.* The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631–642 (2003).

46. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113,** 643–655 (2003).

47. Simeone, A. et al. Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. *Nat. Lett.* **346,** 763–766 (1990).

48. Kondziolka, D. *et al.* Transplantation of cultured human neuronal cells for patients with stroke. *Neurology* **55,** 565–569 (2000).

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

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

51. Evans, M. J., Bradley, A., Kuehn, M. R. & Robertson, E. J. 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 (1985).

52. Doetschman, T. *et al.* Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330,** 576–578 (1987).

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

54. Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336,** 688–690 (1988).

55. Kinoshita, K. *et al.* GABPα regulates Oct-3/4 expression in mouse embryonic stem cells. *Biochem. Biophys. Res. Commun.* **353,** 686–691 (2007).

56. Xu, R.-H. *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat. Biotechnol.* **20,** 1261–1264 (2002).

57. Heldin, C. H., Miyazono, K. & Ten Dijke, P. TGF-beta signaling from cell membrane to nucleus through SMAD proteins. *Nature* **390,** 465–471 (1997).

58. Schier, A. F. Nodal Signaling in Vertebrate Development. *Annu. Rev. Cell Dev. Biol.* **19,** 589–621 (2003).

59. James, D. TGF /activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132,** 1273–1282 (2005).

60. Macías-Silva, M. *et al.* MADR2 is a substrate of the TGFβ receptor and its phosphorylation is required for nuclear accumulation and signaling. *Cell* **87,** 1215–1224 (1996).

61. Kretzschmar, M., Liu, F., Hata, A., Doody, J. & Massagué, J. The TGF-β family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase. *Genes Dev.* **11,** 984–995 (1997).

62. Kingsley, D. M. The TGF-β superfamily: New members, new receptors, and new genetic tests of function in different organisms. *Genes and Development* **8,** 133–146 (1994).

63. Xu, R. H. *et al.* NANOG Is a Direct Target of TGFβ/Activin-Mediated SMAD Signaling in Human ESCs. *Cell Stem Cell* **3,** 196–206 (2008).

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

65. Kessler, D. S. & Melton, D. a. Vertebrate embryonic induction: mesodermal and neural patterning. *Science (80-.).* **266,** 596–604 (1994).

66. D'Amour, K. A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat. Biotechnol.* **23,** 1534–1541 (2005).

67. Conlon, F. L. *et al.* A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120,** 1919–1928 (1994).

68. Varlet, I., Collignon, J., Norris, D. P. & Robertson, E. J. Nodal signaling and axis formation in the mouse. *Cold Spring Harb Symp Quant Biol* **62,** 105–113 (1997).

69. Dunn, N. R. Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. *Development* **131,** 1717–1728 (2004).

70. Smith, J. R. *et al.* Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Dev. Biol.* **313,** 107–117 (2008).

71. Singh, A. M. *et al.* 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 (2012).

72. Hayashi, Y. *et al.* BMP4 induction of trophoblast from mouse embryonic stem cells in defined culture conditions on laminin. *Vitr. Cell. Dev. Biol. - Anim.* **46,** 416–430 (2010).

73. Teo, A. K. K. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev.* **25,** 238–250 (2011).

74. Kurek, D. *et al.* Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Reports* **4,** 114–128 (2015).

75. Steventon, B., Araya, C., Linker, C., Kuriyama, S. & Mayor, R. Differential requirements of BMP and Wnt signalling during gastrulation and neurulation define two steps in neural crest induction. *Development* **136,** 771–779 (2009).

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

77. Hao, J. *et al.* Dorsomorphin, a selective small molecule inhibitor of BMP signaling, promotes cardiomyogenesis in embryonic stem cells. *PLoS One* **3,** (2008).

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

79. Sasai, Y. *et al.* Xenopus chordin: A novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell* **79,** 779–790 (1994).

80. Lee, H. *et al.* Directed Differentiation and Transplantation of Human Embryonic Stem Cell-Derived Motoneurons. *Stem Cells* **25,** 1931–1939 (2007).

81. Morizane, A., Doi, D., Kikuchi, T., Nishimura, K. & Takahashi, J. Small-molecule inhibitors of bone morphogenic protein and activin/nodal signals promote highly efficient neural induction from human pluripotent stem cells. *J. Neurosci. Res.* **89,** 117–126 (2011).

82. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27,** 275–280 (2009).

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

84. Turner, N. & Grose, R. Fibroblast growth factor signalling: From development to cancer. *Nature Reviews Cancer* **10,** 116–129 (2010).

85. Vallier, L., Alexander, M. & Pedersen, R. A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* **118,** 4495–4509 (2005).

86. Eiselleova, L. *et al.* A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells. *Stem Cells* **27,** 1847–1857 (2009).

87. Li, J. *et al.* MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. *Differentiation* **75,** 299–307 (2007).

88. Armstrong, L. *et al.* The role of PI3K/AKT, MAPK/ERK and NFκβ signalling in the maintenance of human embryonic stem cell pluripotency

and viability highlighted by transcriptional profiling and functional analysis. *Hum. Mol. Genet.* **15,** 1894–1913 (2006).

89. Na, J., Furue, M. K. & Andrews, P. W. Inhibition of ERK1/2 prevents neural and mesendodermal differentiation and promotes human embryonic stem cell self-renewal. *Stem Cell Res.* **5,** 157–169 (2010).

90. Chen, Y. G., Li, Z. & Wang, X. F. Where PI3K/Akt meets smads: The crosstalk determines human embryonic stem cell fate. *Cell Stem Cell* **10,** 231–232 (2012).

91. Hori, Y. *et al.* Growth inhibitors promote differentiation of insulin-producing tissue from embryonic stem cells. *Proc. Natl. Acad. Sci.* **99,** 16105–16110 (2002).

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

93. Yamaguchi, T. P., Harpal, K., Henkemeyer, M. & Rossant, J. fgfr-1 is required for embryonic growth and mesodermal patterning during mouse gastrulation. *Genes Dev.* **8,** 3032–3044 (1994).

94. Morrison, G. M. *et al.* Anterior Definitive Endoderm from ESCs Reveals a Role for FGF Signaling. *Cell Stem Cell* **3,** 402–415 (2008).

95. Yu, P., Pan, G., Yu, J. & Thomson, J. A. FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation. *Cell Stem Cell* **8,** 326–334 (2011).

96. Huelsken, J. The Wnt signalling pathway. *J. Cell Sci.* **115,** 3977–3978 (2002).

97. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).

98. Davidson, K. C. *et al.* Wnt/ -catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. *Proc. Natl. Acad. Sci.* **109,** 4485–4490 (2012).

99. Blauwkamp, T. A., Nigam, S., Ardehali, R., Weissman, I. L. & Nusse, R. Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. *Nat. Commun.* **3,** 1070 (2012).

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

101. Chan, Y. S. *et al.* Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* **13,** 663–675 (2013).

102. Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504,** 282–286 (2013).

103. Ware, C. B. *et al.* Derivation of naive human embryonic stem cells. *Proc. Natl. Acad. Sci.* **111,** 4484–4489 (2014).

104. Varelas, X. *et al.* The Hippo pathway effectors TAZ and YAP in development, homeostasis and disease. *Development* **141,** 1614–1626 (2014).

105. Mo, J.-S. *et al.* The Hippo signaling pathway in stem cell biology and cancer. *EMBO Rep.* **15,** 642–656 (2014).

106. Kim, M. & Jho, E. H. Cross-talk between Wnt/β-catenin and Hippo signaling pathways: A brief review. *BMB Reports* **47,** 540–545 (2014).

107. Hansen, C. G., Moroishi, T. & Guan, K. L. YAP and TAZ: A nexus for Hippo signaling and beyond. *Trends in Cell Biology* **25,** 499–513 (2015).

108. Sawada, A. *et al.* Redundant Roles of Tead1 and Tead2 in Notochord Development and the Regulation of Cell Proliferation and Survival. *Mol. Cell. Biol.* **28,** 3177–3189 (2008).

109. Varelas, X. *et al.* The Crumbs Complex Couples Cell Density Sensing to Hippo-Dependent Control of the TGF-β-SMAD Pathway. *Dev. Cell* **19,** 831–844 (2010).

110. Pan, D. The Hippo Signaling Pathway in Development and Cancer. *Dev. Cell* **19,** 491–505 (2017).

111. Varelas, X. *et al.* TAZ controls Smad nucleocytoplasmic shuttling and regulates human embryonic stem-cell self-renewal. *Nat. Cell Biol.* **10,** 837–848 (2008).

112. Hiemer, S. E., Szymaniak, A. D. & Varelas, X. The transcriptional regulators TAZ and YAP direct transforming growth factor β-induced tumorigenic phenotypes in breast cancer cells. *J. Biol. Chem.* **289,** 13461–13474 (2014).

113. Beyer, T. A. *et al.* Switch enhancers interpret TGF-β and hippo signaling to control cell fate in human embryonic stem cells. *Cell Rep.* **5,** 1611–1624 (2013).

114. Alarcón, C. *et al.* Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF-β Pathways. *Cell* **139,** 757–769 (2009).

115. Imajo, M., Miyatake, K., Iimura, A., Miyamoto, A. & Nishida, E. A molecular mechanism that links Hippo signalling to the inhibition of Wnt/β-catenin signalling. *EMBO J.* **31,** 1109–1122 (2012).

116. Varelas, X. *et al.* The Hippo Pathway Regulates Wnt/β-Catenin Signaling. *Dev. Cell* **18,** 579–591 (2010).

117. Bascuñán-Gamboa, K. A., Araya-Quezada, M. & Pérez-Bravo, F. MicroRNAs: an epigenetic tool to study celiac disease. *Rev. Esp. Enferm. Dig.* **106,** 325–333 (2014).

118. Tsutsumi, R. *et al.* YAP and TAZ, hippo signaling targets, act as a rheostat for nuclear SHP2 function. *Dev. Cell* **26,** 658–665 (2013).

119. Takahashi, A. *et al.* SHP2 Tyrosine Phosphatase Converts Parafibromin/Cdc73 from a Tumor Suppressor to an Oncogenic Driver. *Mol. Cell* **43,** 45–56 (2011).

120. Konsavage, W. M., Kyler, S. L., Rennoll, S. A., Jin, G. & Yochum, G. S. Wnt/β-catenin signaling regulates yes-associated protein (YAP) gene expression in colorectal carcinoma cells. *J. Biol. Chem.* **287,** 11730–

11739 (2012).

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

122. Waddington, C. H. *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser.* (London: George Allen & Unwin, Ltd., 1957).

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

124. Enver, T. *et al.* Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. *Hum Mol Genet* **14,** 3129–3140 (2005).

125. Hu, M. *et al.* Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11,** 774–785 (1997).

126. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448,** 196–199 (2007).

127. Takashima, Y. *et al.* Erratum: Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human (Cell (2014) 158 (1254-1269)). *Cell* **162,** 452–453 (2015).

128. Guo, G. *et al.* Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136,** 1063–1069 (2009).

129. Tonge, P. D., Shigeta, M., Schroeder, T. & Andrews, P. W. Functionally defined substates within the human embryonic stem cell compartment. *Stem Cell Res.* **7,** 145–153 (2011).

130. Hough, S. R., Laslett, A. L., Grimmond, S. B., Kolle, G. & Pera, M. F. A continuum of cell states spans pluripotency and lineage commitment in human embryonic stem cells. *PLoS One* **4,** (2009).

131. Hough, S. R. *et al.* Single-cell gene expression profiles define self-renewing, pluripotent, and lineage primed states of human pluripotent stem cells. *Stem Cell Reports* **2,** 881–895 (2014).

132. Laslett, A. L. *et al.* Transcriptional analysis of early lineage commitment in human embryonic stem cells. *BMC Dev Biol* **7,** 12 (2007).

133. Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. *Dev. Cell* **10,** 615–624 (2006).

134. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).

135. Thomson, J. A. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science (80-.).* **282,** 1145–1147 (1998).

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

137. Mendjan, S. *et al.* NANOG and CDX2 pattern distinct subtypes of human mesoderm during exit from pluripotency. *Cell Stem Cell* **15,** 310–325 (2014).

138. Chen, G. *et al.* Chemically defined conditions for human iPS cell derivation and culture. *Nat. Methods* **8,** 424–429 (2011).

139. Amit, M. *et al.* Clonally Derived Human Embryonic Stem Cell Lines Maintain Pluripotency and Proliferative Potential for Prolonged Periods of Culture. *Dev. Biol.* (2000). doi:10.1006/dbio.2000.9912

140. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.* **25,** 681–686 (2007).

141. Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256,** 495–497 (1975).

142. Williams, B. P. *et al.* Biochemical and genetic analysis of the Oka blood group antigen. *Immunogenetics* **27,** 322–329 (1988).

143. Wright, A. *et al.* Mapping the stem cell state: eight novel human embryonic stem and embryonal carcinoma cell antibodies. *Int J Androl* **34,**

e175–87; discussion e187–8 (2011).

144. Shevinsky, L. H., Knowles, B. B., Damjanov, I. & Solter, D. Monoclonal antibody to murine embryos defines a stage-specific embryonic antigen expressed on mouse embryos and human teratocarcinoma cells. *Cell* **30,** 697–705 (1982).

145. Kannagi, R. *et al.* Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells. *EMBO J.* **2,** 2355–61 (1983).

146. Andrews, P. W., Banting, G., Damjanov, I., Arnaud, D. & Avner, P. Three Monoclonal Antibodies Defining Distinct Differentiation Antigens Associated with Different High Molecular Weight Polypeptides on the Surface of Human Embryonal Carcinoma Cells. *Hybridoma* **3,** 347–361 (1984).

147. Lippmann, E. S., Estevez-Silva, M. C. & Ashton, R. S. Defined human pluripotent stem cell culture enables highly efficient neuroepithelium derivation without small molecule inhibitors. *Stem Cells* **32,** 1032–1042 (2014).

148. Kunisada, Y., Tsubooka-Yamazoe, N., Shoji, M. & Hosoya, M. Small molecules induce efficient differentiation into insulin-producing cells from human induced pluripotent stem cells. *Stem Cell Res.* **8,** 274–284 (2012).

149. Neely, M. D. *et al.* DMH1, a highly selective small molecule BMP inhibitor promotes neurogenesis of hiPSCs: Comparison of PAX6 and SOX1 expression during neural induction. *ACS Chem. Neurosci.* **3,** 482–491 (2012).

150. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9,** 676–682 (2012).

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

152. Bhatia, S., Pilquil, C., Roth-Albin, I. & Draper, J. S. Demarcation of Stable

Subpopulations within the Pluripotent hESC Compartment. *PLoS One* **8,** (2013).

153. Cui, L. *et al.* Spatial distribution and initial changes of SSEA-1 and other cell adhesion-related molecules on mouse embryonic stem cells before and during differentiation. *J. Histochem. Cytochem.* **52,** 1447–1457 (2004).

154. Furusawa, T., Ohkoshi, K., Honda, C., Takahashi, S. & Tokunaga, T. 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 Embryo1. *Biol. Reprod.* **70,** 1452–1457 (2004).

155. Gokhale, P. J. *et al.* Culture Adaptation Alters Transcriptional Hierarchies among Single Human Embryonic Stem Cells Reflecting Altered Patterns of Differentiation. *PLoS One* **10,** (2015).

156. Wright, A. J. & Andrews, P. W. Surface marker antigens in the characterization of human embryonic stem cells. *Stem Cell Res* **3,** 3–11 (2009).

157. Ng, E. S., Davis, R., Stanley, E. G. & Elefanty, A. G. 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. *Nat. Protoc.* **3,** 768–776 (2008).

158. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: Cluster analysis of microarray data. *Bioinformatics* **18,** 207–208 (2002).

159. Badcock, G., Pigott, C., Goepel, J. & Andrews, P. W. The Human Embryonal Carcinoma Marker Antigen TRA-1-60 Is a Sialylated Keratan Sulfate Proteoglycan The Human Embryonal Carcinoma Marker Antigen TRA-1-60 Is a Sialylated Keratan Sulfate Proteoglycan. *Cancer Res.* 4715–4719 (1999).

160. Oka, M. *et al.* CD9 is associated with leukemia inhibitory factor-mediated maintenance of embryonic stem cells. *Mol. Biol. Cell* **13,** 1274–1281 (2002).

161. Nakamura, Y., Iwamoto, R. & Mekada, E. Expression and distribution of CD9 in myelin of the central and peripheral nervous systems. *Am. J. Pathol.* **149,** 575–583 (1996).

162. Park, K. R. *et al.* CD9 is expressed on human endometrial epithelial cells in association with integrins alpha(6), alpha(3) and beta(1). *Mol. Hum. Reprod.* **6,** 252–7 (2000).

163. Polanco, J. C. *et al.* Enrichment and Purging of Human Embryonic Stem Cells by Detection of Cell Surface Antigens Using the Monoclonal Antibodies TG30 and GCTM-2. *J. Vis. Exp.* **6,** (2013).

164. Vega Crespo, A., Awe, J. P., Reijo Pera, R. & Byrne, J. a. Human skin cells that express stage-specific embryonic antigen 3 associate with dermal tissue regeneration. *Biores. Open Access* **1,** 25–33 (2012).

165. Tippett, P., Andrews, P. W., Knowles, B. B., Solter, D. & Goodfellow, P. N. Red Cell Antigens P (Globoside) and Luke: Identification by Monoclonal Antibodies Defining the Murine Stage-Specific Embryonic Antigens -3 and -4 (SSEA-3 and SSEA-4). *Vox Sang.* **51,** 53–56 (1986).

166. Barnes, R. M., Firulli, B. A., Conway, S. J., Vincentz, J. W. & Firulli, A. B. Analysis of the Hand1 cell lineage reveals novel contributions to cardiovascular, neural crest, extra-embryonic, and lateral mesoderm derivatives. *Dev. Dyn.* **239,** 3086–3097 (2010).

167. Vodyanik, M. A. *et al.* A mesoderm-derived precursor for mesenchymal stem and endothelial cells. *Cell Stem Cell* **7,** 718–729 (2010).

168. Morikawa, Y. Extra-embryonic vasculature development is regulated by the transcription factor HAND1. *Development* **131,** 2195–2204 (2004).

169. Koeffler, H. P., Billing, R., Lusis, A. J., Sparkes, R. & Golde, D. W. An undifferentiated variant derived from the human acute myelogenous leukemia cell line (KG-1). *Blood* **56,** 265–73 (1980).

170. Lubbert, M., Herrmann, F. & Koeffler, H. P. Expression and regulation of myeloid-specific genes in normal and leukemic myeloid cells. *Blood* **77,** 909 LP – 924 (1991).

171. Siemerink, M. J. *et al.* CD34 marks angiogenic tip cells in human vascular endothelial cell cultures. *Angiogenesis* **15,** 151–163 (2012).

172. Yang, J. *et al.* CD34+ cells represent highly functional endothelial progenitor cells in murine bone marrow. *PLoS One* **6,** (2011).

173. Zhang, X. *et al.* Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7,** 90–100 (2010).

174. Sansom, S. N. *et al.* The level of the transcription factor Pax6 is essential for controlling the balance between neural stem cell self-renewal and neurogenesis. *PLoS Genet.* **5,** (2009).

175. Sullivan, K. F. & Cleveland, D. W. Identification of conserved isotype-defining variable region sequences for four vertebrate beta tubulin polypeptide classes. *Proc. Natl. Acad. Sci.* **83,** 4327–4331 (1986).

176. Caccamo, D. V *et al.* An immunohistochemical study of neuropeptides and neuronal cytoskeletal proteins in the neuroepithelial component of a spontaneous murine ovarian teratoma. Primitive neuroepithelium displays immunoreactivity for neuropeptides and neuron-associated beta-tu. *Am. J. Pathol.* **135,** 801–13 (1989).

177. Potterf, S. B., Furumura, M., Dunn, K. J., Arnheiter, H. & Pavan, W. J. Transcription factor hierarchy in Waardenburg syndrome: Regulation of MITF expression by SOX10 and PAX3. *Hum. Genet.* **107,** 1–6 (2000).

178. Plouhinec, J. L. *et al.* Pax3 and Zic1 trigger the early neural crest gene regulatory network by the direct activation of multiple key neural crest specifiers. *Dev. Biol.* **386,** 461–472 (2014).

179. Honoré, S. M., Aybar, M. J. & Mayor, R. Sox10 is required for the early development of the prospective neural crest in Xenopus embryos. *Dev. Biol.* **260,** 79–96 (2003).

180. Artus, J., Piliszek, A. & Hadjantonakis, A. K. The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17. *Dev. Biol.* **350,** 393–404 (2011).

181. Qu, X. Bin, Pan, J., Zhang, C. & Huang, S. Y. Sox17 facilitates the differentiation of mouse embryonic stem cells into primitive and definitive endoderm in vitro. *Dev. Growth Differ.* **50,** 585–593 (2008).

182. Fisher, J. B., Pulakanti, K., Rao, S. & Duncan, S. A. GATA6 is essential for endoderm formation from human pluripotent stem cells. *Biol. Open* **6,** 1084–1095 (2017).

183. Davis, R. P. *et al.* Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak-like cells and enables isolation of primitive hematopoietic precursors. *Blood* **111,** 1876–1884 (2008).

184. Chabot, J. R., Pedraza, J. M., Luitel, P. & Van Oudenaarden, A. Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature* **450,** 1249–1252 (2007).

185. Snapp, E. L. Fluorescent proteins: a cell biologist's user guide. *Trends Cell Biol.* **19,** 649–655 (2009).

186. Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. & Lehrach, H. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343,** 617–622 (1990).

187. Kispert, A., Koschorz, B. & Herrmann, B. G. The T protein encoded by Brachyury is a tissue-specific transcription factor. *EMBO J.* **14,** 4763–72 (1995).

188. Dobrovolskaia-Zavadskaia, N. Sur la mortification spontanée de la queue chez la souris nouveau-née et sur l'existence d'un caractère (facteur) héréditaire, non-viable. *CR des séances la société Biol.* **XC VII,** (1927).

189. Bollag, R. J. *et al.* An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T locus. *Nat. Genet.* **7,** 383–389 (1994).

190. Chelsey, P. Lethal Action in the Short-Tailed Mutation in the House Mouse. *Exp. Biol. Med.* **29,** 437–438 (1932).

191. Pennimpede, T. *et al.* In vivo knockdown of Brachyury results in skeletal

defects and urorectal malformations resembling caudal regression syndrome. *Dev. Biol.* **372,** 55–67 (2012).

192. Smith, J. C., Price, B. M. J., Green, J. B. A., Weigel, D. & Herrmann, B. G. Expression of a xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67,** 79–87 (1991).

193. Wilkinson, D. G., Bhatt, S. & Herrmann, B. G. Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* **343,** 657–659 (1990).

194. Chesley, P. Development of the short-tailed mutant in the house mouse. *J. Exp. Zool.* **70,** 429–459 (1935).

195. Herrmann, B. G. & Kispert, A. The t-genes in embryogenesis. *Trends Genet.* **0010,** 280–286 (1994).

196. Rashbass, P., Cooke, L. A., Herrmann, B. G. & Beddington, R. S. P. A cell autonomous function of Brachyury in T/T embryonic stem cell chimaeras. *Nature* **353,** 348 (1991).

197. Wilson, V., Rashbass, P. & Beddington, R. S. Chimeric analysis of T (Brachyury) gene function. *Development* **117,** 1321 LP – 1331 (1993).

198. Vujovic, S. *et al.* Brachyury, a crucial regulator of notochordal development, is a novel biomarker for chordomas. *J. Pathol.* **209,** 157–165 (2006).

199. King, T., Beddington, R. S. P. & Brown, N. A. The role of the brachyury gene in heart development and left-right specification in the mouse. *Mech. Dev.* **79,** 29–37 (1998).

200. Loebel, D. A. ., Watson, C. M., De Young, R. A. & Tam, P. P. . Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Developmental Biology* **264,** 1–14 (2003).

201. Aramaki, S. *et al.* A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev. Cell* **27,** 516–529 (2013).

202. Chatfield, J. *et al.* Stochastic specification of primordial germ cells from mesoderm precursors in axolotl embryos. *Development* **141,** 2429–2440 (2014).

203. Miettinen, M. *et al.* Nuclear brachyury expression is consistent in chordoma, common in germ cell tumors and small cell carcinomas, and rare in other carcinomas and sarcomas: An immunohistochemical study of 5229 cases. *Am. J. Surg. Pathol.* **39,** 1305–1312 (2015).

204. Tada, S. Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture. *Development* **132,** 4363–4374 (2005).

205. Vallier, L. *et al.* Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways. *PLoS One* **4,** (2009).

206. Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. & Keller, G. Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432,** 625–630 (2004).

207. Pereira, L. A. *et al.* Brachyury and related tbx proteins interact with the Mixl1 homeodomain protein and negatively regulate Mixl1 transcriptional activity. *PLoS One* **6,** (2011).

208. Latinkic, B. V & Smith, J. C. Goosecoid and mix.1 repress Brachyury expression and are required for head formation in Xenopus. *Development* **126,** 1769–1779 (1999).

209. Izumi, N., Era, T., Akimaru, H., Yasunaga, M. & Nishikawa, S.-I. Dissecting the Molecular Hierarchy for Mesendoderm Differentiation Through a Combination of Embryonic Stem Cell Culture and RNA Interference. *Stem Cells* **25,** 1664–1674 (2007).

210. Kalisz, M., Winzi, M., Bisgaard, H. C. & Serup, P. Even-Skipped Homeobox 1 controls human ES cell differentiation by directly repressing Goosecoid expression. *Dev. Biol.* **362,** 94–103 (2012).

211. Kubo, A. Development of definitive endoderm from embryonic stem cells

in culture. *Development* **131,** 1651–1662 (2004).

212. Gouon-Evans, V. *et al.* BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nat. Biotechnol.* **24,** 1402–1411 (2006).

213. Kita-Matsuo, H. *et al.* Lentiviral vectors and protocols for creation of stable hESC lines for fluorescent tracking and drug resistance selection of cardiomyocytes. *PLoS One* **4,** (2009).

214. Faial, T. *et al.* Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development* **142,** 2121–2135 (2015).

215. Koch, F. *et al.* Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev. Cell* **42,** 514–526.e7 (2017).

216. Arnold, S. J. *et al.* Brachyury is a target gene of the Wnt/β-catenin signaling pathway. *Mech. Dev.* **91,** 249–258 (2000).

217. Turner, D. A., Rué, P., Mackenzie, J. P., Davies, E. & Martinez Arias, A. Brachyury cooperates with Wnt/β-catenin signalling to elicit primitive-streak-like behaviour in differentiating mouse embryonic stem cells. *BMC Biol.* **12,** (2014).

218. Bernardo, A. S. *et al.* BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* **9,** 144–155 (2011).

219. Mendjan, S. *et al.* NANOG and CDX2 Pattern Distinct Subtypes of Human Mesoderm during Exit from Pluripotency. *Cell Stem Cell* **15,** 310–325 (2014).

220. Gokhale, P. J., Giesberts, A. M. & Andrews, P. W. Brachyury is expressed by human teratocarcinoma cells in the absence of mesodermal differentiation. *Cell Growth Differ.* **11,** 157–62 (2000).

221. Bernemann, C. *et al.* Distinct developmental ground states of epiblast

stem cell lines determine different pluripotency features. *Stem Cells* **29,** 1496–1503 (2011).

222. Ghule, P. N. *et al.* Reprogramming the pluripotent cell cycle: restoration of an abbreviated G1 phase in human induced pluripotent stem (iPS) cells. *J. Cell. Physiol.* **226,** 1149–1156 (2011).

223. Hay, D. C., Sutherland, L., Clark, J. & Burdon, T. Oct-4 Knockdown Induces Similar Patterns of Endoderm and Trophoblast Differentiation Markers in Human and Mouse Embryonic Stem Cells. *Stem Cells* **22,** 225–235 (2004).

224. Morrisey, E. E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev.* **12,** 3579–3590 (1998).

225. Hyslop, L. *et al.* Downregulation of NANOG Induces Differentiation of Human Embryonic Stem Cells to Extraembryonic Lineages. *Stem Cells* **23,** 1035–1043 (2005).

226. Seguin, C. A., Draper, J. S., Nagy, A. & Rossant, J. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell Stem Cell* **3,** 182–195 (2008).

227. Lolas, M., Valenzuela, P. D. T., Tjian, R. & Liu, Z. Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. *Proc. Natl. Acad. Sci.* **111,** 4478–4483 (2014).

228. Chan, S. S. K. *et al.* Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. *Cell Stem Cell* **12,** 587–601 (2013).

229. Wittler, L. *et al.* Expression of Msgn1 in the presomitic mesoderm is controlled by synergism of WNT signalling and Tbx6. *EMBO Rep.* **8,** 784–789 (2007).

230. Mfopou, J. K. *et al.* Efficient definitive endoderm induction from mouse embryonic stem cell adherent cultures: A rapid screening model for differentiation studies. *Stem Cell Res.* **12,** 166–177 (2014).

231. Kim, P. T. W. & Ong, C. J. Differentiation of definitive endoderm from mouse embryonic stem cells. *Results Probl. Cell Differ.* **55,** 303–319 (2012).

232. Irie, N. *et al.* SOX17 is a critical specifier of human primordial germ cell fate. *Cell* (2015). doi:10.1016/j.cell.2014.12.013

233. Liu, Y. *et al.* Sox17 is essential for the specification of cardiac mesoderm in embryonic stem cells. *Proc. Natl. Acad. Sci.* **104,** 3859–3864 (2007).

234. Stefanovic, S. *et al.* Interplay of Oct4 with Sox2 and Sox17: A molecular switch from stem cell pluripotency to specifying a cardiac fate. *J. Cell Biol.* **186,** 665–673 (2009).

235. Aksoy, I. *et al.* Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J.* **32,** 938–953 (2013).

236. Wolf, X. A., Serup, P. & Hyttel, P. Three-dimensional immunohistochemical characterization of lineage commitment by localization of T and FOXA2 in porcine peri-implantation embryos. *Dev. Dyn.* **240,** 890–897 (2011).

237. Tamplin, O. J. *et al.* Microarray analysis of Foxa2 mutant mouse embryos reveals novel gene expression and inductive roles for the gastrula organizer and its derivatives. *BMC Genomics* (2008). doi:10.1186/1471-2164-9-511

238. Gouti, M. *et al.* In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLoS Biol.* **12,** (2014).

239. Eiselleova, L. *et al.* Comparative study of mouse and human feeder cells for human embryonic stem cells. *Int. J. Dev. Biol.* **52,** 353–363 (2008).

240. Naujok, O., Diekmann, U. & Lenzen, S. The Generation of Definitive Endoderm from Human Embryonic Stem Cells is Initially Independent from Activin A but Requires Canonical Wnt-Signaling. *Stem Cell Rev. Reports* **10,** 480–493 (2014).

241. Giacomelli, E. *et al.* Three-dimensional cardiac microtissues composed of

cardiomyocytes and endothelial cells co-differentiated from human pluripotent stem cells. *Development* **144,** 1008–1017 (2017).

242. Garcia-Gonzalo, F. R. & Belmonte, J. C. I. Albumin-associated lipids regulate human embryonic stem cell self-renewal. *PLoS One* **3,** (2008).

243. Qin, H. *et al.* YAP Induces Human Naive Pluripotency. *Cell Rep.* **14,** 2301–2312 (2016).

244. Kleinman, H. K. *et al.* Isolation and Characterization of Type IV Procollagen, Laminin, and Heparan Sulfate Proteoglycan from the EHS Sarcoma. *Biochemistry* **21,** 6188–6193 (1982).

245. Kleinman, H. K. & Martin, G. R. Matrigel: Basement membrane matrix with biological activity. *Semin. Cancer Biol.* **15,** 378–386 (2005).

246. Orkin, R. W. *et al.* A murine tumor producing a matrix of basement membrane. *J. Exp. Med.* **145,** 204–20 (1977).

247. Hughes, C. S., Postovit, L. M. & Lajoie, G. A. Matrigel: a complex protein mixture required for optimal growth of cell culture. *Proteomics* **10,** 1886–1890 (2010).

248. Gattazzo, F., Urciuolo, A. & Bonaldo, P. Extracellular matrix: A dynamic microenvironment for stem cell niche. *Biochimica et Biophysica Acta - General Subjects* **1840,** 2506–2519 (2014).

249. Vukicevic, S. *et al.* Identification of multiple active growth factors in basement membrane matrigel suggests caution in interpretation of cellular activity related to extracellular matrix components. *Exp. Cell Res.* **202,** 1–8 (1992).

250. Yao, S. *et al.* Long-term self-renewal and directed differentiation of human embryonic stem cells in chemically defined conditions. *Proc. Natl. Acad. Sci.* **103,** 6907–6912 (2006).

251. Tong, Z. *et al.* Towards a defined ECM and small molecule based monolayer culture system for the expansion of mouse and human intestinal stem cells. *Biomaterials* **154,** 60–73 (2018).

252. Schvartz, I., Seger, D. & Shaltiel, S. Vitronectin. *Int. J. Biochem. Cell Biol.* **31,** 539–544 (1999).

253. Boron, W. F. & Boulpaep, E. L. Medical Physiology. *Postgrad. Med. J.* **51,** 683–683 (1975).

254. Braam, S. R. *et al.* Recombinant Vitronectin Is a Functionally Defined Substrate That Supports Human Embryonic Stem Cell Self-Renewal via αVβ5 Integrin. *Stem Cells* **26,** 2257–2265 (2008).

255. Rowland, T. J. *et al.* Roles of Integrins in Human Induced Pluripotent Stem Cell Growth on Matrigel and Vitronectin. *Stem Cells Dev.* **19,** 1231–1240 (2010).

256. Heng, B. C. *et al.* Translating Human Embryonic Stem Cells from 2-Dimensional to 3-Dimensional Cultures in a Defined Medium on Laminin- and Vitronectin-Coated Surfaces. *Stem Cells Dev.* **21,** 1701–1715 (2012).

257. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459,** 262–265 (2009).

258. Hayashi, Y. & Furue, M. K. Biological Effects of Culture Substrates on Human Pluripotent Stem Cells. *Stem Cells International* **2016,** (2016).

259. Xu, C. *et al.* Feeder-free growth of undifferentiated human embryonic stem cells. *Nat. Biotechnol.* **19,** 971–974 (2001).

260. Ludwig, T. E. *et al.* Feeder-independent culture of human embryonic stem cells. *Nat. Methods* **3,** 637–646 (2006).

261. Chen, B. *et al.* Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat. Chem. Biol.* **5,** 100–107 (2009).

262. Mo, M. L. *et al.* Inhibition of the Wnt palmitoyltransferase porcupine suppresses cell growth and downregulates the Wnt/β-catenin pathway in gastric cancer. *Oncol. Lett.* **5,** 1719–1723 (2013).

263. Berge, D. ten *et al.* Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat. Cell Biol.* **13,** 1070–1077 (2011).

264. Turner, D. A. *et al.* Wnt/β-catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells. *Development* **141,** 4243–4253 (2014).

265. Savory, J. G. A. *et al.* Cdx2 regulation of posterior development through non-Hox targets. *Development* **136,** 4099–4110 (2009).

266. Hendrickx, M., Van, X. H. & Leyns, L. Anterior-posterior patterning of neural differentiated embryonic stem cells by canonical Wnts, Fgfs, Bmp4 and their respective antagonists. *Dev. Growth Differ.* **51,** 687–698 (2009).

267. Gouti, M. *et al.* A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev. Cell* **41,** 243–261.e7 (2017).

268. Skelton, R. J. P. *et al.* SIRPA, VCAM1 and CD34 identify discrete lineages during early human cardiovascular development. *Stem Cell Res.* **13,** 172–179 (2014).

269. Graf, T. & Stadtfeld, M. Heterogeneity of Embryonic and Adult Stem Cells. *Cell Stem Cell* **3,** 480–483 (2008).

270. Crisan, M. & Dzierzak, E. The many faces of hematopoietic stem cell heterogeneity. *Development* **143,** 4571–4581 (2016).

271. Evseenko, D. *et al.* Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *Proc Natl Acad Sci U S A* **107,** 13742–13747 (2010).

272. Chu, L.-F. *et al.* Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.* **17,** 173 (2016).

273. Li, Y. *et al.* Transcriptome analysis reveals determinant stages controlling human embryonic stem cell commitment to neuronal cells. *J. Biol. Chem.* **292,** 19590–19604 (2017).

274. Azzolin, L. *et al.* Role of TAZ as mediator of wnt signaling. *Cell* **151,** 1443–1456 (2012).

275. Konsavage, W. M. & Yochum, G. S. Intersection of Hippo/YAP and Wnt/B-catenin signaling pathways. *Acta Biochimica et Biophysica Sinica* **45,** 71–79 (2013).

276. Heallen, T. *et al.* Hippo pathway inhibits wnt signaling to restrain cardiomyocyte proliferation and heart size. *Science (80-.).* **332,** 458–461 (2011).

277. Ohno, Y. *et al.* Distinct iPS cells show different cardiac differentiation efficiency. *Stem Cells Int.* 11 (2013). doi:10.1155/2013/659739

278. Osafune, K. *et al.* Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat. Biotechnol.* **26,** 313–315 (2008).