The University Of Sheffield.

Identification, Characterisation and Manipulation of Substates of Human Pluripotent Stem Cells with Potential Mesoderm Bias

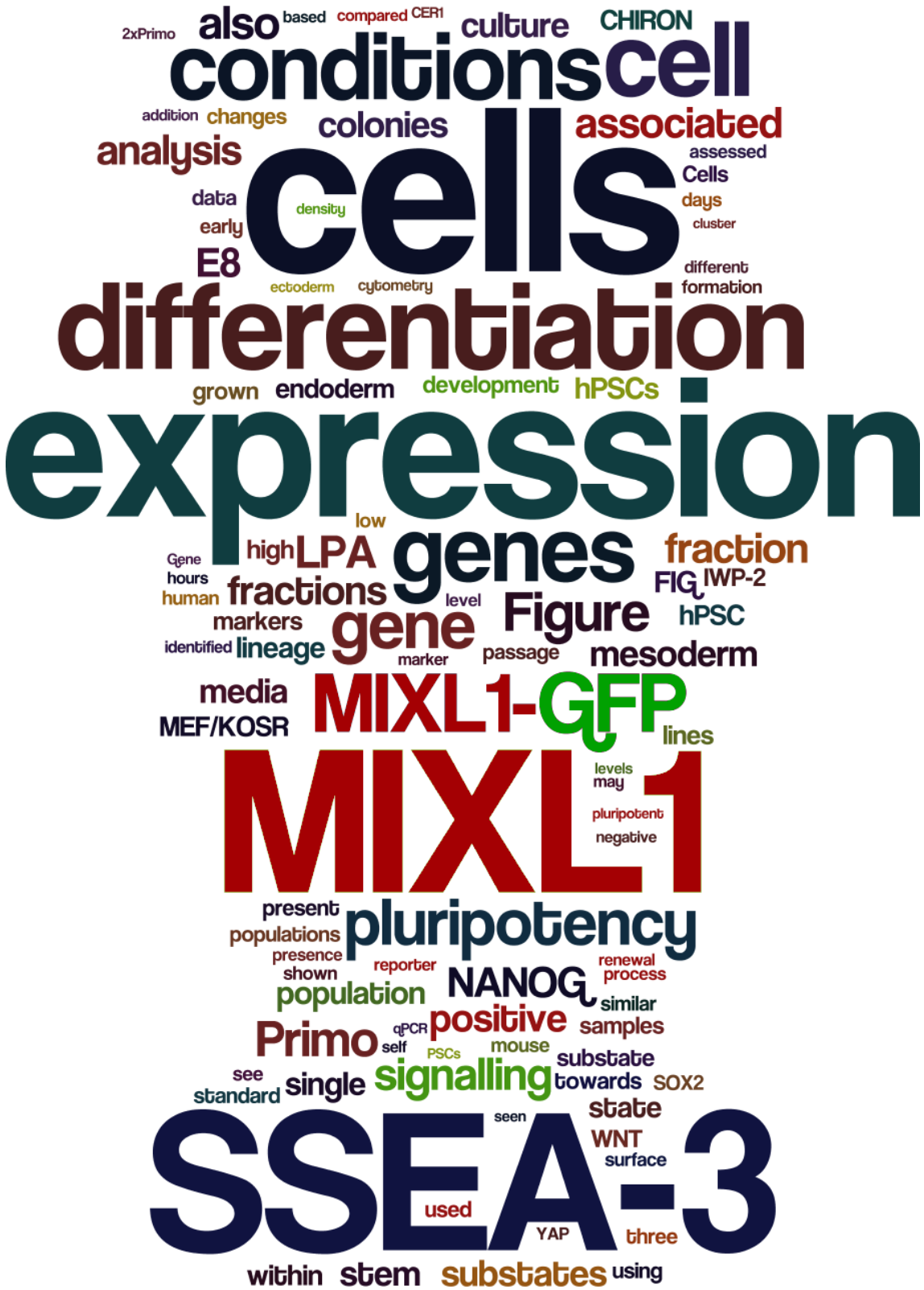
By:

Dylan James Edward Stavish

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

The University of Sheffield
Faculty of Science
Department of Biomedical Science

July 2018



Thesis Word Cloud

The top 100 most used words in this thesis, the size of each word is relative to the number of times used. To give you an overall idea of what you are about to read and introduce the concept of word clouds.

Acknowledgements

Wow, where do I start? I could feasible write another 10,000 words just for this section. My PhD didn't start in the best circumstances and during my very first month my father passed away. Many people helped me through this difficult time and as I finish this journey some of the pain returns. The person I would like to share this work with the most is not here, yet I take solace in knowing I completed this journey with Dr. Phil Stavish in my heart to emerge, hopefully, as Dr. Dill Stavish.

After the "rocky" start to my PhD, I began to really enjoy my work, the project and all the members of the lab. I can honestly say that my PhD was one of the best periods of my life. I had the good fortune of being involved in a large European Consortium. Although this entailed many, many, many, reports, it also involved many, many, many trips to wonderful destinations across Europe. Throughout my project I also got to collaborate with some amazing labs and amazing scientists, many of whom helped me along the way, to name but a few, Charlotta Boiers and Tariq Enver.

I am the last of a group of students which started at the same time, so now I must say my thanks to all the lab members past and present who kept me going. Firstly, the beer pong playas, Tom Frith(White Kobe, my ideas man), Tom Allison(Alli-GATAr-6, my cholesterol confidant), Jim Hackland (Jimbo, my thought provoker BLA, TDi), Oliver Thomson(Crollie, crossword king). To other members past and present, David Preskey (my bro, worms master but the world's sorest loser), James Mason (the excel wizard), Adam Hirst (The Game Maker), Jason Halliwell (Jase, if it ain't fried chicken I'm not interested), Chris Price (Pricey Boy, the Prince of Cakes). I also want to thank those who gave me "constructive" criticism and discussed science with me as an equal Paul Gokhale, Ivana Barbaric, Anestis Tsakiridis, and especially post viva, Andrew Furley.

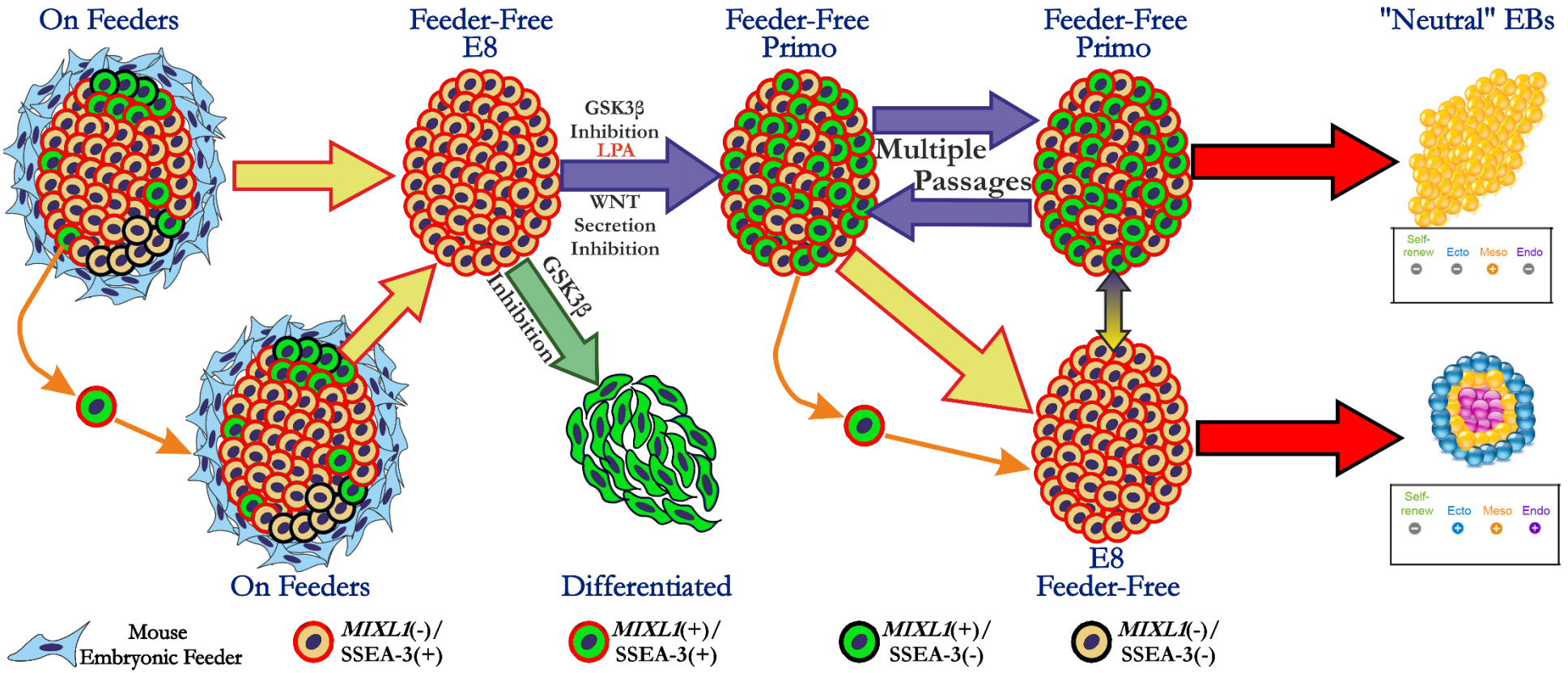
Aside from the Lab members I also want to thank my family for supporting through this. To Chloe and Phoebe, thank you for picking up the slack at all the family functions when I needed to go to the lab. Also, thanks for always being there for me whenever I needed to talk. To my mum, Jacquelyn Stavish, thank you for always believing in me even at times when I doubted myself. To my "brothers" Kris King and Ryan Pearson, thank you for being there when I needed you. Also, thanks for literally hundreds of FIFA games early in my PhD and also for no games near the end, I didn't need the distraction.

A special note for Ingrid Saldana, "mi cielo", thank you for all the support during these crucial late stages on my PhD. Together, we have shared each other's burdens and collectively lightened the load. You are an incredibly brave person and you give me strength.

My biggest regret in my PhD comes in the form of many a modern person's regrets, Cholesterol. You see, my other project throughout this PhD involved the wonderful world of cholesterol and its impact on pluripotent cells. Alas, time and commitments got in the way of things my work on cholesterol is on the cutting room floor, for now… I did manage to sneak it in to my final medium recipe though! My extensive work on human Naïve PSC also didn't make the cut.

Finally, I must thank Professor Peter Andrews. 6 years ago, I sat in your office a passionate but struggling scientist and you took a chance on me when no one else seemed to be willing. After two years as a research technician, I was honoured to be awarded a PhD position with you. I cannot truly say whether I have amassed more knowledge from or appreciation for Professor Andrews. I can however, undeniably say that I am a better scientist because of his tutelage.

Graphical Abstract



Abstract:

Human Pluripotent Stem Cells (PSC) exist in heterogeneous populations when grown in standard culture. The population can be separated into substates based on metabolism, antigen expression and gene expression. We utilised a human PSC reporter line for *MIXL1* coupled with stem cell surface antigen, SSEA-3, to identify substates within standard stem cell culture. The *MIXL1*(+)/SSEA-3(+) substate contains a stem cell population, as defined by the establishment of self renewing clonal lines which retain their pluripotent potential, whilst exhibiting functional differences in regards to differentiation on a population and single cell basis. Bulk and single cell transcriptomics from standard culture and a bulk time course of directed differentiation was performed on the substates defined by *MIXL1*/SSEA-3 expression and the gene expression changes reveal a continuum from pluripotency through lineage priming and finally to differentiation. While the *MIXL1*(**+**)/SSEA-3(**+**) fraction is apparent when grown on a mouse embryonic feeder layer, it is virtually non-existent in the defined culture system of E8 and vitronectin. Utilising this defined system, we sought to balance pro-self renewal signals and pro-differentiation signals to recreate a *MIXL1*(**+**)/SSEA-3(**+**) substate. We used Lysophophatidic Acid (LPA) to attenuate the differentiation effects of GSK3β inhibition by CHIR99021. As within standard culture the newly generated *MIXL1*(**+**)/SSEA-3(**+**) substate contains a stem cell population, verified through the production of clonal lines. The newly generated fractions also exhibit a functional bias in terms of their differentiation potential, in particular neutral Embryoid Bodies producing enhanced mesoderm populations. We performed bulk and single cell transcriptomics on the generated fractions to map them back to the populations seen in standard conditions and the time course of differentiation. Cells correlated with the populations seen in standard culture and early differentiation but crucially expressed an active pluripotency gene network. Optimisation of our “Primo” medium has enabled us to maintain cells in this lineage biased state for multiple passages, with a normal karyotype and pluripotency-associated surface marker expression. Cells from Primo medium can also be transitioned back into standard culture conditions with transcriptional changes reverting back to normal human PSC expression, highlighting the interconversion between states of human PSC.

Table of Abbreviations

Abbreviation	Meaning
(e)GFP	(enhanced) Green Fluorescent Protein
2nd Only	Secondary only antibody staining
BMP	Bone Morphogenetic Protein
BRA	BRACHYURY
CHIRON	CHIR99021 (GSK3β inhibitor)
Ct value	Cycle threshold value
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
E8V	Essential 8 medium with Vitronectin matrix
EB	Embryoid Body
EC	Embryonal Carcinoma
EpiSC	Epiblast-like stem cell
ESC	Embryonic Stem Cells
FACS	Fluorescence Activated Cell Sorting
FCS	Fetal Calf Serum
FGF	Fibroblast Growth Factor
GSK3β	Glycogen Synthase Kinase 3 beta
ICM	Inner Cell Mass
iPS	Induced Pluripotent Stem cell
KO-DMEM	Knock-out Dulbecco's Modified Eagle Medium
KOSR	Knock Out Serum Replacement
LIF	Leukaemia Inhibitory Factor
LPA	Lysophosphatidic acid
M+3-	*MIXL1*(**+**)/SSEA-3(-)
M+3+	*MIXL1*(**+**)/SSEA-3(**+**)
M-3+	*MIXL1*(-)/SSEA-3(**+**)
MEF	Mouse Embryonic Feeder
MEF/KOSR	Growth conditions on MEFs with KOSR containing medium
nucYAP	YAP localised to nucleus
PBS	Phosphate Buffered Saline
PSC	Pluripotent Stem Cells
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNA-seq	Ribonucleic Acid sequencing
Rocki	ROCK inhibitor (Y-27632)
TGFβ	Transforming Growth Factor Beta

Table of Contents

Tables

Table of Figures.

1. Chapter 1: Introduction

Human Embryonic Stem Cells (ESC) have been heralded as a revolutionary tool in regenerative medicine since their first derivation from human embryos in 1998 (Thomson et al., 1998). These cells seemingly retained the enhanced differentiation potential of the early embryonic cells, which have the ability to form all three germ layers of the body and subsequently all somatic cells. This aspect, coupled with their ability to self renew indefinitely, created an attractive platform for the generation of particular cell types or tissues to treat individuals suffering from disease or physical damage. However, the use of human ESC in regenerative medicine has not translated as quickly to real world applications as many predicted. Complications with their use vary from ethical issues to safety concerns. The most pertinent safety issues pertained to immunological compatibility and genetic stability. With the discovery that forced expression of key pluripotency-associated genes could reprogram a somatic cell to an ES like state, termed induced pluripotent stem (iPS) cells (Takahashi et al., 2007; Takahashi and Yamanaka, 2006), and the development of non-integrating reprogramming methods (Fusaki et al., 2009; Warren et al., 2010) , some of the largest obstacles to translational use were circumvented. The establishment of stable pluripotent stem cells (PSC), both ESC and iPS, has reinvigorated the field of PSC for regenerative medicine. Testament to this, recent developments have seen retinal pigment cells derived from human ESC used in clinical trials to treat age related macular degeneration (da Cruz et al., 2018).

Cellular differentiation, the process by which cells transition from one cell type to another, is a vital aspect of most biological systems. This process allows for the specification of functionally discrete cell types and the generation of multifaceted tissues within an organism. Some cells can have the ability to become only one cell type, termed unipotent, whereas PSC can produce a multitude of other cell types, and thus an increased variety of fate decisions. Both intrinsic and extrinsic factors can influence the fate decisions of PSC and, of course, by doing so these fate decisions can dramatically impact the development and maintenance of an organism, in vivo. In vitro, these can affect the differentiation of PSC in a similar manner.

If PSC are to achieve the early aspirations of the field innovators there has to be a focus on the molecular mechanisms that guide these fate decisions. In depth research has to be undertaken to generate robust differentiation protocols designed to produce specific cells or tissues for targeted regenerative applications. Directed differentiation of PSC to specific cell types has been achieved for many lineages. The efficiency of these differentiation protocols, however, is not ideal, particularly from the perspective of translation medicine, as protocols can be hampered with low cell numbers or non-specific differentiation. The plethora of fate decisions accessible to PSC gives them significant potential for regenerative medicine but this aspect, maybe

surprisingly, is also one of the greatest concerns with their use. Inefficient differentiation protocols can result in contamination of undesired cell types within the population, which may have dangerous implications when introduced into the target tissue of prospective patients. The research described in this thesis was undertaken as part of a wider project, PluriMes, with an overall aim of guiding human PSC towards clinically relevant mesodermal derivatives for use in regenerative medicine. My work within this project sought to optimise the initial fate decision of differentiating human PSC towards a mesodermal fate.

Much like PSC, adult stem cells are multipotential cells with the ability to form many cell types albeit with a reduced differentiation potential compared to PSC. The maintenance of many tissues such as skin, intestinal tract, and skeletal muscle are dependent on a stem cell population which provides a progenitor pool for the multiple cell types which make up the tissue. These key characteristics position stem cells at pivotal developmental events as well as at sites for the continual maintenance of tissues in the developed organism. Adult stem cells are maintained in niches within tissues which regulate their stem cell identity through signalling and environmental maintenance. When appropriately activated by intrinsic and extrinsic cues cells will leave their niche and subsequently commit to a new lineage. Commitment describes a physiologically irreversible cell fate decision, restricting the cell to a new cell type. Prior to commitment cells may undergo early specification events which impact upon their final cell fate decision.

Populations of stem cells that exhibit similar characteristics and are broadly classified as the same cell type can nevertheless be heterogenous, with individual cells exhibiting differences in gene expression, surface marker expression or even metabolism. These heterogeneous differences are transient as cells inter-convert between states. Subsequently, when viewed as an average over time these differences can be masked in analysis. Since stem cells have many cellular fates, self renewal or differentiation to distinct cell types, the differences underlining the heterogeneity could impact on the fate decisions which cells undergo. The heterogeneity seen within these populations may be a result of differences in the early lineage specification events which cells may be undergoing.

The lineage commitment of Haematopoietic stem cells (HSC) has been heavily investigated. Analysis of bulk populations has previously masked the heterogeneity within these populations, but recent developments in single cell analysis have highlighted this and the downstream implications of these heterogeneous differences. Many models of early lineage specification of HSCs have been proposed in the past, but more recent work speaks to a stochastic rather than deterministic model of lineage specification and future commitment (Nimmo et al., 2015; Pina et al., 2012; Pina et al., 2015; Teles et al., 2013). These stem cell dynamics may be shared

amongst other stem cells including PSC and could offer insight into how PSC are allowed to explore these lineage-specified states without committing to differentiation.

The pluripotent cells of the early embryo go through a process of coordinated development towards differentiated lineages. Investigation into early mammalian development, particularly that of the mouse, has provided a substantial knowledge base for understanding the mechanisms that control early fate decision and downstream differentiation of PSC in an embryo setting. Although there are substantial differences between human and mouse embryos, which also extend to the PSC derived from them, there are conserved elements between the species during this early stage in development. The early specification and subsequent commitment of PSC is controlled, in part, by changing dynamics to the niche in which the cells reside. Cytokines, chemokines, and growth factors present in the niche can begin to alter the gene expression of cells as they respond to the environmental cues. Heterogeneity can arise through differences in the activation of lineage gene expression by these environmental factors. Understanding the signalling required to control the maintenance of the stem cell identity, the early specification of a lineage and the eventual commitment of a cell provides insight into how a particular stem cell state might be maintained in vitro. Whilst exploring these concepts, my overall goal was to manipulate the environment in which we grow PSC, affecting their early lineage specification and ultimately biasing their fate decisions prior to commitment, more specifically biasing their fate towards mesodermal cell types.

1.1. The Mammalian Embryo and Potency

The cellular makeup of the human body is vast, both in size and in its diversity. Recent estimates indicate the average human body contains over 35 trillion cells (Bianconi et al., 2013). These cells can be broadly classified into several hundred cell types that differ drastically from each other in morphology, gene expression pattern and function. Generation of this vast array of cells, requires a carefully orchestrated process of embryo development that begins with a single cell, the fertilised zygote. The zygote contains genetic information from the father's sperm and mother's ovum, which provides the blueprint for development, but environmental cues play a significant role in guiding the zygote through proper development. Therefore, cellular differentiation potential is highest at the earliest stages of development.

The word potency describes a cell's inherent ability to differentiate into other cell types and can be broadly split into three categories, totipotency, pluripotency, and multipotency. Cells transition through these states of potency during mammalian embryo development, of which

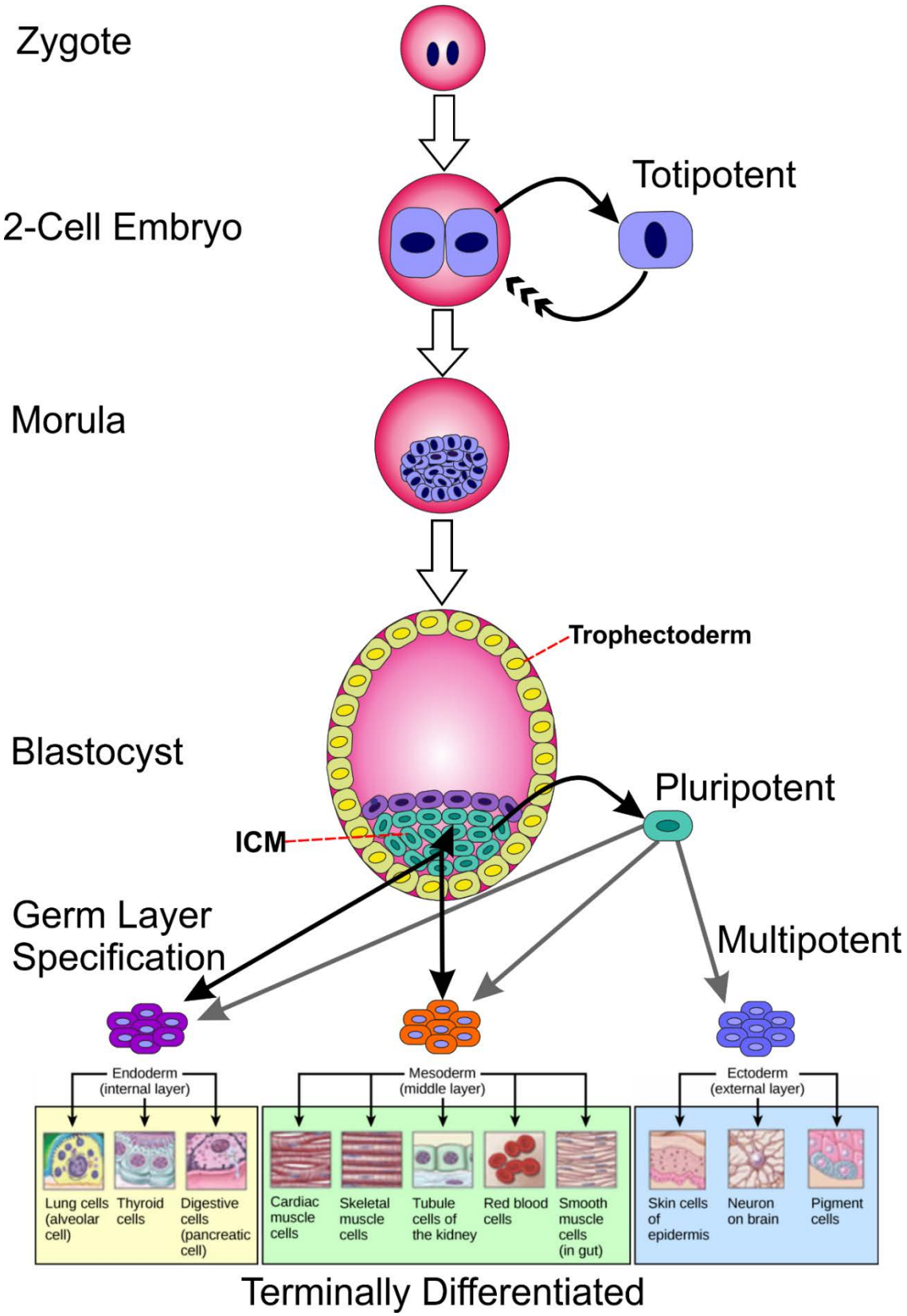

Zygote
2-Cell Embryo
Totipotent
Morula
Trophectoderm
Blastocyst
Pluripotent
ICM
Germ Layer Specification
Multipotent
Endoderm (internal layer)
Mesoderm (middle layer)
Ectoderm (external layer)
Lung cells (alveolar cell)
Thyroid cells
Digestive cells (pancreatic cell)
Cardiac muscle cells
Skeletal muscle cells
Tubule cells of the kidney
Red blood cells
Smooth muscle cells (in gut)
Skin cells of epidermis
Neuron on brain
Pigment cells
Terminally Differentiated

Figure 1-1 Stages of cellular potency during the development of the mouse embryo

The fertilised mouse zygote undergoes progressive cleavage events until it develops into a small cluster of cells named the morula. Cells at the 2-cell stage of cleavage are totipotent and when separated at this stage are able to regenerate an entire morula and subsequently a fully developed embryo. The cells of the morula begin to become specified as they become the blastocyst, outer cells forming the trophoblast and inside cells forming an inner cell mass. Cells within the inner cell mass are pluripotent and give rise to all three germ layers. When separated at this stage cells retain this pluripotent potential in vitro. As development of the embryo continues pluripotency is lost during gastrulation as three germ layers, endoderm, mesoderm and ectoderm, are specified. After germ layer specification cells are now multipotent, restricted to progeny of a specific germ layer but able to differentiate into many cell types within that lineage. Eventually the cells can become terminally differentiated progeny with restricted or limited differentiation ability.

mouse development has been the most investigated (Figure 1-1). Critical stages in mouse embryo development have been shown to mirror similar events in other mammalian systems, including humans. Our research interests are particularly focused on human development but nevertheless the mouse offers a well investigated mammalian paradigm for our work, providing insight into cellular states of potency.

In the first stages of mouse embryo development, as the zygote undergoes cleavage to form small cluster of cells, the individual cells, or blastomeres, remain totipotent. Totipotency is defined by a cell's ability to form a complete embryo consisting not only of all derivatives of the three germ layers, the ectoderm, endoderm and mesoderm, but also the extra embryonic tissues required for the accurate development of the entire organism. Extraembryonic tissues such as the trophoblast and yolk sac have crucial roles in development, including providing nutrients to the embryo, ensuring stable implantation in to the uterus and healthy development of the placenta.

As development progresses, the morula forms as cells begin their transition from totipotency to pluripotency, occupying a transitionary state some have termed plenipotent (Condic, 2014). The loss of complete totipotency has been indicated to occur after the 2-cell stage of cleavage in mouse, but up to the 8-cell stage in sheep, cow and monkey (reviewed in (Suwinska, 2012). The morula becomes the blastocyst, as cells undergo their first moments of distinct specialisation. Trophoblast cells arise from the outer cells of the morula and form the wall of the blastocyst in

which the inner cavity retains a cluster of cells termed the inner cell mass (ICM). Cells from the ICM already show vast differences from their differentiated trophoblast counterparts. Trophoblast cells have restricted potency and are limited to the formation of extra embryonic lineages. The cells within the inner cell mass will go on to form all of the somatic cells that comprise the human body. At this stage trophoblast cells are multipotent as they can only differentiate into a few cell types but the cells of the ICM are pluripotent as they can differentiate into any of the three main germ layers.

As the mouse embryonic development continues the cells in the ICM are specified towards distinct germ layers during the process of gastrulation. During this process, cells transition from a pluripotent state to a multipotent state, restricted to one of the three lineages. After specification into endoderm, ectoderm and mesoderm, the embryo continues to develop with cells continuing to progress towards the plethora of terminally differentiated cells which contribute to the entire organism. Broadly speaking endoderm cells contribute to respiratory and digestive organ formation, mesoderm to muscular and cardiovascular systems and ectoderm to the nervous system.

1.2. Pluripotency; states present in vivo and those captured in vitro

Lewis Wolpert is widely quoted as saying “It is not birth, marriage or death but gastrulation which is truly the most important time in your life” (Slack, 1984). He said this to highlight the importance of studying early embryo development. Pluripotency can be seen as the point at which all avenues of differentiation towards any somatic cell are still possible, while gastrulation is the process by which the fate of pluripotent cells becomes restricted as they transition towards the different specialised germ layers. Therefore, to paraphrase Shakespeare and pay homage to Lewis Wolpert, I would say, “it is not at the end of university but rather inside the pluripotent inner cell mass when the world is truly your oyster”. Much like Wolpert, I say this to highlight the importance of studying early embryo development but also to highlight the pivotal role of pluripotent cells during this critical period of development and how these cells, expanded in vitro, have massive potential for understanding biological events of the early embryo and future use in regenerative medicine.

The pluripotent states seen in vivo are present only briefly during the development of the early embryo. Many years of research have been undertaken to attempt to capture these states and retain their pluripotent potential within a proliferating self renewing population. Harnessing the self renewal abilities of PSC, these momentary pluripotent states can be expanded, seemingly indefinitely, making them an attractive prospect for developmental research and regenerative medicine.

1.2.1. Mouse Pluripotent States

The first instances of propagation of a pluripotent cell state in vitro arose through the explantation of cells from the malignant counterpart of teratomas, teratocarcinomas. While teratomas contain cells from all three germ layers, teratocarcinomas also contain a population of self renewing, pluripotent cells termed embryonal carcinoma (EC) cells or cancer stem cells (Kleinsmith and Pierce, 1964). The genetic mutations that resulted in the formation of cancer had captured a pluripotent identity within some of the cells arising in the tumour.

The tumour location and incidence indicated that these tumours arose from cells in early development. Evans and Kaufman (1981) assessed whether the inner cell mass of the early mouse embryo contained cells similar to ECs which could be expanded in vitro. They found that cells from the inner cell mass of day 2.5 mouse blastocysts could be expanded on mitotically inactivated STO feeder cells. These cells resembled EC lines but, crucially, maintained a normal diploid karyotype (Evans and Kaufman, 1981). The term "embryonic stem cells" was later proposed by Martin, 1981, who was able to derive a similar population using teratocarcinoma conditioned media. Later analysis showed that the gene expression of these cells correlates well with the inner cell mass of the early embryo (Martin Gonzalez et al., 2016). The derivation of mouse ESC now provided an in vitro platform to investigate the earliest stages of mammalian development.

The work of Austin Smith highlighted the ability of Leukaemia Inhibtory Factor (LIF) to promote self renewal and maintenance of mouse ESC when grown in vitro (Smith et al., 1988). Further to this, BMP4 secreted by the mouse embryonic feeder layer on which these cells are grown, also helped to promote self renewal of these cells (Qi et al., 2004). The pluripotent identity was maintained by a gene regulatory network (GRN) which contained genes such as Pou5f1, Nanog and Sox2. LIF was shown to promote this pluripotent GRN through LIF/STAT3 signalling (Hall et al., 2009). Initially mouse PSC could only be derived from "permissive" strains of mice and other similar rodent counterparts such as rat could not generate ES lines using serum and LIF conditions. Growing mouse ESC in serum and LIF conditions revealed these propagated cells actually formed a heterogenous population and cells could be separated with expression of key transcription factors such as STELLA (Hayashi et al., 2008), or NANOG (Smith et al., 2017). Investigations into these cells also demonstrated that these states interconverted within this heterogeneous population. The addition of a GSK3β inhibitor and a MEK/ERK inhibitor in feeder free cultures not only stabilised a more homogenous population of mouse ESC but also allowed lines to be established from non-permissive mice and rats (Ying et al., 2008). The formulation of this "2iLIF" medium highlights two concepts; firstly, that correct signal manipulation can "trap"

pluripotent cells in vitro and, secondly, that optimal conditions can maintain a more homogenous population.

Pluripotent cells exist only briefly during embryo development as signalling cues drive their differentiation to the three main germ layers, endoderm, ectoderm and mesoderm. In the mouse they emerge in the inner cell mass but they are not fully committed towards a lineage until gastrulation is finished. Yet previously it was thought that only the inner cell mass of the early blastocyst contained derivable pluripotent stem cells. However, Tesar et al and Brons et al (2007), demonstrated that pluripotent cells could be derived from the post-implantation epiblast of the mouse embryo. Implantation of the embryo, produces a pronounced change in the PSC that can be derived from the embryo. These cells, termed Epi-Stem cells (EpiSC), are morphologically distinct from their conventional mouse ES counterparts. Mouse ESC typically grow in tightly compact round colonies, whereas these EpiSC grow as flat colonies.

Thus, although the term pluripotency is specific, it can still encompass these distinctly separate states in development. Mouse PSC have two broad classifications based on their relation to embryonic development. Cells derived from pre-implantation embryos are deemed "Naïve" and post implantation as "Primed" (Nichols, 2009). Classical mouse ESC, derived from pre-implantation embryos, are considered to be naive, whereas cells derived from the early epiblast of the post-implantation embryo, EpiSC, are deemed primed. While both Naïve and Primed mouse PSC can generate all the somatic cells in the body, hence are pluripotent, they differ significantly in variety of ways. For instance, the conditions with which to propagate "Primed" EpiSC differ significantly from mouse ESC, requiring FGF/Activin A signalling rather than LIF. The cells also respond differently to BMP4 addition, ERK inhibition and GSK3β inhibition, promoting differentiation rather than self renewal. Gene expression analysis highlights that common pluripotent transcription factors such as *Pou5f1, Nanog* and *Sox2* are shared but a collection of differentially expressed genes can distinguish these two populations. Mouse ESC also exhibit chimerisation ability, this is when mouse ESC injected into the early epiblast contributed to the cellular makeup of the final organism. This functional aspect is distinctly lacking in mouse EpiSC.

1.2.2. Differences between mouse and human embryo development

While many elements of mammalian development are conserved, species differences exist even at the earliest stages. Pre-implantation mouse and human embryos are very similar in structure, yet the inner cell mass exhibits some differences in gene expression at this stage (Figure 1-2A) (Blakeley et al., 2015; Boroviak et al., 2015; Kameda and Thomson, 2005). Post implantation, the differences in the structure of the mouse and human embryo are more apparent (Figure 1-2B). The inner cell mass becomes the epiblast. In the mouse this forms as a cylindrical structure. On top of this cylinder, an ectodermal cone forms and the cells within this structure provide a source of many cytokines present in the embryo. The epiblast of the post-implantation embryo of humans and other primates forms as a flat disc shape rather than cylindrical. These embryos also lack the ectodermal cone seen in mouse embryos.

Gastrulation occurs after the formation of the primitive streak. The primitive streak forms as a groove along the epiblast disc in primates or the anterior side of the epiblast cylinder in the mouse. While the structures differ slightly the function of the primitive streak remains the same and is the primary site of gastrulation. Cells which ingress through the primitive streak are patterned into the early mesoderm and endoderm. The cells that remain, which do not pass through the streak, are patterned into early ectoderm.

There are many reasons why human embryos, particularly the events post implantation, cannot be as well investigated as they have been in the mouse. Aspects of development have shown evolutionary conservation, but nevertheless species divergence has led to differences arising during the evolution of these species. The question then becomes, what has diverged and what has been conserved? Do the cells of the post-implantation epiblast in humans maintain a pluripotent identity like the mouse? Are the pluripotent states seen in the mouse present in human? Many of these questions cannot be investigated by in vivo assessment, so the question becomes, can in vitro models answer these questions?

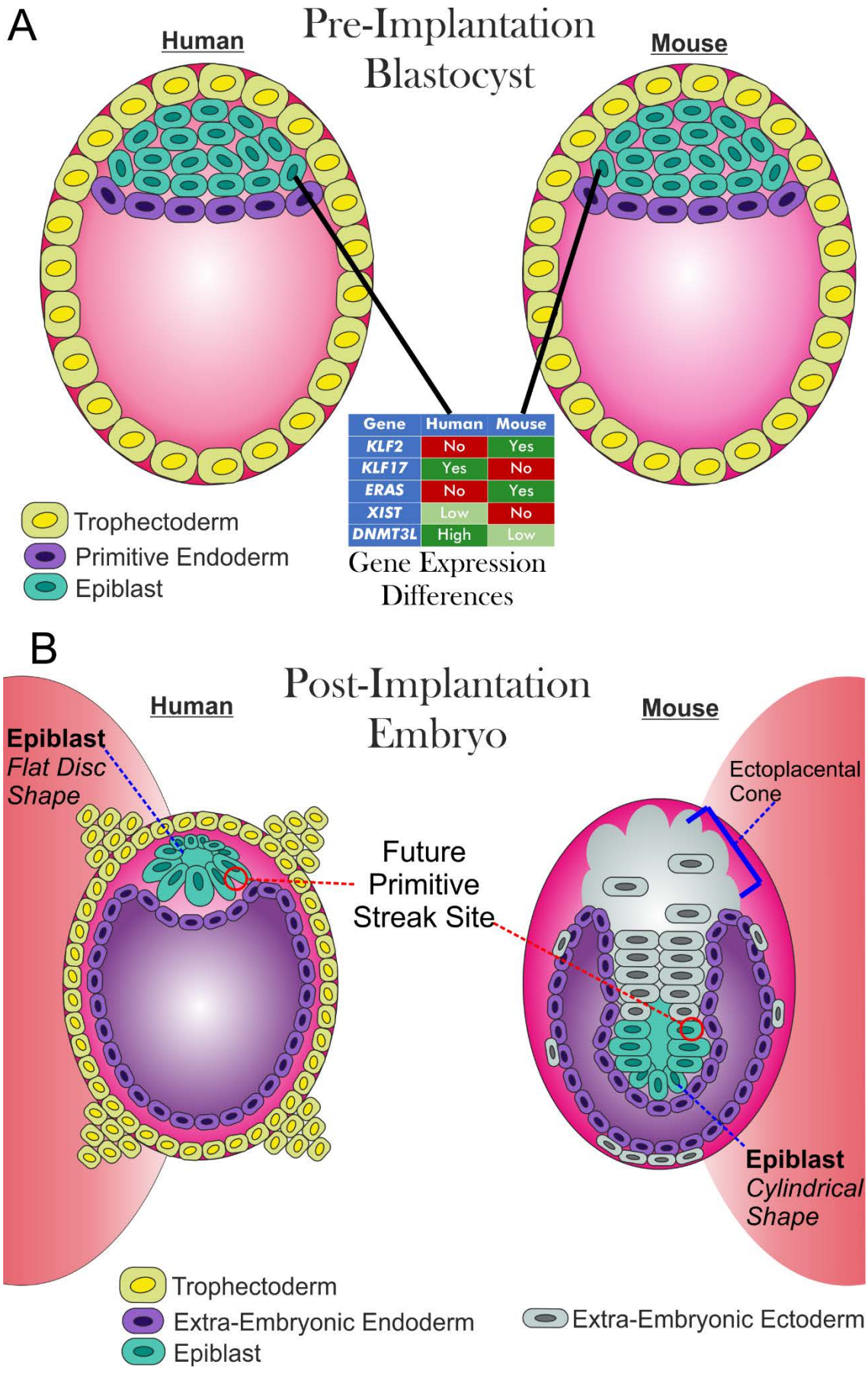

A
Pre-Implantation Blastocyst
Human
Mouse
Gene Human Mouse
KLF2 No Yes
KLF17 Yes No
ERAS No Yes
XIST Low No
DNMT3L High Low
Gene Expression Differences
Trophectoderm
Primitive Endoderm
Epiblast
B
Post-Implantation Embryo
Human
Mouse
Epiblast
Flat Disc Shape
Future Primitive Streak Site
Ectoplacental Cone
Epiblast
Cylindrical Shape
Trophectoderm
Extra-Embryonic Endoderm
Epiblast
Extra-Embryonic Ectoderm

Figure 1-2 Embryo development in human and mouse

A) At the Pre-implantation stage the blastocyst of both human and mouse has a very similar structure. Outer cells are made up of trophectoderm and the inner cell mass has segregated into the primitive endoderm and early epiblast. The Epiblast at this stage shows differences in gene expression between the two species. **B)** At the post implantation stage the embryos of mouse and human have differences in structure and cellular makeup. In the human, the epiblast develops in a flat disc shape, whereas the mouse epiblast develops as a cylindrical shape. The mouse embryo also forms an ectoplacental cone which is not present in the human. The future primitive streak forms at the site indicated with red circles.

1.2.3. Human Pluripotent states captured in vitro

The development of in vitro systems to propagate mouse PSC was a long process, spanning over 30 years of work. The work highlighted that two discrete states of pluripotency can be propagated in vitro, "Naïve" and "Primed". Work to develop similar systems for human PSC evolved in tandem with mouse albeit with a delayed start in investigation. Following the establishment of mouse ESC in 1981 (Evans and Kaufman, 1981), groups attempted to isolate a similar population from human blastocysts. In 1998, Thomson et al, isolated a population of self-renewing cells grown on a layer of mouse embryonic feeders. While these cells exhibited similar differentiation capacity to mouse ESC they were distinctly different, morphologically. Mouse ESC typically grew in tightly compact round colonies, whereas these human ESC grew as flat colonies. The conditions to grow the cells was also drastically different, relying on FGF signalling to maintain pluripotency rather than LIF.

The lines established by Thomson et al, expressed many of the same surface markers identified in human EC cells (Andrews et al., 1984a; Andrews et al., 1982; Andrews et al., 1984b; Kannagi et al., 1983; Shevinsky et al., 1982), including Alkaline Phosphatase, SSEA-3, SSEA-4, TRA-1-60, and TRA-1-81; they were also negative for SSEA-1. Mouse PSC had an inverse of this surface marker expression, being positive for SSEA-1 and negative for the other markers (Solter and Knowles, 1978; Thomson et al., 1998). Functionally human ESC had relatively low cloning efficiency compared to traditional mouse ESC. The transcriptome of the cells shared some common expression with mouse ESC but differences were apparent (Ginis et al., 2004).

Initially it was thought that all of these differences might be related to a species difference between mouse and humans. However, the discovery of mouse EpiSC (Brons et al., 2007; Tesar et al., 2007) began to question that notion. EpiSC grew in a similar manner and required similar growth conditions to human ESC, rather than classical mouse PSC. Functionally mouse EpiSC and

human ESC, share some characteristics including the induction of differentiation by BMP4, ERK inhibition and GSK3β inhibition. They also both have low cloning efficiency and low chimerisation ability (Brons et al., 2007; Tesar et al., 2007). This led many to believe that some aspects of human and mouse PSC differ not due to a species difference but due to the stage of embryonic development with which the cells relate.

Human PSC correlate closer to mouse EpiSC than mouse ESC, but they are not deemed equivalent. Firstly, their embryonic origin differs, with human ESC being derived from the ICM of pre-implantation embryos and mouse EpiSC being derived from the epiblast of the post-implantation embryo. Secondly, while they have similar gene expression patterns, a collection of genes associated with mouse naivety ZFP42, TEAD4, FBOX15, NROB1 and KLF4 were expressed at similar levels in the human blastocyst and human ESC grown in culture (Yan et al., 2013). SOX2, NANOG and FOXD3 also showed high expression in human ESC, whereas they have lower expression in mouse "primed" cells. Also, even though the expression of lineage specific gene expression has been identified in human ESC in vitro(Allison et al., 2018b; Gokhale et al., 2015; Hough et al., 2009; Hough et al., 2014), it does not appear as readily or abundantly as in mouse EpiSC (Kojima et al., 2014; Tsakiridis et al., 2014). These studies indicate that human ESC might represent something between these two states albeit closer to EpiSC.

Delineating what are inherent species differences or simply different states of pluripotency is difficult to ascertain and can be inappropriately asserted. An example which highlights how interpretation of these data could speak to either hypothesis relates to early lineage specification in human and mouse embryos. The primitive endoderm (hypoblast) forms from differentiation of some cells of the inner cell mass. In the mouse this process is dependent on FGF signalling, but in human embryos inhibition of FGF signalling did not affect the formation of the primitive endoderm (Kuijk et al., 2012; Roode et al., 2012). Therefore, one could argue that this is a species difference between the cells and their response to FGF, but it could also be argued that the cells are the same but the human embryo exploits another pathway to induce this differentiation, so the difference arises from the embryo rather than the PSC themselves. Equally it could be argued, as the propagation of a human "naïve" state has proved difficult, that FGF signalling is required to exit the naïve state in the mouse which may not be apparent in the human embryo development.

Naïve and Primed states of pluripotency are clearly specified for the mouse system but normal human ESC seem to correspond to the primed state (Nichols and Smith, 2009b). Recent developments of methods to convert human ESC to naïve like state (Takashima et al., 2014; Theunissen et al., 2014) or to derive them into that state directly from embryos (Guo et al., 2016) appear to distinguish that these two states can exist at least in vitro for human pluripotent

stem cells. These “Naïve” human PSC share similarities with their naïve mouse counterparts but also have distinct differences such as a lack of chimerisation ability, which may imply that these cells relate to a further stage of embryo development. Although these naïve growth conditions were designed to maintain cells in a Naïve state, they are not all equivalent with comparisons between them showing vast differences both functionally and transcriptomically (Blakeley et al., 2015; Guo et al., 2016; Theunissen et al., 2016; Warrier et al., 2017) and their relevance to human development is a question that remains unanswered (Rugg-Gunn, 2017).

In both mouse and human systems, we have also seen that pluripotent potential can be restored to somatic cells by a reprogramming process. First shown in mouse and then in human, the work of Shinya Yamanka demonstrated that pluripotency could be induced by overexpression key pluripotency-associatedgenes such as Pou5f1, Sox2, Klf4 and Myc (Takahashi et al., 2007; Takahashi and Yamanaka, 2006) . These iPS cells were equivalent to the conventional ESC derived from either mouse or human embryos. Expanding on this approach, groups demonstrated that overexpression of key naïve genes, namely *NANOG* and *KLF2* could facilitate the reprogramming of “primed” PSC to a more “naïve” state, in both mouse(Hall et al., 2009; Silva et al., 2009) and human(Takashima et al., 2014). These techniques demonstrate a process that is not seen in vivo, but nevertheless provide important tools and insight into the pluripotent state. More specifically they highlight critical genes within the pluripotency gene regulatory network and the potent action of said genes during the reversion of a cell fate.

1.3. Heterogeneity in Pluripotent Stem Cells with respect to lineage markers expression.

The earliest events of cell heterogeneity within the inner cell mass of mouse embryos occurs at embryonic stage E3. At this stage cells begin express genes associated with epiblast or primitive endoderm (Artus et al., 2011; Chazaud et al., 2006; Kurimoto et al., 2006). The expression at this stage is non-uniformed and has downstream implications as to the fate of these cells, whether they form epiblast or primitive endoderm (Artus et al., 2011; Chazaud et al., 2006). This stage of development is crucial for downstream development and pertubations of this process can result in poor development as the ICM is unable to differentiate towards primitive endoderm (Lorthongpanich et al., 2013; Yamanaka et al., 2010).

These events happen during critical periods of normal embryo development *in vivo* but similar events have also been characterised *in vitro.* When growing human or mouse PSC in culture, cells exhibit similar characteristics that, when viewed as an average over time, implies all the cells are the same. Nevertheless, populations of PSC frequently appear to be heterogeneous

with respect to their expression of a wide variety of markers, including specific genes and cell surface antigen markers and can be classified into different substates (e.g. Hough et al., 2009; Martinez Arias and Brickman, 2011; Toyooka et al., 2008, Enver et al., 2009). When mouse ES are cultured in serum and LIF conditions some early signs of primitive endoderm formation can be seen. Canham et al, 2010 employed a system using a fluorescent reporter line to identify subpopulations of mouse ESC that were expression both pluripotency-associated markers and markers of primitive endoderm. The study found that a small population of cells expressed Hex, a homeobox transcription factor associated with early primitive endoderm formation (Canham et al., 2010).

EpiSC often present a more heterogeneous population than do mouse ESC, particularly when these are grown in 2iLif. EpiSC have been characterised both transcriptionally and functionally, showing that they correspond to the anterior primitive streak of the developing embryo (Kojima et al., 2014). The expression of lineage specific markers has been identified in EpiSC cultures (Brons et al., 2007; Kojima et al., 2014; Tesar et al., 2007). A substate expressing early primitive streak markers, including T and Mixl1, has been identified previously (Tsakiridis et al., 2014). These cells expressed pluripotency-associatedmarkers such as Sox2, Oct4 and Nanog, and were able to self renew, but exhibited increased mesendodermal differentiation when induced. The emergence of this substate was driven by endogenous WNT signalling within the culture, assessed by using WNT inhibition (Tsakiridis et al., 2014). Their bias in differentiation indicated that this cells are in a “lineage primed” state (Nichols and Smith, 2009a).

In the human system, human PSC have been separated into substates based upon the heterogeneous expression of the surface antigens (Enver et al., 2005; Hough et al., 2009; Hough et al., 2014). The expression of SSEA-3 has long since been used as a surface marker to identify pluripotent cells both in EC and ESC (Shevinsky et al., 1982), while not completely infallible previous studies have shown loss of SSEA-3 coincided with the loss of pluripotency and decreased self-renewal ability (Enver et al., 2005; Fenderson et al., 1987; Tonge et al., 2011). Much like the mutations seen in embryonal carcinoma lines, mutations occurring during the culture of human PSC might trap cells in a particular substate. Enver et al, 2005, described the patterns of gene expression in substates of human PSC defined by expression of SSEA-3, and their alteration in culture adapted, genetic variants of the same cells (Enver et al 2005). Culture adapted cells exhibited pluripotent gene expression and clonogenic potential even when SSEA-3 expression had been lost. This work demonstrated how substate dynamics could be altered by genetic mutations, and that defining a substate was particularly context dependent.

Many reports indicate that undifferentiated human PSC, like other stem cells, form a heterogeneous populations and that when occupying different substates the stem cells may exhibit different functional properties particularly with respect to their differentiation potential (For review, see Enver et al 2009). The existence of lineage biased substates in the pluripotent embryonal carcinoma line, NTERA2, was inferred by Tonge et al, 2010, but it was not possible to identify directly cells in different lineage biased substates. Again, using SSEA-3 as a marker of pluripotency and adding assessment of a human ES line, Tonge went on to show that high expression of SSEA-3 correlated with increased self renewal capacity where as low expression correlated with increased neural differentiation potential (Tonge et al., 2011).

While this work demonstrated that functionally distinct substates existed in human PSC culture, it could not speak as to whether the initial differences went beyond the surface antigen expression. A number of other studies have addressed the patterns of gene expression in human PSC and how they differ between substates of cells defined in various ways. Fearing that bulk analysis might mask the presence of substates, many groups began to assess human PSC at a single cell level. Single cells have previously been isolated from different substates of human PSC identified by SSEA-3 expression and their gene expression was assessed (Gokhale et al., 2015). Some apparently undifferentiated cells (SSEA-3 positive) nevertheless expressed genes associated with endoderm differentiation, *GATA6*, *GATA4* and *SOX17.* This suggested that some of these cells were already primed for endoderm differentiation while still within the stem cell compartment. Hough et al, 2014 have reported on the gene expression patterns of substates defined by differential expression of cell surface antigens CD9, Epcam and GCTM2. They were able to identify three substates which exhibited gene expression patterns indicative of pristine pluripotency, early mesendoderm and early ectoderm. Such studies have highlighted the differences in gene expression in substates defined by surface antigens and have begun to provide insights from which gene regulatory networks can ultimately be constructed. They can also provide indications of new markers that could be used to refine substate definition, as well as indicating potential lineage bias and relationships to cells in the early embryo.

The findings of both Hough et al, 2014 and Gokhale et al, 2015 identified cells expressing the early endoderm marker *GATA6*. Further to this, much like the results of Canham et al (2010) with mouse ESC, Allison et al, 2018 utilised a human ES line carrying fluorescent reporter for *GATA6* and stained for a pluripotency-associated surface marker, to identify and characterise an endodermal biased substate of human PSC. This substate was found regularly in standard culture systems and exhibited self renewal capacity but displayed a bias in its differentiation potential, favouring endoderm formation. Gene expression analysis of the substate also revealed co-expression of pluripotency and mesendodermal gene regulatory networks.

In both the mouse and human systems, heterogeneity was present with respect to endoderm and mesoderm associated gene expression but what was a key finding in these studies was the ability for cells to inter-convert easily between the substates (Allison et al., 2018a; Canham et al., 2010; Hough et al., 2014; Tsakiridis et al., 2014). This established the concept that cells can occupy a lineage primed space without committing to differentiation but bias the outcome of differentiation towards a particular lineage when tested to do so.

1.4. Continuum of Pluripotency.

The states of pluripotency identified in vivo and those propagated in vitro could be classified as discrete populations. The heterogeneity identified within these populations speaks to a broader classification, where cells can reside in different position along a continuum of pluripotency (Figure 1-3). Cells inter-convert between these states but at a given time the population contains functionally distinct substates of PSC.

Until the much more recent development of 2iLif culture system (Ying et al., 2008), traditionally mouse PSC were grown in media containing ES qualified serum and LIF (Leukaemia Inhibitory Factor) (Smith et al., 1988). Using 2iLif, they established a much more homogenous population (Ying et al., 2008) and they correlated to an earlier stage in embryo development than cells cultured in serum (Martin Gonzalez et al., 2016). Mouse PSC grown in serum, have a different transcriptional profile, correlating embryonic stage E4.5. They also exhibit functional differences, with a poor ability to generate chimeras when injected into an early mouse embryo, a stark contrast to mouse PSC maintained in 2iLIf (Martin Gonzalez et al., 2016).

These studies reveal a continuum of pluripotency. Once cells have transitioned from a totipotent state, to a pluripotent state, the cells begin on a trajectory that progressively narrows their window of developmental potential. This first transition can be seen through the loss of chimerisation ability, when injected into the early embryo of the mouse (Martin Gonzalez et al., 2016), demonstrating a somewhat restricted potency. This effect happens prior to the complete loss of the naïve identity but is further compounded after the transition into primed pluripotency, EpiSC (Brons et al., 2007; Tesar et al., 2007).

The next transition comes in the form of lineage priming, where cells can exhibit both downregulation of pluripotency genes and expression of early lineage markers, such as *T, Lefty2* and *Mixl1* (Han et al., 2010; Kojima et al., 2014). It is important to know that the process of lineage priming restricts pluripotency but does not remove pluripotency. This dynamic can be difficult to understand: whilst cells are within the pluripotent stem cell compartment, they inherently retain the ability to form all three germ layers, but their propensity to do so can be

altered by lineage priming events. Pluripotent cells can inter-convert between these states, regaining functional properties thought to be lost such as chimerisation ability (Han et al., 2010).

Moving cells between these states of pluripotency can be as simple as changing the medium in which they are grown, including altering signalling molecules present or matrix on which cells grow (Martin Gonzalez et al., 2016; Silva et al., 2008). The transition from primed back to naïve, however, is a more difficult task requiring the exogenous expression of key Naïve related genes (Hall et al., 2009; Silva et al., 2009).

The development of human Naïve media does highlights the ability to move human PSC in between discrete substates through the use of chemical inhibitors and signalling molecules. In particular, the use of GSK3β inhibition within these media is particularly interesting. GSK3β inhibition, releases β-catenin and allows its translocation into the nucleus, a process similar to the cascade present in active WNT signalling (Huelsken and Behrens, 2002). WNT signalling, and subsequently the use of GSK3β inhibition, has been heavily associated with human PSC differentiation (Davidson et al., 2012). GSK3β inhibition, is often an important component of naïve media formulations, but through addition of other inhibitors and signalling molecules, the pluripotency network remains active in the presence of this inhibition (Chan et al., 2013a; Gafni et al., 2013; Qin et al., 2016; Takashima et al., 2014; Theunissen et al., 2014), highlighting the ability to attenuate traditional differentiation cues, through signalling manipulation.

When we culture PSC in a medium, we are attempting to propagate the cells indefinitely in a state that is momentary in development. In theory, the cells are “trapped” in a particular state of pluripotency and through environmental manipulation we can alter the state of pluripotency to which the cells relate. In the mouse and human PSC field, we have seen development of naïve culture systems, which aim to trap cells in an earlier stage of development, but whether we can trap cells further along this developmental axis, remains to be seen. The heterogenous population that exist in standard PSC culture reveals co-expression of gene regulator networks corresponding to both pluripotency and early lineage specification within some PSC. Specifically, for our work, I want to know whether a substate exists within the heterogenous population which co-expresses early mesodermal and pluripotency-associated genes. Further to this, through signalling manipulation, I may be able to induce and maintain this co-expression, trapping cells further in this developmental axis and subsequently narrowing cells developmental potential towards a given lineage

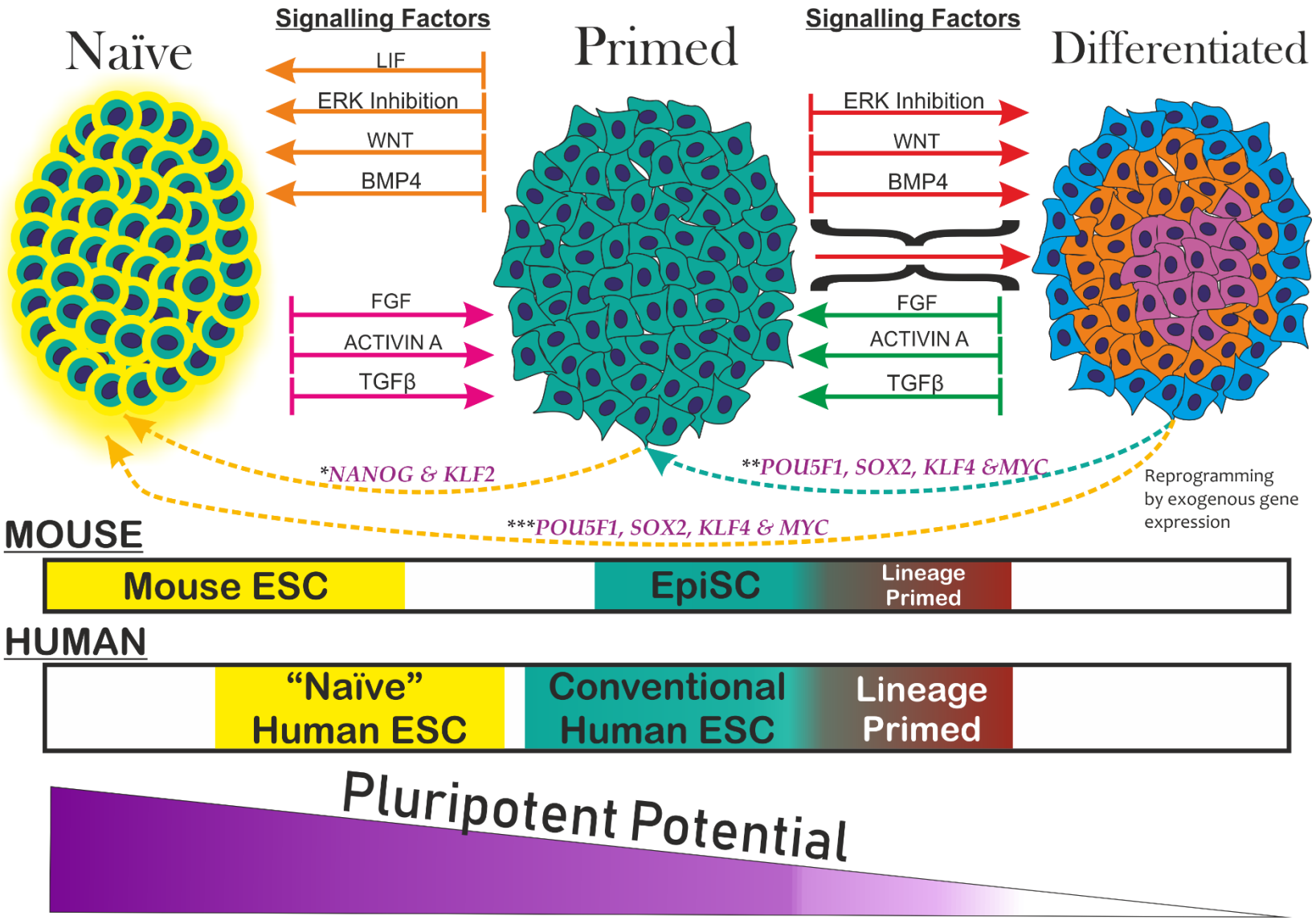

Naïve
Signalling Factors
LIF
ERK Inhibition
WNT
BMP4
FGF
ACTIVIN A
TGFβ
Primed
Signalling Factors
ERK Inhibition
WNT
BMP4
FGF
ACTIVIN A
TGFβ
Differentiated
*NANOG & KLF2
**POU5F1, SOX2, KLF4 &MYC
***POU5F1, SOX2, KLF4 & MYC
Reprogramming by exogenous gene expression
MOUSE
Mouse ESC
EpiSC
Lineage Primed
HUMAN
"Naïve" Human ESC
Conventional Human ESC
Lineage Primed
Pluripotent Potential

Figure 1-3 Continuum of pluripotency and the signalling governing state maintenance or transition

PSC can exist in many states but broadly can be classified into “Naïve” and “Primed”. Cells transition through these states towards differentiation, with an ever-narrowing pluripotent potential. In the mouse, and in part in human, factors in orange maintain the naïve state and prevent the transition to primed. Factors in pink, promote the transition to primed from naïve. The same factors, in green, maintain the primed state and inhibit differentiation. Factors in red, promote transition to differentiation and eventually the loss of pluripotency, the same is true when these factors are added to those in green. Exogenous gene expression can reprogram differentiated cells towards induced pluripotent cells, correlating to different states of pluripotency. Along this pluripotency developmental axis, mouse ESC represent the highest in pluripotent potential, highlighted by their blastocyst chimerisation ability. “Naïve” Human ESC exhibit similar traits to mouse ESC cells but lack this ability, so reside further along this axis. The differences between Naïve human ESC and conventional human ESC is not as pronounced as mouse ESC and EpiSC, thus are closer in this developmental axis. Both EpiSC and conventional human ESC are considered to be primed PSC. As EpiSC and human ESC form a heterogenous population in culture, these cells span a larger fraction of the pluripotent developmental axis with some cells appearing as lineage primed but crucially not committed to differentiation.

*Demonstrated in mouse and human (Hall et al., 2009; Silva et al., 2009; Takashima et al., 2014). **Demonstrated in human(Takahashi et al., 2007). ***Demonstrated in mouse(Takahashi and Yamanaka, 2006).

1.5. Project Aims:

My overall work is focused in the lineage primed part of the pluripotency continuum within human PSC. Unlike previous work in this area my focus is on a mesodermal biased state and whether it might already exist in a substate of our standard human PSC culture. Further to this, I want to test whether, in a defined system, I can induce a similar state that can be maintained through signalling manipulation. The aims below describe the overall approach to achieving this goal:

1. *Utilise a combination of fluorescent reporter human PSC lines and surface antigen expression for both the PSC state and early mesodermal derivatives to identify sub-states within standard stem cell culture systems. Further to this, evaluate the functional differences between the identified sub-states.*

   In this aim I attempt to exploit the apparent heterogeneity that exists within standard human PSC cultures to identify a stem cell substate which exhibits differences ideally functionally but possibly just transcriptionally. A key aspect is that the substate can readily inter-convert and re-establish a stem cell population. This would confirm the concept that human PSC can exist in a state that expresses early mesodermal associated markers.

2. *Determine the key factors and cellular signals that regulate the creation, self-renewal and differentiation of the sub-state by comparing different standard PSC media/matrix combinations and passaging techniques. Furthermore, to evaluate whether a given sub-state can be trapped and expanded by manipulating the signals the cells receive.*

   This aim builds on the work of aim one and I assessed how media/matrix can affect the substates identified in aim one. Combining the known knowledge of signalling and its effect on human pluripotent stem cells, we asked whether we could balance pro-self renewal and pro-differentiation signals to trap cells in a substate further down the developmental axis.

3. *Utilise both bulk and single cell transcriptomic approaches to evaluate the differences between the substates identified in standard culture systems and newly engineered substates generated in defined medium.*

   The final aim speaks to where the cells we have identified from aims one and two, reside in our continuum of pluripotency. Transcriptomic techniques can segregate populations, the use of single cell approaches can further resolve the heterogeneity seen. We can ascertain the relationship of the states identified and also relate the data to early

differentiation to gauge where these states reside in the transition from pluripotency to committed differentiated progeny.

2. Chapter 2: Materials and Methods

2.1. Cell Lines

HES3, human embryonic stem cell line, a gift from Adam Hirst. First described in Cooper et al, 2002 derived using the methods described in Reubinoff et al., 2000.

HES3 *MIXL1*-GFP, a gift from Andrew Elefanty, Monash University, Australia, reporter human ESC line developed in Davis et al, 2008.

H9 T-Venus, a gift from Roger Pedersen and Daniel Ortmann, published in Mendjan et al 2014. This reporter was generated in the human embryonic stem cell line, H9 (WA09) line, published in Thomson et al, 1998.

n2102Ep, human Embryonal Carcinoma cell line (Andrews et al., 1982), used for titration of antibodies for pluripotency-associated surface markers.

Mouse Embryonic Feeders (MEF) used in MEF/KOSR conditions were extracted from the CF1 mouse strain. MEFs were mitotically inactivated through mitomycin-C treatment for two hours, and frozen in DMEM 10% serum, 10% DMSO.

2.2. Human Embryonic Stem Cell Culture

Both feeder and feeder free systems, specified in the text, were used throughout the project, in the culturing of human ESC. For either system, flask/plates of human ESC were grown in humidified incubators at 37°C and 5% CO2. Our feeder culture system uses Knock Out Serum Replacement (KOSR) medium (Table 2.2) and is referred to as MEF/KOSR conditions. KOSR is a albumin and lipid rich supplement designed to replace traditional fetal calf serum within human ESC media. For feeder free systems I used three matrices Matrigel (corning), Geltrex (ThermoFisher) and Vitronectin (Stem cell Technologies). This was combined with either mTESR medium (Stem Cell Technologies) or E8 medium (made in house, adapted from Chen et al, 2011).

2.2.1. Coating growth vessels

2.2.1.1. MEF coating:

Vessels were coated with 0.1% Gelatin/PBS and incubated at room temperature for 30 minutes. Defrosted MEFs were resuspended in DMEM with 10% FCS, after aspirating the gelatin from the flasks/plates MEF were seeded at 10,000 cells/cm2. MEFs flasks were incubated at 37°C until use within four to five days of plating, however the day after plating was the most common.

2.2.1.2. Vitronectin coating:

Vitronectin XF (Stem Cell Technologies) was resuspended in PBS(w/o Ca**+**, Mg**++**), at a 1:50 dilution. Diluted vitronectin solution was added to flasks/plates and incubated for 30-60

minutes at room temperature. Vessels were either used directly or stored at 4°C for no longer than one week.

2.2.1.3. Geltrex/Matrigel coating:

Concentrated Geltrex/Matrigel vials were defrosted on ice. 500µL aliquots were placed in cryovials and frozen at -80°C. When Geltrex/Matirgel was need an aliquot was thawed by adding 1ml of cold DMEM into the cryovial. The solution was diluted further to 1:60 by adding DMEM up to 30ml. Vessels were coated with diluted Geltrex/Matrigel, 1ml per 6 well and left to set at room temperature for 2 hours. Vessels were either used directly or stored at 4°C for no longer than one week.

2.2.2. Preparation of E8 and E6 Medium.

Essential 8 (E8) or E6 media was made in house with a recipe adapted from Chen et al, 2011. One key difference in the in-house media from the recipe published by Chen et al, 2011 is the replacement of standard glutamine with GlutaMax (Thermofisher). Glutamax is a thermostable form of glutamine, which is bound to alanine to increase stability. Large batches of 50X E8 supplements were prepared and frozen as 10ml aliquots at -20°C. For 50X supplements components were added to DMEM/F12 without glutamine and phenol red (Sigma, D6434). Defrosted 10ml aliquots were added to 490ml of DMEM/F12 without glutamine and filtered using a stericup (Millipore) 0.22µM filter. For imaging E8 aliquots were added to DMEM/F12 without glutamine and phenol red. Table 2.1 details the concentration of components in the 50x supplements and subsequent 1x final media.

Table 2.1 Composition of E6 and E8 Media

	Component	50X concentrate	Final concentrations per 1 Litre of E8	Company	Catalogue Number
E6	DMEM/F12	-	-	Sigma	D6421 Or D6434 for Imaging Media
	L-ascorbic acid	3200mg/L	64mg/L	Sigma	A8960
	Sodium selenium	700ug/L	14ug/L	Sigma	S5261
	Insulin	970mg/L	19.4mg/L	ThermoFisher	A11382IJ
	$NaHCO_3$	27.15g/L	543mg/L	Sigma	S5761
	Transferrin	535mg/L	10.7mg/L	Sigma	T0665
	Glutamax	50X	10ml/L	ThermoFisher	35050038
E8	FGF2	5mg/L	100ug/L	Peprotech	100-18B
	TGFB1	100ug/L	2ug/L	Peprotech	100-21

2.2.3. Preparation of KOSR Medium.

KOSR medium was made by combining the components in Table 2.2 and filtered using a stericup (Millipore) 0.22µM filter.

Table 2.2 Composition of Knockout Serum Replacement (KOSR) containing medium

Component	Final Volume	Company	Catalog Number
KO DMEM	400ml	ThermoFisher	10829018
KOSR 100ml	100ml	ThermoFisher	10828028
Non Essential Amino Acids	5ml	ThermoFisher	11140050
1mM L-Glutamine	5ml	ThermoFisher	25030081
0.1mM 2-Mercaptoethanol	1mL	ThermoFisher	31350010
4ng/mL human FGF2	500µL	Peprotech	100-18B

2.2.4. LPA containing media preparation

The LPA containing media is first made as a 10X stock, 10X BCL (BSA, Cholesterol, LPA). This is then diluted as needed. For production of trapping media CHIRON, a GSK3β inhibitor, is added at 3µM. After optimisation steps IWP-2, a porcupine inhibitor, was added at 1µM. 4.8 µM LPA was chosen as the starting concentration as it was calculated as the level of LPA in 15% KOSR human ESC media (Garcia-Gonzalo and Izpisúa Belmonte, 2008). For culture after optimisation 2X(0.96µM) and 1X(0.48µM) were mostly commonly used, in further text these are referred to as 1X or 2X Primo. Other concentrations were used during assessment of LPAs effect in the presence of CHIRON. CHIR99021 (Tocris, #4423) and IWP-2 (Tocris, #3533) was resuspended in DMSO(Sigma, #D2650) at 10mM and 5mM, respectively. Cholesterol (Synthechol, Sigma, #C1231) was resuspended at 20mM in 100% ethanol. Oleoyl-L-α-**lysophosphatidic acid** sodium salt (LPA) (Sigma, #L7260) was resuspended in PBS (w/o Ca**+**, Mg**++**), with 0.1% Fatty Acid free BSA (Probumin, Millipore, #810664) at 122 µM. Components were added to E8 medium at concentration in Table 2.3 for a 10X supplements. 10X supplements were then diluted in E8 medium to desired concentrations. Continued passage of cells under trapping conditions proved difficult after 3-4 passages, however this was alleviated by the removal of 2-mercaptoethanol.

Table 2.3 Composition of Primo Medium at various concentrations

			Concentration						
Media	Component	Company	**10X Stock (4.8µM LPA)**	**2X (0.96µM LPA)**	**1X (0.48µM LPA)**	2.4µM LPA	1.2µM LPA	0.6µM LPA	0.3µM LPA
BCL	E8 (**Table 2.1**)	Homemade	-	-	-	-	-	-	-
	Fatty Acid free BSA	Millipore	1%	0.2%	0.1%	0.5%	0.25%	0.125 %	0.0625 %
	*2-mercaptoethanol	Gibco	100nM	*20nM	10nM	50nM	25nM	12.5 nM	6.25 nM
	Cholesterol	Sigma-Aldrich	20µM	4 µM	2µM	10µM	5µM	2.5µM	1.25 µM
	LPA	Sigma-Aldrich	4.8µM	0.96µM	0.48µM	2.4 µM	1.4 µM	0.6 µM	0.3µM
+									
Primo	CHIR99021	Tocris	3µM	3µM	3µM	3µM	3µM	3µM	3µM
	IWP-2	Tocris	1µM	1µM	1µM	1µM	1µM	1µM	1µM

*2-mercaptoethanol was removed for multiple passaging, and not included in the final formulation of the medium.

2.2.5. Routine Passaging:

Cells for MEF/KOSR and Matrigel/mTESR were manually passaged with a similar technique. Media was aspirated from flasks/plates and Collagenase IV (company) was added. Flasks were then incubated for 7 minutes at 37°C to lift colonies. The collagenase was then aspirated and fresh media added before scraping the colonies with a glass/plastic pipette. Dissected colonies were resuspended in fresh media and divided amongst new flask at a ratio on average between 1:3-1:6. Cells for Vitronectin/E8 were passaged with a non-enzymatic disassociation solution ReLeSR (Stem Cell Technologies). Media was aspirated from flasks/plates and cells were washed once with PBS before the addition of ReLeSR (Stem Cell Technologies). ReleSR was left on the cells for ~30 seconds and then aspirated. Flasks were left to incubate at room temperature for 4-6 minutes before the addition of fresh media. Flasks were gently tapped to dislodge colonies and the media was gently pipetted up and down to break colonies into smaller aggregates. Dissected colonies were resuspended in fresh media and divided amongst new flask at a ratio on average between 1:3-1:6. Cells grown in trapping/Primo medium were passaged in the same manner as Vitronectin/E8 cultures.

2.2.6. Human ESC Freezing:

Cells were frozen differently depending on the condition they were grown from Vitronectin/E8 or MEF/KOSR. Cells were first harvested using the method described in the routine passage section (2.2.5). Cells were centrifuged for 3 minutes at 1000rpm and the supernatant was removed. Cells were then resuspended in Stem Cell Banker (Takara) and aliquoted into 1.5ml cryovials. To prevent water crystal formation and improve viability cryovials were placed in a freezing container filled with Isopropanol. This ensures a gradual temperature decrease of 1 degree per min when the container is placed into a -80°C freezer. After 24 hours at -80°C, vials were transferred to liquid nitrogen for long term storage.

2.2.7. Human ESC thawing:

Cells were thawed differently depending on what condition they had been growing in prior to freezing. Cryovials were removed from liquid nitrogen and 1ml of warm media, either KOSR or E8 media, was added to the vial. Once the cells were defrosted they were added to 10ml of fresh media in a 15ml falcon tube which was subsequently centrifuged for 3 minutes at 1000rpm. After pelleting the cells, the supernatant was removed and cells were resuspended in warm media, either KOSR or E8 media. The cells were then added to plates either coated with MEFs or Vitronectin, which ever matched their condition prior to freezing. To improve recovery, 10µM Y-27632(ROCKi) (Tocris, #1254) was sometimes added to medium.

2.3. Single cell Suspension.

When single cells were required, media was aspirated from the flask and cells were washed once with PBS. Cells were then treated with either 0.25% Trypsin in EDTA for 1-2 minutes, TrypLE Select (Thermo Fisher Scientific, 12563029) for 5 minutes or Accutase (Thermofisher, #A1110501) for 10 mins, all at 37°C. Media was added to the flasks, if using Trypsin or TrypLE media contained 10% FCS to neutralise the trypsin, when using accutase, standard growth media was used as accutase is neutralised by dilution. Media containing cells were than centrifuged at 1000rpm for 3mins to pellet cells.

2.4. Cell counting:

After cells had been dissociated by one of the methods described in 2.3 and diluted to an appropriate volume with human ESC media. If cells were to be plated immediately after resuspension, 10µL was added to 10µL of trypan blue (Sigma, #T8154), to aid in counting of only live cells. Then 10µL of cell suspension was added to a haemocytometer and after counting four corner grids the average was taken. If trypan blue had been added to the suspension the average is doubled. This value is multiplied by 10,000 and equates to the concentration of cells

per ml. Finally, the concentration is multiplied by the total volume of media to give a final total cell number.

2.5. Immunostaining

2.5.1. Antibodies

A wide range of antibodies were used throughout the project from many sources. All antibodies made from In-house hybridomas, for surface markers, were titrated against n2102Ep cells growing in 10% FCS DMEM. Titration range from 1:2 to 1:100, was assessed by flow cytometer and the optimal dilution was selected based on the median fluorescent intensities.

For intracellular markers, from companies, antibodies were titrated on human PSC lines either in self renewal conditions or post differentiation, depending on the antibody target. The manufacturers recommendation was chosen as the titration range mid-point and I used more and less concentrated preparation to assess the titre. Antibodies were also tested on samples that should be negative, ie. Anti-NANOG on differentiated samples and anti-BRACHYURY in self-renewal conditions. This process removed antibodies that gave a false positive signal. Secondary antibodies were titrated against know positive primary antibodies, such as the pan-human marker, TRA-1-85.

Table 2.4 Primary Antibodies

Antibody	Antigen	Iso-type	Summary	Host Species	Dilution	Reference	Source
BF4	BF4	IgM	sialated, non-ceramide	Mouse Monoclonal	1:5	Wright et al 2011	In-House Hybridoma
CH8	CD9	IgG	non-sialated, non-ceramide	Mouse Monoclonal	1:20	Wright et al 2011	In-House Hybridoma
63-AG8	P3X	IgG1	Negative Control	Mouse Monoclonal	1:10	Kohler and Milstein 1975	In-House Hybridoma
MC631-2C2	SSEA-3	IgM	globoseries glycoplipid	Mouse Monoclonal	1:10	Shevinsky et al 1982	In-House Hybridoma
813-70	SSEA-4	IgG3	globoseries glycoplipid	Mouse Monoclonal	1:100	Kannagi et al 1983	In-House Hybridoma
THY1	THY1	IgG	CD90	Mouse Monoclonal	1:10	Henderson et al., 2002 Andrews PW, 1983	In-House Hybridoma
TRA-1-60s	TRA-1-60s	IgM	sialated, non-ceramide	Mouse Monoclonal	1:5	Andrews et al 1984a	In-House Hybridoma

						Wright et al 2011	
TRA-1-81	TRA-1-81	IgM	non-sialated, non-ceramide	Mouse Monoclonal	1:10	Andrews et al 1984a Wright et al 2011	In-House Hybridoma
TRA-1-85	TRA-1-85	IgG	CD147	Mouse Monoclonal	1:20	Williams et al 1988	In-House Hybridoma
TRA-2-49	TRA-2-49	IgG	Alkaline Phosphotase	Mouse Monoclonal	1:10	Andrews et al 1984c	In-House Hybridoma
Anti-NANOG	NANOG	IgG	(D73G4) XP® #4903	Rabbit Monoclonal	1:400	-	Cell Signalling Technology
Anti -SOX2	SOX2	IgG	(D6D9) XP® #3579	Rabbit Monoclonal	1:400	-	Cell Signalling Technology
Anti-OCT4A	OCT4A	IgG	(C52G3) #2890	Rabbit Monoclonal	1:800	-	Cell Signalling Technology
ANTI-BRACHYURY	BRACHYURY	IgG	AF2085-SP	Goat Polyclonal	1:500	-	R&D Systems
Anti-PAX6	PAX6	IgG	Ab195045	Rabbit Monoclonal	1:200	-	Abcam
Anti-TBX6	TBX6	IgG	AF4744	Goat Polyclonal	1:200	-	R&D Systems
ANTI-YAP	YAP	IgG	(63.7): sc-101199	Mouse Monoclonal	1:200	-	Santa Cruz Biotechnology

Table 2.5 Secondary Antibodies

Antibody	Fluorophore	Summary	Dilution	Source
Goat anti Mouse Affinipure IgG+IgM (H+L)	AlexaFluor 647	115-605-044-JIR	1:200	Stratech (Jackson ImmunoResearch)
Goat anti Rabbit Affinipure IgG (H+L)	AlexaFluor 594	111-585-003-JIR	1:200	Stratech (Jackson ImmunoResearch)
Donkey Anti Goat IgG (H+L) AF594	AlexaFluor 594	A-11058	1:200	Thermofisher
Donkey Anti Rabbit IgG (H+L) 647	AlexaFluor 647	A-31573	1:200	Thermofisher

2.5.2. Fixation

When plates were to be immunostained they were first fixed with 4% PFA. Media was aspirated from the wells and then washed once with PBS (w/o Ca**+**, Mg**++**), before 4% PFA was added to fix the cells. Samples were incubated at room temperature for 15 mins, the 4% PFA was removed and plates were subsequently washed with PBS. PBS was added to the wells to prevent them from drying out and plates were stored at 4°C until future staining.

2.5.3. Permeabilisation

If cells were to be stained for intracellular markers they were permeabilised with 0.5% Triton X-100. PBS was aspirated from fixed wells and permeablisation buffer was added. Permeablisation buffer consisted of 10% FCS, 0.1% BSA and 0.5% Triton in PBS (w/o Ca**+**, Mg**++**). Wells were incubated at room temperature for 15 minutes. Permeabilsation buffer was removed and wells were washed with PBS.

2.5.4. Blocking

After Fixation (2.5.2) and Permeabilisation (2.5.3) cells were incubated in Blocking buffer for 1 hour at room temperature. Blocking Buffer consisted of 10% FCS and 0.1% BSA in PBS (w/o Ca**+**, Mg**++**). After blocking, buffer was removed and PBS was added and stored at 4°C until samples were to be stained.

2.5.5. Intracellular staining

Primary and secondary antibodies were diluted separately at the relevant concentration depicted in Table 2.4 and 2.5 in blocking buffer (2.5.4), Hoescht 33342 (ThermoFisher, #H3570) was added at 1:1000 to the diluted secondary antibody solution. Diluted primary antibody was added to wells and plates were incubated overnight at 4°C on an orbital shaker. After primary antibody incubation wells were washed once with blocking buffer before diluted secondary antibody with Hoescht was added. Plates were incubated for 2 hours at 4°C on an orbital shaker. After secondary antibody incubation cells were washed twice with PBS and wells were filled with PBS. Plates were either imaged immediately or stored in sealed bags at 4°C until they were imaged. Plates were imaged using the INCell analyser 2200 (GE Healthcare). Quantifying expression of images was performed by pipelines designed in CellProfiler (Carpenter et al., 2006).

2.5.6. Antibody staining for Fluorescent flow cytometry analysis.

Single cell suspensions were harvested as described in 2.3 and counted as in 2.4 and centrifuged at 1000 rpm for 3 minutes. Cells were resuspended in DMEM (without phenol red) and 10% FCS at a density of $1x10^7$ per mL. 100µL of the samples was dispensed into 5ml tubes and antibodies were added at the appropriate dilution, Table 2.4. After addition of the

primary antibody cells were incubated at 4°Cfor 30 minutes. Cells were then washed with DMEM/FCS and centrifuged at 1000 rpm for 3 minutes. After pelleting, the cells are resuspended in 200µL of DMEM/FCS. Secondary antibody was added at dilution indicated in

Table 2.5 and incubated at 4°C for 30 minutes. Cells were then washed with DMEM/FCS, centrifuged for 3 minutes at 1000rpm and resuspended in fresh DMEM/FCS for analysis by flow cytometry.

To set baselines for *MIXL1*-GFP and negative secondary 647, unlabelled HES3 line was harvested (2.5.6) and stained as in 2.5 for P3X. P3X is an IgG1 antibody which is secreted from the parent myeloma which all in house antigens were derived. P3X shows minimal reactivity to human cells (Kohler and Milstein, 1975). Positive gates were set according to HES3 P3X negative controls. Samples were also stained for P3X to assess non-specific binding. All flow cytometry analysis contained P3X samples for baseline setting (Figure 2-1)

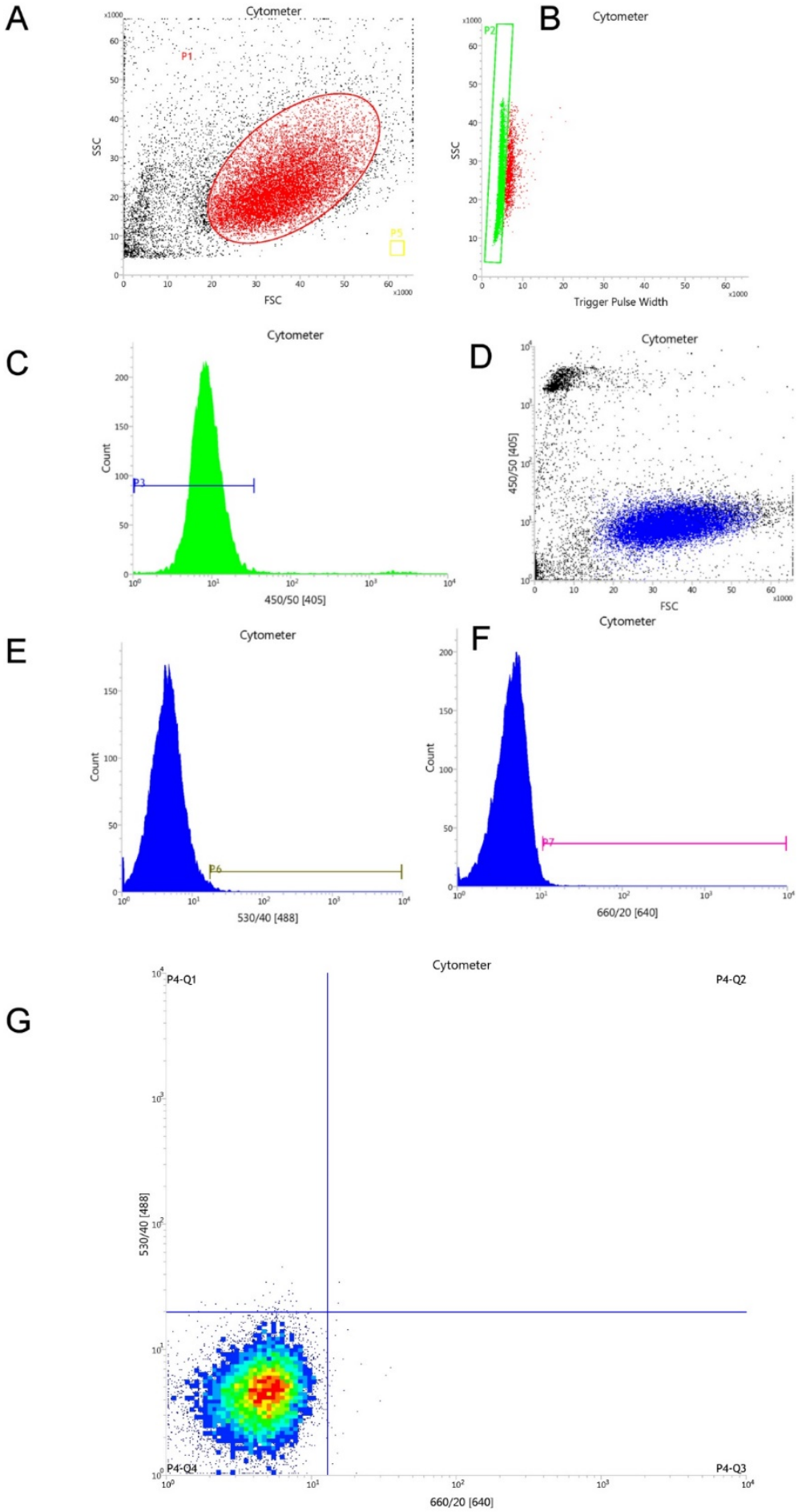

A
Cytometer
P1
P5
SSC
FSC
x1000
B
Cytometer
P2
SSC
Trigger Pulse Width
x1000
C
Cytometer
P3
Count
450/50 [405]
D
Cytometer
450/50 [405]
FSC
x1000
E
Cytometer
P6
Count
530/40 [488]
F
Cytometer
P7
Count
660/20 [640]
G
Cytometer
P4-Q1
P4-Q2
P4-Q4
P4-Q3
530/40 [488]
660/20 [640]

Figure 2-1 Flow Cytometry Gating

A) Scatter plot of Forward Scatter (FSC) versus Side scatter (SSC) from cells running through the flow cytometer. P1 gate in red identifies the cell population. **B)** Doublet discrimination, Trigger Pulse width versus SSC, P2 in green identifies single cells. **C)** Histogram plot for Live/Dead discrimination. DAPI staining is analysed with excitation by the violet laser 405nm and emission filter 450/50. P3 in blue identifies live cells. **D)** Scatter plot of FSC versus 450/50 emission, to check the P3 gate (blue) of live cells. **E)** Histogram of wildtype HES3 to set baseline of GFP expression, P6 gate marks positive cells for GFP expression. **F)** Histogram of wildtype HES3 stained with P3X and Alexafluor 647 goat anti mouse secondary, P7 gate marks positive cells for marker expression. **G)** Density plot of 660/20 (marker expression) versus 540/40 (GFP expression) from wildtype HES3 stained with P3X and Alexafluor 647 goat anti mouse secondary. Quadrant gates set to flank this double negative population.

2.6. Flourescence activated cell sorting (FACS)

The staining method for sorting was the same as in 2.5.6 however, DMEM/FCS was replaced for standard KOSR medium. Performing staining and wash steps using KOSR medium in place of DMEM/FCS helps to improve viability of cells pre and post sorting.

2.6.1. Bulk Cell sorts

Following laser alignment, Accudrop beads were run through the machine to set the drop delay for sorting. Cells were analysed on the machine and sorting gates were set within the population of interest. The gates were positioned to allow a suitable margin between to populations to ensure accurate separation. Cells were sorted into the appropriate vessels and post sort the samples were reanalysed on the flow cytometer. Only samples that had high efficiency percentages were used in further experiments (**Figure 2-2)**. For experiments such as for RNA-sequencing cells were deposited straight into lysis buffer, precluding their reanalysis. For these experiments equivalent numbers of cells were sorted immediately before into PBS(w/o Ca**+**, Mg**++**) and reanalysed to estimate the sorting efficiency of the subsequent samples sorted into lysis buffer.

2.6.2. FACS into 96 well plates.

Sorts were performed on the BD FACS Jazz. To ensure accurate deposition of single cells into individual wells of the 96 well plate firstly droplets were sorted to align the stream to the middle of the wells. After this I used a mixture of Red (Accudrop) and Green (Big Bang) fluorescent beads to determine whether the sorter was separating populations accurately. Beads were gated using the BD Sortware program then a row of Red beads and row of Green beads were

sorted as single beads. This was performed as an index sort just as the subsequent cell sorts would be. 96 well plates were then imaged on the InCell Analyzer at 4x with both the FITC and Cy5 channels to verify the correct number, colour and placement of the deposited beads was correct (Figure 2-3).

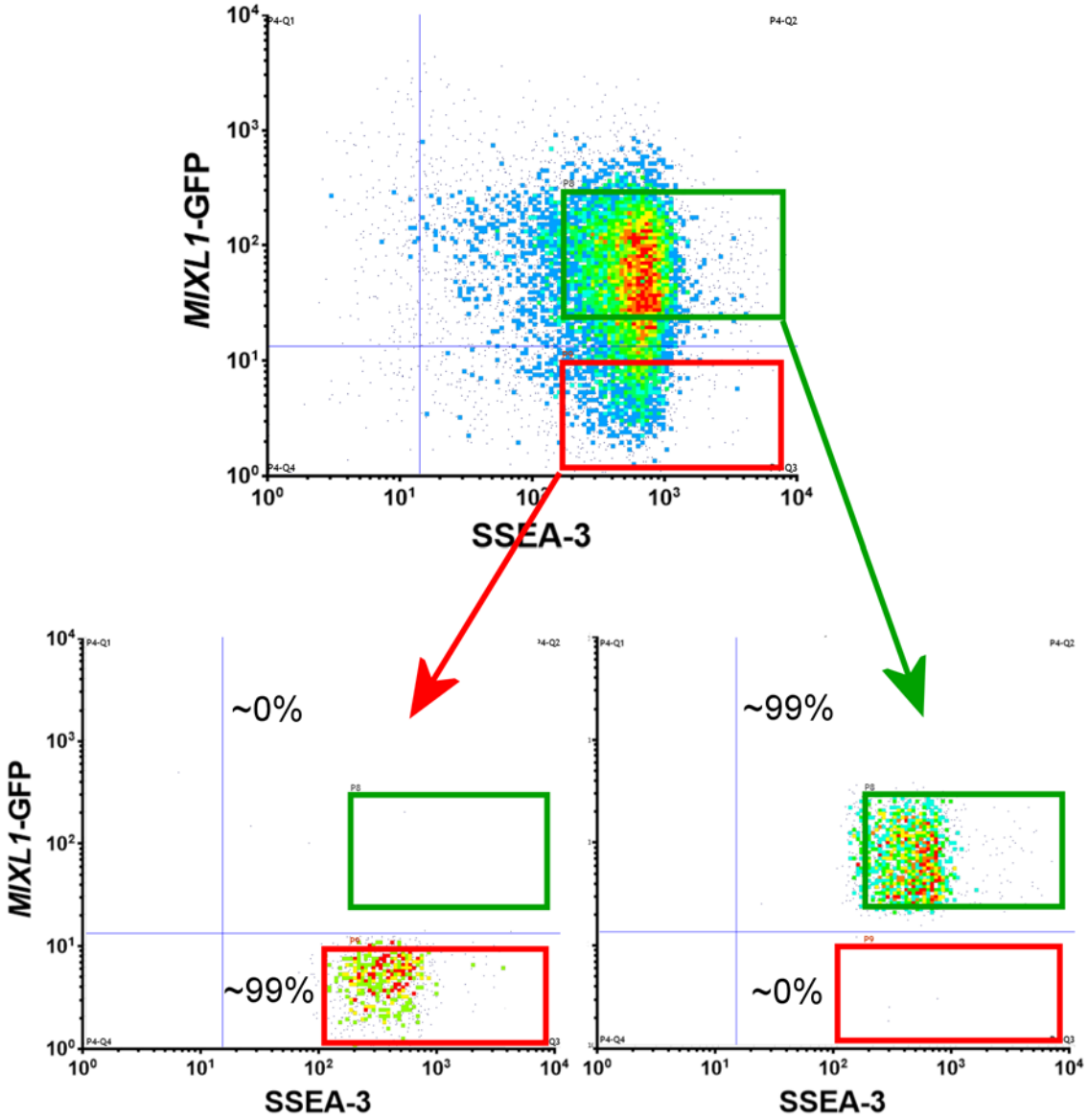


Figure 2-2 Post FACS Analysis

Representative example of post FACS analysis. Flow cytometry density plots of HES3 *MIXL1*-GFP growing in 2x Primo conditions. Sorting gates were positioned inside the population of interest with suitable margins between populations to prevent overlaps. The red gate indicates position of sorted populations of *MIXL1*(-)/SSEA-3(**+**) and green *MIXL1*(**+**)/SSEA-3(**+**) cells. Post sort analysis displays high percentage efficiency of sorted populations.

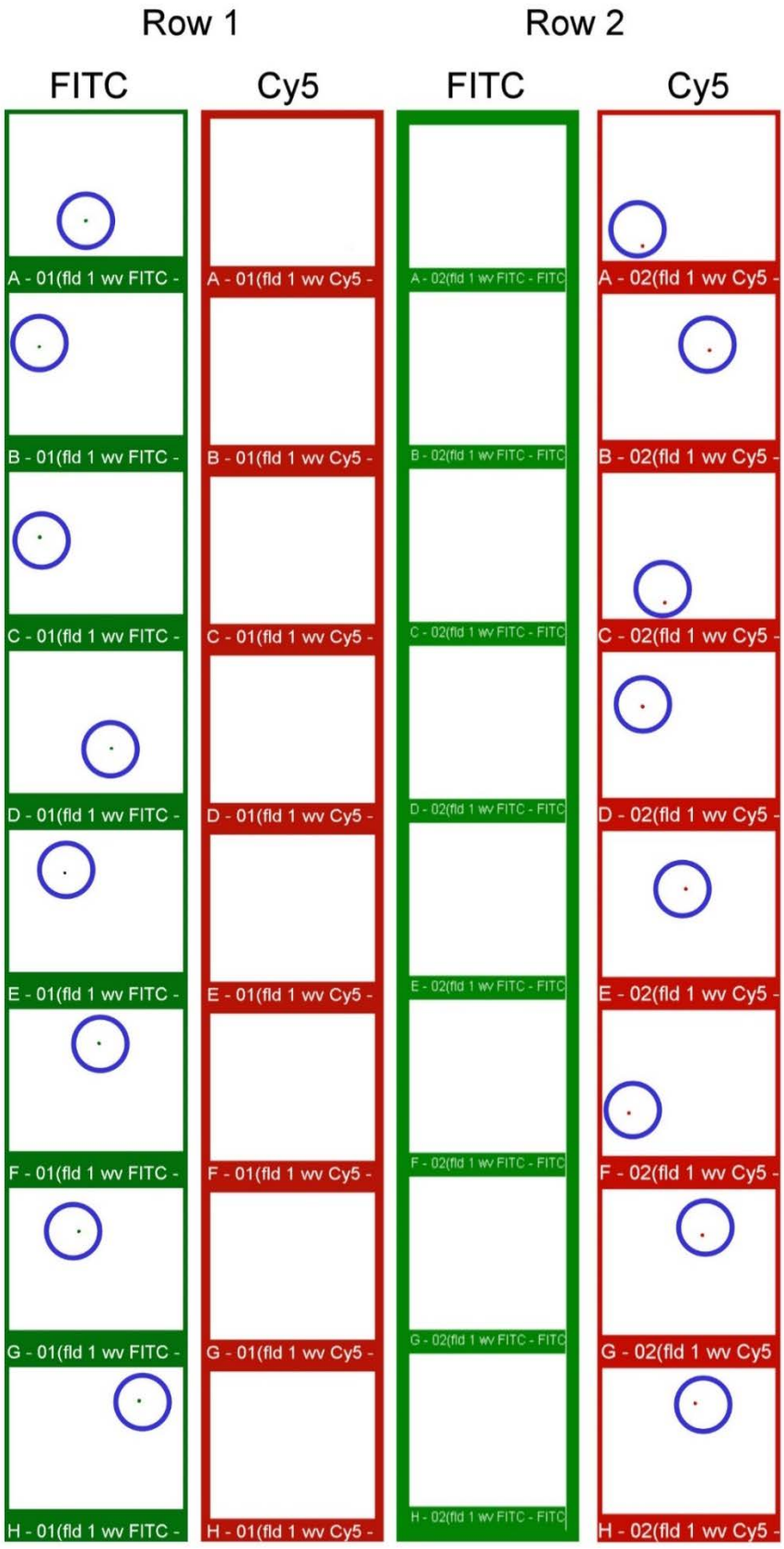


Figure 2-3 Single Bead Sorts into 96 well plates

Fluorescent image analysis of FITC and Cy5 images from the first two rows of a 96 well plate, Row one contains single sorted green beads and row 2 contains single sorted red beads. The Blue circle highlights the bead, image colours have been inverted for visual ease in printed publications. Single green beads were identified in row one and single red beads were identified in row two, indicating accurate single bead segregation and deposition.

2.7. Single Cell Cloning.

96 well plates were coated with gelatin and a layer of mouse embryonic feeders. Cells were harvested using accutase (see 2.3) and passed through a 70µM filter (Millipore) to remove larger cell aggregates. After staining as described in 2.5.6, DAPI (ThermoFisher, #62248) was added at 1:10,000 and used for live/dead discrimination. After gating on the BD sortware program the desired population was sorted as single cells directly into 96 well plates. The sort was indexed to retain information regarding the *MIXL1*-GFP and SSEA-3 expression levels. For single cell cloning the medium differed from standard culture. For this I used Dylan's Knockout Assisted Medium (DKAM), in brief this comprises of a 50/50 mix between standard HES medium and mTESR medium (Stem Cell Technologies) and the addition of 20µM Synthechol (Sigma). During initial plating 10µM Rock Inhibitor, Y-27632, was added to the medium. Immediately after sorting into the wells the plates were centrifuged at 1000 rpm for 1 minute to aid attachment of the cells. After two days the medium was replaced with fresh DKAM medium to remove the ROCK inhibitor. Colonies were left to develop over 9-12 days before passaging the wells which looked to contain typical human PSC colonies. Colonies were passaged from 96 well plate into a 48 well plate by manual scrapping with a p200 tip, then aspirating and dispensing the dissected colony into one well of a 48 well plate. DKAM medium was used to grow clones until lines appeared to be growing stably, often up to the third passage.

2.8. Live TRA-1-81 staining.

In order to assess the progress of the clonal line formation I performed live staining for a pluripotency-associated surface marker TRA-1-81. Lines were assessed after the first passage into 48 well plates. TRA-1-81 antibody was added to warm DKAM medium at 1:10 dilution and incubated at 37°C for 30 minutes. Wells were then washed twice with KOSR medium before medium containing goat anti mouse alexa fluor 647 antibody was added to each well, cells were incubated at 37°C for 30 minutes. Wells were then washed once with KOSR medium and twice with FluoroBrite DMEM (ThermoFisher). FluoroBrite DMEM is specially formulated for live fluorescent imaging, as it lacks phenol red and other fluorescent components. Wells were then imaged on the INCell analyser on brightfield, FITC and Cy5 channels. Wells were tiled at 4x magnification for time purposes and at 37°C to preserve viability. While fluoroBrite DMEM is optimal for imaging fluorescent cells it is not conducive to pluripotent growth so the time the cells were kept in this medium was limited as much as possible. After imaging the medium was replaced with fresh DKAM medium and returned to 37°C incubator.

2.9. Clonogenics

2.9.1. Standard Clonogenic assay.

Clonogenic assays were performed on cells after FACS. Cells grown either on MEF or Vitronectin were dissociated with accutase (see 2.3). After antibody staining (see 2.5.6.) DAPI was added at and used for live/dead discrimination. The cells were analysed and subsequently sorted on the BD FACS Jazz. Cells were sorted into 5ml of DMEM (without phenol red) and 10% FCS with 10µM Y-27632 added. After sorting cells were counted (see 2.4) and centrifuged for 3 mins at 1000 rpm to pellet the cells. Cells were resuspended in KOSR medium containing 10µM Y-27632 and 50µg/mL gentamycin and plated at a density of ~500 cells/cm2. 24 hours post plating the medium was replaced with standard KOSR medium to remove the Rock inhibitor and cell debris. Colonies were left to grow for 4 days at 37°C. Wells were then washed once with PBS (w/o Ca**+**, Mg**++**), and then 4% PFA was added to fix the colonies. Samples were incubated at room temperature for 15 mins, the 4% PFA was removed and plates were subsequently washed with PBS. PBS was added to the wells to prevent them from drying out and plates were stored at 4°C until future staining.

2.10. Differentiation Assays

Cells were taken from unsorted or bulk sorted populations, cells were resuspended in E8 medium plus 10µM Rocki. 30,000 cells were plated into each well of a 24 well plate in E8 medium plus Rock inhibitor. Cells were plated in E8 medium to aid plating survival, as plating directly into differentiation medium proved problematic. The following day the medium was changed for either E8, E6, E6 **+** 3µM CHIRON or E8 **+** 3µM CHIRON. E8 medium promotes self-renewal of pluripotent stem cells, whereas E6 medium lacks FGF2 and TGFβ, critical components for human PSC self renewal. E6 medium should allow for cells to randomly differentiate without large amounts of cell death. E6 medium alone may not however confer to a neutral condition as it has been shown to promote ectodermal differentiation (Lippmann et al., 2014). Due to the lack of additional cytokines or small molecules E6 alone is referred to as “gentle ectoderm” differentiation in the text. The addition of CHIRON to E6 or E8 were added to promote mesodermal differentiation. Medium was replaced every two days and cells were left to grow for 6 days. When the expression of *MIXL1*-GFP was set to be monitored over the days in culture plates were put in the Biostation Incubator (Nikon). A 5X5 grid was imaged at 10x magnification in the centre of the wells every 6 hours. Cells were imaged on both brightfield and FITC channels, subsequent images were analysed on the CL- quant v3.10 software.

2.11. Embryoid Bodies

2.11.1. Formation

To assess the trilineage potential of either established clonal lines or particular *MIXL1* positive substates I used an approached which entailed the formation of Embryoid Bodies (EB) under "neutral" conditions, in this context neutral simply indicates that no exogenous cytokines or chemicals were added to guide differentiation. Cells were either used directly from flasks or after they had been FACS sorted for a particular population. In either situation cells were centrifuged at 1000rpm for 3 minutes before being resuspended in APEL 2 medium (Stem Cell Technologies) containing 10µM ROCKi. Cells were resuspended at 3,000 cells per 50µL. 50µL of cells were added to the inner 60 wells of a non-adherent Grenier U bottom 96 well plate. The outer 36 wells were filled with PBS as to prevent the inner wells from drying out. After adding the cells plates were centrifuged at 1000rpm for 3 minutes, to pellet cells. Plates were incubated at 37°C, %CO2 for 7 days.

2.11.2. Harvesting

After 7 days, the medium and EBs were collected from all 60 wells and transferred into a 15ml falcon tube. The wells are then washed with 5 mL of PBS and transferred to the 15mL tube to collect any EBs that had been left behind. The sample is then left for 15 minutes at room temperature to allow EBs to sediment by gravity. The supernatant is then removed and 1mL of Trizol is added with a 1000µL pipette tip. The solution is pipetted up and down vigorously to break up EBs. The solution is then incubated for 3 minutes before another session of vigorous pipetting. Samples are then transferred to -80°C until RNA extraction.

2.11.3. RNA Extraction

RNA was extracted from Trizol treated samples. Trizol samples were defrosted on ice before 200µl of chloroform was added per ml of trizol sample. Samples were then vortexed vigorously until solution appeared milky pink in colour. Samples were incubated at room temperature for 10 minutes before centrifugation at 14,000g for 30 minutes at 4°C, to facilitate phase separation. After this the clear aqueous phase was transferred in to an RNA binding column from a Norgen Total RNA Purification Kit (Norgen, 37500) and processed as per manufacturer's instructions. RNA was eluted into 50µl of elution buffer. RNA samples were then analysed for RNA concentration using a Nanodrop lite (Thermofisher).

2.11.4. RNA to cDNA Conversion

Prior to cDNA conversion, samples were first treated with DNA-free DNA Removal kit (Thermofisher, AM1906). This removes contaminating genomic DNA to ensure better qPCR results. Samples were diluted to 200µg per mL and processed as per manufacturer's

instructions. After DNA removal remaining samples were diluted to 1µg in 225µL and 25µL was dispensed into all wells of an 8 well 0.2mL PCR tube strip. 25µL of reaction master mix was added from the High Capacity cDNA Reverse Transcription Kit (Thermofisher, 4368814) to each well. Samples were then loaded into a PCR machine (QuantStudio 12k Flex, Applied Biosystems) and run on the following cycle: 10 Minutes 25°C, 120 Minutes 37°C, 5 minutes 85°C and hold at 4°C.

2.11.5. Human PSC Scorecards

Human PSC scorecards were purchased from ThermoFisher(A15870). These scorecards are designed to assess the trilineage potential of human PSC lines, based on the work from Alex Meissner (Bock et al., 2011). The scorecard analyses each sample for 96 genes, related to pluripotency, endoderm, mesoderm, ectoderm and some housekeeping genes as well. Purified cDNA samples are diluted as per manufacturer's instructions and 10µL is added to each well, which contains desiccated primer and probe for a given gene. The scorecards are then loaded and run on the QuantStudio Flex 12K Thermocycler. The cycle conditions are shown below in Table 2.6. After the run had finished, data was uploaded to Thermofisher for analysis using their software (https://www.thermofisher.com/uk/en/home/life-science/stem-cell-research/taqman-hpsc-scorecard-panel/scorecard-software.html).

Table 2.6 Human PSC scorecard qPCR Cycle Conditions

Step	Temperature	Time	Cycles
Hold	50°C	2 minutes	-
Hold	95°C	10 minutes	-
Melt	95°C	15 seconds	40
Aneal/Extend	60°C	1 minute	

2.12. RNA Sequencing

2.12.1. Sorting for Bulk RNA Sequencing

After setting up the streams on the BD FACs Jazz or FACS Aria we first begin by sorting 10,000 cells from a given population into an eppendorf tube containing 200µL DMEM/FCS. 1,000 cells are then reanalysed from this sorted fraction to ensure accurate sorting has taken place. Immediately after a successful reanalysis, 10,000 cells from the given population are sorted into an eppendorf tube containing 800µL Trizol. Post sort, samples are sealed and stored at -80°C until extraction.

2.12.2. RNA Analysis of Bulk RNA Sequencing

RNA samples were processed, sequenced and aligned to the human reference genome by colleagues at University College London (UCL). In brief, RNA extraction and cDNA prepared using the SMARTer v3.0 cDNA synthesis kit (Clontech). RNA-Seq was performed on an Illumina Nextseq 500 machine. The Raw data was processed by Tophat (Kim et al., 2013a) and Cufflink/Cuffdiff (Roberts et al., 2011a; Roberts et al., 2011b; Trapnell et al., 2013; Trapnell et al., 2010). Reads were aligned to human genome reference consortium grch38. Results of RNA sequencing were transferred back to me in Sheffield in two formats; BAM, alignment files and FPKM (Fragments Per Kilobase per Million Reads) values. Gene expression data was then processed by Seqmonk software (Babraham-Bioinformatics, 2018) for quality control and to produce heat maps and clustering.

2.13. Single Cell Transcriptomics

2.13.1. Sorting for single cell qPCR or RNA-seq

To ensure accurate deposition of single cells into individual wells of the 96 well plate firstly droplets were sorted to align the stream to the middle of the wells. Green beads were sorted as single beads into 96 well plates. These were viewed on an EVOS microscope at 10x on the FITC channel to verify the correct number and placement of the deposited beads was correct. Cells were stained (see 2.5.6) for SSEA-3 expression using monoclonal antibody clone MC631-2C2. They were then stained with a secondary antibody goat anti mouse IgG + IgM Heavy and light chain Alexa Fluor 647 (Jackson ImmunoResearch). Samples were sorted using the flow cytometer, BD FacsAria. DAPI or Propidium Iodide was added and used for live/dead discrimination. Sorting population gates were set in the flow cytometry software, FACS DIVA, based on GFP expression (488nm laser), and SSEA-3 expression (635nm laser). Single cells from each selected substate were then directly deposited into 4µl lysis buffer in a 96 well plate (Table 2.7):

Lysis Buffer for single cell qPCR 1 plate:(96 wells) 10%NP-40 IGEPAL (sigma) 17µl, 10mM dNTP 2.8µl, 0.1M DTT 10µl, RNase inhibitor (RNaseOUT, Invitrogen) 5.3µl, DEPC or nuclease-free water 390µl.

Table 2.7 Sort Layout for Single Cell qPCR:

	1	2	3	4	5	6
A	SC no RT	SC no RT	SC no RT	SC	SC	SC
B	SC	SC	SC	SC	SC	SC
C	SC	SC	SC	SC	SC	SC
D	SC	SC	SC	SC	SC	SC
E	SC	SC	SC	SC	SC	SC
F	SC	SC	SC	SC	SC	SC
G	SC	SC	SC	10	10	10
H	NTC	NTC	NTC	20	20	20

SC= Single Cell, no RT= No Reverse Transcription, NTC = Non-Template Control. 10 and 20 cells were added as controls. Red wells are from target population, Green wells are from the SSEA-3**+** population and Purple wells are from the double negative population.

After cells were sorted, plates were sealed and then transferred to -80°C for storage.

2.13.2. Single cell qPCR.

qPCR primer and probes were ordered through ThermoFisher Scientific Taqman Gene Expression Assay design website. Primer and Probe list is present in appendix, **Table 9.1.**

Samples were put through a pre-amplification step using SSIII/Taq Cell Direct kit (ThermoFIsher), cycle conditions shown in Table 2.8

Table 2.8 Cycle Conditions for Reverse Transcription and Pre amplification of Single Cell RNA for qPCR.

RT-PCR thermal condition

RT	50°C	60min	
Inactivation of RT/activation of Taq	95C	2 min	
Specific target amplification	95C	15sec	25 cycles total **22 cycle total for bulk**
	60°C	4min	
END	25c	10 sec	

Amplified samples are mixed with the appropriate probe and loaded in the IFC chip then run on the BioMark HD machine. A schematic of the overall process is shown in **Figure 2-4.**

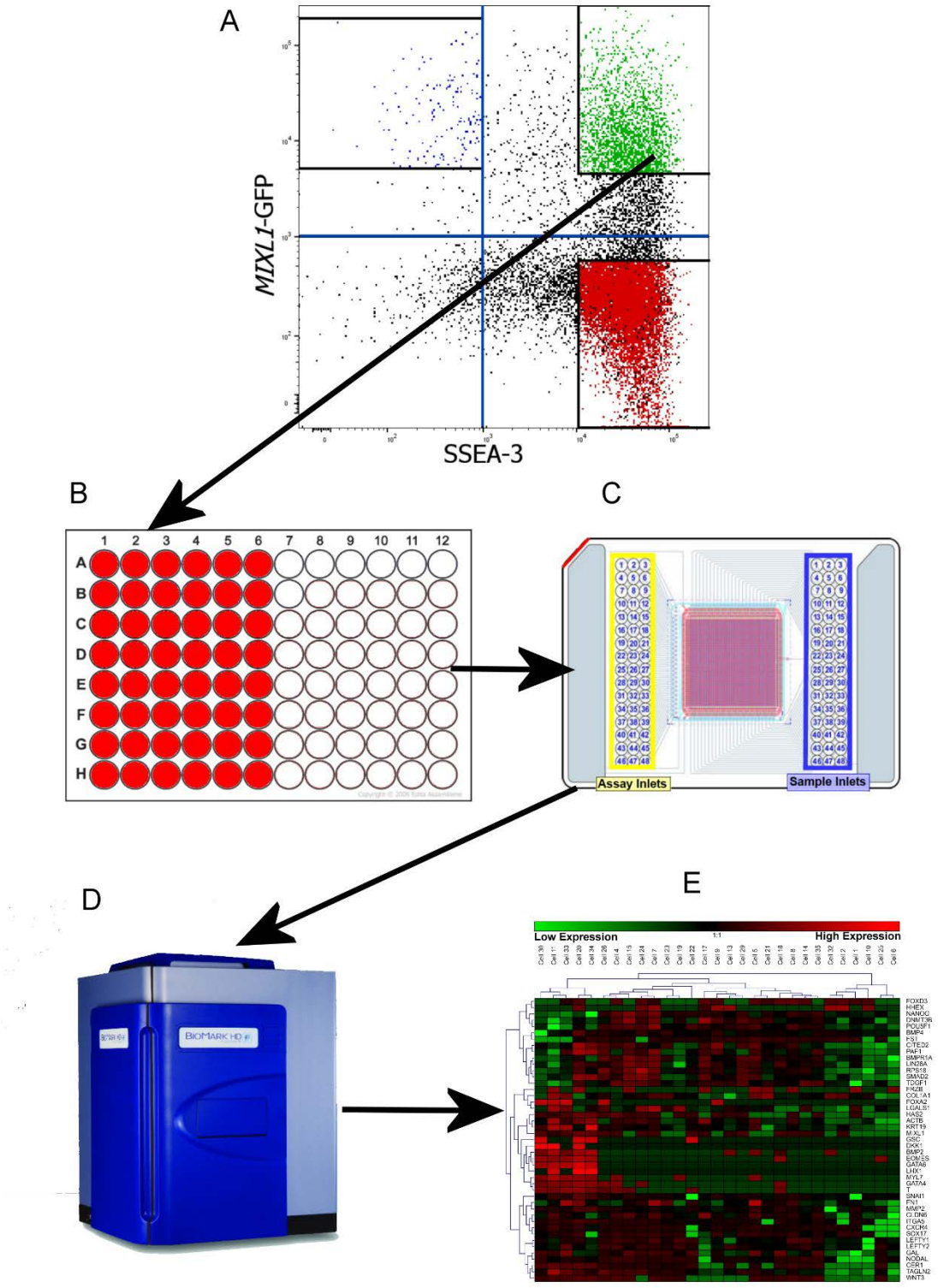


Figure 2-4 Schematic of Single Cell Analysis Pipeline

Cells sorted using BD FACS Aria according to theses gates depicted in **A)**. **B)** Single cells deposited into a 96 well PCR plate for preamplification. **C)** Samples and probes added to Fluidigm chip. **D)** Chip run through the BioMark HD qPCR machine. **E)** Data processed afterwards to produce relevant analysis.

2.13.3. Single cell qPCR analysis

Single cell qPCR analysis was performed by three pieces of software. Prior to analysis cells were screened and cells with particularly high CT values for the housekeeping genes ACTB and RPS18 were removed. Depending on the application being used, some pre-processing was performed. For SCExV software (Lang et al., 2015) 999 values were left unadjusted. For Monocle software (Trapnell et al., 2014) 999 values were set to 40, as this is the maximum number of cycles on the PCR machine. For Genesis software (Sturn et al., 2002) 999 values were removed and global Z-scores were computed for all remaining values.

Global Z-scores are calculated using this formula:

(Ct value – Ct mean of all values) / Standard Deviation of all Ct.

2.13.3.1. Monocle Single Cell Analysis

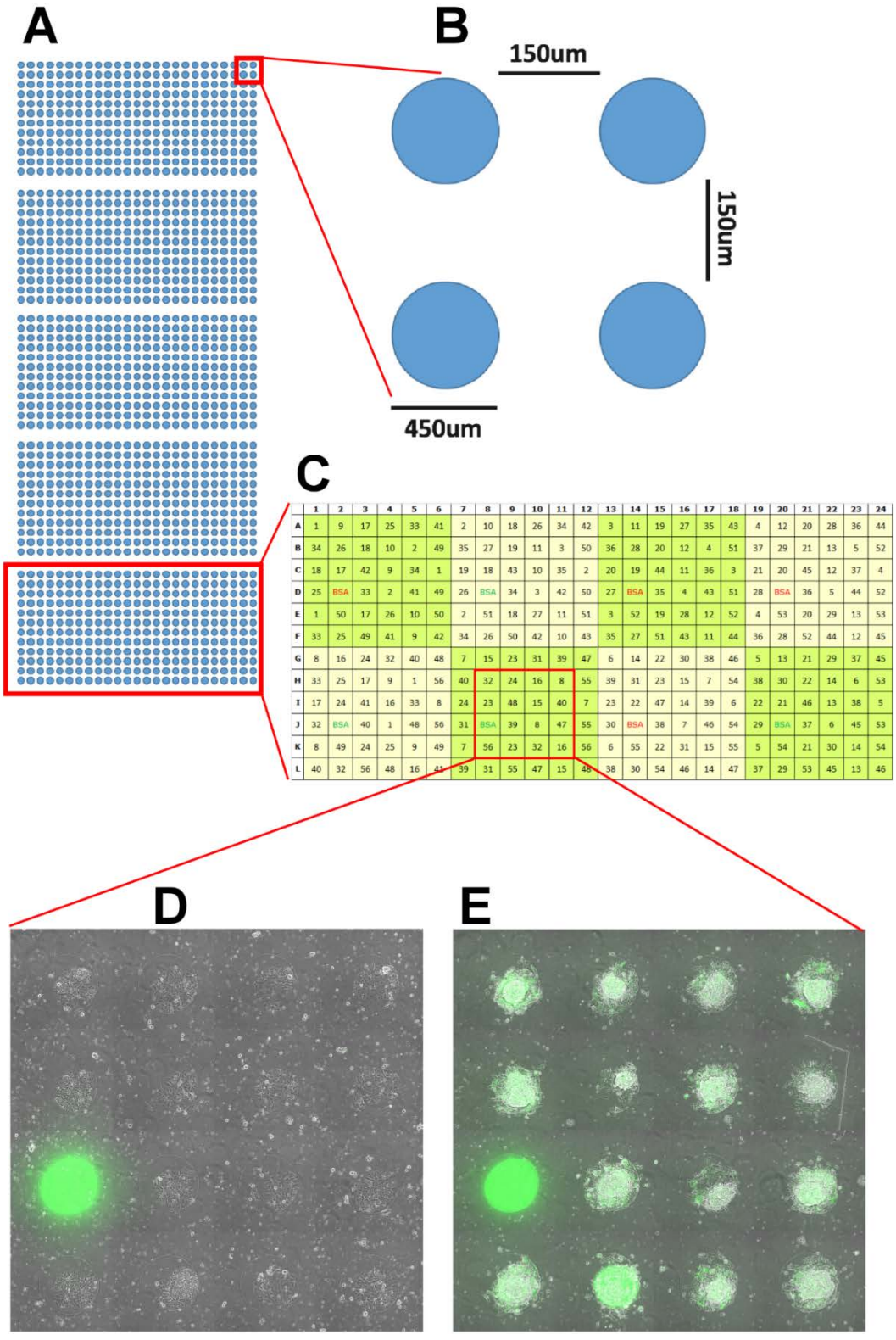
The Monocle2 package from Trapnell laboratory (Qiu et al., 2017a; Qiu et al., 2017b; Trapnell et al., 2014) was used to model the differentiation process. The Monocle2 package is designed for RNA sequencing data but can be used on qPCR data with some changes in parameters. The script used for Monocle analysis is attached in the Appendix. Two basic parameters that had to be changed for qPCR data is to firstly change the data expression family to gaussianff() and for many graphs to reverse the scales so that lower Ct values appear higher on plots and graphs. In brief, cell types were specified based on gene expression of *MIXL1* and *NANOG* and cells from the *MIXL1*(-)/SSEA-3(**+**) fraction were chosen as the start of pseudotime. Monocle then used gene expression values from each cell to order the cells in pseudotime. I was then able to plot Monocle's analysis in various plots and heatmaps, including cell trajectory trees and tSNE plots.

2.14. Microwell matrix assessment.

To assess whether a particular matrix combination was more suited to the biasing conditions we utilised a technique of microwells coated with combinations of proteins (Table 2.9). The microwells were provided by Andrea Manfrin, of Mathias Lutolf's group at École polytechnique fédérale de Lausanne (EPFL). The wells were created using the techniques described in their paper Gobaa et al, 2011. Microwell plates were stored in PBS at 4°C until they were used. Seven Polyethylene glycol (PEG) microwell arrays of 288 microwells are stamped into 3 wells of a NUNC (167063 Nunclon 4-Well). The five middle arrays were loaded with the protein combination in Table 2.9.

To prepare plates for cells they were first pre-warmed at 37°C for 30 minutes still containing PBS. After pre-warming PBS was aspirated from the plates and re-placed with E8 medium, then incubated for another 30 minutes at 37°C. Whilst the plate was warming up cells were prepared for plating.

Cells were harvested using Tryple and counted as detailed in 2.4. The cells were passed through a 70µM cell strainer to remove larger cell aggregates. Once harvested cells were also washed twice in clear E8 medium to remove any residual Tryple. Cells were then resuspended at 2 million cells per 5mL in clear E8 medium containing 10µM ROCKi and 1:100 Pen/Strep (Thermofisher). 5mL cell suspension was added to each well of the 4 well plate containing the microwells. The cells were left to plate overnight. The following day the medium was removed and plates were washed twice with PBS to remove cell debris. 5mL of medium was then added to each well, one well with clear imaging E8, one well with clear 1x Primo and one well with clear 2x Primo, all containing 1:100 Pen/strep. Plate was placed in the Biostation incubator to be imaged regularly over 3 days, monitoring both cell morphology and GFP expression (Figure 2-5). The medium was replaced every day with two PBS washes between medium change to remove cell debris. After 3 days cells were fixed (see 2.5.2 and 2.5.3) and stained (2.5.5) for NANOG and Brachyury. The wells were then imaged on the InCell Analyzer (GE Healthcare) on Brightfield, DAPI, FITC, Texas Red and Cy5 channels. Images were analysed by CellProfiler (Carpenter et al., 2006) and data plotted using GraphPad Prism 7.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	1	9	17	25	33	41	2	10	18	26	34	42	3	11	19	27	35	43	4	12	20	28	36	44
B	34	26	18	10	2	49	35	27	19	11	3	50	36	28	20	12	4	51	37	29	21	13	5	52
C	18	17	42	9	34	1	19	18	43	10	35	2	20	19	44	11	36	3	21	20	45	12	37	4
D	25	BSA	33	2	41	49	26	BSA	34	3	42	50	27	BSA	35	4	43	51	28	BSA	36	5	44	52
E	1	50	17	26	10	50	2	51	18	27	11	51	3	52	19	28	12	52	4	53	20	29	13	53
F	33	25	49	41	9	42	34	26	50	42	10	43	35	27	51	43	11	44	36	28	52	44	12	45
G	8	16	24	32	40	48	7	15	23	31	39	47	6	14	22	30	38	46	5	13	21	29	37	45
H	33	25	17	9	1	56	40	32	24	16	8	55	39	31	23	15	7	54	38	30	22	14	6	53
I	17	24	41	16	33	8	24	23	48	15	40	7	23	22	47	14	39	6	22	21	46	13	38	5
J	32	BSA	40	1	48	56	31	BSA	39	8	47	55	30	BSA	38	7	46	54	29	BSA	37	6	45	53
K	8	49	24	25	9	49	7	56	23	32	16	56	6	55	22	31	15	55	5	54	21	30	14	54
L	40	32	56	48	16	41	39	31	55	47	15	48	38	30	54	46	14	47	37	29	53	45	13	46

Figure 2-5 Schematic of Microwell Matrix Assessment

A) Diagram of microwells: A 288 PEG microwell grid is stamped five times per well of a four well plate. **B)** Each microwell has a diameter of 450µm and are spaced evenly apart, 150µm. **C)** Table of condition number (according to Table 2.9 Microwell Protein Combinations and Concentrations) placement within the grid. **D, E)** Representative images of region highlighted, the green fluorescent channel overlaid onto brightfield image, images taken at day 0 **(D)** and day 3**(E).**

Table 2.9 Microwell Protein Combinations and Concentrations

Combinations	PROTEIN 1	Concentration PROTEIN 1 (µg/ml)	PROTEIN 2	Concentration PROTEIN 2 (µg /ml)	PROTEIN 3	Concentration PROTEIN 3 (µg /ml)
1	Vitronectin	87.5	Collagen I	350	-	-
2	Vitronectin	87.5	Collagen III	350	-	-
3	Vitronectin	87.5	Collagen IV	350	-	-
4	Vitronectin	87.5	Collagen V	350	-	-
5	Vitronectin	87.5	Collagen VI	350	-	-
6	Vitronectin	87.5	Laminin 211	35	-	-
7	Vitronectin	87.5	Laminin 221	35	-	-
8	Vitronectin	87.5	Laminin 521	35	-	-
9	Vitronectin	87.5	Fibronectin	700	-	-
10	Vitronectin	87.5	E-Cadherin	35	-	-
11	Vitronectin	87.5	PBS	-	-	-
12	Vitronectin	87.5	Collagen I	175	Collagen III	175
13	Vitronectin	87.5	Collagen I	175	Collagen IV	175
14	Vitronectin	87.5	Collagen I	175	Collagen V	175
15	Vitronectin	87.5	Collagen I	175	Collagen VI	175
16	Vitronectin	87.5	Collagen I	175	Laminin 211	17.5
17	Vitronectin	87.5	Collagen I	175	Laminin 221	17.5
18	Vitronectin	87.5	Collagen I	175	Laminin 521	17.5
19	Vitronectin	87.5	Collagen I	175	Fibronectin	350
20	Vitronectin	87.5	Collagen I	175	E-Cadherin	17.5
21	Vitronectin	87.5	Collagen III	175	Collagen IV	175
22	Vitronectin	87.5	Collagen III	175	Collagen V	175
23	Vitronectin	87.5	Collagen III	175	Collagen VI	175
24	Vitronectin	87.5	Collagen III	175	Laminin 211	17.5
25	Vitronectin	87.5	Collagen III	175	Laminin 221	17.5
26	Vitronectin	87.5	Collagen III	175	Laminin 521	17.5
27	Vitronectin	87.5	Collagen III	175	Fibronectin	350
28	Vitronectin	87.5	Collagen III	175	E-Cadherin	17.5
29	Vitronectin	87.5	Collagen IV	175	Collagen V	175
30	Vitronectin	87.5	Collagen IV	175	Collagen VI	175
31	Vitronectin	87.5	Collagen IV	175	Laminin 211	17.5
32	Vitronectin	87.5	Collagen IV	175	Laminin 221	17.5
33	Vitronectin	87.5	Collagen IV	175	Laminin 521	17.5
34	Vitronectin	87.5	Collagen IV	175	Fibronectin	350
35	Vitronectin	87.5	Collagen IV	175	E-Cadherin	17.5
36	Vitronectin	87.5	Collagen V	175	Collagen VI	175
37	Vitronectin	87.5	Collagen V	175	Laminin 211	17.5
38	Vitronectin	87.5	Collagen V	175	Laminin 221	17.5
39	Vitronectin	87.5	Collagen V	175	Laminin 521	17.5
40	Vitronectin	87.5	Collagen V	175	Fibronectin	350

Combinations	PROTEIN 1	Concentration PROTEIN 1 (µg/ml)	PROTEIN 2	Concentration PROTEIN 2 (µg /ml)	PROTEIN 3	Concentration PROTEIN 3 (µg /ml)
41	Vitronectin	87.5	Collagen V	175	E-Cadherin	17.5
42	Vitronectin	87.5	Collagen VI	175	Laminin 211	17.5
43	Vitronectin	87.5	Collagen VI	175	Laminin 221	17.5
44	Vitronectin	87.5	Collagen VI	175	Laminin 521	17.5
45	Vitronectin	87.5	Collagen VI	175	Fibronectin	350
46	Vitronectin	87.5	Collagen VI	175	E-Cadherin	17.5
47	Vitronectin	87.5	Laminin 211	17.5	Laminin 221	17.5
48	Vitronectin	87.5	Laminin 211	17.5	Laminin 521	17.5
49	Vitronectin	87.5	Laminin 211	17.5	Fibronectin	350
50	Vitronectin	87.5	Laminin 211	17.5	E-Cadherin	17.5
51	Vitronectin	87.5	Laminin 221	17.5	Laminin 521	17.5
52	Vitronectin	87.5	Laminin 221	17.5	Fibronectin	350
53	Vitronectin	87.5	Laminin 221	17.5	E-Cadherin	17.5
54	Vitronectin	87.5	Laminin 521	17.5	Fibronectin	350
55	Vitronectin	87.5	Laminin 521	17.5	E-Cadherin	17.5
56	Vitronectin	87.5	Fibronectin	350	E-Cadherin	17.5
BSA	BSA-Alexa 555		-	-	-	-
BSA	BSA-Alexa 488		-	-	-	-

3. Chapter 3: Characterisation of human PSC substates present in Standard Culture Systems.

3.1. Introduction

3.1.1. *MIXL1* function and expression patterns in development.

The formation of the primitive streak is a crucial part of gastrulation and indeed vertebrae development. The primitive streak encompasses the cells which will form two of the three main germ layers, mesoderm and endoderm (Mikawa et al., 2004). Due to their common progenitor state, the precursor to both of these lineages is termed, mesendoderm (Technau and Scholz, 2003). In the mouse epiblast, cells pass through the primitive streak and the mesendoderm is patterned into mesoderm and definitive endoderm (Kinder et al., 1999).

While a handful of MIX/BIX protein family members contribute to the formation of the primitive streak and subsequent endoderm and mesoderm patterning in xenopus and zebrafish (Sahr et al., 2002), only one member of this family is present in chicken (Stein et al., 1998) , mouse (Pearce and Evans, 1999) and human (Robb et al., 2000), Mix Like 1 (CMIX, Mixl1, MIXL1). Mixl1 is a transcription factor with multiple targets related to early mesendoderm formation (Pereira et al., 2012; Zhang et al., 2009). Specific targets included genes associated with endoderm and mesoderm, Gsc, Sox17, Cer1, Nkx2 and Tbx6 (Pereira et al., 2012; Zhang et al., 2009).

Knockout and overexpression studies have highlighted *Mixl1*'s importance to the formation of mesoderm and endoderm, both *in vitro* and *in vivo*. *Mixl1* deficient mice embryos present serious defects in mesoderm and endoderm formation and display arrested development by the early somite stage (Hart et al., 2002). *In vitro* studies using mouse PSC show compromised endoderm formation for *Mixl1* null cell lines during differentiation, whereas enforced *Mixl1* expression promotes the endoderm formation (Lim et al., 2009; Pereira et al., 2012). Nevertheless, its expression is crucial for mesendoderm formation (Hart et al., 2002) and subsequently conditional activation of *Mixl1* has been shown to accelerate mesoderm formation in mouse PSC (Zhang et al., 2009).

Mixl1's direct relationship with Gsc and Brachyury is more complicated, with a regulatory triangle being proposed (Latinkic and Smith, 1999). Mixl1, Gsc and Brachyury show similar patterns of expression during the initial formation of the primitive streak (Blum et al., 1992; Latinkic and Smith, 1999). However, as cells progress through the streak and begin to pattern into mesoderm and endoderm, Mixl1 and Gsc expression persists in endoderm, whilst Brachyury persists in mesoderm. Gsc is a transcriptional repressor and shown to be a direct target of Mixl1. In xenopus models, it was thought that after transcriptional activation by Mix 1, xGsc repressed

the expression of Xbra (Latinkic and Smith, 1999). In differentiating mouse PSC, however, Brachyury and other T-box factors have been shown to interact directly with Mixl1 and repress its transactivation ability, particularly its activation of Gsc (Pereira et al., 2011).

The coordination of these three genes is an important mechanism in the patterning of mesendoderm into mesoderm and endoderm. Whilst Mixl1's expression shows a stronger correlation with endoderm, it plays a vital role in the formation of mesendoderm and subsequently the efficient formation of mesoderm.

3.1.2. *MIXL1* during differentiation of Human ESC

The exact expression of *MIXL1* in the developing human primitive streak is unknown, and much of our knowledge is inferred from animal models (Blum et al., 1992; Latinkic and Smith, 1999; Sahr et al., 2002). The derivation of human embryonic stem (ES) cells by Thomson et al, in 1998 , allowed us to monitor gene expression changes during differentiation and infer these patterns to *in vivo* human development. Aided by the development of a fluorescent reporter line by Davis et al, 2008, *MIXL1* expression has been well investigated in terms of its expression in differentiation of human ESC. The paper has received 145 citations with many of these utilising the reporter to study *MIXL1* expression during differentiation.

Some examples of the use of the reporter include monitoring its expression during differentiation towards downstream mesoderm derivatives such as cardiomyocytes (Lian et al., 2012) and also definitive endoderm (Loh et al., 2014). Loh et al, 2016 utilised the *MIXL1* reporter to investigate the pairwise choices cells make during mesodermal differentiation. At day one of mesoderm induction, they identify both anterior and mid primitive streak like population which expressed high *MIXL1*-GFP expression. Single cell RNA-sequencing of these populations revealed upregulation of *MIXL1* and *T* (BRACHYURY) within these day 1 primitive streak like populations.

Directed differentiation of human PSC often has a complex signalling infrastructure and is highly specific for certain cell lineages. The complexity is further expanded when a temporal aspect is added to the differentiation down these lineages. Using the *MIXL1*-GFP reporter line, Jackson et al, 2010, showed that differentiating cells pass through a temporal window which affects their responsiveness to the signalling molecules. Signalling cues are dose dependant, reliant on specific co-signalling and organised temporally. The addition of signalling molecules in the incorrect dose or order can lead to a completely different cell lineage. Addition of Activin A prior to BMP4 would help to maintain the stem cell state instead of differentiation towards mesoderm. This is reflective of the spatial, temporal patterning of cells across a gradient of

signalling during both gastrulation and further embryogenesis (Row and Kimelman, 2009; Zernicka-Goetz, 2002). Any directed differentiation protocol basically tries to recapitulate this process of gastrulation to a particular lineage and further specification to a terminally differentiated cell type.

MIXL1 has exhibited upregulation during early differentiation towards mesodermal lineages of PSC both in vivo and in vitro. In the context of mouse EpiSC, its expression appears prior to the loss of pluripotent identity yet the level of expression has been shown to impact the differentiation potential of these cells (Kojima et al., 2014; Tsakiridis et al., 2014). The expression of early differentiation genes such as *GATA6* has been assessed within a substate of human ESC cells (Allison et al., 2018b). *MIXL1* expression has not been assessed within this context, it is possible that amongst the heterogenous population of human ESC in vitro, a substate might exist which expresses *MIXL1*. Subsequently this expression could have functional ramifications with regard to their self renewal and differentiation potential. I sought to investigate in standard human PSC culture conditions using the same *MIXL1* reporter line described previously (Davis et al., 2008).

3.2. Results

3.2.1. Identification of substates expressing *MIXL1*-GFP in MEF/KOSR culture conditions.

I used the surface marker SSEA-3 (Shevinsky et al., 1982) as a tool to monitor the pluripotent population within cultures. The expression of *MIXL1* in early differentiation of human PSC towards mesodermal derivatives (Loh et al., 2016) identified it as a candidate gene to assess for early lineage priming. I utilised the *MIXL1* reporter line (Davis et al., 2008) and coupled this with SSEA-3 staining. Flow cytometry analysis of HES3-*MIXL1* cells grown in MEF/KOSR reveals a population of *MIXL1*-GFP/SSEA-3 double positive cells (Figure 3-1). However, the fraction although often present its abundance is variable in MEF/KOSR conditions, ranging from 0% up to 20% of the culture. Variation in culture conditions can impact on this proportion including, density of the MEF layer, passaging ratios and survival post passage.

3.2.2. Variability of substates presence in standard conditions.

I assessed whether the *MIXL1*(**+**)/SSEA-3(**+**) substate showed expression or variability in expression in other standard culture conditions. When I transitioned the HES3 *MIXL1*-GFP line into defined feeder free systems using mTESR medium on matrigel matrix or E8 medium on a matrix of vitronectin protein (E8V) the *MIXL1*(**+**)/SSEA-3(**+**) substate was significantly reduced. E8V was the most consistent with this double positive substate remaining virtually non-existent across multiple samples (Figure 3-2).

As with *MIXL1*-GFP expression, different conditions can also have a profound effect on the distribution of SSEA-3 expression in culture of human PSC. The distribution of SSEA-3 in MEF/KOSR cultures is often a heterogeneous distribution ranging from negative to very high expressing (Tonge et al., 2011). The distribution of SSEA-3 in E8V conditions is slightly different with a shift of both extremes towards the middle. This equates to a lower number of negative cells but also a loss of the very high expressing SSEA-3 fraction (Figure 3-3). This is evident when looking at the average distribution between three samples of HES3 on MEF/KOSR and E8V conditions (Figure 3-3D). E8V does offer a more consistent distribution between samples, shown in the similarity table (Figure 3-3C). This table details the percentage similarity between the distribution across samples. Samples taken from E8V show higher similarity ranging from 57% up to 75% whereas the similarity between MEF/KOSR samples similarity ranges from 49% to 55% (Figure 3-3C). Highlighting the variability of not only *MIXL1*-GFP expression in MEF/KOSR but also SSEA-3 expression. Assessing the average distribution of SSEA-3 in these two culture conditions demonstrates the shift in intensity towards mid-level in E8V conditions (Figure 3-3D).

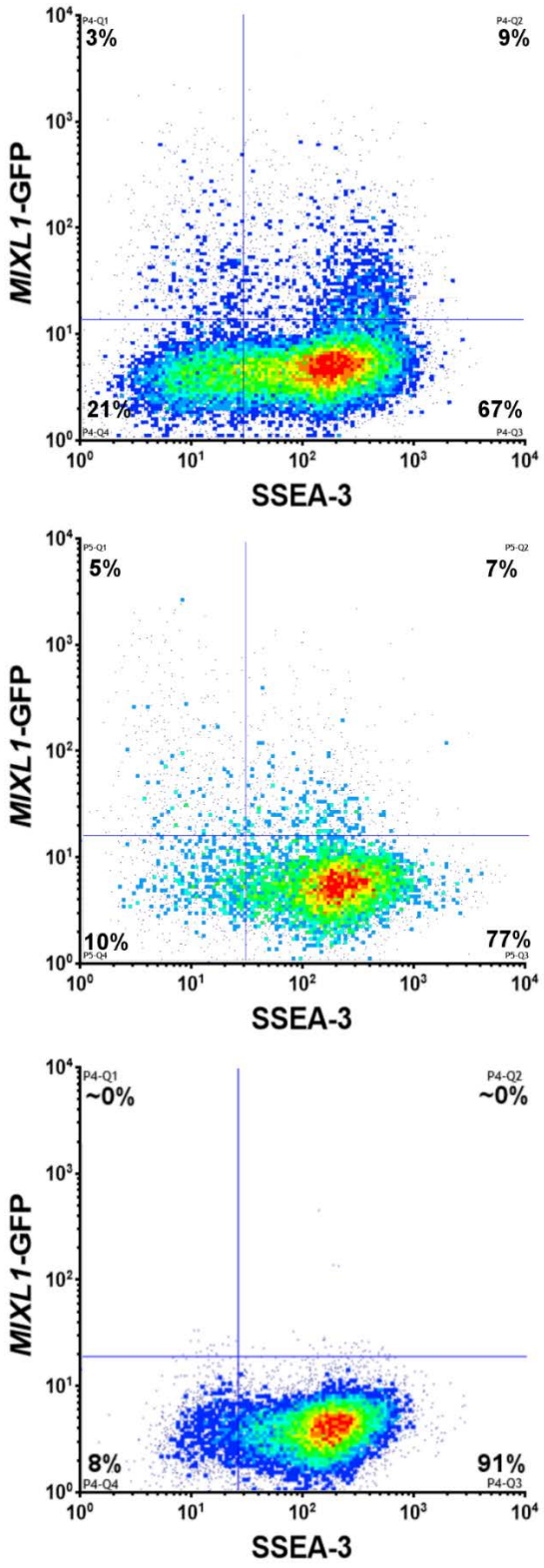


Figure 3-1 *MIXL1*-GFP / SSEA-3 Co-expression in MEF/KOSR Conditions

Flow cytometry density plots of *MIXL1*-GFP versus SSEA-3 from cells grown in MEF/KOSR conditions, showing variable expression.

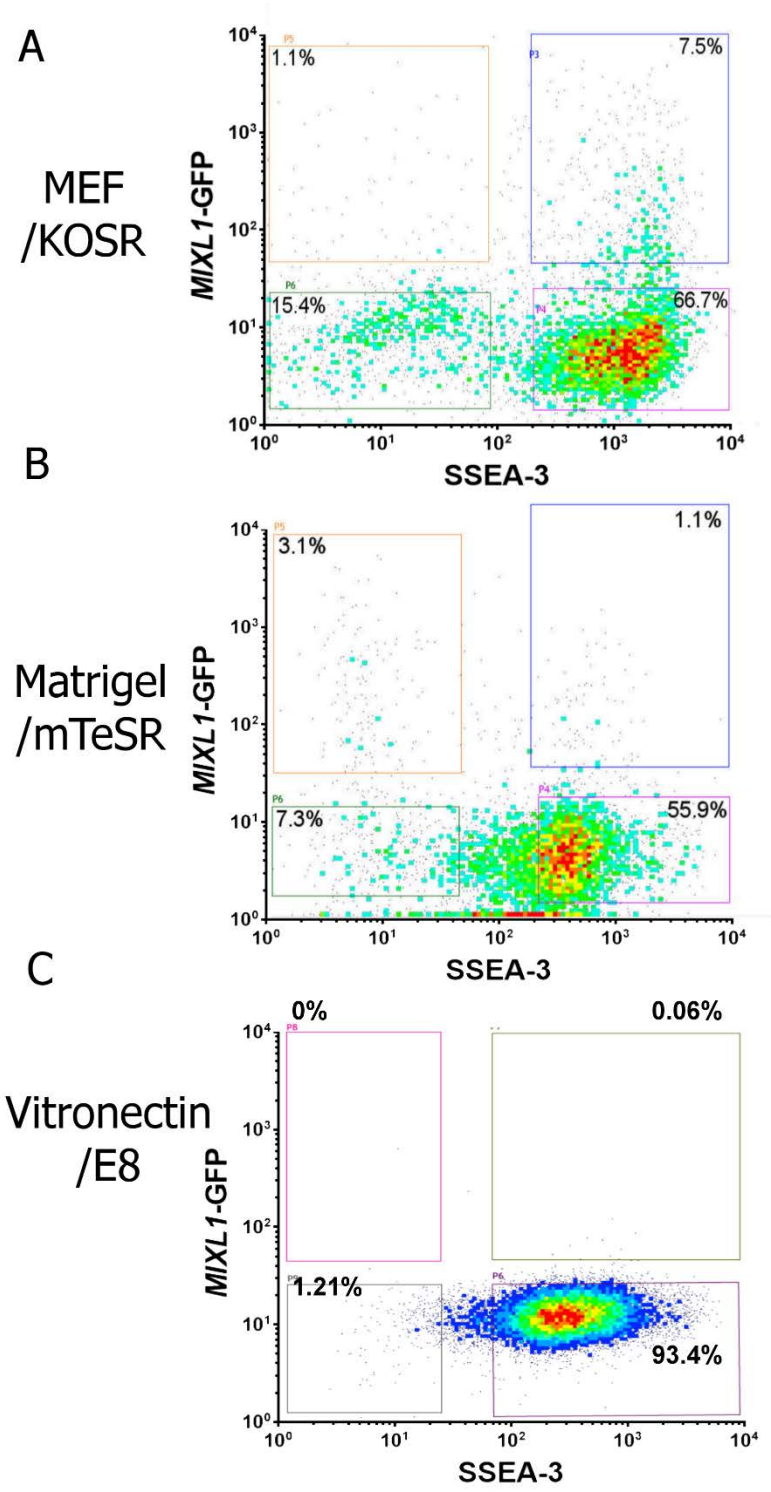
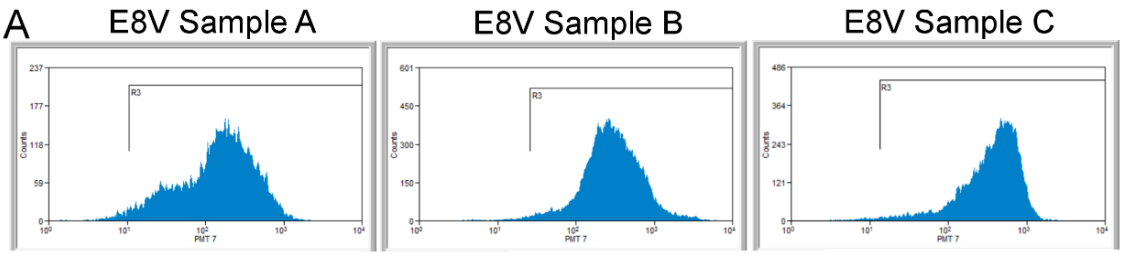


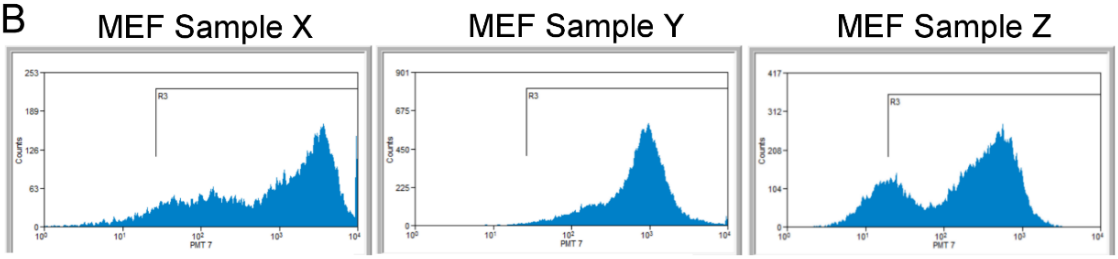
Figure 3-2 Coexpression of the stem cell marker, SSEA-3 and *MIXL1* reporter in different conditions that support self renewal.

A) A flow cytometry density plot of *MIXL1*-GFP versus SSEA-3 from cells grown in MEF/KOSR conditions. **B)** A flow cytometry density plot of *MIXL1*-GFP versus SSEA-3 from cells grown in Matrigel/mTeSR conditions. **C)** A flow cytometry density plot of *MIXL1*-GFP versus SSEA-3 from cells grown *in* vitronectin/E8 conditions.

A



B



C

	E8V Sample A	E8V Sample B	E8V Sample C	MEF Sample X	MEF Sample Y	MEF Sample Z
E8V Sample A		75.70%	73.12%	51.47%	56.94%	78.01%
E8V Sample B	75.70%		57.43%	40.33%	43.29%	61.92%
E8V Sample C	73.12%	57.43%		41.74%	71.86%	68.61%
MEF Sample X	51.47%	40.33%	41.74%		53.11%	49.64%
MEF Sample Y	56.94%	43.29%	71.86%	53.11%		54.95%
MEF Sample Z	78.01%	61.92%	68.61%	49.64%	54.95%	

D

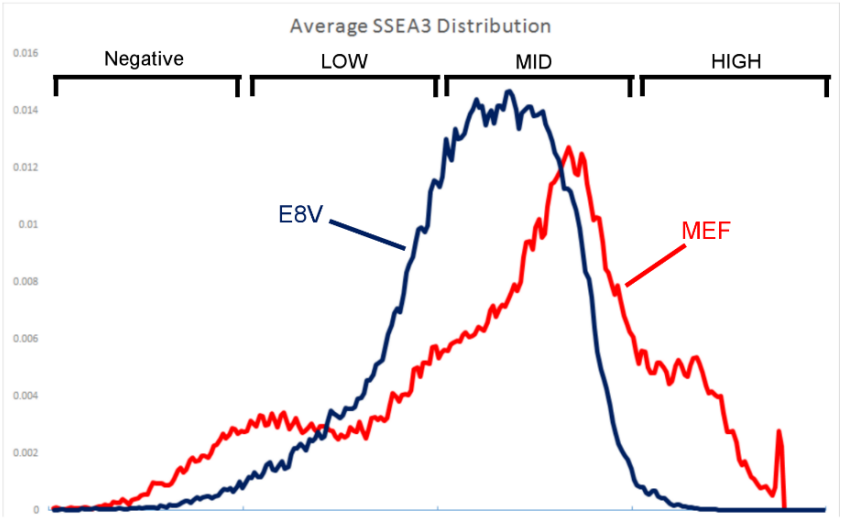


Figure 3-3 Effects of Media and Matrix Changes on SSEA-3 Expression

A) SSEA-3 flow cytometry histograms from three cultures of HES3 human PSC grown in E8V conditions. **B)** SSEA-3 flow cytometry histograms from three cultures of HES3 human PSC grown in MEF/KOSR conditions. **C)** Similarity matrix of the plots in **A)** and **B)**. Percentages indicate the percentage similarity in the distribution of SSEA-3 between samples. Boxes are coloured on a scale according percentage similarity green indicates lowest similarity, red indicates highest similarity. **D)** Average SSEA-3 distribution from **A)** E8V conditions (blue) and **B)** MEF/KOSR conditions (red).

3.2.3. Functional testing of substates defined by *MIXL1*-GFP and SSEA-3 expression.

After identifying the *MIXL1*(**+**)/SSEA-3(**+**) substate in MEF/KOSR, I assessed whether the expression of *MIXL1*-GFP had any functional consequences for the cells in respect to their ability to self renew or differentiate. I sorted *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(-)/SSEA-3(**+**) as bulk populations and plated them into feeder free conditions overnight (E8/Geltrex). The following morning medium was changed for the appropriate conditions, self renewal (E8), gentle ectoderm induction (E6) and mesoderm induction (E6 **+** 3µM CHIRON) (Lippmann et al., 2014). I elected for a gentle approach to differentiation as to see subtle differences between the population and ascertain any bias in the differentiation.

I sorted by FACS three samples from MEF/KOSR conditions (Figure 3-4A), plated in three conditions and left to grow for 6 days. Immunofluorescent analysis of the self renewal conditions, E8, using an anti-NANOG antibody revealed a high percentage of NANOG positive cells from both *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(-)/SSEA-3(**+**) populations (Figure 3-4 B, C). Cells in this condition appeared to grow as colonies as well for both substates. There was a small proportion of *MIXL1*(**+**) cells present from the *MIXL1*(**+**)/SSEA-3(**+**) but these cells mainly co-expressed NANOG. There was also an increase in NANOG/*MIXL1* double negative cells from this population, this may indicate a further differentiated population that has lost NANOG and *MIXL1* expression (Figure 3-4 C).

Immunofluorescent analysis of the ectoderm conditions, E6, using an anti-PAX6 antibody revealed a high percentage of PAX6 positive cells from both *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(-)/SSEA-3(**+**) populations (Figure 3-4 E, F). Again, there was a small proportion of *MIXL1*(**+**) cells present from the *MIXL1*(**+**)/SSEA-3(**+**) but these cells mainly co-expressed PAX6. There was an increase in PAX6/*MIXL1* double negative cells from the *MIXL1*(-)/SSEA-3(**+**) population, again this could be indicative of a further differentiated population (Figure 3-4 F).

Under mesoderm conditions there was a strong induction of *MIXL1*-GFP and immunofluorescence analysis revealed high expression of the mesoderm marker TBX6 after staining with anti-TBX6 antibody (Figure 3-4 H, I). This was apparent for both populations and widespread morphology changes were apparent early in the differentiation.

Wells were imaged as a 7x7 array of images at 10x in the same positions in each well. Based on the expression of the relevant markers for each condition it appears both fractions perform similarly in all conditions. The overall cell numbers in the analysed images for each condition highlights the difference between the fractions (Figure 3-4 D, G, J). *MIXL1*(**+**)/SSEA-3(**+**) show increased variability between samples, with sample 3 performing particularly poorer in

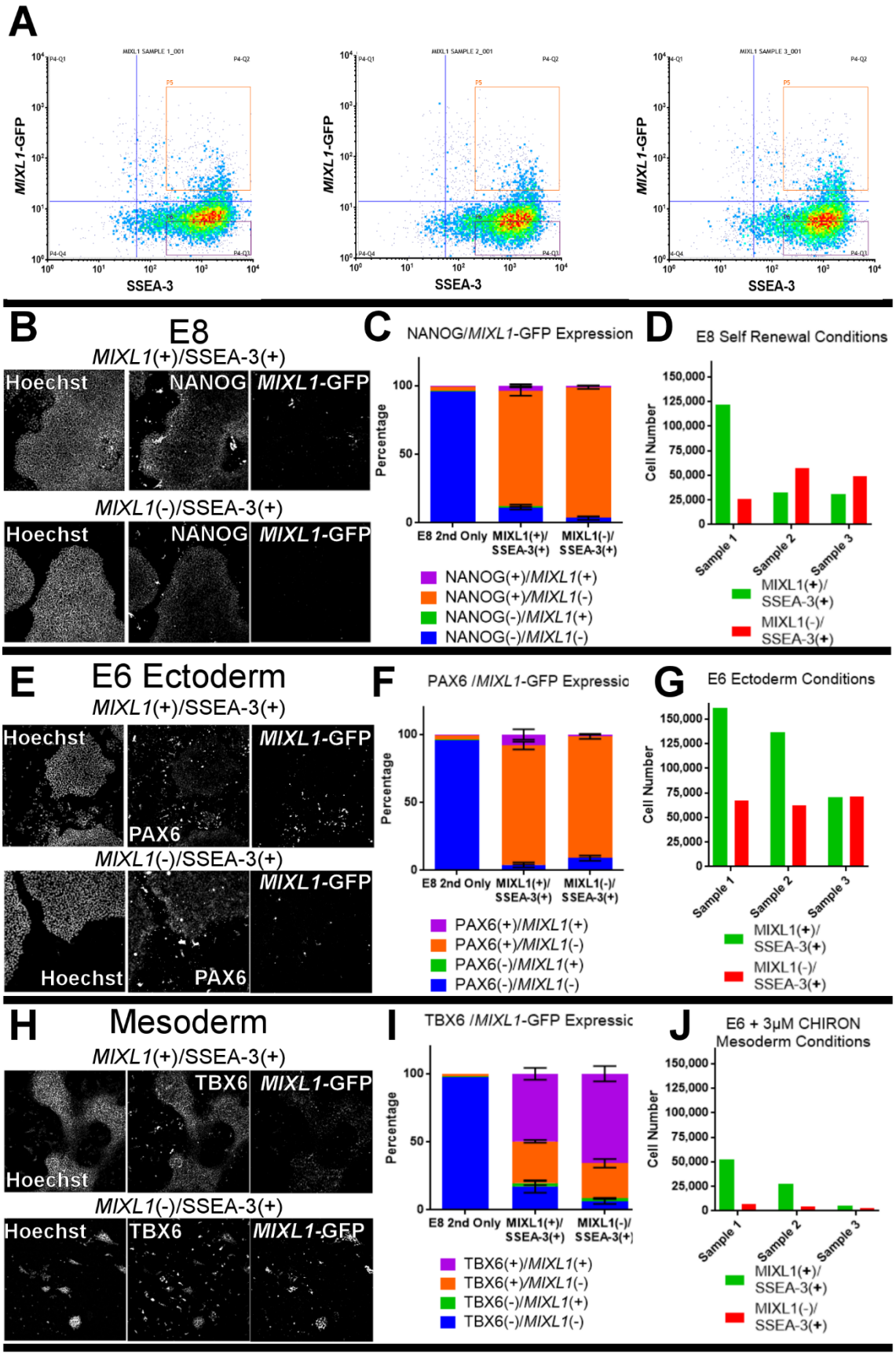

A
MIXL1-GFP
SSEA-3
B
E8
MIXL1(+)/SSEA-3(+)
Hoechst
NANOG
MIXL1-GFP
MIXL1(-)/SSEA-3(+)
C
NANOG/MIXL1-GFP Expression
Percentage
E8 2nd Only
MIXL1(+)/ SSEA-3(+)
MIXL1(-)/ SSEA-3(+)
NANOG(+)/MIXL1(+)
NANOG(+)/MIXL1(-)
NANOG(-)/MIXL1(+)
NANOG(-)/MIXL1(-)
D
E8 Self Renewal Conditions
Cell Number
Sample 1
Sample 2
Sample 3
MIXL1(+)/ SSEA-3(+)
MIXL1(-)/ SSEA-3(+)
E
E6 Ectoderm
MIXL1(+)/SSEA-3(+)
PAX6
MIXL1(-)/SSEA-3(+)
F
PAX6 /MIXL1-GFP Expressic
PAX6(+)/MIXL1(+)
PAX6(+)/MIXL1(-)
PAX6(-)/MIXL1(+)
PAX6(-)/MIXL1(-)
G
E6 Ectoderm Conditions
H
Mesoderm
MIXL1(+)/SSEA-3(+)
TBX6
MIXL1(-)/SSEA-3(+)
I
TBX6 /MIXL1-GFP Expressic
TBX6(+)/MIXL1(+)
TBX6(+)/MIXL1(-)
TBX6(-)/MIXL1(+)
TBX6(-)/MIXL1(-)
J
E6 + 3μM CHIRON Mesoderm Conditions
K
Secondary Only for (B) and (E)
Cy5
MERGED
L
Secondary Only for (H)
TexasRed

Figure 3-4 Bulk Plating of *MIXL1* / SSEA-3 substates

A) Flow cytometry density plots of *MIXL1*-GFP versus SSEA-3 of three samples from cells grown in MEF/KOSR conditions. Red gate indicates sorting gate for *MIXL1*(**+**)/SSEA-3(**+**) and blue gate indicates sorting gate for *MIXL1*(-)/SSEA-3(**+**) cells. **B, E, H)** Immunofluorescence images of cells grown under the three conditions E8 **(B),** E6 ectoderm **(E)** and mesoderm **(H)** from *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(-)/SSEA-3(**+**) fractions. **C, F, I)** Stacked Bar charts of quantified percentages from immunofluorescence analysis, percentage of cells positive for *MIXL1*-GFP and stained marker, NANOG **(C),** PAX6 **(F)** and TBX6 **(I). D, G, J)** Stacked Bar charts of total cell numbers from immunofluorescence analysis under the three conditions E8 **(D),** E6 ectoderm **(G)** and mesoderm **(J),** *MIXL1*(**+**)/SSEA-3(**+**) in green and *MIXL1*(-)/SSEA-3(**+**) in red (Bars are mean ± SD). **K, L)** Secondary only staining control, **(K)** control for **B)** and **E)** and **(L)** control for **H).**

ectoderm and mesoderm conditions, and sample 1 performing exceptionally well in all conditions. Variability between samples appears far less pronounced in the *MIXL1*(-)/SSEA-3(**+**) fractions.

MIXL1(-)/SSEA-3(**+**) fraction had good cell growth under self renewal conditions and even more so in ectoderm conditions, but particularly poor cell numbers in the mesoderm conditions (Figure 3-4 D, G, J). By comparison *MIXL1*(**+**)/SSEA-3(**+**) fractions had increased cell numbers in most conditions, except for samples 2 and 3 under self renewal conditions which had less cells than the *MIXL1*(-)/SSEA-3(**+**) fraction. The *MIXL1*(**+**)/SSEA-3(**+**) fraction seemingly performed better, with respect to surviving cell numbers when assessed in terms of differentiation, both to ectoderm and mesoderm, but particularly pronounced in mesoderm conditions (Figure 3-4 J). For sample one the *MIXL1*(**+**)/SSEA-3(**+**) fraction had over 50,000 cells in the analysed images, in contrast the *MIXL1*(-)/SSEA-3(**+**) fraction had less than 10,000 cells.

The mesoderm induction condition had the poorest cell survival of all the conditions but it is also the harshest differentiation approach. The self-renewal condition differs from MEF/KOSR conditions which the cells are sorted from but obviously is designed to support human PSC cell growth. The ectoderm condition induces differentiation simply by a removal of pluripotency signalling molecules FGF and TGFβ. Whereas mesoderm induction removes these factors whilst also driving differentiation using GSK3β inhibition.

The cell number difference between samples could be inherent to the variation within the population but could also be the result of a more technical issue. The *MIXL1*(**+**)/SSEA-3(**+**) population is a relatively small population, so sorting and plating in excess of 30,000 cells for each condition can be time consuming. The amount of time cells are left in a single cell suspension and also the physical strains of sorting could decrease the viability of the cells post sort.

3.2.4. Clonogenic Assessment of *MIXL1*-GFP expressing Cells

On a population basis there was a difference in the survival of *MIXL1*(+)/SSEA-3(+) cells when tested to differentiate, compared to *MIXL1*(-)/SSEA-3(+) cells. Since the effect was most pronounced in mesoderm induction conditions it might imply a bias towards this lineage. Similar to the experiments of Allison et al, 2018, to assess the bias at a single cell level we performed a clonogenic assay of the *MIXL1*(+)/SSEA-3(+) and *MIXL1*(-)/SSEA-3(+) sorted fractions. Unlike the previous bulk experiment, this involved plating the cells back into MEF/KOSR conditions as feeder-free systems do not support efficient clonogenic growth. I then screened the resulting colonies for expression of a pluripotency marker, OCT4 and a mesoderm marker, BRACHYURY (BRA) by immunocytochemistry (Figure 3-5A). Colonies could be split into 4 groups, depending on positive or `negative expression of OCT4 and BRACHYURY.

There were significant differences in the colonies derived from the *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(-)/SSEA-3(**+**) sorted fractions. The *MIXL1*(**+**)/SSEA-3(**+**) fraction exhibited higher abundance of OCT4**+**/BRA**+** colonies than OCT4**+**/BRA- colonies, whereas this relationship was inverse in the *MIXL1*(-)/SSEA-3(**+**) fraction with OCT4**+**/BRA- colonies being the most abundant (Figure 3 5C).

I also assessed at the expression of *MIXL1*-GFP in the resulting colonies and coupled that with the in situ staining data of OCT4 and BRACHYURY. From this, the colonies were further segregated into 8 categories based on positive or negative OCT4/Brachyury/*MIXL1* expression. Both the *MIXL1*(**+**)/SSEA-3(**+**) and the *MIXL1*(-)/SSEA-3(**+**) substates demonstrated heterogeneity in respect to expression of these markers. However, for the *MIXL1*(-)/SSEA-3(**+**) substate, the most abundant colony were ones that expressed OCT4, but not *MIXL1* or BRACHYURY (Figure 3 5B). Whilst the presence of colonies that expressed all three markers was much lower. In contrast, the *MIXL1*(**+**)/SSEA-3(**+**) substate exhibited an opposite distribution with the largest proportion of colonies being positive for all three markers with fewer OCT4 only colonies being present (Figure 3 5B).

The percentage of BRACHYURY positive cells per colony demonstrated a shift in the distribution between the two substates. The *MIXL1*(**+**)/SSEA-3(**+**) fraction had an increased number of colonies with high percentages of BRACHYURY expression. Although the *MIXL1*(-)/SSEA-3(**+**) did have some colonies with relatively high BRACHYURY expression, the number of those colonies was less than the *MIXL1*(**+**)/SSEA-3(**+**) (Figure 3 5D).

Figure 3 5E shows the percentage of BRACHYURY or *MIXL1* positive cells from the total number of cells within colonies identified. The *MIXL1*(+)/SSEA-3(+) substate has a higher percentage of both BRACHYURY and *MIXL1*-GFP positive cells compared to the *MIXL1*(+)/SSEA-3(+) substate.

To assess whether the differences in colonies might be a consequence of differences in cloning efficiency between the substates, I counted the number of OCT4 positive colonies in each well. The cloning efficiency between the two substates was comparable (Figure 3-5F).

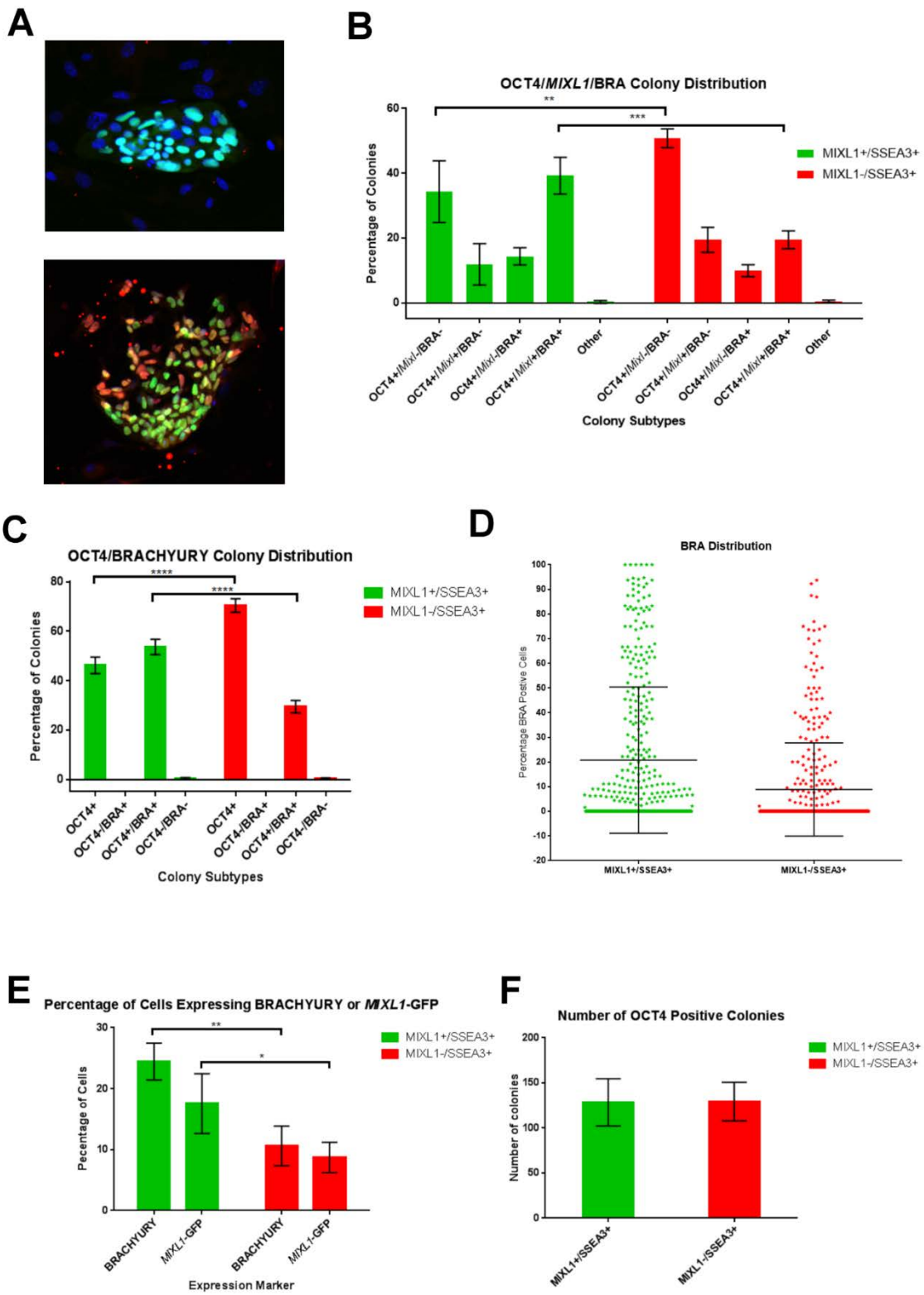


Figure 3-5 Clonogenic Lineage Bias Assessment

Single cells from both *MIXL1*(**+**)/SSEA-3(**+**)(green bars) and *MIXL1*(-)/SSEA-3(**+**)(red bars) were plated at a density of 500 cells per cm^2 in self-renewal conditions and after 5 days the resulting colonies were stained for OCT4 and BRACHYURY(BRA). **A)** Immunofluorescent images of OCT4(green) only colony and OCT4 colony expressing BRACHYURY(red), Hoechst (blue). **B)** Incorporates the expression of *MIXL1*-GFP and shows the percentage of colonies that were identified as either OCT4**+**/*MIXL1*-/BRA-, OCT4**+**/*MIXL1***+**/BRA-, OCT4**+**/*MIXL1*-/BRA**+**, OCT4**+**/*MIXL1***+**/BRA**+**, OCT4- /*MIXL1***+**/BRA-, OCT4-/*MIXL1*-/BRA**+**, OCT4-/*MIXL1***+**/BRA**+** or OCT4-/*MIXL1*-/BRA-. **C)** The percentage of colonies that were identified as either OCT4**+**/BRA-, OCT4**+**/BRA**+**, OCT4-/BRA- or OCT4/BRA**+** (Bars are mean ± SD). **D)** Percentage of BRACHYURY positive cells in OCT4 positive colonies for each fraction. **E)** Percentage of total number of cells within colonies identified as BRACHYURY or *MIXL1*-GFP positive. **F)** Number of OCT4 positive colonies identified for each fraction tested. N= three biological repeats.

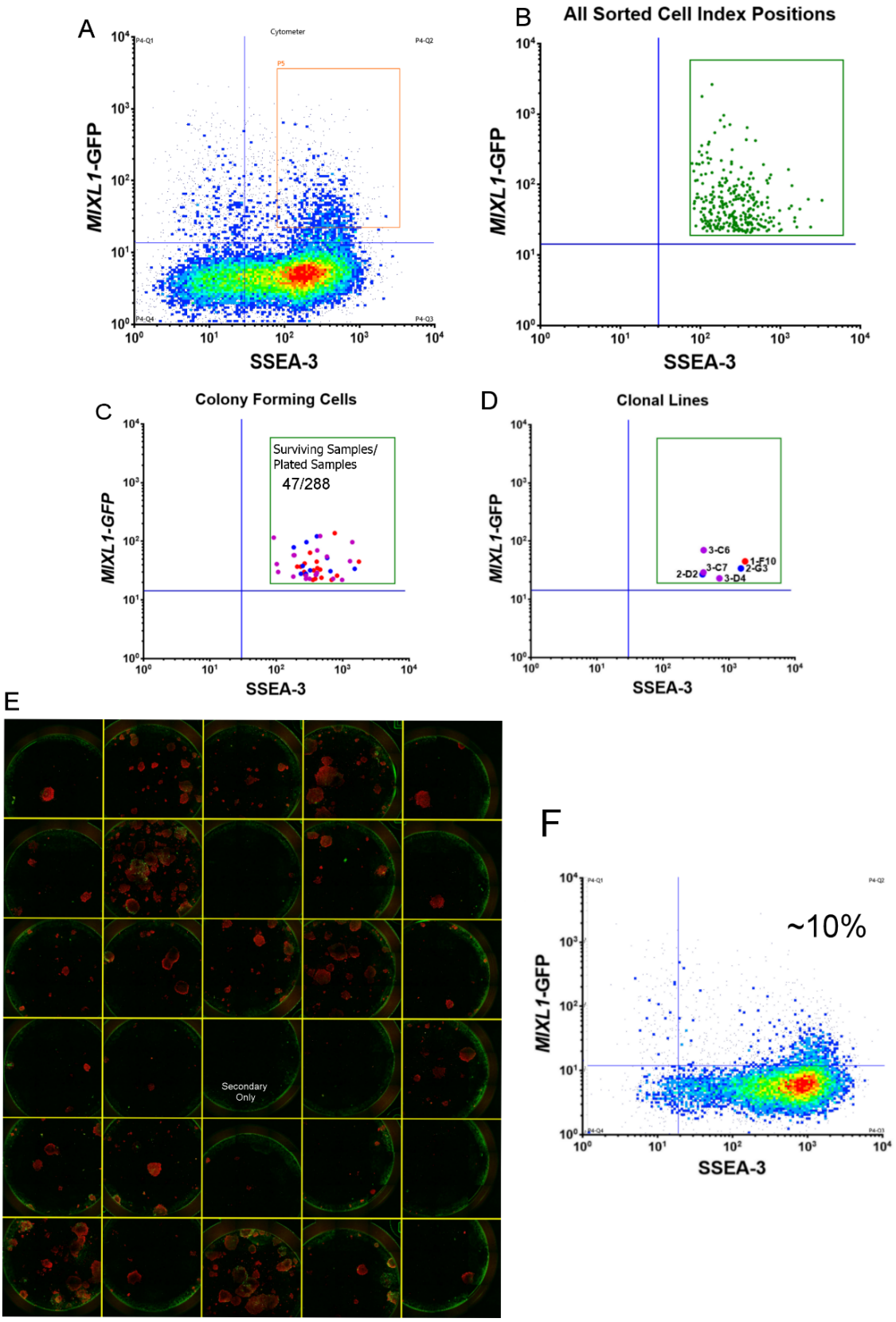

A
Cytometer
P4-Q1
P4-Q2
P5
P4-Q4
P4-Q3
MIXL1-GFP
SSEA-3
B
All Sorted Cell Index Positions
MIXL1-GFP
SSEA-3
C
Colony Forming Cells
Surviving Samples/
Plated Samples
47/288
MIXL1-GFP
SSEA-3
D
Clonal Lines
3-C6
1-F10
3-C7
2-G3
2-D2
3-D4
MIXL1-GFP
SSEA-3
E
Secondary
Only
F
~10%
MIXL1-GFP
SSEA-3

Figure 3-6 Single Cell cloning of *MIXL1*/SSEA-3 positive cells from HES3 *MIXL1*-GFP

Single cells from MIXL1(+)/SSEA-3(+) fraction were sorted into single wells of 96 well plate. After ~10 days wells containing stem cell like colonies were passaged further. After positive staining for TRA-1-81, six colonies were chosen at random to establish clonal lines, these lines exhibit similar MIXL1/SSEA-3 expression as the starting population. **A)** Density dot plot of *MIXL1*-GFP expression against the surface marker SSEA-3 expression. Region 5 (P5) gate was used to specify *MIXL1*/SSEA-3 double positive fraction and cells from this region were sorted into three 96 well plates containing a MEF layer and DKAM medium, containing 10µM Rock inhibitor. **B)** *MIXL1*/SSEA-3 scatter plot shows the indexed positions of all sorted *MIXL1*(**+**)/SSEA-3(**+**) cells, coloured by the plate with which they relate (plate one = red, plate two = blue and plate three = purple). **C)** *MIXL1*/SSEA-3 scatter plot shows the indexed positions of the 47 cells which formed colonies. **D)** *MIXL1*/SSEA-3 scatter plot shows the indexed positions of the 6 cells that formed established clonal lines. **E)** Live TRA-1-81 staining fluorescent images of colonies derived after the first passage into a 48 well plate at 4x, TRA-1-81(RED) and *MIXL1*-GFP (GREEN). **F)** Density dot plot of *MIXL1*-GFP expression against the surface marker SSEA-3 expression from clonal line HES3 *MIXL1*-GFP 3-C6. Clonal lines re-establish similar *MIXL1*-GFP/SSEA-3 distribution as the starting populations.

3.2.5. Cloning of *MIXL1*-GFP Positive Cells

The cells from *MIXL1*(**+**)/SSEA-3(**+**) fraction appeared to form stem cell colonies, indicated by the generation of OCT4 positive colonies. Whether the colonies could grow beyond 5 days and maintain stem cell properties was unknown. To assess whether the population contained a stem cell population capable of stable self renewal I produced clonal lines from the *MIXL1*(**+**)/SSEA-3(**+**) population. I performed indexed single cell sorting into 96 well plates with feeders, Figure 3-6 chronicles this process. I sorted 288 *MIXL1*(**+**)/SSEA-3(**+**) single cells into 96 well plates (Figure 3-6B) and from these stem cell like colonies were obtained in 47 wells, indexed data reveals the *MIXL1*-GFP and SSEA-3 expression levels from the initial sorted cells (Figure 3-6C). From these, 44 colonies were passaged further of which 29 of those survived. At this point I performed live staining for another pluripotency-associatedsurface marker TRA-1-81 (Andrews et al., 1984a; Wright et al., 2011), all surviving clones were positive for TRA-1-81 staining (Figure 3-6E). Further to this, I randomly selected six clones to expand and generate clonal lines, their initial indexed MIXL1/SSEA-3 expression positions are shown (Figure 3-6D).

When assessing the *MIXL1*-GFP and SSEA-3 expression of the established cell lines there is a recapitulation of the expression profile of the original culture. Rather than retaining the *MIXL1*

positive status of the starting cell, cultures exhibited normal *MIXL1*-GFP levels seen in MEF/KOSR conditions. The distribution of SSEA-3 also displayed similar distribution to the starting culture (Figure 3-6F). When transferred to E8V conditions the *MIXL1*-GFP proportion decreased to none (data not shown).

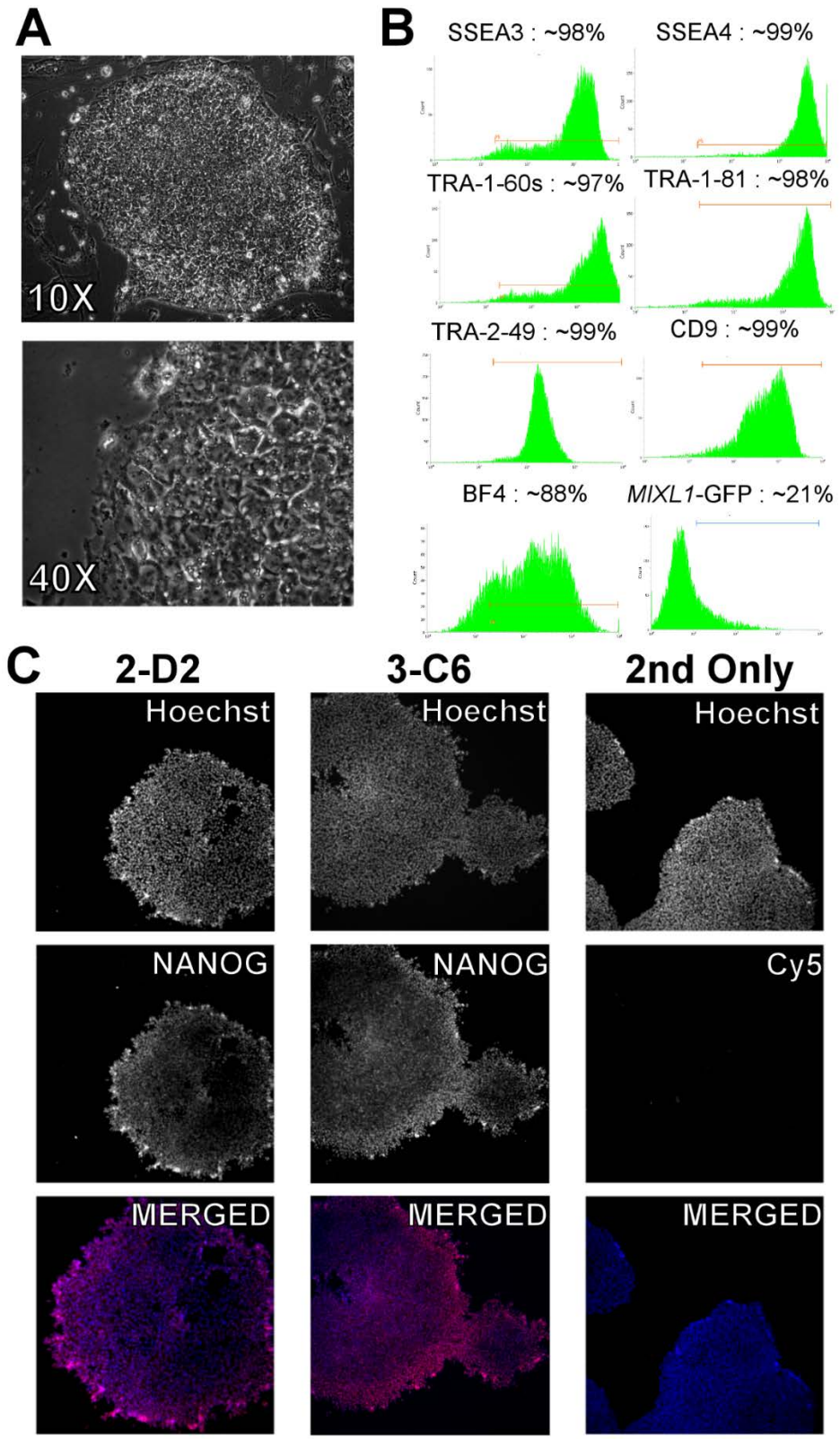


Figure 3-7 Characteristation of Clonal Lines in Self Renewal Conditions

Clonal lines established from MIXL1(+)/SSEA-3(+) single cells were assessed by morphology, surface antigen expression and NANOG expression, all aspects were consistent observations seen with standard human PSC. **A)** Phase contrast images of HES3 *MIXL1*-GFP clone 2-D2 growing in MEF/KOSR conditions with taken at 10x and 40x magnification. The line displays normal human PSC cell and colony morphology. **B)** Flow Cytometry Histograms of surface antigen and reporter expression from clonal line HES3 *MIXL1*-GFP 3-C6. Stem Cell associated antigens SSEA-3, SSEA4, TRA-1-60s, TRA-1-81, TRA-2-49, CH8 and BF4 were highly positive. *MIXL1*-GFP expression returned to a low percentage. **C)** Immunofluorescent analysis of NANOG expression in HES3 *MIXL1*-GFP clones 2-D2 and 3-C6 growing in E8V conditions. Merged images display Hoechst (Nuclei) in blue and NANOG positive cells in red.

3.2.6. Characterisation of clonal lines

The clonal lines established from *MIXL1*(**+**)/SSEA-3(**+**) cells displayed growth and SSEA-3 expression consistent with human PSC, but had to be further assessed since SSEA-3 is associated with pluripotency but not a definitive marker of pluripotency as it appears in other tissues (Vega Crespo et al., 2012). Clonal lines exhibited normal human PSC cell morphology with a high nuclear to cytoplasmic ratio and colony morphology growing as round colonies (Figure 3-7A). When assessed with a panel of pluripotency-associated surface markers (BF4, CD9, SSEA-3, SSEA-4, TRA-1-60s, TRA-1-81 and TRA-2-49), lines displayed high levels of expression for all markers (Figure 3-7B). The expression of *MIXL1*-GFP returned to levels seen in standard culture conditions. Intracellular staining for the pluripotency-associated marker NANOG, reveals high expression in the colonies of these clonal lines (Figure 3-7C).

The ability to generate all three lineages, by definition, is an inherent property of pluripotent stem cells. I assessed whether clonal lines could generate cells expressing markers corresponding to each germ layer, ectoderm, endoderm and mesoderm. Our bulk sorts of the *MIXL1*(**+**)/SSEA-3(**+**) fraction indicated the cells could generate ectoderm (PAX6 positive) and mesoderm (TBX6 positive) cells. Nevertheless, it was important to validate that clonal lines expanded from the population retained this differentiation potential. Clonal lines were assessed by the directed differentiation protocols I used previously designed for ectoderm, endoderm and mesoderm. Each line was able to generate cells positive for the mesoderm marker TBX6 (Figure 3-8BA), the ectoderm marker PAX6 (Figure 3-8B) and the endoderm marker SOX17 (Figure 3-8C), in the relevant differentiation condition. *MIXL1*-GFP expression was detected abundantly in the mesoderm and endoderm forming conditions, but not present under ectoderm conditions mirroring its expression *in vivo* (Figure 3-8B).

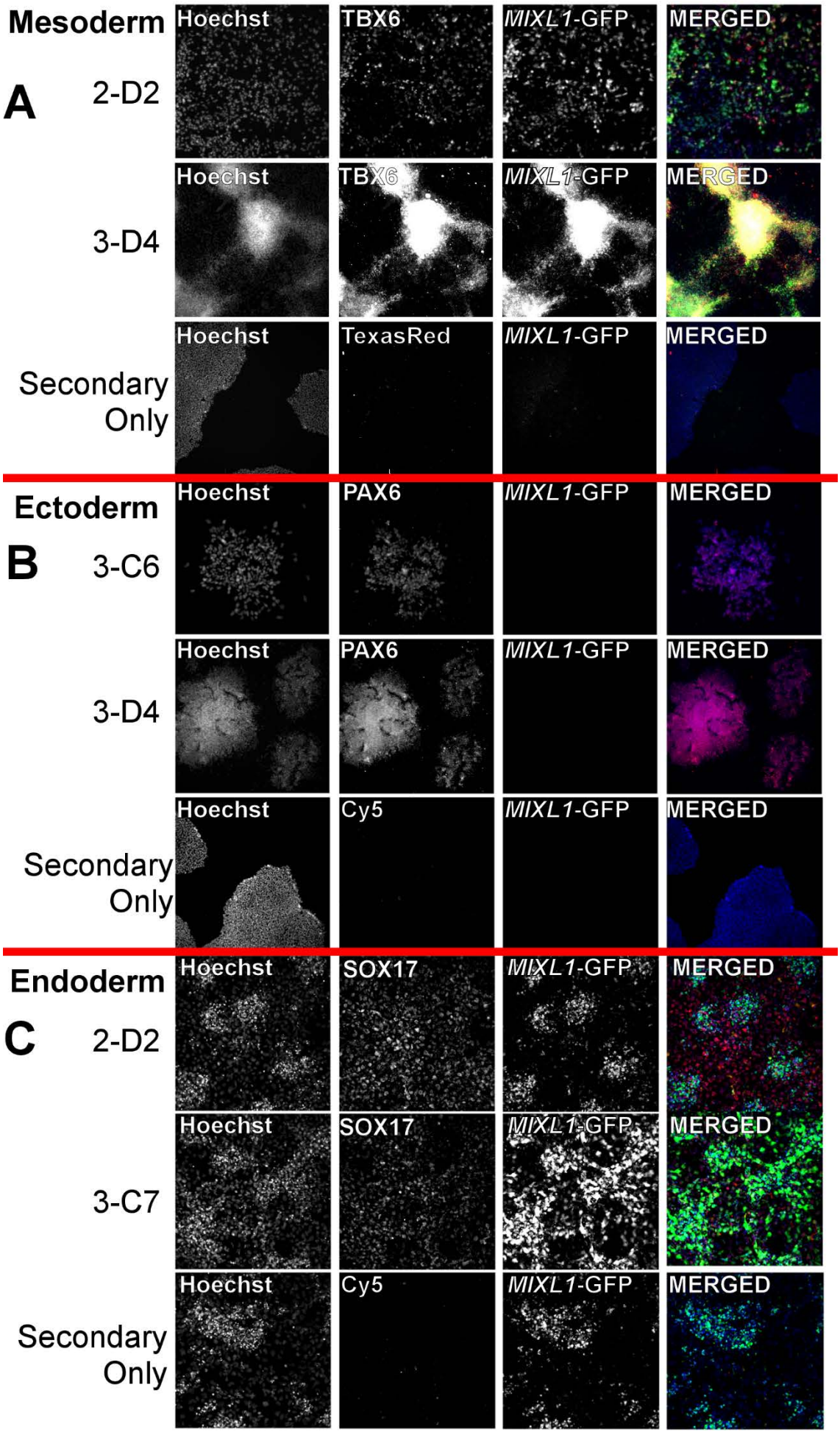

Mesoderm
A
2-D2
Hoechst
TBX6
MIXL1-GFP
MERGED
3-D4
Hoechst
TBX6
MIXL1-GFP
MERGED
Secondary Only
Hoechst
TexasRed
MIXL1-GFP
MERGED
Ectoderm
B
3-C6
Hoechst
PAX6
MIXL1-GFP
MERGED
3-D4
Hoechst
PAX6
MIXL1-GFP
MERGED
Secondary Only
Hoechst
Cy5
MIXL1-GFP
MERGED
Endoderm
C
2-D2
Hoechst
SOX17
MIXL1-GFP
MERGED
3-C7
Hoechst
SOX17
MIXL1-GFP
MERGED
Secondary Only
Hoechst
Cy5
MIXL1-GFP
MERGED

Figure 3-8 Clonal Lines Generated from *MIXL1*(+)/SSEA-3(+) Cells form cells corresponding to three germ layers

Clonal lines established from *MIXL1*(**+**)/SSEA-3(**+**) single cells were plated and induced to differentiate, into Mesoderm (Panel **A)**, Ectoderm (Panel **B)** and Endoderm (Panel **C)**. 4 Clonal lines were assessed for each condition, 2-D2, 3-C6, 3-C7 and 3-D4, each panel display representative immunofluorescent images for two clones. **A-C)** Immunofluorescence analysis of Hoechst and *MIXL1*-GFP, coupled with either TBX6 **(A),** PAX6**(B)** or SOX17**(C)** expression. Merged image colours, Hoechst (Blue), *MIXL1*-GFP(Green), TBX6 (Yellow) and PAX6 or SOX17(Red).

3.3. Discussion

The conditions in which human PSC are cultured can vary and the effect that has on particular substates of human PSC has not been fully evaluated. I have shown that even in standard human PSC conditions the presence of particular substates can be altered. When moving into defined systems such as E8/Vitronectin we can observe changes based on antigen expression. In terms of SSEA-3 expression I get a consistently positive fraction but a loss in the very high SSEA-3 expressing cells. SSEA-3 expression has been used to identify human PSC substates with increased cloning efficiency (Tonge et al., 2011) and what implications the loss of the high expressing fraction is yet to be investigated.

Heterogeneity is a known and apparent feature of human PSC culture systems and the expression level of antigens, such as SSEA-3, has been used to assess this heterogeneity (Tonge et al., 2011). Functional attributes can be related to SSEA-3 expression levels on human PSC with high expressing cells having increased clonogenic potential, while low expressing cell have a higher propensity to differentiate. In standard culture, human PSC occupy SSEA-3 expression states ranging from low to very high, this can also be seen in colonies. Interestingly, this may have a correlation with differentiation efficiency of cells within a colony. It has previously been observed that cells at the periphery of the colony appear to have a greater propensity to differentiate towards a mesendoderm state than the cells in the centre of the colony (Rosowski et al., 2015 (Rosowski et al., 2015). This low expression of SSEA-3 may indicate a substate of human PSC with a higher propensity to differentiate and thus why differentiation of the colony does not take place in a uniformed fashion.

Human PSC form a heterogenous population in standard culture and it has been previously shown that sub-populations of cells grown in MEF/KOSR conditions express mesendodermal genes (Allison et al., 2018a; Gokhale et al., 2015; Hough et al., 2014). Utilising the HES3 *MIXL1*-GFP reporter, created by Davis et al, 2008, coupled with the surface marker SSEA-3 I assessed the presence of *MIXL1*(**+**)/SSEA-3(**+**) cells in standard cultures. While variable in proportion I found that human PSC grown in MEF/KOSR conditions often contained a *MIXL1*(**+**)/SSEA-3(**+**) population. Moving to the defined feeder-free system there was a loss of *MIXL1*-GFP expression and cells were almost completely *MIXL1*(-)/SSEA-3(**+**).

The presence of MEFs or the signalling molecules they secrete appears to be crucial for the formation of the *MIXL1*(**+**)/SSEA-3(**+**) substate. BMP4 has been revealed to be one of the most abundant factor secreted from MEFs (Qi et al., 2004). BMP4 has been used to initiate differentiation during mesoderm induction of human PSC (Jackson et al., 2010). BMP4 drives differentiation towards mesoderm in part, by upregulation of secreted WNT ligands (Kurek et al., 2015). It could be that this low dose of BMP4 is able to initiate some low level of expression

of differentiation genes but insufficient to overcome the TGFβ and FGF2 signalling established pluripotency network. Equally other factors present in KOSR medium have been shown to help maintain pluripotency, particularly the lipids bound to albumin (Garcia-Gonzalo and Izpisúa Belmonte, 2008).

MIXL1 has been identified as an integral part of the gene expression changes during mesoderm differentiation (Loh et al., 2016). The temporal expression of *MIXL1* being so early in differentiation indicates a period of time in which pluripotent and differentiation gene networks are coexpressed. SSEA-3 is often used as a marker of pluripotency, and one of the most sensitive to differentiation, but their relationship is not absolute and differentiating human PSC can still express SSEA-3 (Draper et al., 2002; Ramirez et al., 2011). It has also been shown that some SSEA-3 negative cells, albeit from genetically abnormal lines, retain stem cell self renewal abilities (Enver et al., 2005). While the expression of the surface marker, SSEA-3 and fluorescent reporter for *MIXL1* might indicate the presence of both networks, whether cells in this double positive state retain stem cell functional properties had never been formally tested.

Single cell cloning of the *MIXL1*(**+**)/SSEA-3(**+**) cells provides the first evidence that this population retains stem cell functionality. Cells from this population form colonies that exhibit normal stem cell antigen expression and morphology. Their *MIXL1* expression is not retained however, as the cultures re-establish a *MIXL1*-GFP/SSEA-3 distribution similar to the starting population. This indicates there is a level of interconversion between these substates and the dynamics of these states. Further to this, the established clonal lines retain the ability to make all three germ layers under directed differentiation conditions.

I was able to conclude that *MIXL1*(**+**)/SSEA-3(**+**) population contained a stem cell population, but whether the expression of *MIXL1* had any functional consequences on the cells had to be assessed. Using a clonogenic assay we assessed the make up of the colonies arising from single cells. Although *MIXL1*(-)/SSEA-3(**+**) cells generated OCT4/BRACHYURY positive colonies in the assay, there was a significant proportion increase generated when *MIXL1*(**+**)/SSEA-3(**+**) cells were plated in the assay.

4. Chapter 4: "Trapping" a Mesoderm Bias human PSC substate.

4.1. Introduction

4.1.1. Signalling, and its effects on human Pluripotent stem cells

Coordinated signalling directs both the emergence, maintenance and subsequent differentiation of pluripotent cells. The signalling pathways can complement or antagonise each other and can be applied sequentially or simultaneously. PSC maintain their pluripotency by regulating the expression of the genes in a pluripotency network. These genes include *OCT4*, *SOX2* and *NANOG*, deemed the master regulators of pluripotency (Loh et al., 2011) The regulation of this network is controlled by environmental cues and cell signalling. The induction of differentiation down any lineage is controlled by signalling pathways. There can be distinct signalling pathways for different lineage as well as an overlap of signalling pathways in a concentration and timing dependant manner (Yu et al., 2011). Understanding the signalling pathways that control stem cell fates is crucial for any aspect of research using these cells, particularly when working on the manipulation of the pluripotent state in vitro.

LIF signalling regulates pluripotency in mouse ESC but it fails to maintain self-renewal of human ESC (Daheron et al., 2004) instead human ESC require FGF signalling (Thomson et al, 1998). FGF2 is one of eighteen proteins from the FGF family of secreted protein. Traditionally FGF growth factors were shown to be involved in processes such as angiogenesis and wound healing. FGF2s action on human ESC, required further interrogation to understand its action on maintaining pluripotency.

FGF2 exhibited stimulation of the mitogen-activate protein kinase (MAPK), mediated through binding to FGF receptor on the surface of humans ESC (Eiselleova et al., 2009). Activation of MAPK signalling pathway imparts a proliferative effect on human ESC. This effect is also exhibited for members of the insulin and epidermal growth factor family, which had no discernible effect on maintaining pluripotency. Thus, a secondary action of FGF binding was thought to maintain the stem cell.

PI3K/AKT and ERK signalling has exhibited activation via FGF signalling and chemical inhibition of this pathway leads to differentiation of human PSC (Li et al., 2007). This also highlights a species difference between mouse and human PSC, as mouse PSC could be propagated in the presence of an ERK inhibitor (Ying et al., 2008). Active PI3K/AKT signalling has been shown to promote SMAD2/3 binding to pluripotency targets, rather than interacting with β-catenin to promote differentiation targets including *MIXL1* (Singh et al., 2012).

TGFβ, Activin A and Nodal are all members of the TGFβ super family and have exhibited an ability to promote stem cell self renewal. Binding of these ligands to TGFβ receptors, ALK4, ALK5 and ALK7 activates signalling via SMAD2/3 proteins. James et al, 2005, first characterised the complex dynamics between TGFβ/Activin/Nodal signalling and SMAD proteins. Mechanistically, the signalling is activated by rapid phosphorylation of SMAD2/3, releasing it from the SARA (Small Anchor for Receptor Activation) complex (Tsukazaki et al., 1998). After release SMAD2/3, forms a trimeric complex with SMAD4. This complex then undergoes nuclear translocation where it can activate downstream targets (Nakao et al., 1997). Amongst the targeted genes was *NANOG*, a major regulator of the pluripotent state (Xu et al., 2008) (Figure 4-1). Although TGFβ is used often in human PSC culture media, both Activin A (Vallier et al., 2005; Xiao et al., 2006) and Nodal (Chen et al., 2011) have been used in its place to maintain pluripotency. Chemical inhibition of this pathway leads to downregulation of *NANOG* and *POU5F1* (OCT4) and subsequent differentiation (James et al., 2005).

Bone Morphogenic Proteins (BMP) are members of the TGFβ superfamily but exhibits contrasting effects on human PSC compared to the other family members. While in mouse PSC, BMP4 promotes self renewal (Qi et al., 2004), in human PSC it is a potent driver of differentiation (Xu et al., 2002a). BMP works in a similar to fashion to other members of the TGFβ superfamily however, after binding to surface receptors leads to the phosphorylation of SMAD1/5. SMAD1/5 also forms a complex with SMAD4, and after translocation into the nucleus binds to downstream targets to promote differentiation (Kurek et al., 2015; Teo et al., 2012) (Figure 4-1). In particular BMP4 signalling leads to a repression of the pluripotency gene network which is further compounded by the activation of endogenous *BMP2* and *BMP4* expression (Teo et al., 2012). The potent action that BMP4 exhibits on the pluripotency gene network has made it a common addition to differentiation protocols of human PSC. It has been demonstrated that in combination with geometric confinement ES colonies treated with BMP4 can generate cells related to all three germ layers (Warmflash et al, 2014).

To promote differentiation, BMP4 signalling activates expression of WNT ligands such as WNT3 which has been shown to be involved in the formation of the primitive streak (Kurek et al., 2015). Canonical WNT signalling, is characterised by the translocation of β-catenin to the nucleus. When WNT signalling is not present, β-catenin is sequestered to the β-catenin destruction complex in the cytoplasm. This complex consists, in part, of the tumor suppressors Axin, adenomatous polyposis coli (APC), Glycogen synthase kinase 3 beta (GSK3β), protein phosphatase 2A (PP2A), and other cofactors (Stamos and Weis, 2013). When WNT signalling is active, GSK3β is inhibited from phosphorylating β-catenin, subsequently β-catenin is released from the destruction complex and enters the nucleus (Huelsken and Behrens, 2002). In the

nucleus β-catenin cooperates with a handful of cofactors such as TCF-LEF, SMAD2/3 and SMAD4, to promote expression of gene targets. In mouse, this canonical WNT signalling is a requirement for the generation of mesoderm *in vivo* and *in vitro* (Lindsley et al., 2006).

Much like LIF and BMP4, WNT signalling also demonstrates a species difference between mouse and human PSC. Whereas WNT signalling and its surrogate, small molecule GSK3β inhibition, leads to differentiation in human PSC (Davidson et al., 2012), it stabilises pluripotency in mouse PSC (Ying et al., 2008). In combination with MEK/ERK inhibition, mouse PSC could be propagated in a 2i culture system also containing LIF. Many groups have tried to implement the same inhibitors and signalling molecules in other conditions to propagate "naïve" human PSC which resemble mouse PSC (Chan et al., 2013b; Gafni et al., 2013; Qin et al., 2016; Takashima et al., 2014; Theunissen et al., 2014). Demonstrating how signal manipulation can alter the state in which PSC reside.

The directed differentiation of mesoderm or endoderm can be particularly challenging because of their entwined intermediate stage, mesendoderm (Technau and Scholz, 2003). The addition of BMP4 alone is sufficient to trigger differentiation of PSC but towards a trophoblast lineage only (Golos et al., 2010; Xu et al., 2002b) addition of other signalling molecules is required for other germ layer lineage specification. FGF and TGFβ/Activin A/Nodal, signalling is crucial for the maintenance of pluripotency but also has implications in the differentiation of PSC. The FGF protein family has been shown to have diverse roles in development, with knock-out animal models for both ligand and receptors often resulting in embryonic lethality reviewed in (Dorey and Amaya, 2010). In the mouse, FGF receptor 1 or 2 deficient mice, display a lack of paraxial mesoderm and defects in visceral endoderm, respectively.

The actions of FGF and Activin A/Nodal on the patterning of mesoderm and endoderm during BMP4 induction of differentiation has been investigated in vitro. (Lowe et al., 2001; Vincent et al., 2003). Lowe et al, 2001 and D'Amour et al, 2005 demonstrated how differing concentrations of Nodal / Activin A concentrations controls the differentiation down endoderm or mesoderm. After BMP4 triggered human PSC differentiation in the presence of FGF, high concentrations of Activin A (100ng/ml) differentiates towards definitive endoderm where as a lower concentration (30ng/ml) differentiates towards mesoderm. Similarly, in human PSC, FGF2 has been shown to switch the outcome of BMP4 induced differentiation towards mesendoderm by sustaining *NANOG* expression (Yu et al., 2011), while WNT and TGFβ signalling demonstrate coordination in the development of mesendoderm (Kempf et al., 2016).

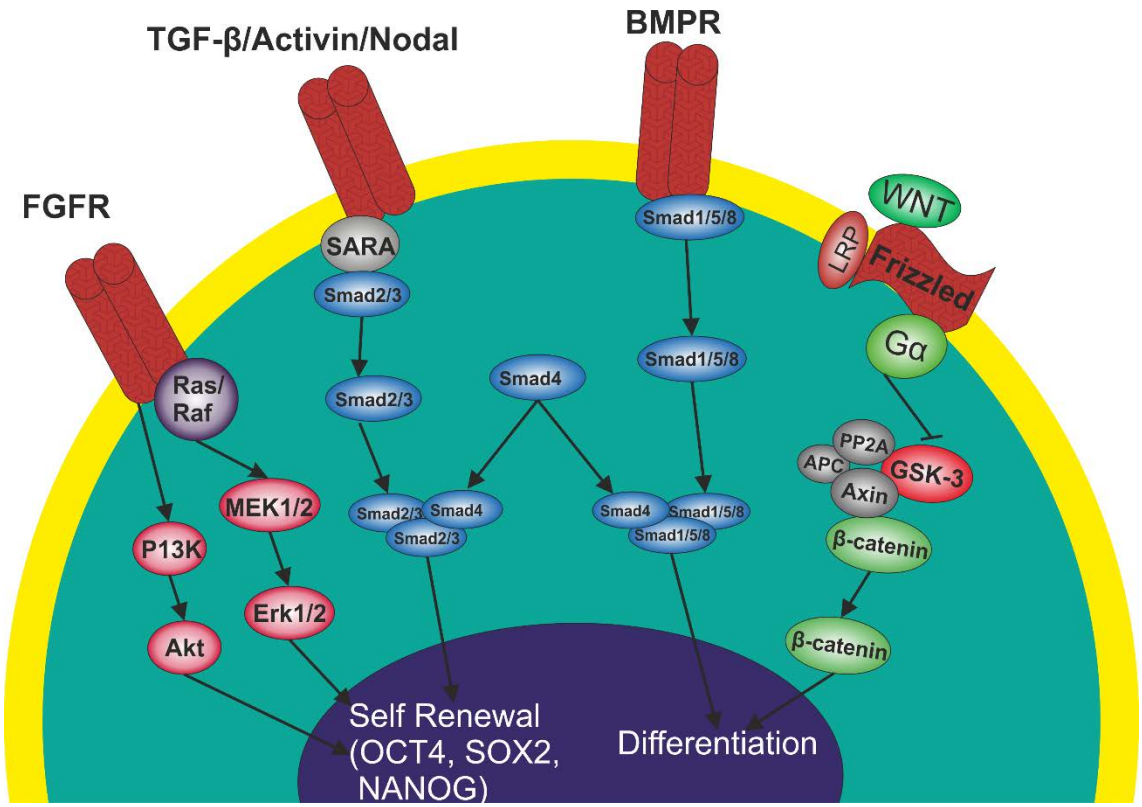


Figure 4-1 Signalling in Human Pluripotent Stem Cells

A schematic diagram of interactions in active signalling pathways in human Pluripotent Stem Cells and their effects on self renewal and differentiation. The FGF, TGF-β/Activin/Nodal, BMP and WNT signalling pathways are shown.

4.1.2. Changes in human PSC growth conditions overtime

Since their first derivation in 1998 (Thomson et al., 1998), many different media and matrices have been developed for the culture of human embryonic stem (ES) cells. In the early days of human ESC research cells were grown on a layer of mouse embryonic feeders (MEF) with media containing bovine serum supplemented with FGF. Serum was soon replaced by a commercial alternative, Knockout Serum Replacement (KOSR, Gibco) (Amit et al., 2000). While more consistent than serum, KOSR is proprietary and its exact formulation is unknown, although elements have been identified through research (Garcia-Gonzalo and Izpisúa Belmonte, 2008). Fully defined medias were established such as mTESR1 (Ludwig et al., 2006) and a further refined version E8 (Chen et al., 2011) which contains the minimal components for human PSC maintenance. The substrate which has been used to grow human PSC on has also developed over the years. The development of defined and minimalistic systems required a move away from a MEF layer, in part this was due to the need to remove the differentiation promoting signals secreted from MEFs such as BMP4 (Qi et al., 2004). MEF free matrices have been adopted

such as the mouse tumour extracellular matrix extracts, matrigel/geltrex and the xeno-free substrates based on recombinant proteins such as vitronectin and laminins.

Central to the development of the application of human PSC in regenerative medicine is an understanding of the genetic and molecular mechanisms that govern lineage specification. The induction of differentiation down any lineage is controlled by signalling pathways dependent upon extracellular factors acting in a concentration and timing dependant manner (Blauwkamp et al., 2012; Yu et al., 2011). Cell interactions with the substrate may also influence human PSC behaviour including differentiation efficiency (Canton et al., 2016; Narayanan et al., 2014; Singh and Schwarzbauer, 2012). However, elucidation of these mechanisms and signalling pathways is compromised when cells are grown in undefined media and substrates.

While the newer defined culture systems have allowed for efficient propagation of human PSC, the effects of these media/matrix combinations on differentiation of human PSC in culture has not been fully assessed. Human PSC exist as a heterogeneous population in culture with substates that may exhibit lineage bias upon induction of differentiation and can be identified by differential antigen and gene expression or by metabolism differences. Understanding the mechanisms that control fate determination during the differentiation of human PSC depends upon understanding the mechanisms that control the dynamics of substate within the undifferentiated stem cell compartment.

By culturing in different growth medium, the bias of human PSC towards particular lineages could be altered in another direction. In this work I saw a loss of our particular substate in defined conditions however, this does not rule out the presence of a substate bias in another direction such as ectoderm. High levels of FGF are present in defined media (Chen et al., 2011; Ludwig et al., 2006) and it has been shown that high levels of FGF promote the neural specification of human ESC (Cohen et al., 2010). Efficient neural induction from cells grown in MEF/KOSR conditions requires dual SMAD inhibition to block both BMP and TGFβ signalling (Chambers et al., 2009). However, from defined medium, E8, the simple removal of FGF and TGFβ results in efficient formation of neural epithelium (Lippmann et al., 2014). This highlights how human PSC growth medium can alter differentiation potential and reflects the loss of mesendoderm associated substates in defined conditions.

4.1.3. The use of small molecules and signalling molecules to cause or prevent differentiation

In addition to FGF other factors such as TGFβ, Nodal and Activin A were also shown to be beneficial to the maintenance of pluripotency (Chen et al., 2011; Xiao et al., 2006). Much like

the maintenance of pluripotency, human PSC differentiation is governed by signalling molecules, in particular BMP and WNT. (D'Amour et al., 2005; Davidson et al., 2012).

Along with traditional signalling molecules there has been an increase in small molecules to block or mimic signalling (Schugar et al., 2007). There are many benefits to using small molecules. They can offer effective targeting of elements of a signalling pathway and produce strong inhibition or activation of the desired pathway. They also often bypass the need for surface receptor ligand binding and impart their action internally, after uptake into the cells. Their specificity can be an issue though as off-target effects can responsible for the observed changes.

Chemical inhibitors have been used successful to block unwanted endogenous signals during differentiation. A notable example of this comes from the development of "dual-SMAD inhibition" neural induction. This approach uses Noggin (BMP antagonist) and SB431542 (small molecule inhibitor of TGFβ signalling) to efficiently induce neural formation of human PSC (Chambers et al., 2009). This has further been refined to replace recombinant Noggin protein with a selective BMP receptor inhibitor DMH1 (Neely et al., 2012).

In dual-SMAD inhibition protocols, DMH1 is used to completely block BMP signalling but it has also been used to regulate BMP signalling level. "Top Down Inhibition" utilises DMH1 to attenuate BMP signalling when the system has been saturated with BMP ligands, maintaining the desired level of BMP activation (Hackland et al., 2017). The system improved the reproducibility to generate neural crest cells from human PSC. The authors proposed a similar model to attenuate WNT signalling, termed "Baseline Activation", which would use a GSK3β inhibitor to mimic WNT signalling but also incorporate a recombinant WNT antagonist (DKK1) to prevent endogenous signalling (Hackland et al., 2017). Both "Top Down Inhibition" and "Baseline activation" use a combination of small molecules and recombinant proteins to control the active signalling present in the cultures.

WNT signalling works in part by inhibiting the function of GSK3β to allow the release of beta-catenin into the nucleus. CHIR99021 (CHIRON) is a highly selective GSK3β inhibitor that has been used to mimic WNT signalling in during human PSC differentiation (Giacomelli et al., 2017; Lian et al., 2014; Loh et al., 2016). Kim et al, 2013, demonstrated that by using an inhibitior of tankyerase, the enzyme responsible for AXIN degradation, they could stabilise AXIN2 and modulate the β-catenin function induced by CHIRON to maintain mouse EpiSC and human PSC in an undifferentiated state (Kim et al., 2013b).

While the classical signalling relationships are well established other factors can affect stem cell maintenance by promoting self-renewal or by repressing differentiation. The lipids bound to

albumin in KOSR were identified as important contributors to stem cell self-renewal, in particular lysophosphatidic acid (**LPA)** had a pronounced effect (Garcia-Gonzalo and Izpisúa Belmonte, 2008).

Blauwkamp et al, 2012 demonstrated how the addition of LPA could attenuate WNT mediated differentiation of human PSC. Treatment by either recombinant WNT proteins or by GSK3β inhibition, which normally induce differentiation could be ablated by addition of 10μM LPA. Further to this work, Qin et al, 2016 demonstrated that LPA maintained pluripotency by the upregulation of YAP expression and proposed an inhibitory effect on beta-catenin. This indicates that LPA plays a role in blocking the differentiation of human PSC.

The maintenance of stem cell self-renewal and blocking of differentiation has been well established for many of these factors. Assessment of cross antagonism of self-renewal and differentiation signals have been less successful and less investigated. The work of Blauwkamp et al, 2012 struggled with a chemically defined system to test their LPA addition in, resulting in poor growth in even the control conditions. Since that time a robust chemically defined human PSC medium, Essential 8 (E8), has been developed (Chen et al., 2011). This medium now provides a platform to accurately assess the effects of specific component addition or combinations.

Utilising this platform, and the knowledge of the signalling pathways which act on human PSC, we wanted to investigate whether combinations of self-renewal and differentiation signals can induce changes in gene expression and expand a particular human PSC substates, for our work a state similar to that of the *MIXL1*(+)/SSEA-3(+) state described in chapter 3. Beyond expansion we wanted to see whether this substate could be trapped, by trapped we mean that the state is consistently maintained over multiple passages. Ultimately, we want a state that exhibits a differentiation bias towards mesoderm but a key aspect of the state would also need to be an ability to revert to an unbiased pluripotent state.

4.2. Results

4.2.1. Endogenous WNT secretion is necessary for the presence of *MIXL1* positive cells in standard culture.

The presence of *MIXL1* positive cells is often variable in MEF/KOSR conditions. I sought to assess the importance of endogenous WNT secretion in the presence of the *MIXL1* positive substates within MEF/KOSR conditions. The porcupine inhibitor, IWP-2 has been used in culture and directed differentiation of human PSC previously (Blauwkamp et al., 2012; Kurek et al., 2015). Through its action on porcupine, IWP-2 inhibits WNT ligand secretion from cells (Chen et al., 2009a). The presence of IWP-2 in culture or during differentiation exhibits a significant decrease in gene expression consistent with mesendoderm formation (Blauwkamp et al., 2012; Kurek et al., 2015). I utilised two of our clonal lines 2-D2 and 3-C6, derived in chapter 3. Cells were passaged onto MEF coated flasks into two sister flasks and fed with standard KOSR medium. IWP-2 was then added at 1μM to one flask, after three days cells were stained for SSEA-3 and analysed by flow cytometry.

While the cells from both flasks maintained high SSEA-3 levels, flasks treated with IWP-2 exhibit a stark reduction of *MIXL1*-GFP expression (Figure 4-2). This implies that endogenous WNT secretion is important for the formation of the *MIXL1* positive substates, which is consistent with previous studies blocking WNT secretion (Blauwkamp et al., 2012; Kurek et al., 2015).

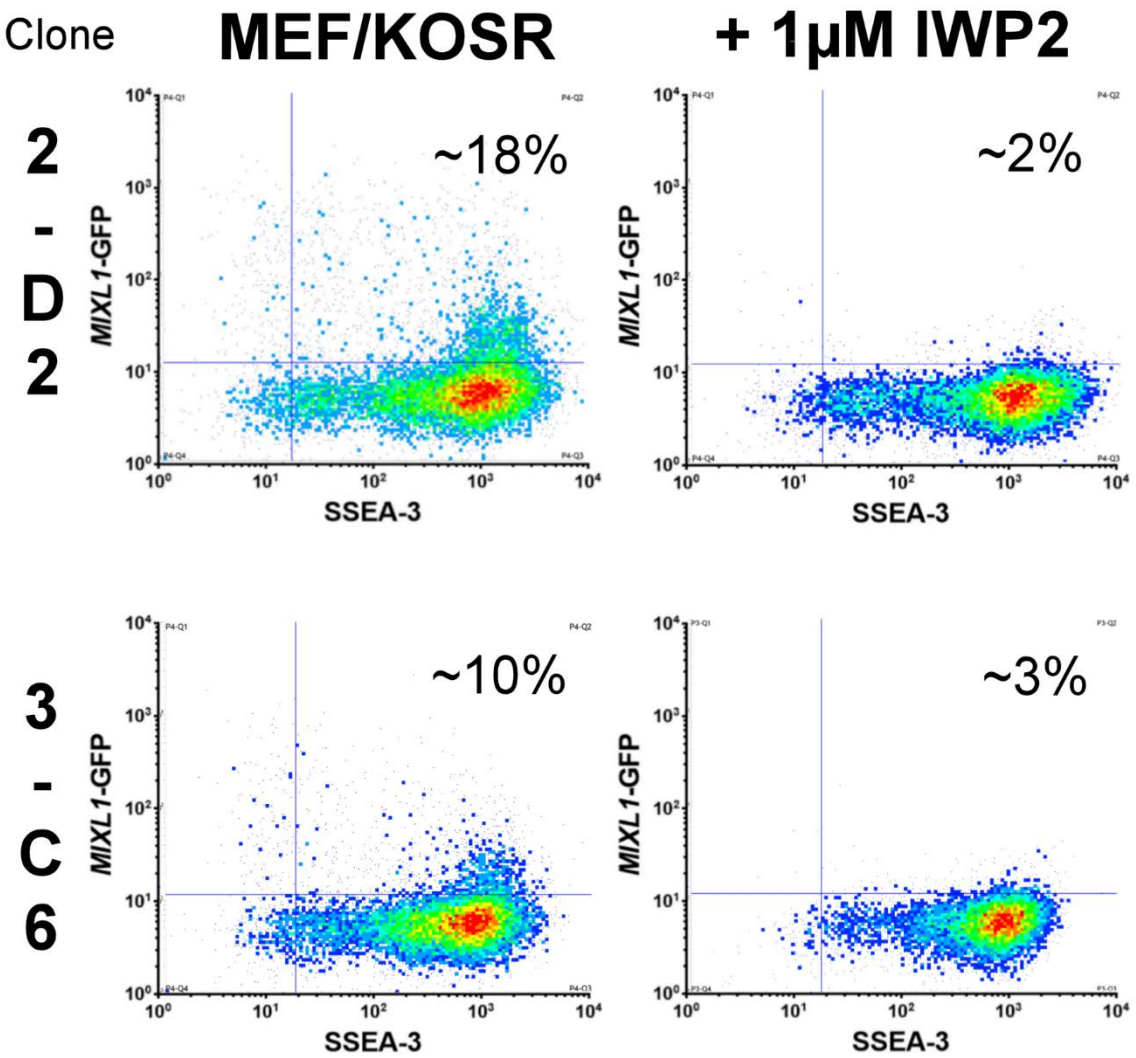


Figure 4-2 Inhibiting WNT Secretion in ES Culture

The addition of IWP-2 into MEF/KOSR conditions decreases the proportion of MIXL1(+)/SSEA-3(+) cells. Flow cytometry density plots of clonal lines 2-D2 and 3-C6 under MEF/KOSR conditions with and without the addition of 1µM IWP-2. X axis displays SSEA-3 expression level, Y Axis *MIXL1*-GFP expression. The percentage of *MIXL1*(**+**)/SSEA-3(**+**) is shown on each density plot.

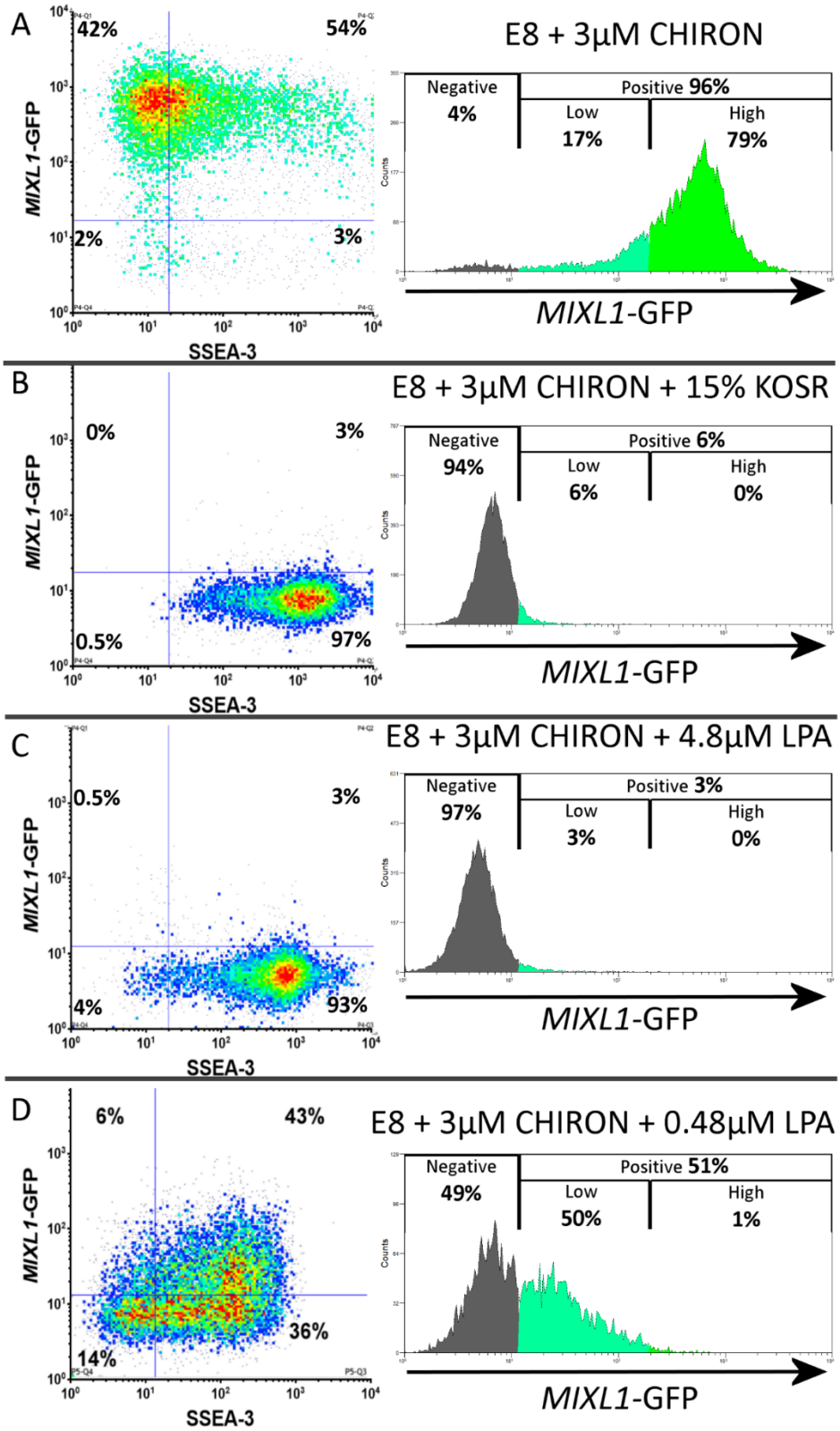


Figure 4-3 *MIXL1*-GFP and SSEA-3 Expression Changes under GSK3β inhibition and counteraction by KOSR and LPA

Each panel A-D contains flow cytometry analysis for HES3 *MIXL1*-GFP growing under different conditions. Flow cytometry density plots of SSEA-3 versus *MIXL1*-GFP and histograms of *MIXL1*-GFP intensity are displayed. **A)** Cells grown in E8 with 3μM CHIRON added. **B)** Cells grown in E8 with 3μM CHIRON and 15% KOSR added. **C)** Cells grown in E8 with 3μM CHIRON and 4.8μM LPA added. **D)** Cells grown in E8 with 3μM CHIRON and 0.48μM LPA added.

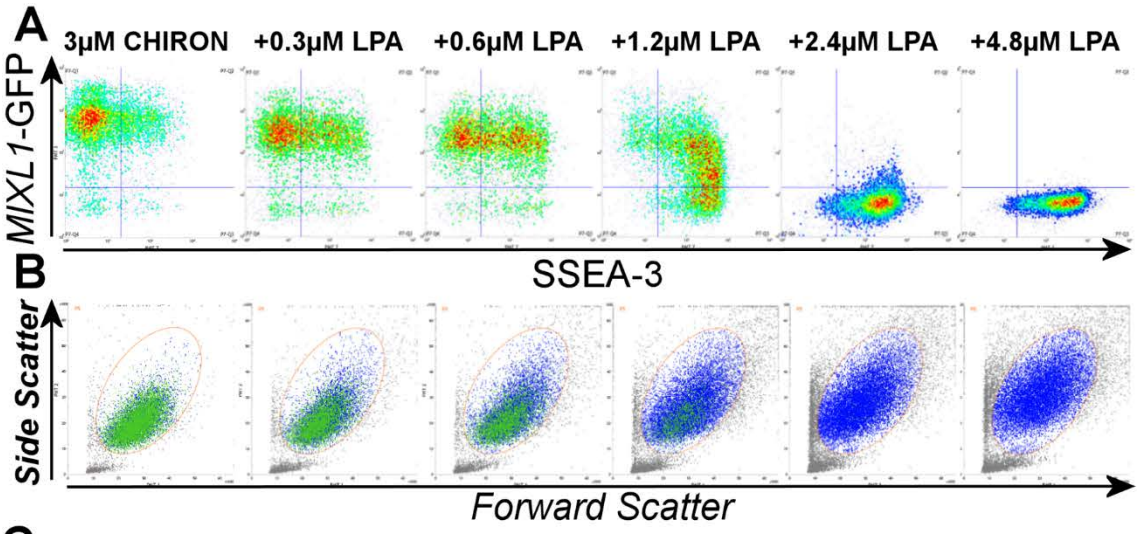
4.2.2. *MIXL1*-GFP induction by addition of GSK3β inhibitor and attenuation by addition of KOSR or LPA.

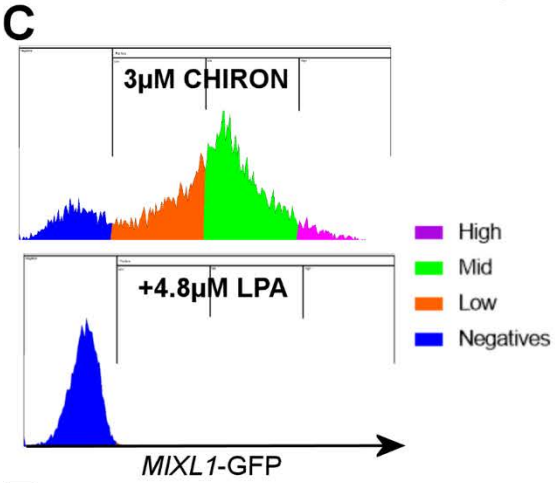
Our work using the porcupine inhibitor IWP-2 indicated that WNT signalling was important in the creation of our *MIXL1* positive substates. The expression of *MIXL1*-GFP is lost in standard E8V cultures but when cultured in the presence of a GSK3β inhibitor, CHIRON, to mimic WNT signalling *MIXL1* expression was induced. When I add CHIRON at 3μM to standard E8V conditions over 3 days I observed a drastic shift towards positive *MIXL1* expression (Figure 4-3A). When I also examined SSEA-3 expression I found that 54% cells co-expressed *MIXL1* and continue to express SSEA-3 (Figure 4-3A), which could suggest that some cells are still undifferentiated stem cells. Cells in this condition, however, exhibited morphology changes similar to those often seen during differentiation of human PSC (data not shown). The *MIXL1*-GFP intensity under these conditions is very high compared to the intensity seen in *MIXL1* positive cells in MEF/KOSR conditions.

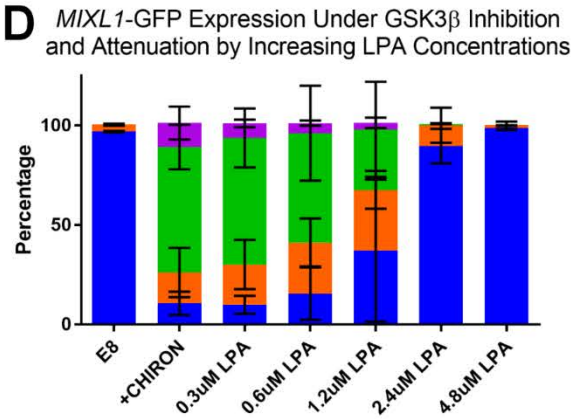
Blaukwamp et al, 2012 demonstrated the addition of KOSR was able to prevent differentiation caused by WNT3A or CHIRON addition to the medium. I sought to see whether the addition of 15% KOSR could effectively block the *MIXL1*-GFP expression induced by CHIRON addition. When cells were grown in E8 medium containing 3μM CHIRON and supplemented with 15% KOSR, *MIXL1*-GFP expression was not induced. Cells also maintained fairly high and uniform expression of SSEA-3 under these conditions (Figure 4-3B).

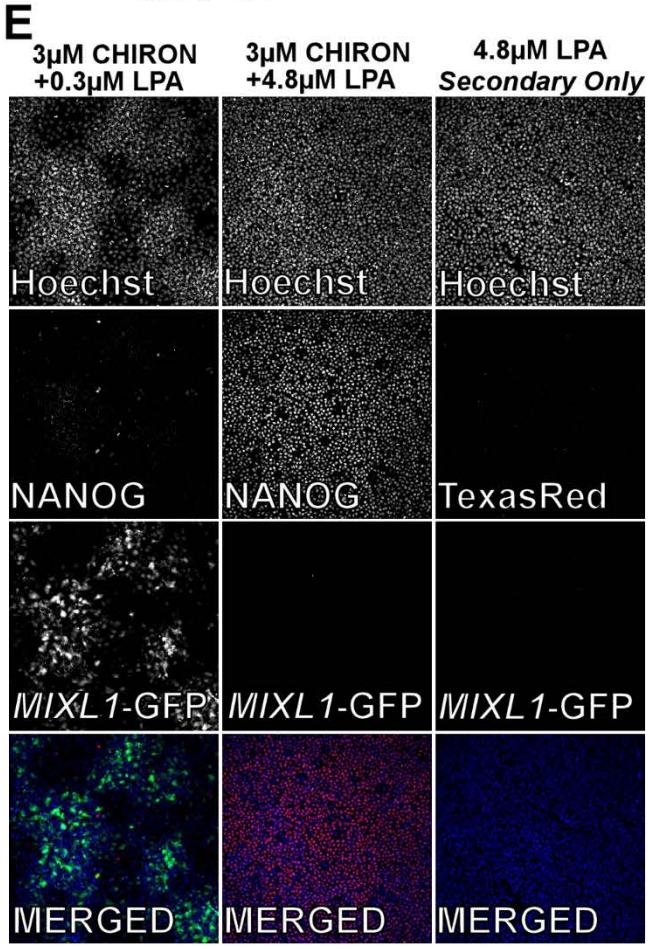
One of the major components of KOSR is Lysophophatidic Acid (LPA). Further to this, it has been shown that LPA can attenuate WNT mediated differentiation of ESC (Blauwkamp et al., 2012). I therefore assessed whether the same level of LPA which is present in 15% KOSR counteracted the CHIRON treated differentiation in a similar manner. This medium comprised of E8 containing CHIRON at 3μM, with 4.8μM LPA. After three days cells were stained for SSEA-3 and analysed by FACS. 4.8μM LPA treated samples showed no *MIXL1*-GFP expression in the presence of 3μM CHIRON, whilst maintain high levels of SSEA-3 (Figure 4-3C).

The induction of MIXL1 is a primary objective in this work, so I chose to decreasing the amount of LPA to 0.48μM which caused an increase in the amount of *MIXL1*-GFP positive cells, to approximately 50%. However, when compared to samples treated with just 3μM CHIRON there was a decrease in the intensity of *MIXL1*-GFP cells, with nearly all cells residing in *MIXL1*-GFP low expression (Figure 4-3D). The LPA appeared to attenuate this *MIXL1* induction, restricting the reporter to low level fluorescence.


A
3μM CHIRON
+0.3μM LPA
+0.6μM LPA
+1.2μM LPA
+2.4μM LPA
+4.8μM LPA
MIXL1-GFP
SSEA-3
B
Side Scatter
Forward Scatter


C
3μM CHIRON
+4.8μM LPA
High
Mid
Low
Negatives
MIXL1-GFP


D
MIXL1-GFP Expression Under GSK3β Inhibition
and Attenuation by Increasing LPA Concentrations
Percentage
100
50
0
E8
+CHIRON
0.3uM LPA
0.6uM LPA
1.2uM LPA
2.4uM LPA
4.8uM LPA


E
3μM CHIRON
+0.3μM LPA
3μM CHIRON
+4.8μM LPA
4.8μM LPA
Secondary Only
Hoechst
Hoechst
Hoechst
NANOG
NANOG
TexasRed
MIXL1-GFP
MIXL1-GFP
MIXL1-GFP
MERGED
MERGED
MERGED

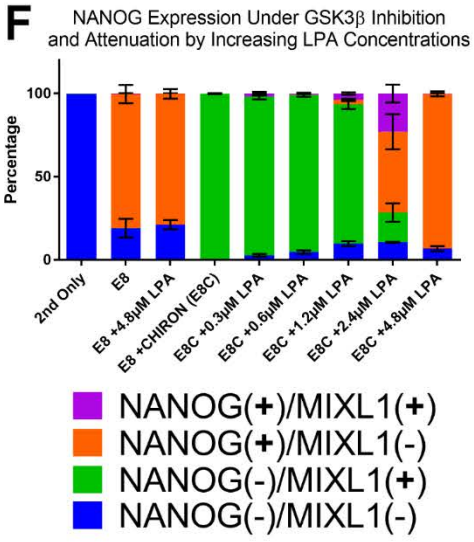

F
NANOG Expression Under GSK3β Inhibition
and Attenuation by Increasing LPA Concentrations
Percentage
100
50
0
2nd Only
E8
E8 +4.8μM LPA
E8 +CHIRON (E8C)
E8C +0.3μM LPA
E8C +0.6μM LPA
E8C +1.2μM LPA
E8C +2.4μM LPA
E8C +4.8μM LPA
NANOG(+)/MIXL1(+)
NANOG(+)/MIXL1(-)
NANOG(-)/MIXL1(+)
NANOG(-)/MIXL1(-)

Figure 4-4 LPA attenuation of CHIRON-induced Differentiation

HES3 *MIXL1*-GFP grown in E8 medium with 3µM CHIRON and increasing levels of LPA **A)** Flow cytometry density plots of displaying SSEA-3 expression versus *MIXL1*-GFP expression. **B)** Flow cytometry scatter plots of forward scatter versus side scatter, green dots indicated cells identified as *MIXL1*(**+**)/SSEA-3(-) from A. **C)** Representative histogram plots of *MIXL1*-GFP intensity from E8 **+** 3µM CHIRON with and without 4.8µM LPA added, coloured according to high, mid, low and negative expression. **D)** Stacked bar charts displaying percentage of *MIXL1*-GFP intensity for each condition. **E)** Immunofluorescence analysis of Hoechst, *MIXL1*-GFP and NANOG of HES3 *MIXL1*-GFP cells E8 with 3µM CHIRON with 0.3µM and 4.8µM LPA and E8 with 4.8µM LPA (Secondary antibody only staining). A merged image of all four channels is present below Hoechst (Blue), *MIXL1*-GFP (Green), NANOG(Red). **F)** Stacked bar charts displaying percentage of *MIXL1*-GFP and NANOG positive and negative cells for each condition. (Bars are mean ± SD, D, F) n= 3 biological repeats).

4.2.3. Levels of LPA affects *MIXL1*-GFP expression intensity when induced by GSK3β Inhibition.

Both 15% KOSR and 4.8μM LPA appeared to completely abolish the expression of CHIRON-induced *MIXL1*-GFP in culture and that lower levels of LPA induced low level *MIXL1* expression. I then set to assess how the level of LPA affects CHIRON-induced *MIXL1*-GFP expression as well as its effect on SSEA-3 expression, cell size and NANOG expression. This would hopefully provide a better insight into the dynamics of using LPA to attenuate CHIRON-induced differentiation, including the cellular characteristics that might change. 100,000 cells were plated in E8V conditions with ROCKi added, to aid single cell survival (Watanabe et al., 2007). Each sample was treated with 3μM CHIRON as this level produced strong *MIXL1*-GFP expression over 3 days. Concentrations of LPA ranged from 0.3μM to 4.8μM were added to the medium. After 3 days, samples were analysed by flow cytometry. The expression levels of *MIXL1*-GFP and SSEA-3 changed with increased LPA concentrations, with *MIXL1*-GFP decreasing and SSEA-3 increasing (Figure 4-4A). As cells begin to differentiate, there are changes to the morphology of the cells, with the cells becoming small and compact (Ramirez et al., 2013), this is reflected in the forward scatter versus side scatter flow cytometry density plots, *MIXL1*(**+**)/SSEA-3(-) cells are labelled in green (Figure 4-4B). Observing the intensity of *MIXL1*-GFP expression there was effectively no detected expression in 4.8μM LPA, as seen before. At 2.4μM there was a small proportion of *MIXL1*-GFP low cells. At 1.2μM there was an increase in low and mid *MIXL1*-GFP expression (Figure 4-4D).

I also performed intracellular NANOG staining on samples treated in the same way (Figure 4-4E). In E8 **+** 3μM CHIRON, I saw a substantial loss of NANOG expression compared to E8 alone or E8 with 4.8μM LPA medium. This loss was apparent up to 1.2μM LPA (Figure 4-4F). LPA caused down regulation of *MIXL1*-GFP expression at 4.8μM and less pronounced down regulation at 2.4μM. In these conditions around 20-30% of cells exhibited co-expression of *MIXL1*-GFP and NANOG (Figure 4-4F).

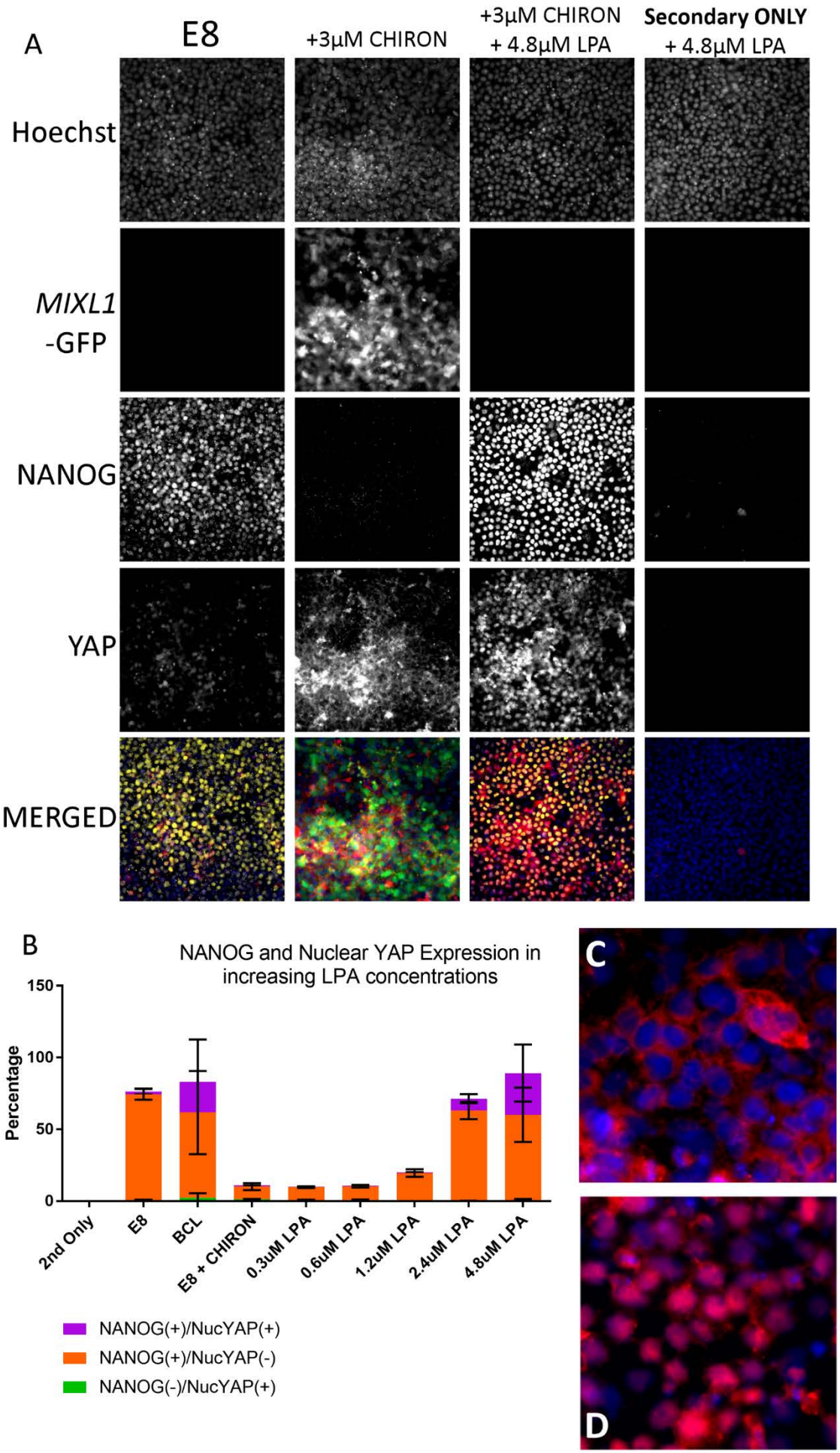

A
E8
+3μM CHIRON
+3μM CHIRON
+ 4.8μM LPA
Secondary ONLY
+ 4.8μM LPA
Hoechst
MIXL1
-GFP
NANOG
YAP
MERGED
B
NANOG and Nuclear YAP Expression in
increasing LPA concentrations
Percentage
150
100
50
0
2nd Only
E8
BCL
E8 + CHIRON
0.3uM LPA
0.6uM LPA
1.2uM LPA
2.4uM LPA
4.8uM LPA
NANOG(+)/NucYAP(+)
NANOG(+)/NucYAP(-)
NANOG(-)/NucYAP(+)
C
D

Figure 4-5 YAP Localisation under GSK3β Inhibition and LPA

A) Immunofluorescence analysis of Hoechst, *MIXL1*-GFP, NANOG and YAP expression of HES3 *MIXL1*-GFP cells in E8 medium alone, E8 with 3μM CHIRON, E8 with 3μM CHIRON and 4.8μM LPA and E8 with 4.8μM LPA (Secondary antibody only staining). A merged image of all four channels is present below Hoechst (Blue), *MIXL1*-GFP (Green), NANOG(Yellow) and YAP (Red). **B)** Stacked bar chart displaying percentage of cells positive and negative for NANOG and Nuclear YAP (NucYap) in the conditions tested (Bars are mean ± SD, n= 3 technical repeats). **C - D)** Immunofluorescence analysis of Hoechst (Blue) and YAP (Red) in **C)** E8 with 3μM CHIRON and 0.3μM LPA and **D)** E8 with 3μM CHIRON and 4.8μM LPA. Nuclear Localisation of YAP increased with increasing levels of LPA addition, even in the presence of 3μM CHIRON. Nuclear YAP cells tended to be NANOG positive.

4.2.4. LPA concentrations affects YAP Cellular Localisation

The effect of LPA on CHIRON-induced *MIXL1*-GFP expression was clear from our analysis, going forward LPA forms a critical factor in our trapping medium but the mechanism of how it prevents differentiation remains elusive. LPA has been shown to be a potent inhibitor of the hippo pathway (Yu et al., 2012). The hippo pathway is activated, in part, by cell to cell contact and in many other system mediates cell proliferation, decreasing as the pathway is activated at high density (Yu and Guan, 2013). When the hippo pathway is active, a cascade of events eventually results in an inhibition the nuclear localisation of the transcription factor YAP. Previous studies have indicated that the nuclear localisation of YAP, can be used as a readout of an inactive hippo pathway and that this may be an important factor in the action of LPA attenuating CHIRON-induced differentiation (Qin et al., 2016). Therefore, I assessed the localisation of YAP in conditions containing 0 - 4.8μM LPA and/or 3μM CHIRON.

In standard E8 conditions the nuclear localisation of YAP was variable, often depending on the density of the area assessed, but overall a very low percentage of cells showed nuclear staining (Figure 4-5A). Analysis of NANOG and *MIXL1*-GFP in E8 conditions revealed high and non-expression, respectively. In E8 medium with 4.8μM LPA, again I saw high NANOG expression and low *MIXL1*-GFP expression (Figure 4-5B). YAP, however, exhibited increased nuclear localisation in the presence of LPA compared to E8 medium alone, increasing to 25%.

E8 medium containing only 3μM CHIRON induced widespread differentiation in the wells, as highlighted by the lack of cells positive for NANOG and the abundance of *MIXL1* positive cells (Figure 4-5A). Yap was distinctly located in the cytoplasm in these cells with less than 2% showing nuclear localisation (Figure 4-5B). At higher levels of LPA in the conditions three things occurred in a progressive manner: first a reduction in *MIXL1*-GFP positive cells, second an increase in NANOG positive cells and finally an increase in YAP nuclear localisation.

Whilst NANOG positive cells were present with and without nuclear YAP, cells in which YAP was nuclear cells tended to be NANOG positive (Figure 4-5B). Conversely, nuclear YAP positive cells displayed low correlation with *MIXL1* positive cells. At low levels of LPA (0.3μM) YAP was cytoplasmic but higher levels (Figure 4-5C), such as 4.8 μM, there was a higher proportion of cells with nuclear YAP even in the presence of CHIRON (Figure 4-5D).

4.2.5. Optimisation of a medium to trap *MIXL1*(+)/SSEA-3(+) cells

After assessing the level of LPA required to attenuate the *MIXL1*-GFP expression induced by CHIRON, I utilised two formulations for further analysis: **1x trapping medium** (0.48μM LPA) and **2x trapping medium** (0.96μM LPA) as under bulk passaging conditions both these levels display induction of a *MIXL1* low level of expression. Self-renewal and propagation through passaging is an inherent function of human PSC. Whether these formulations could maintain the state post passage had to be assessed. For controls and assessment of reversion in the presence of LPA I used the same two formulations without the addition of CHIRON, named "**BCL**" after the three main components, BSA, Cholesterol and LPA added to E8 medium.

While 1x trapping medium was able to induce low level expression of *MIXL1*-GFP and maintain SSEA-3 expression (Figure 4-6A), its ability to maintain this level post passage was variable (Figure 4-6B). When looking at other pluripotency-associated surface markers, BF4 and CD9 (Wright et al., 2011), there was a decrease in levels post passage. Although the culture does not lose expression of CD9, there is a drop in intensity, this intensity shift has been noted by others with respect to early mesoderm differentiation (Hough et al., 2009). When the cells were passaged into E8 medium containing only LPA, cells retained high levels of surface marker expression and exhibited relatively no *MIXL1*-GFP expression (Figure 4-6C)

I thought the post passage issue in the trapping medium might be an increasing build-up of endogenous WNT ligands secreted by cells undergoing early differentiation. The level of LPA was optimised to combat the effect of the CHIRON. However, increased endogenous WNT signalling may overcome this LPA effect. To control for this, I opted to implement a system, similar to the Baseline Activation (BLA) method described in Hackland et al, 2017. Unlike BLA, rather than blocking WNT ligands with the recombinant protein DKK1, I used a chemical inhibitor of Porcupine, IWP-2, to block the secretion of WNT ligands. IWP-2 has been used to block endogenous WNT signalling in human ESC previously (Blauwkamp et al., 2012).

I added IWP-2 at levels between 0.25 and 2μM to the trapping medium. The colony growth within the first 3 days was comparable in all conditions (Figure 4-7). Assessment of *MIXL1*-GFP and surface marker expression after 3 days reveals the effect of IWP-2 addition. The percentage of cells exhibiting low *MIXL1*-GFP expression decreased in all IWP-2 containing trapping

conditions. In contrast, in the control conditions containing 3µm CHIRON, the addition of IWP-2 had very little effect on the *MIXL1*-GFP expression (Figure 4-8B)

At day 3 SSEA-3 and CD9 expression had shifted towards low/negative in all samples containing just 3µM CHIRON, with IWP-2 having very little effect on these changes (Figure 4-8C). Samples cultured in trapping medium retained higher expression levels of these two markers and at

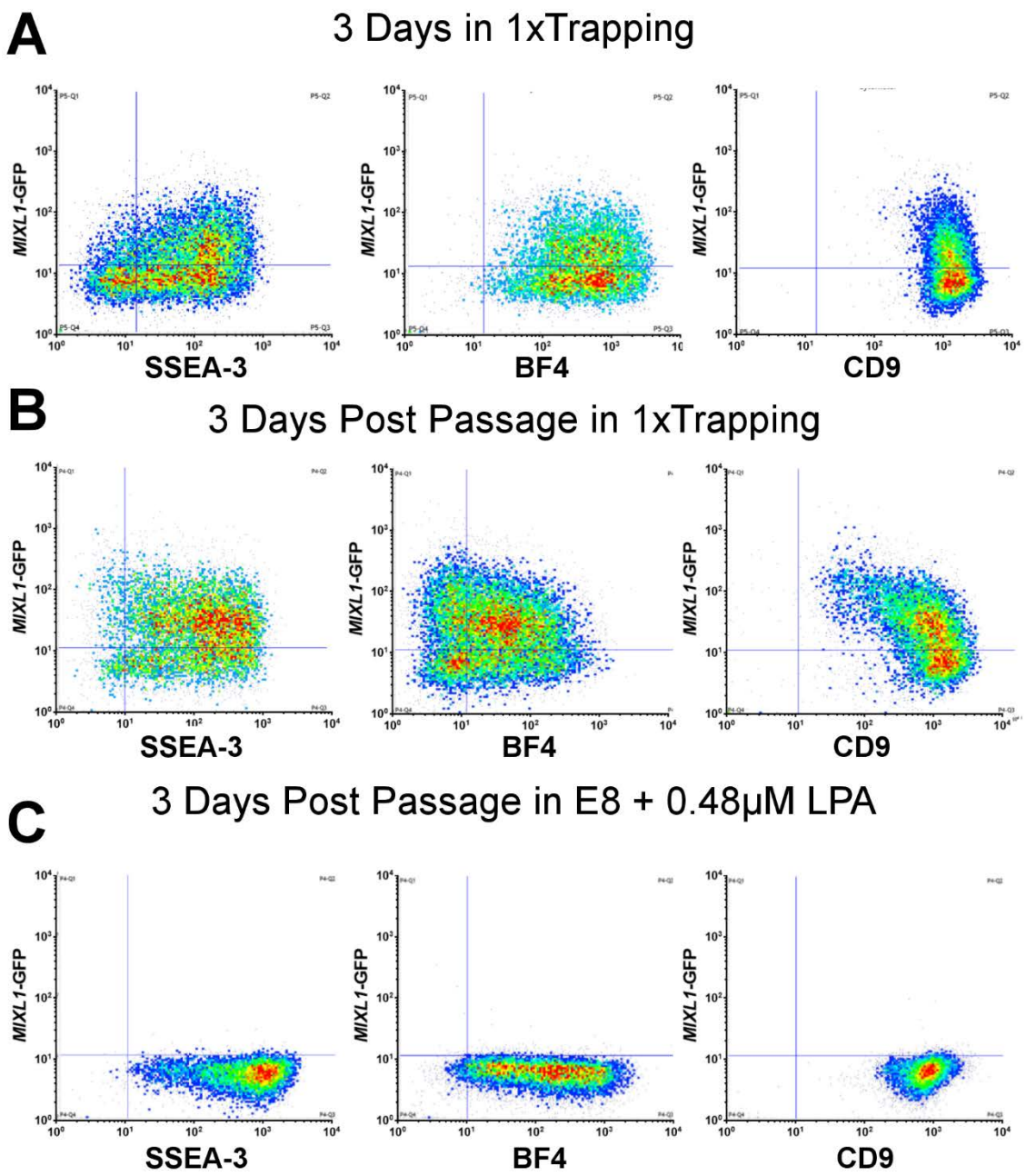


Figure 4-6 1xTrapping Analysis of *MIXL1* and Surface Markers Expression

A -C) Flow cytometry density plots of surface markers, SSEA-3, BF4 and CD9 versus *MIXL1*-GFP after **(A)** 3 days in 1xTrapping, **(B)** 3 days post the first passage in 1xTrapping and **(C)** 3 days post the first passage in E8 with 0.48µM LPA added.

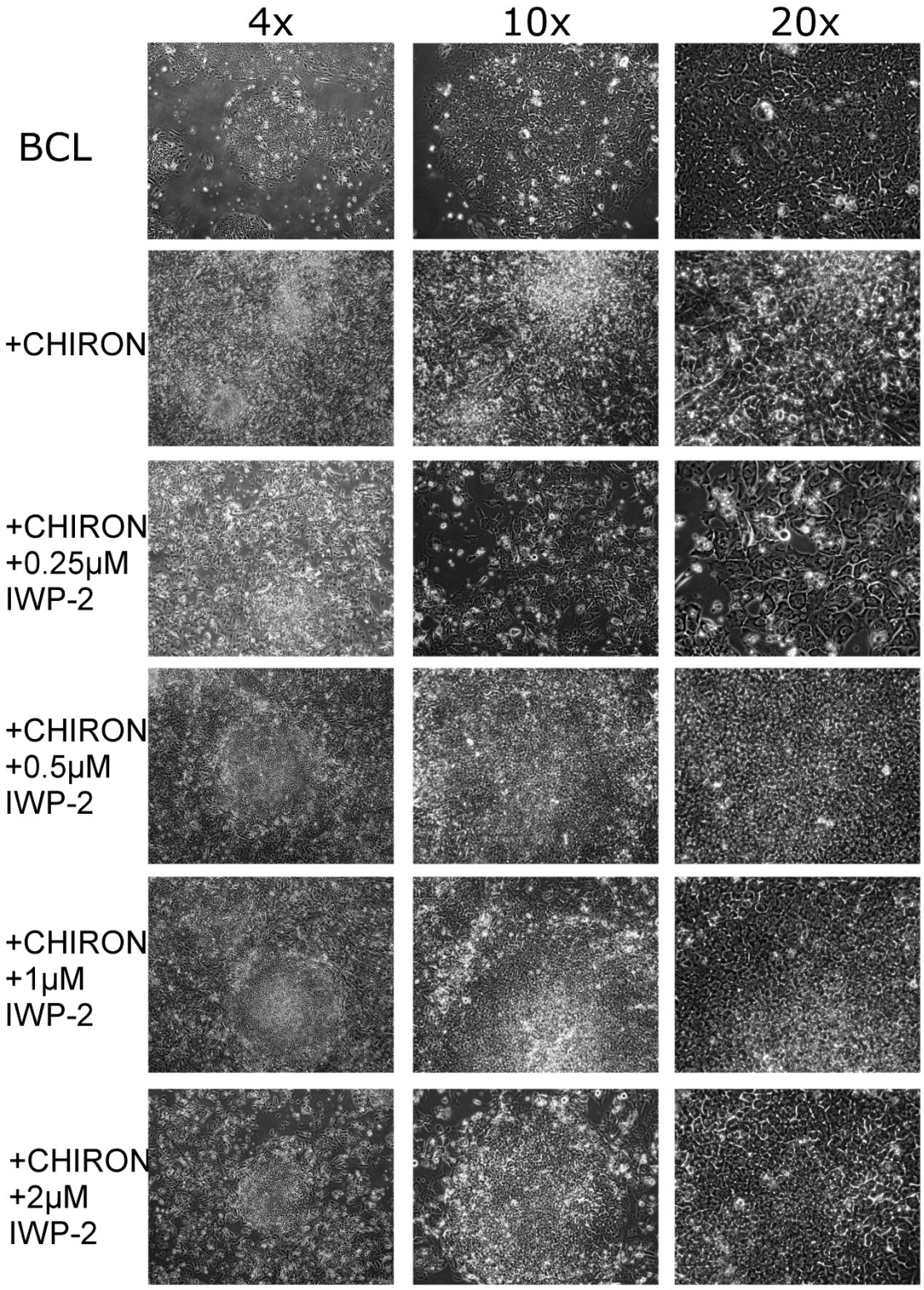


Figure 4-7 Cell and Colony Morphology in the Presence of IWP-2

Phase contrast images at 4x, 10x and 20x of HES3 *MIXL1*-GFP cells in E8 with LPA (BCL), 1xTrapping (+CHIRON) and 1xTrapping with 0.25µM, 0.5µM, 1µM and 2µM IWP-2 at day 4.

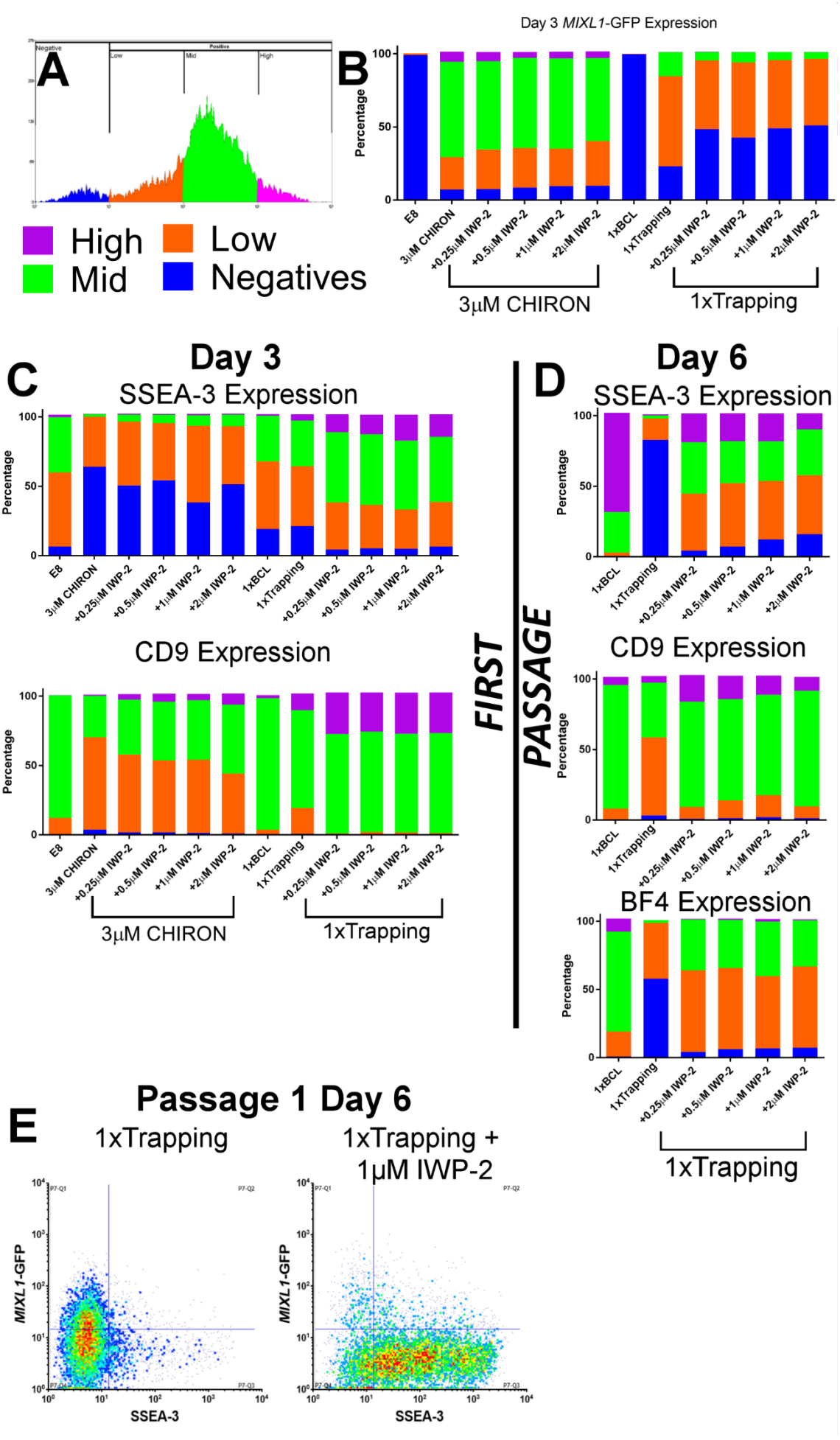

A
Negative
Positive
Low
Mid
High
High
Low
Mid
Negatives
B
Day 3 MIXL1-GFP Expression
Percentage
100
50
0
E8
3μM CHIRON
+0.25μM IWP-2
+0.5μM IWP-2
+1μM IWP-2
+2μM IWP-2
1xBCL
1xTrapping
+0.25μM IWP-2
+0.5μM IWP-2
+1μM IWP-2
+2μM IWP-2
3μM CHIRON
1xTrapping
C
Day 3
SSEA-3 Expression
CD9 Expression
3μM CHIRON
1xTrapping
FIRST PASSAGE
D
Day 6
SSEA-3 Expression
CD9 Expression
BF4 Expression
1xTrapping
E
Passage 1 Day 6
1xTrapping
1xTrapping + 1μM IWP-2
MIXL1-GFP
SSEA-3
P7-Q1
P7-Q2
P7-Q4
P7-Q3

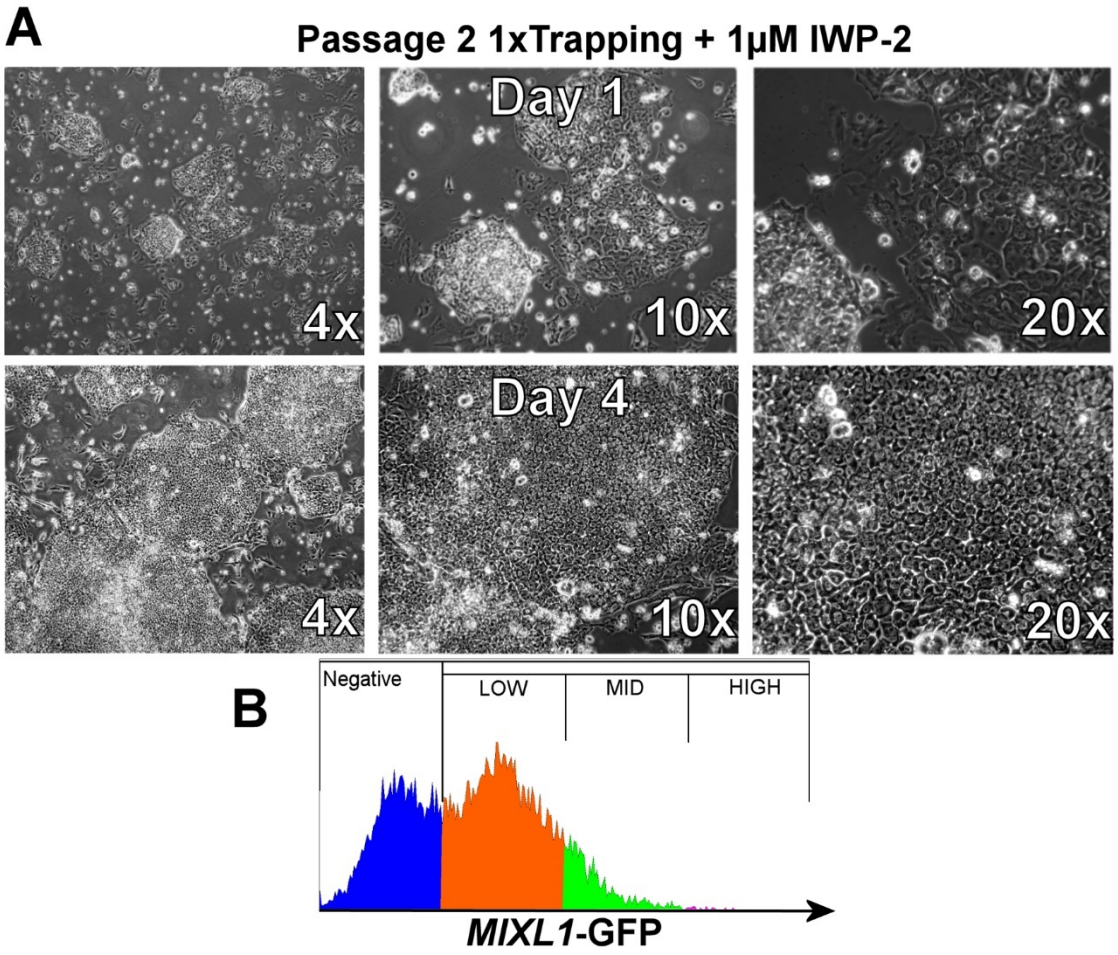
Figure 4-8 Inhibition of WNT Secretion Maintains Pluripotency Surface Marker Expression

In order to assess the effect of WNT secretion inhibition in our system I monitored *MIXL1*-GFP expression and surface marker expression pre and post passage. Samples treated with E8 medium with 3uM CHIRON were relatively unaffected by the addition of IWP-2, gaining high expression of *MIXL1* and low expression of pluripotency-associated surface markers by day 3. Samples in 1xTrapping media exhibited marginal benefits in maintaining surface marker expression at day 3 in the presence of IWP-2 but post passage (day 6) only samples grown in the presence of IWP-2 were able to maintain this expression level. **A)** Flow cytometry histogram of *MIXL1*-GFP intensity of cells grown in E8 with 3μM CHIRON present. Plot is coloured according to High, Mid, Low and Negative *MIXL1*-GFP expression. **B-D)** Stacked Bar Charts displaying the percentage High, Mid, Low and Negative expression of *MIXL1* and surface antigen markers by cells grown in 3μM CHIRON and 1xTrapping with increasing concentrations of IWP-2 (Bars are mean ± SD). **B)** *MIXL1*-GFP after 3 days, **C)** SSEA-3 and CD9 after 3 days, **D)** SSEA-3, CD9, and BF4 after 6 days (3 days post passage). **E)** Flow cytometry density plots of SSEA-3 versus *MIXL1*-GFP of cells 6 days (3 days post passage) from 1xTrapping and 1xTrapping with 1μM IWP-2 added.

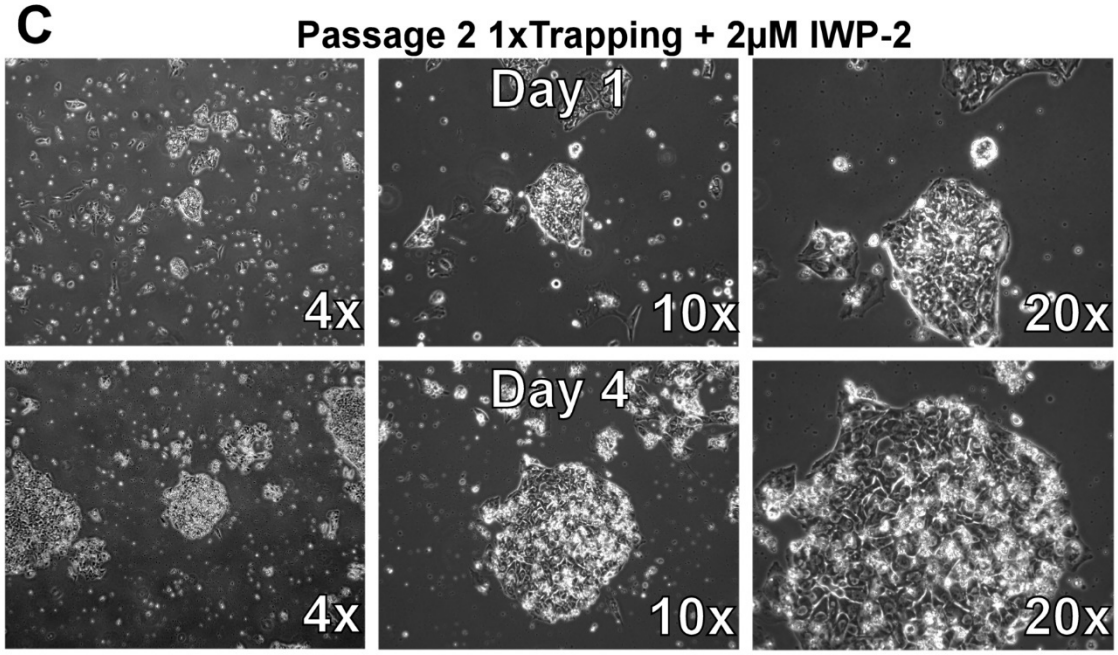
this stage samples treated with IWP-2 had a marginal increase in expression of these two surface markers (Figure 4-8C).

Cells in trapping conditions were passage further and assessed again after another 3 days in culture (6 days total). Expression of SSEA-3, CD9 and another pluripotency-associated marker, BF4, were markedly decreased in 1x trapping medium. The presence of IWP-2 even at 0.25μM, helped to maintain high levels of SSEA-3, BF4 and CD9 post passage (Figure 4-8D) At 1μM IWP-2 there was a decrease in level of *MIXL1* low cells in these cultures, but strong SSEA-3 expression compared to 1xTrapping alone (Figure 4-8E).

Cells growing in both 1μM and 2μM IWP-2 exhibited good colony morphology after the first passage, seemingly better than in the presence of 0.25 and 0.5μM IWP-2. Therefore, the cells were passage further again in the same conditions, 1xTrapping with 1 or 2μM IWP-2. In the presence of 1μM IWP-2 the cells displayed good colony morphology and growth rate over 4 days post passage (Figure 4-9 A). With 2μM IWP-2 I saw a dramatic decrease in proliferation rate (Figure 4-9C), most likely because Porcupine has a WNT-independent effect on proliferation (Covey et al., 2012). Both 1μM and 2μM had a large proportion of *MIXL1*-GFP low cells present in the cultures (Figure 4-9B, D), but because of the proliferation differences, 1μM IWP-2 was selected for the optimised formulation of the Trapping medium.


A
Passage 2 1xTrapping + 1μM IWP-2
Day 1
4x
10x
20x
Day 4
4x
10x
20x

B
Negative
LOW
MID
HIGH
MIXL1-GFP


C
Passage 2 1xTrapping + 2μM IWP-2
Day 1
4x
10x
20x
Day 4
4x
10x
20x

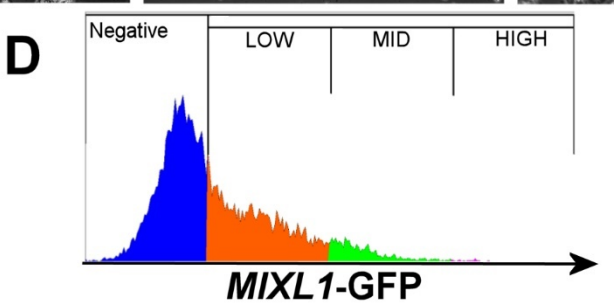

D
Negative
LOW
MID
HIGH
MIXL1-GFP

Figure 4-9 Colony Growth and *MIXL1*-GFP expression in the presence of IWP-2

A, C) Phase contrast images at 4x, 10x and 20x of cells after the second passage in 1xTrapping with 1µM (**A)** and 2µM (**C)** IWP-2 at day 1 and day 4 post passage. **B, D)** Flow cytometry histograms of *MIXL1*-GFP intensity after the second passage in 1xTrapping with 1µM **(B)** and 2µM **(D)** IWP-2 at day 4 post passage.

4.2.6. Post passage maintenance of *MIXL1*(+)/SSEA-3(+) population and reversion to *MIXL1*(-)/SSEA-3(+).

After assessing the use of IWP-2 in the culture system, I now had an optimised version of the trapping medium. I decided to name this new formulation **"Primo"** pronounced ˈpriːməʊ', its formulation at both 1x and 2x is presented in Table 4.1. Since *MIXL1* is an early primitive streak marker, our medium takes the start of its name from the **"Prim",** in primitive streak and the **"O"** from the optimisation process used to develop it. The formulation of BCL is also presented in Table 4.1 and lacks the CHIRON and IWP-2 used in Primo formulation. Using this new Primo medium I first looked at how cells can interconvert between *MIXL1* positive and negative states. Cells were grown in 1xPrimo for 3 days and then passaged in bulk into four conditions: 1xPrimo, E8 containing LPA, E8, and E8 with 1µM IWP-2. Cells were then allowed to grow for 3 days in respective medium before analysis. Immunofluorescent analysis reveals a down regulation of *MIXL1*-GFP expression (Figure 4-10, Figure 4-11) in the BCL, E8 and E8 with IWP-2 conditions, compared to the Primo conditions.

Table 4.1 Primo Formulation

			Concentration	
Media	Component	Company	**2X (0.96µM LPA)**	**1X (0.48µM LPA)**
BCL	E8 (**Table 2.1**)	Homemade	-	-
	Fatty Acid free BSA	Millipore	0.2%	0.1%
	2-mercaptoethanol	Gibco	20nM	10nM
	Cholesterol	Sigma-Aldrich	4 µM	2µM
	LPA	Sigma-Aldrich	0.96µM	0.48µM
+				
Primo	CHIR99021	Tocris	3µM	3µM
	IWP-2	Tocris	1µM	1µM

Intracellular staining for two crucial pluripotency markers, NANOG and SOX2 revealed high expression in all conditions (Figure 4-10, Figure 4-11) There was also a decrease in *MIXL1*-GFP positive cells in conditions not containing any CHIRON. However, the fraction of positive cells in this condition was smaller than expected at this time. All conditions did, however, have a persistent proportion of cells that were neither NANOG/SOX2 nor *MIXL1*-GFP positive (Figure 4-10, Figure 4-11). This fraction might represent a differentiated fraction towards a non *MIXL1* positive state or perhaps a further differentiated state in which *MIXL1*-GFP expression has been turned off.

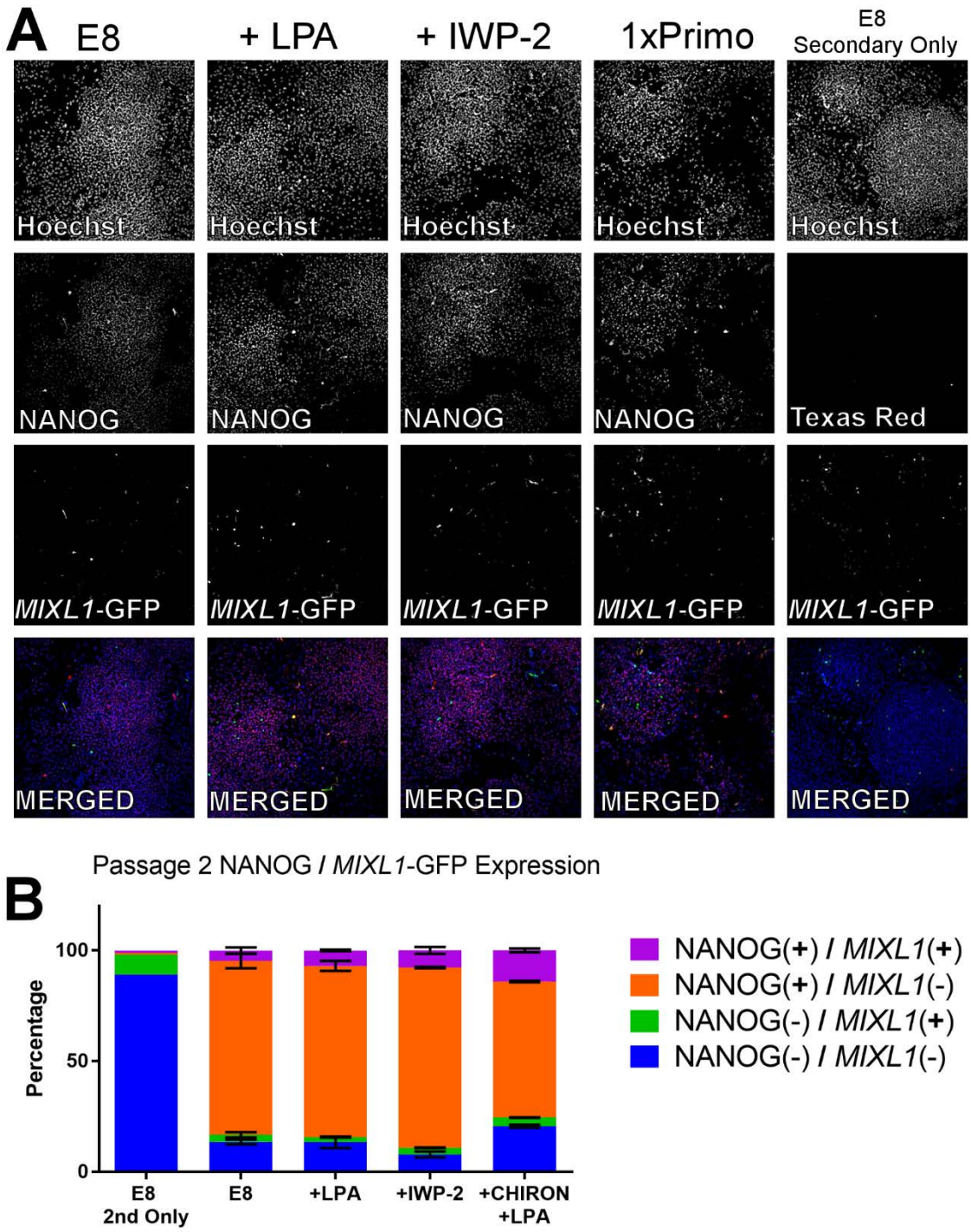
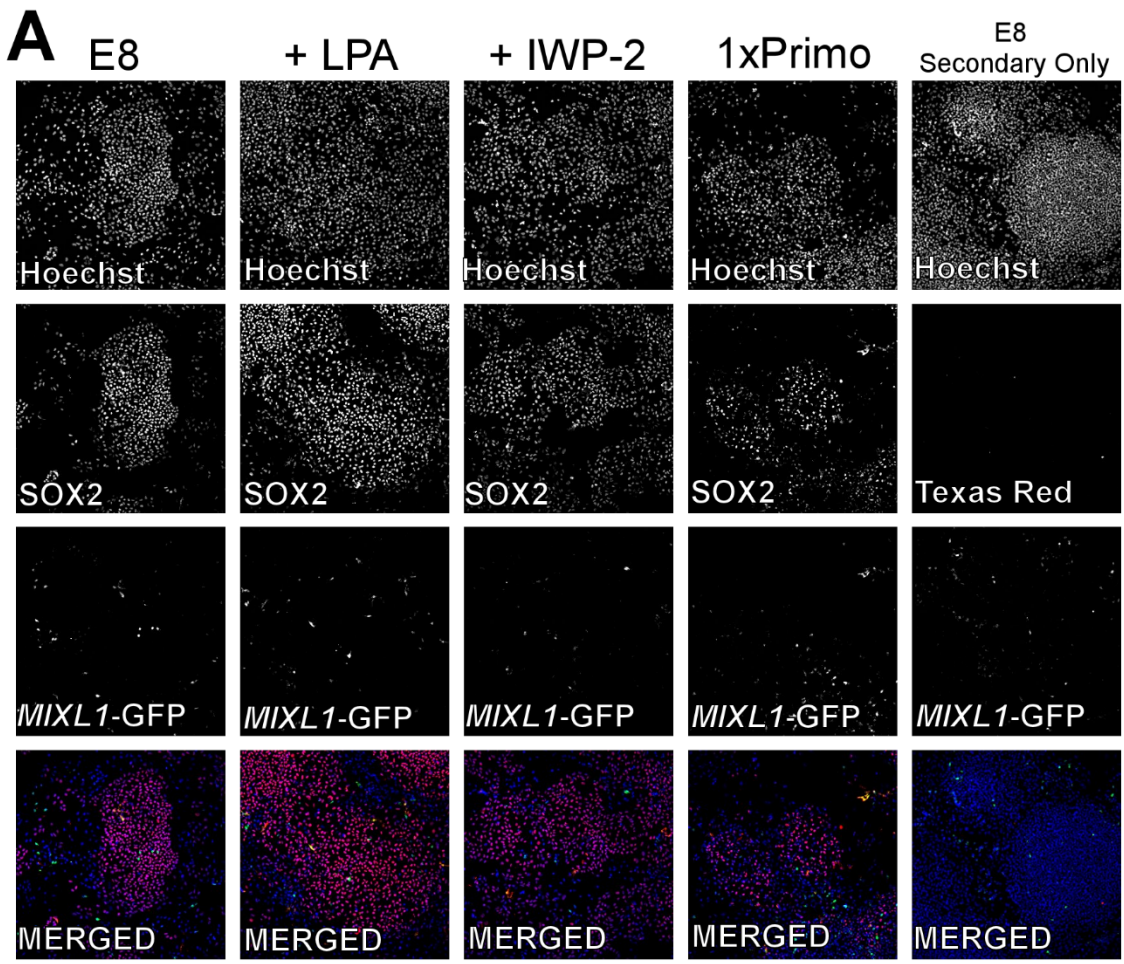


Figure 4-10 Post Passage NANOG expression in cells grown in Primo Medium

A) Immunofluorescence analysis of Hoechst, *MIXL1*-GFP, and NANOG expression of HES3 *MIXL1*-GFP cells in E8, E8 with 0.48µM LPA (**+**LPA), E8 with 1µM IWP-2(**+**IWP-2), 1xPrimo and E8 (Secondary antibody only staining) for 3 days post to 1 passage in 1xPrimo. A merged image of all four channels is present below Hoechst (Blue), *MIXL1*-GFP (Green), and NANOG(red). **B)** Stacked bar chart displaying percentage of cells positive and negative for NANOG and *MIXL1*-GFP in the conditions tested (Bars are mean ± SD, n= 3 technical repeats).



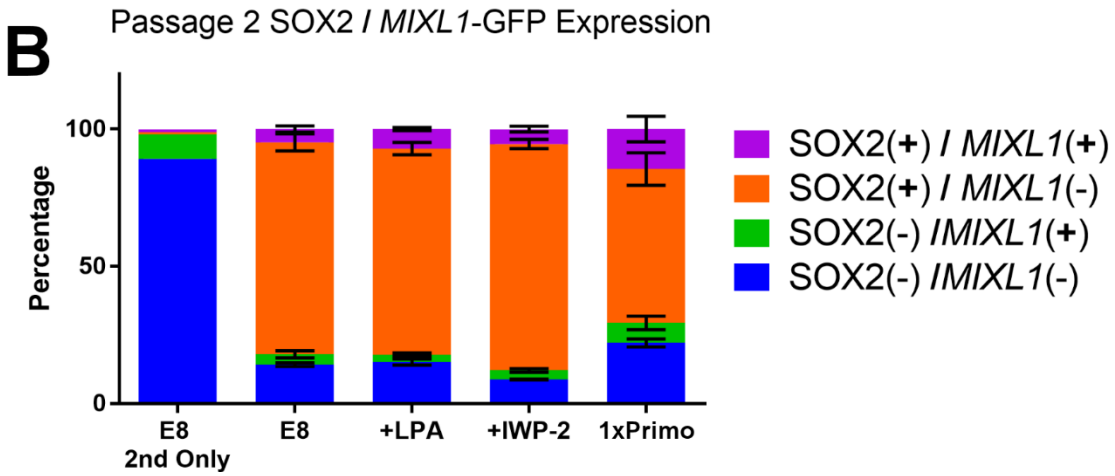
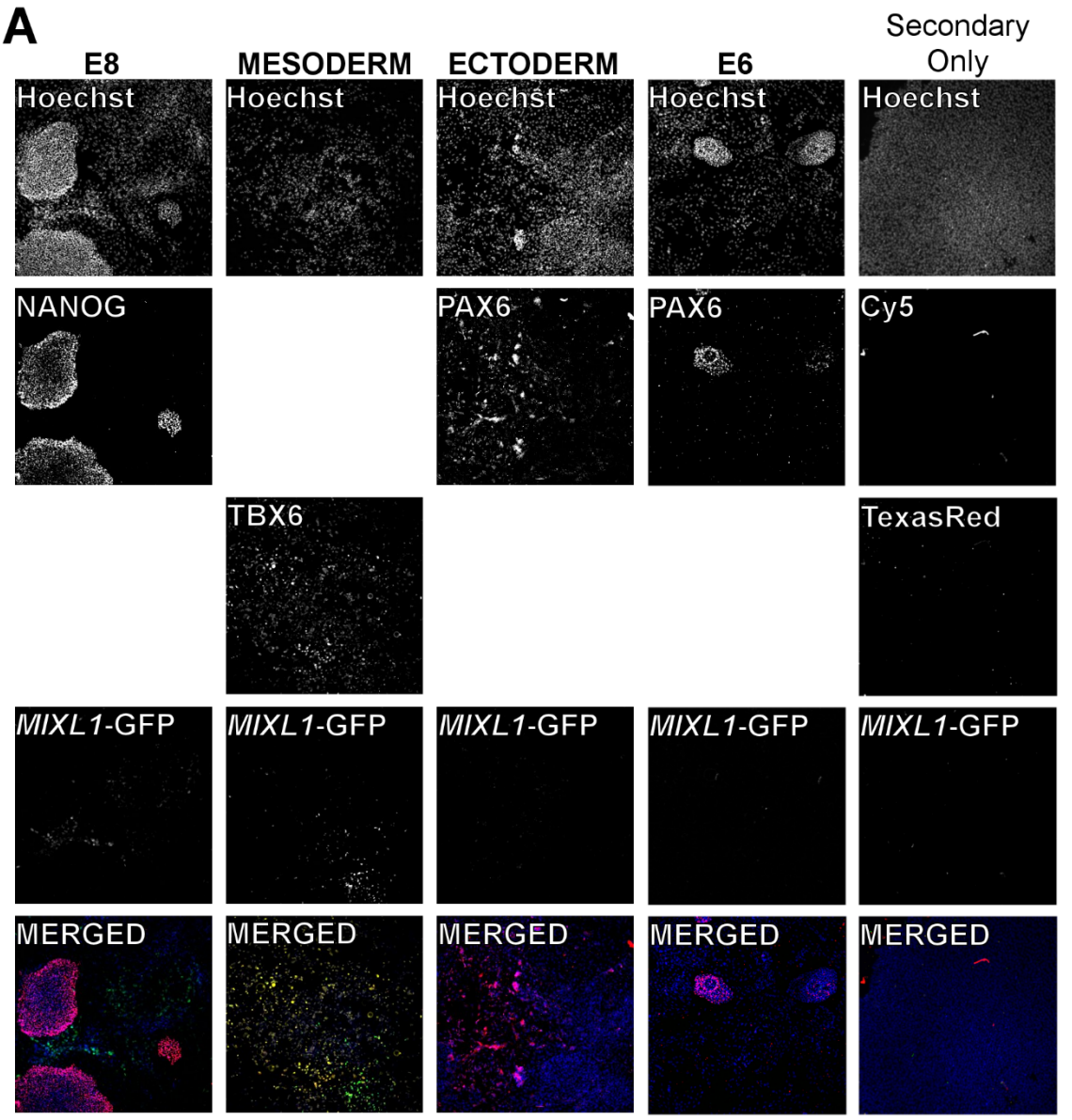
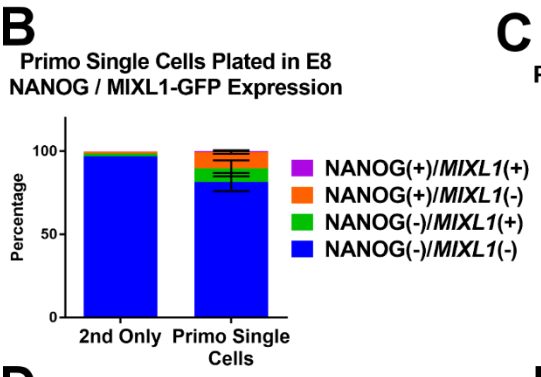
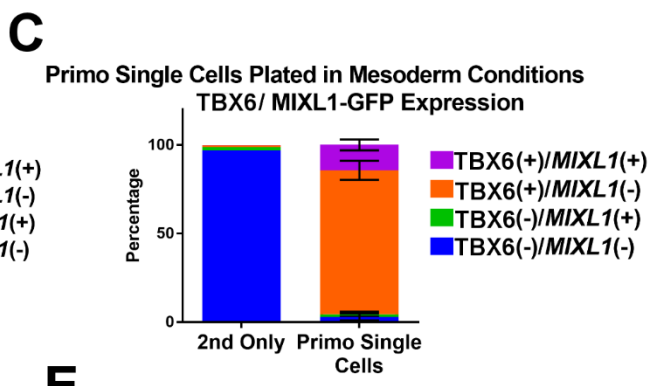


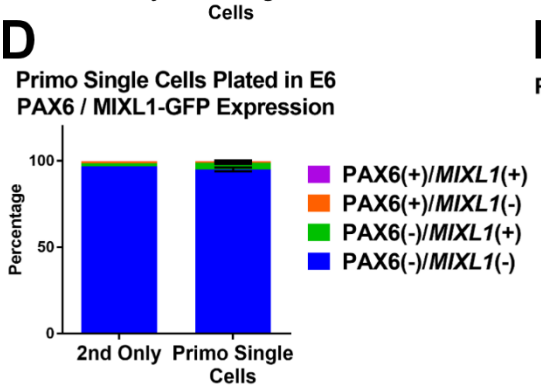
Figure 4-11 Post Passage SOX2 expression in cells grown in Primo Medium

A) Immunofluorescence analysis of Hoechst, *MIXL1*-GFP, and SOX2 expression of HES3 *MIXL1*-GFP cells in E8, E8 with 0.48µM LPA (**+**LPA), E8 with 1µM IWP-2(**+**IWP-2), 1xPrimo and E8 (Secondary antibody only staining) for 3 days post to 1 passage in 1xPrimo. A merged image of all four channels is present below Hoechst (Blue), *MIXL1*-GFP (Green), and SOX2(red). **B)** Stacked bar chart displaying percentage of cells positive and negative for SOX2 and *MIXL1*-GFP in the conditions tested (Bars are mean ± SD, n= 3 technical repeats).


A
E8
MESODERM
ECTODERM
E6
Secondary Only
Hoechst
NANOG
PAX6
Cy5
TBX6
TexasRed
MIXL1-GFP
MERGED


B
Primo Single Cells Plated in E8
NANOG / MIXL1-GFP Expression
Percentage
100
50
0
2nd Only
Primo Single Cells
NANOG(+)/MIXL1(+)
NANOG(+)/MIXL1(-)
NANOG(-)/MIXL1(+)
NANOG(-)/MIXL1(-)


C
Primo Single Cells Plated in Mesoderm Conditions
TBX6/ MIXL1-GFP Expression
Percentage
100
50
0
2nd Only
Primo Single Cells
TBX6(+)/MIXL1(+)
TBX6(+)/MIXL1(-)
TBX6(-)/MIXL1(+)
TBX6(-)/MIXL1(-)


D
Primo Single Cells Plated in E6
PAX6 / MIXL1-GFP Expression
Percentage
100
50
0
2nd Only
Primo Single Cells
PAX6(+)/MIXL1(+)
PAX6(+)/MIXL1(-)
PAX6(-)/MIXL1(+)
PAX6(-)/MIXL1(-)

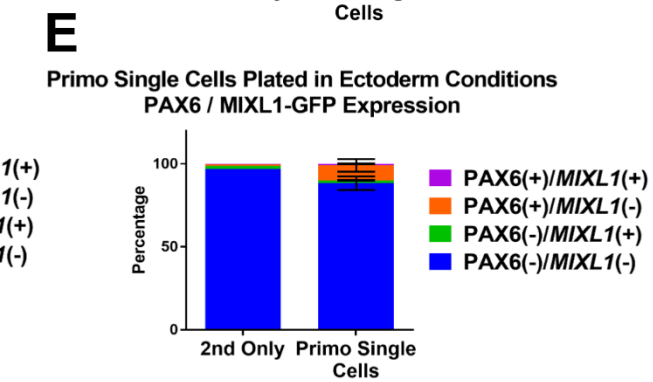

E
Primo Single Cells Plated in Ectoderm Conditions
PAX6 / MIXL1-GFP Expression
Percentage
100
50
0
2nd Only
Primo Single Cells
PAX6(+)/MIXL1(+)
PAX6(+)/MIXL1(-)
PAX6(-)/MIXL1(+)
PAX6(-)/MIXL1(-)

Figure 4-12 Single Cell Plating Limits Reversion and Differentiation Potential

A) Immunofluorescence analysis of Hoechst, *MIXL1*-GFP, marker expression HES3 *MIXL1*-GFP cells in E8 medium (NANOG stained), mesoderm induction medium (TBX6 stained), ectoderm induction medium (PAX6 stained) and E6 medium (PAX6 stained). HES3-*MIXL1* grown in E8 medium was used for secondary antibody only staining. Merged images are present below Hoechst (Blue), *MIXL1*-GFP (Green), TBX6(Yellow) and NANOG (Red). **B, C, D, E)** Stacked bar chart displaying percentage of cells positive and negative for **B)** NANOG and *MIXL1*-GFP in E8 Conditions, **C)** TBX6 and *MIXL1*-GFP in mesoderm conditions, **D)** PAX6 and *MIXL1*-GFP in ectoderm conditions, and **E)** PAX6 and *MIXL1*-GFP in E6 conditions (Bars are mean ± SD, n= 3 biological repeats).

4.2.7. Single cell plating from Primo conditions

I noticed a disparity between the levels of LPA required to maintain human PSC pluripotency markers between experiments that treated bulk cultures (Figure 4-3) and experiments that had been plated as single cells prior to treatment (Figure 4-4). With the Primo medium now optimised I assessed whether cells could be passaged as single cells, and exhibit the same reversion back to *MIXL1*(-)/SSEA-3(**+**) state and maintain expression of NANOG. Separately cells were also taken as single cells for assessment in differentiation protocols for mesoderm ectoderm (Dual SMAD inhibition) and E6.

In order to promote survival, 10µM of Rocki was added to the medium for 24 hours post plating. The addition of Rocki, has a pronounced effect on cell morphology, causing elongated and spiky cell bodies. This is normally rapidly reversed upon its removal from the medium. Cells from Primo conditions, however exhibited this spiky morphology long after the Y-27632 had been removed even in E8 conditions. This morphology difference looked to be indicative of differentiation. Assessing NANOG and *MIXL1*-GFP expression in the cells after 6 days (Figure 4-12A) of growth there was increased *MIXL1*-GFP expression and low percentage of NANOG positive cells. NANOG positive cells were restricted to small colonies and the elongated, spiky cells were NANOG negative, in line with the notion the cells were differentiated. The majority of cells were negative for both markers (Figure 4-12B).

Plating in mesoderm conditions demonstrated the lowest cell numbers of all the conditions. Assessment of TBX6 and *MIXL1*-GFP expression demonstrated most cells were TBX6 positive with a small portion of these coexpressing *MIXL1*-GFP (Figure 4-12C). E6 and ectoderm (dual SMAD inhibition) were stained and assessed for PAX6 expression (Figure 4-12A). Both conditions demonstrated good cell survival. PAX6 and *MIXL1*-GFP expression was not abundant in either condition (Figure 4-12 D, E), with a small proportion of PAX6 only cells present in the ectoderm condition (Figure 4-12E).

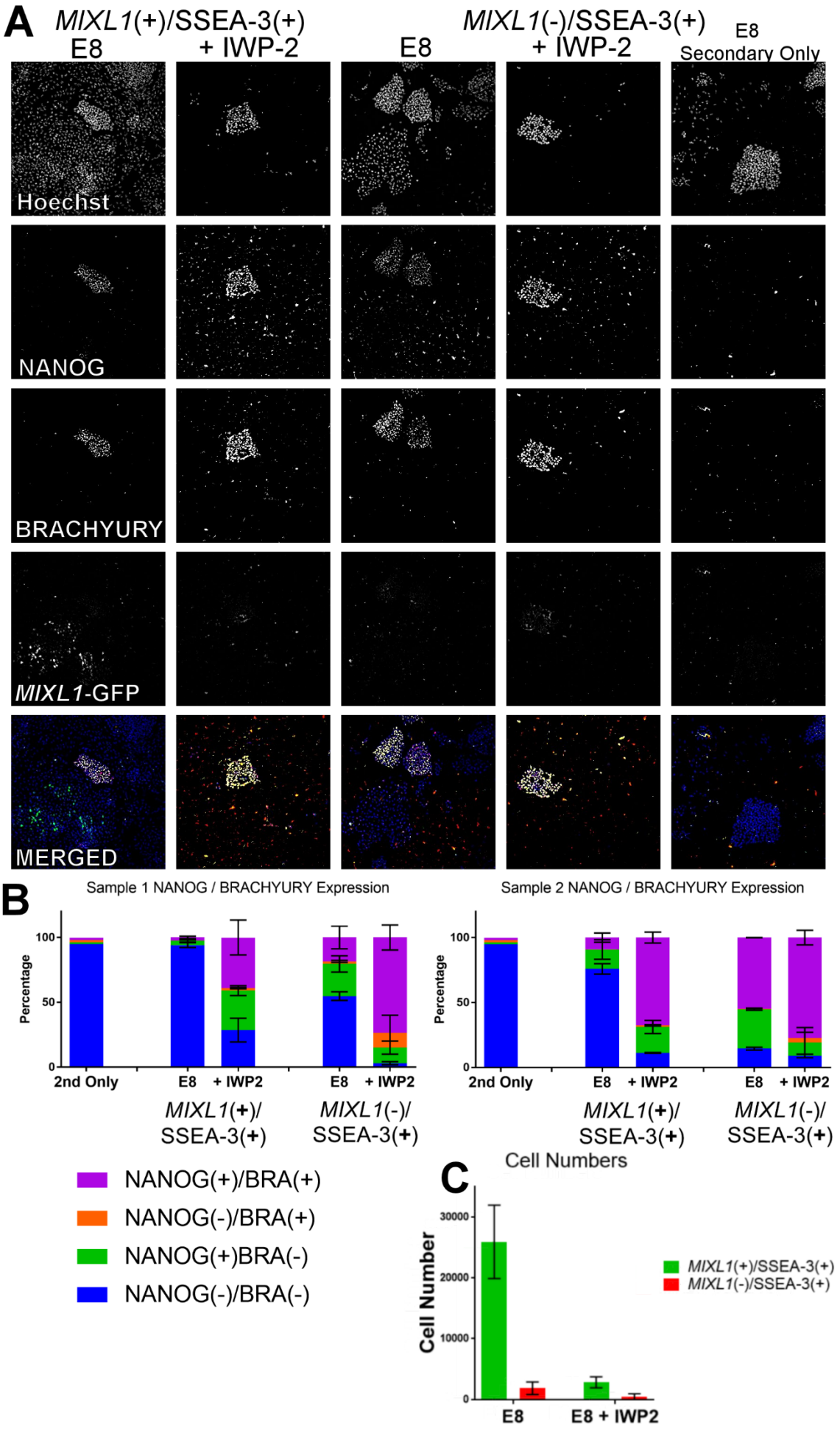

A
MIXL1(+)/SSEA-3(+)
E8
+ IWP-2
E8
MIXL1(-)/SSEA-3(+)
+ IWP-2
E8
Secondary Only
Hoechst
NANOG
BRACHYURY
MIXL1-GFP
MERGED
B
Sample 1 NANOG / BRACHYURY Expression
Sample 2 NANOG / BRACHYURY Expression
Percentage
100
50
0
2nd Only
E8
+ IWP2
E8
+ IWP2
MIXL1(+)/
SSEA-3(+)
MIXL1(-)/
SSEA-3(+)
NANOG(+)/BRA(+)
NANOG(-)/BRA(+)
NANOG(+)BRA(-)
NANOG(-)/BRA(-)
C
Cell Numbers
Cell Number
30000
20000
10000
0
E8
E8 + IWP2
MIXL1(+)/SSEA-3(+)
MIXL1(-)/SSEA-3(+)

Figure 4-13 *MIXL1* Positive and Negative Cells from Primo Medium display survival and self renewal differences.

A) Immunofluorescence analysis of Hoechst, *MIXL1*-GFP, NANOG and BRACHYURY expression of sorted populations of *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(-)/SSEA-3(**+**) from Primo conditions, plated in E8 medium alone, E8 with 1μM IWP-2 and E8 medium (Secondary antibody only staining). A merged image of all four channels is present below Hoechst (Blue), *MIXL1*-GFP (Green), BRACHYURY(Yellow) and NANOG (Red). **B)** Stacked bar chart displaying percentage of cells positive and negative for NANOG and BRACHYURY in the conditions tested (Bars are mean ± SD, n=two technical repeats from two biological repeats).

Plating as single cells into E8 led to a large proportion of cells differentiating, I investigated whether this effect was apparent from both *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(-)/SSEA-3(**+**) populations from Primo conditions. I also investigated as to whether the differentiation could be prevented by the addition on IWP-2 into the medium. Single cells were sorted from each fraction after 3 days in Primo conditions and 30,000 cells were plated per 24 well. Cells were plated in E8 medium containing 10μM Rocki, with and without 1μM IWP-2.

As I saw with single cell plating of unsorted cells, there was apparent differentiation in wells plated in E8 medium containing 10μM of Rocki, from both *MIXL1*(**-**)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**) plated cells (Figure 4-13A). Immunofluorescent analysis of NANOG and BRACHYURY revealed a large proportion of cells to be NANOG and BRACHYURY negative (Figure 4-13B). There was some *MIXL1*-GFP expression within this cell population (Figure 4-13A). Although a large proportion of cells were NANOG negative in this condition, small NANOG positive colonies were also present in the wells. The number of these colonies was higher in the samples which came from *MIXL1*(-)/SSEA-3(**+**) over *MIXL1*(**+**)/SSEA-3(**+**) cells (Figure 4-13B).

In wells which had IWP-2 added, the number of NANOG negative and BRACHYURY negative cells was drastically reduced (Figure 4-13B) but so was the total cell number (Figure 4-13C). Again, I saw small NANOG positive colonies within the wells, but relatively none of the NANOG(-)/*MIXL1*-GFP(**+**) differentiated cells seen previously (Figure 4-13A). Although NANOG was maintained in these colonies, a large proportion of cells still had BRACHYURY expression as well (Figure 4-13). The reduction in cell numbers appears to be a selective effect of the IWP-2 addition, leading to selection of pristine NANOG positive colonies and death of the NANOG(-)/*MIXL1*(**+**) differentiated progeny.

4.2.8. Matrix Assessment.

The matrix on which human PSC grow on can impact their self renewal and differentiation abilities (Chen et al., 2007). While I have used both Vitronectin and Geltrex successfully with the Primo medium, we wanted to assess whether another matrix combination may be better in trapping this biased state. In particular we were focused on whether a certain matrix could overcome the pronounced differentiation we see when cells are plated as single cells. We utilised a microwell format, described in **2.14,** to assess 56 combinations of 10 proteins.

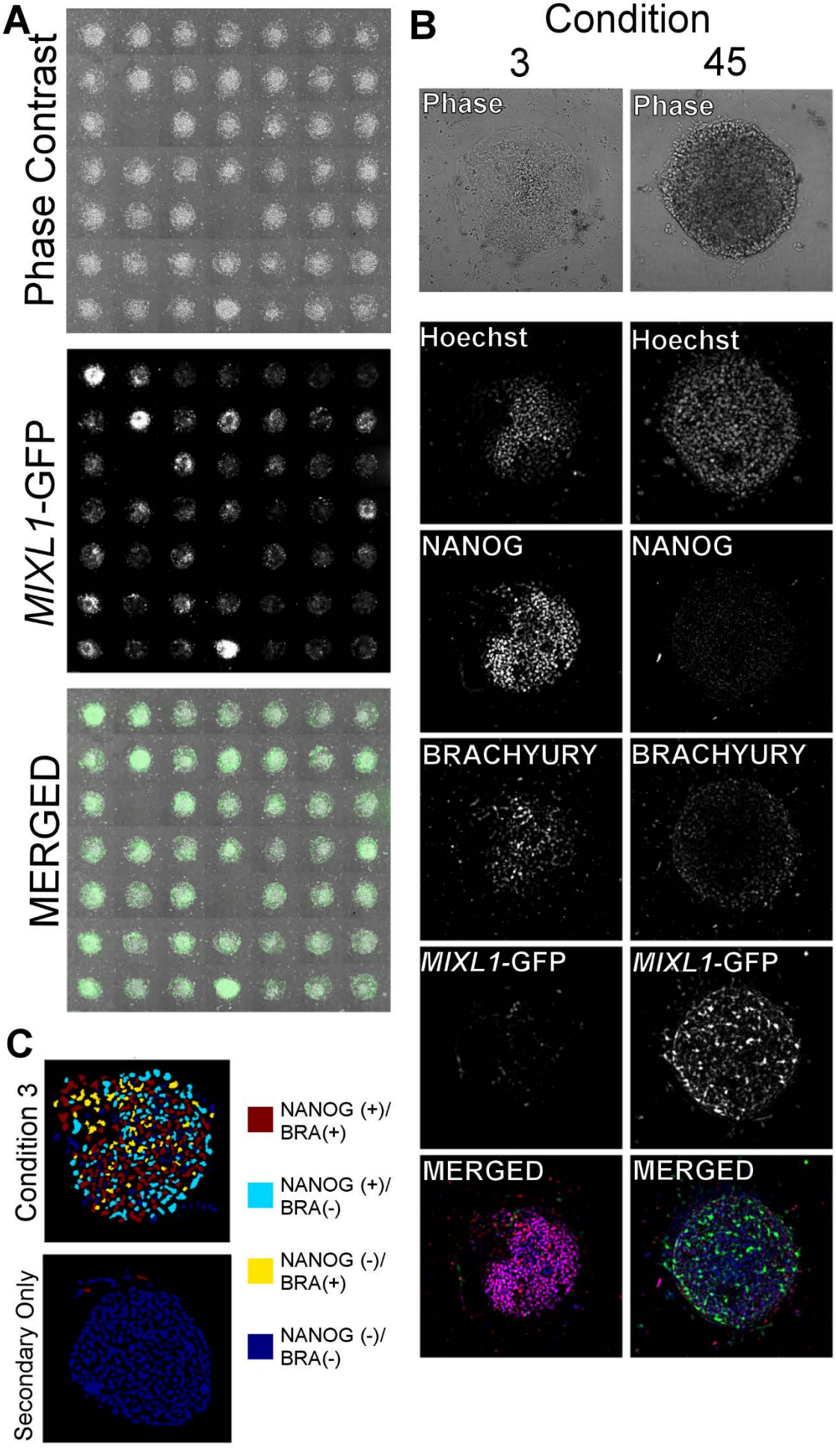
The expression of *MIXL1*-GFP and colony morphology were monitored through live cell imaging (Figure 4-14A). What was apparent during this live imaging was 1xPrimo conditions seemed to not be working as well as I had seen previously. The *MIXL1*-GFP intensity was higher than anticipated and cell morphology appeared more differentiated. This observation was further validated after immunostaining revealed very few NANOG positive cells for all protein combinations when using 1xPrimo (data not shown). Cells in 2xPrimo performed much better, again this was noticed during the live imaging prior to staining. The immunostaining revealed some colonies had higher number of NANOG positive cells than other (Figure 4-14B).

While some colonies exhibited a retention of NANOG and co-expression of BRACHYURY (Figure 4-14C) there wasn't a particular trend between matrix and this retention of NANOG. What appeared to be more critical was the density of the cells within the well. Although I attempt to plate cells at a uniformed density, inevitably in this system some wells will have more cells plating down than another. The intrinsic properties of the protein combinations result in differing cell adherence to the matrix, with human ESC attaching better to certain proteins. Effectively, assessment using this system penalised protein combinations with increased cell adherence, which is counter intuitive to purpose of this experiment. We want to find a matrix which allows for the propagation of our biased state so adherence is a beneficial characteristic, however, in this geometrically confined setup, increased adherence results in differing cell densities between conditions. Regardless, we were able to ascertain the importance of cell density in trapping these states and that increased confluency could lead to differentiation. Further investigation into the dynamic of cell density and differentiation is required to fully understand this relationship, which may highlight the activation of the Hippo pathway by cell to cell contacts as an important aspect.

Figure 4-14 Microwell Matrix Assessment

(Figure on next page)

A) Phase contrast, green fluorescence and merged image of microwell colonies at hour 72 of culture in 2xPrimo. **B)** Representative Phase contrast and immunofluorescent analysis of microwell colonies


A
Phase Contrast
MIXL1-GFP
MERGED
B
Condition
3
45
Phase
Phase
Hoechst
Hoechst
NANOG
NANOG
BRACHYURY
BRACHYURY
MIXL1-GFP
MIXL1-GFP
MERGED
MERGED
C
Condition 3
Secondary Only
NANOG (+)/ BRA(+)
NANOG (+)/ BRA(-)
NANOG (-)/ BRA(+)
NANOG (-)/ BRA(-)

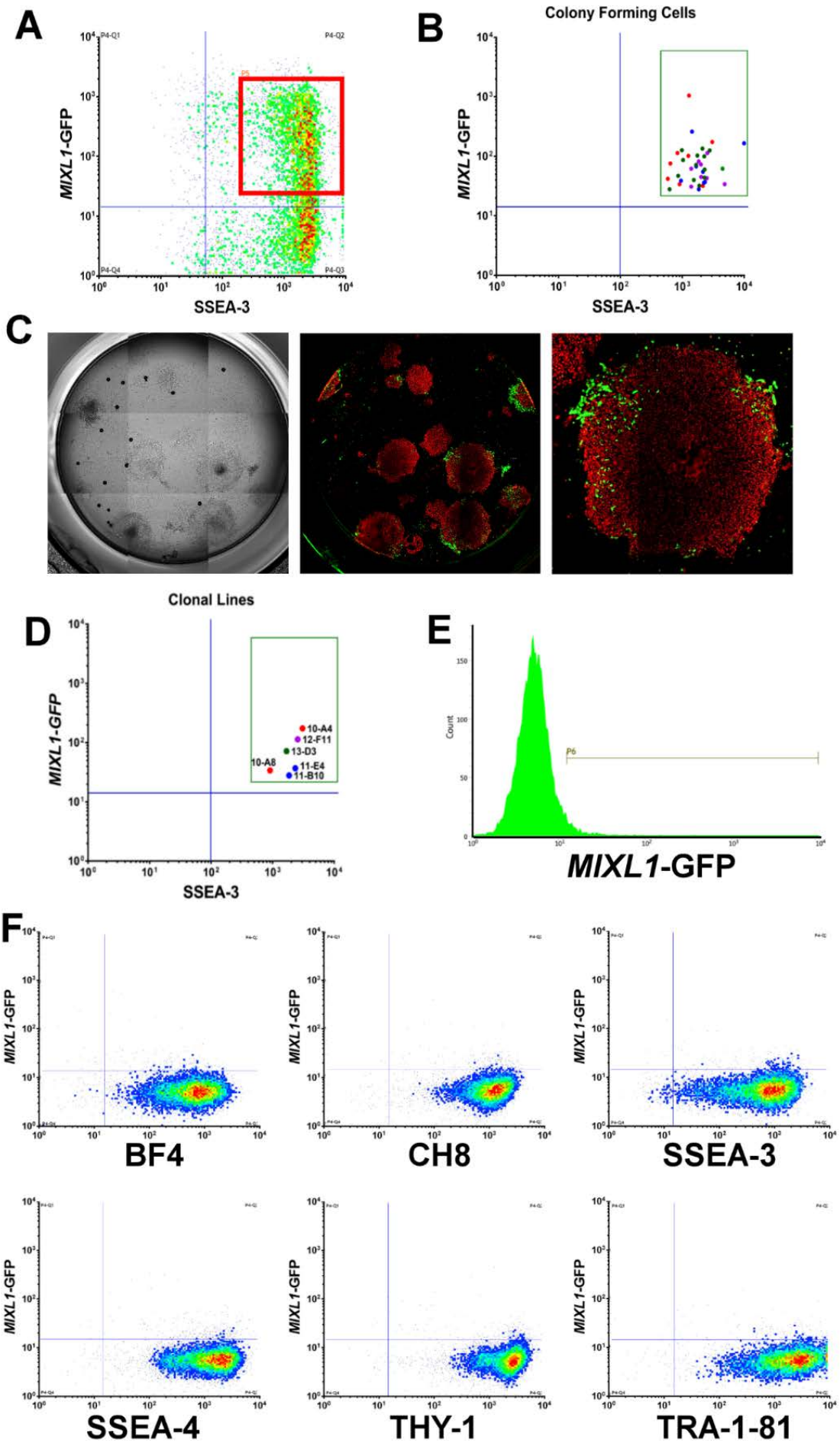

A
B
Colony Forming Cells
MIXL1-GFP
SSEA-3
C
D
Clonal Lines
10-A4
12-F11
13-D3
10-A8
11-E4
11-B10
E
Count
P6
MIXL1-GFP
F
BF4
CH8
SSEA-3
SSEA-4
THY-1
TRA-1-81

Figure 4-15 Single Cell Cloning of Primo *MIXL1*(+)/SSEA-3(+) Cells

A) Flow Cytometry density plots of SSEA-3 versus *MIXL1*-GFP from HES3 *MIXL1* cells grown in 2xPrimo conditions. Red box indicates the sorting gate for the single cell cloning. **B)** Flow cytometry scatter plot of indexed position of colony forming cells. **C)** Representative colony growth post first passage, phase contrast and immunofluorescence images of live TRA-1-81 staining (RED) and *MIXL1*-GFP expression (Green). **D)** Flow cytometry scatter plot of indexed position of clonal lines. **E)** Flow cytometry histogram of *MIXL1*-GFP intensity of clone 10-A4 grown in E8 conditions. **F)** Flow cytometry density plots of clone 10-A4 grown in E8 conditions for surface markers, BF4, CH8(CD9), SSEA-3, SSEA-4, THY1 and TRA-1-81 CH8 versus *MIXL1*-GFP.

4.2.9. Re-cloning of *MIXL1* positive cells from Primo conditions.

As in 3.2.4 I sought to find whether the *MIXL1*(**+**)/SSEA-3(**+**) cells generated in Primo conditions contain cells that could establish long term clonal lines of undifferentiated pluripotent stem cells. I performed indexed single cell sorting into 96 well plates with feeders, Figure 4-15 chronicles this process. I sorted 384 *MIXL1*(**+**)/SSEA-3(**+**) (Figure 4-15A) single cells into 96 well plates and, from these, obtained stem cell like colonies in 38 wells. Indexed data revealed the *MIXL1*-GFP and SSEA-3 expression levels from the initial sorted cells (Figure 4-15B). All 38 colonies were passaged further of which 24 of those survived and were positive for TRA-1-81 staining (Figure 4-15C). Further to this, six clonal lines were established and their initial index position are shown (Figure 4-15D).

After expansion in MEF/KOSR conditions and transition into E8V conditions lines were assessed for their pluripotency-associated surface marker expression and *MIXL1*-GFP expression. All lines displayed high expression levels of the pluripotency-associated markers, BF4, CD9, SSEA-3, SSEA-4, TRA-1-81 (Figure 4-15F). Lines were assessed after passage 5 post single cell cloning and all displayed relatively no *MIXL1*-GFP expression (Figure 4-15E).

For this experiment I decided to re-clone, clone 2-D2 from the original MEF/KOSR clones, in order to highlight the interconversion these cells can undergo. Firstly, the *MIXL1* induction in MEF/KOSR conditions, cloning and establishment of *MIXL1* negative population highlighted in **chapter 3**, then *MIXL1* induction in a defined system followed again by cloning and establishment of *MIXL1* negative population. Demonstrating that these cells have interconverted between *MIXL1* positive and negative states, both in standard culture and through media manipulation.

4.2.9.1. Characterisation of second generation clonal lines

The lines generated from the *MIXL1*(+)/SSEA-3(+) fractions displayed normal human PSC colony morphology growth and SSEA-3 expression as well as relatively low *MIXL1*-GFP expression. I sought to further characterise these clonal lines by examining further surface markers, gene expression and trilineage potential in a "Neutral" Embryoid Body (EB) assess.

Three clonal lines were assessed using human PSC scorecards before and after EB formation. Lines before EB formation, in self renewal conditions E8V, demonstrate a pluripotent signature by qPCR analysis with high expression of pluripotency-associated genes and relatively low expression of genes related to any of the three germ layers. Neutral embryoid bodies display the converse of the self renewal samples, with down regulation of pluripotency-associated genes and an upregulation of genes associated with all three germ layers, ectoderm, mesoderm and endoderm (Figure 4-16). This indicates that these clonal lines are tripotent, a defining characteristic of a pluripotent stem cell.

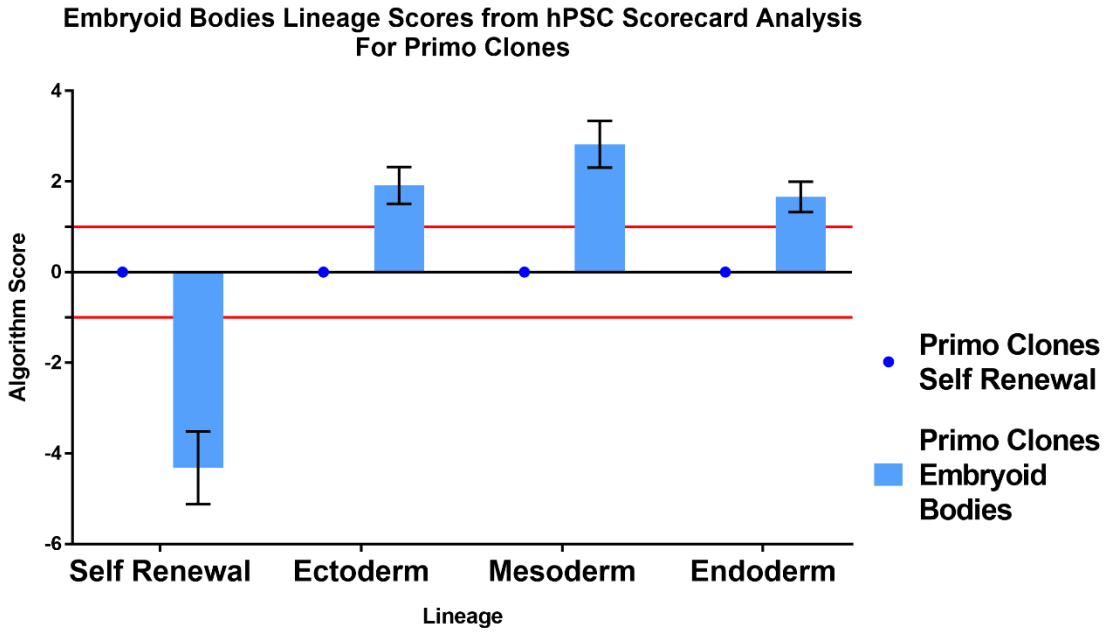


Figure 4-16 Primo Clones Tri-lineage Differentiation Potential

A bar chart displaying the algorithm score for each sample, for self renewal and three lineages, ectoderm, mesoderm and endoderm. Primo Clones in self renewal conditions are compared to day 7 embryoid bodies. The algorithm score is calculated based on the qPCR values for genes of a given lineage, the 0 baseline is based on the average value of undifferentiated samples. EB samples have been normalised to their undifferentiated counterparts. (Bars are mean ± SD, n=single repeat of three independent clones)

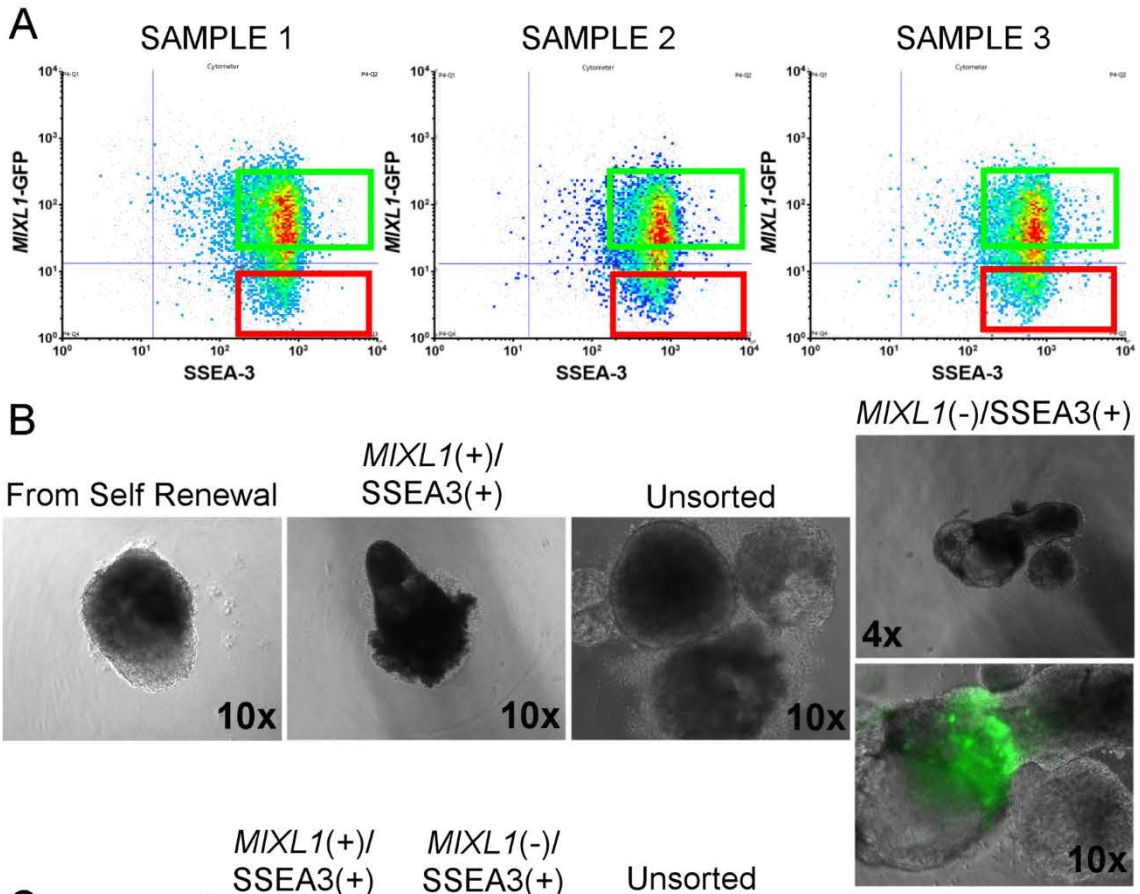

A
SAMPLE 1
SAMPLE 2
SAMPLE 3
MIXL1-GFP
SSEA-3
B
From Self Renewal
MIXL1(+)/ SSEA3(+)
Unsorted
MIXL1(-)/SSEA3(+)
10x
10x
10x
4x
10x
C
MIXL1(+)/ SSEA3(+)
MIXL1(-)/ SSEA3(+)
Unsorted
SAMPLE 1
SAMPLE 2
SAMPLE 3
1) S1 M+3+ EB
2) S1 M-3+ EB
3) S1 U...RTED EB
4. S2 M+3+ EB
5. S2 M-3+ EB
6. S2 U...RTED EB
7. S3 M+3+ EB
8. S3 M-3+ EB
9. S3 U...RTED EB
Self-renew
Ecto
Meso
Endo
D
Embryoid Bodies Lineage Scores from hPSC Scorecard Analysis
Algorithm Score
Self Renewal
Ectoderm
Mesoderm
Endoderm
Lineage
MIXL1 Self Renewal
Primo Clones EB
MIXL1(+)/SSEA-3(+) Primo EB
MIXL1(-)/SSEA-3(+) Primo EB
Unsorted Primo EB

Figure 4-17 Differentiation Potential under "Neutral" Conditions

A) Flow cytometry density plots of three samples of HES3 *MIXL1*-GFP growing in 2x Primo conditions. Sorting gates in red and green indicate position of sorted populations of *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**) cells, respectively. **B)** Phase contrast images of EBs generated from Self renewal conditions, unsorted Primo conditions, Primo *MIXL1*(-)/SSEA-3(**+**) and Primo *MIXL1*(**+**)/SSEA-3(**+**) sorted. Images magnification is specified in the image. *MIXL1*-GFP expression is overlaid in green in the 10x picture of EB from Primo *MIXL1*(-)/SSEA-3(**+**) fraction. **C)** Analysis output from Thermofisher's human PSC scorecard analysis software, displays which lineage signatures were identified. Negative signatures are depicted with grey circles containing a "-" sign. Positive signatures are depicted with coloured circles containing a "**+**" sign. Borderline signatures are depicted with a coloured ring. **D)** A bar chart displaying the algorithm score for each sample, for self renewal and three lineages, ectoderm, mesoderm and endoderm. The algorithm score is calculated based on the qPCR values for genes of a given lineage, the 0 baseline is based on the average value of undifferentiated samples. EB samples have been normalised to their undifferentiated counterparts (bars are mean ±SD, n= three biological repeats).

4.2.10. Assessment of the differentiation bias of human PSC grown in Primo medium

The clone lines of undifferentiated human PSC established from the Primo *MIXL1*(**+**)/SSEA-3(**+**) fraction generated EB containing signatures of all three germ layers, demonstrating the potential for this state to generate trilineage competent human PSC. Whilst this demonstrates that cells can revert back to a more pristine pluripotent state, I now assessed whether cells taken directly from Primo conditions exhibited trilineage potential or a particular lineage bias. I utilised the same "Neutral" EB approach to assess the cells, assessing unsorted cultures as well as the *MIXL1* positive and negative fractions separately.

I grew three separate cultures of HES3 *MIXL1*-GFP in 2xPrimo conditions. We then sorted *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**) from these cultures after 3 days. In all three samples I had high proportion of *MIXL1*(**+**)/SSEA-3(**+**) cells in the cultures and crucially low intensity of *MIXL1*-GFP expression. SSEA-3 also showed high expression. I sorted from the depicted gates and also used an unsorted population to plate for Neutral EB differentiation

(Figure 4-17 A). 3,000 cells were added to each well of a 96 well plate and left to incubate for 7 days. All samples formed EBs of similar size, although on average EBs derived from *MIXL1*(-)/SSEA-3(**+**) were slightly larger and sometimes elongated (Figure 4-17B). Due to their 3D structure monitoring *MIXL1*-GFP expression was difficult but could be seen occasionally, as depicted in Figure 4-17B.

The EBs were harvested for qPCR analysis using the human PSC scorecard from Thermofisher. After running the scorecards, data was analysed using Thermofisher's designed software. The scorecards are designed to predict trilineage potential of human PSC. Analysis for our samples indicated mesoderm signature from all samples coming from Primo conditions, in most cases this was the only positive signature in the analysis (Figure 4-17C). This mesoderm signature was not limited to the *MIXL1*(**+**)/SSEA-3(**+**) fraction either but also present in the *MIXL1*(-)/SSEA-3(**+**) and unsorted fractions. The third sample also had positive signatures for both ectoderm and endoderm in some in unsorted and *MIXL1*(-)/SSEA-3(**+**) fractions, but not ectoderm positive from the *MIXL1*(**+**)/SSEA-3(**+**) fraction.

The human PSC scorecard has a threshold by which it identifies a sample as positive but assessing the algorithm scores revealed an even stronger lineage difference. Figure 4-17D depicts the samples algorithm score for each lineage and self renewal, values have been normalised to undifferentiated control. The 0 point on the graph indicates the alignment score of undifferentiated "self-renewal" sample. EBs derived from standard self renewal conditions exhibited down regulation of self renewal associated genes and an upregulation of genes associated with all three germ layers.

EBs derived from Primo conditions display a similar down regulation of self renewal associated genes and somewhat similar up regulation of endoderm associated genes. However, there was very little upregulation of genes associated with ectoderm and very high upregulation of genes associated with mesoderm (Figure 4-17D). While the alignment score for EBs from self renewal conditions was on average close to 3, all the fractions from Primo conditions generated mesoderm scores close to 7. Under these "Neutral" conditions it appears that cells grown in Primo conditions have an enrichment of mesoderm within their EBs.

4.2.11. Multiple passages in Primo medium.

After optimisation of the medium and the demonstrated ability for cells to revert to a *MIXL1*(-)/SSEA-3(**+**) state, I sought to assess if cells could be maintained in Primo medium for multiple passages. I started with a relatively low passage number HES3-*MIXL1* line with a normal karyotype. I elected to use the 2xPrimo formulation for this, as I saw better maintenance of SSEA-3 post passage using this formulation. The formulation for the first 3 passages still contained 2-mercaptoethanol, and after 3 passages I assessed the *MIXL1* and SSEA-3 expression. Observing colony formation at this stage showed two distinct types of colonies, some of which represented normal stem cell like colonies and others with a very dense centre and seemingly differentiated cells around the periphery (Figure 4-18A). Flow cytometry analysis revealed a mixed population with SSEA-3 positive and negative fractions, and also a range of *MIXL1* positive cells(Figure 4-18B). I sorted single cells from the *MIXL1*(**+**)/SSEA-3(**+**) fraction at passage 3 and a selection of these cells formed colonies.(Figure 4-18C). Intracellular immunostaining demonstrated that I could still obtain *NANOG* positive colonies, from cells at passage 3 (Figure 4-18D).

After 3 passages the 2-mercaptoethanol was removed and I immediately noticed much better growth in this condition. Colonies observed after passage 4 exhibited a much better colony morphology compared to passage 3, with colonies appearing similar to traditional human PSC colonies (Figure 4-19A). The cells were passaged further in this medium, without 2-mercaptoethanol. At day one of passage 5 I could see survival of small human PSC colonies post passage (Figure 4-19B). Previously in medium containing 2-mercaptoethanol, there was some induction of death of particularly small colonies and a selective pressure for larger colonies to survive. After 3 days growth, the culture was stained for SSEA-3 and assessed. Cells were 99% positive for SSEA-3 and a *MIXL1*(**+**)/SSEA-3(**+**) was still present at 25% of the culture (Figure 4-19C)

Observing this *MIXL1*-GFP expression at passage 5 I assessed the ability to revert the cells to a *MIXL1*(-)/SSEA-3(**+**) by passaging cells into 2xBCL (0.96µM LPA) medium and E8 with 1µM IWP-2. After 3 days of growth in these conditions I reassessed the *MIXL1*-GFP and SSEA-3 expression. Cultures grown in these conditions showed a reduction of *MIXL1*-GFP expression to less than 1% and a maintenance of high SSEA-3 expression (Figure 4-19D-E).

The cells were continually passaged in 2xPrimo and at passage 7, I assessed the expression of a panel of pluripotency-associated surface markers. Cells were stained for BF4, CD9, SSEA-3, SSEA-4, THY1 and TRA-1-81. All antigens showed high levels of expression, over 95%. The culture also still exhibits a *MIXL1* positive population, approximately 20% (Figure 4-20).

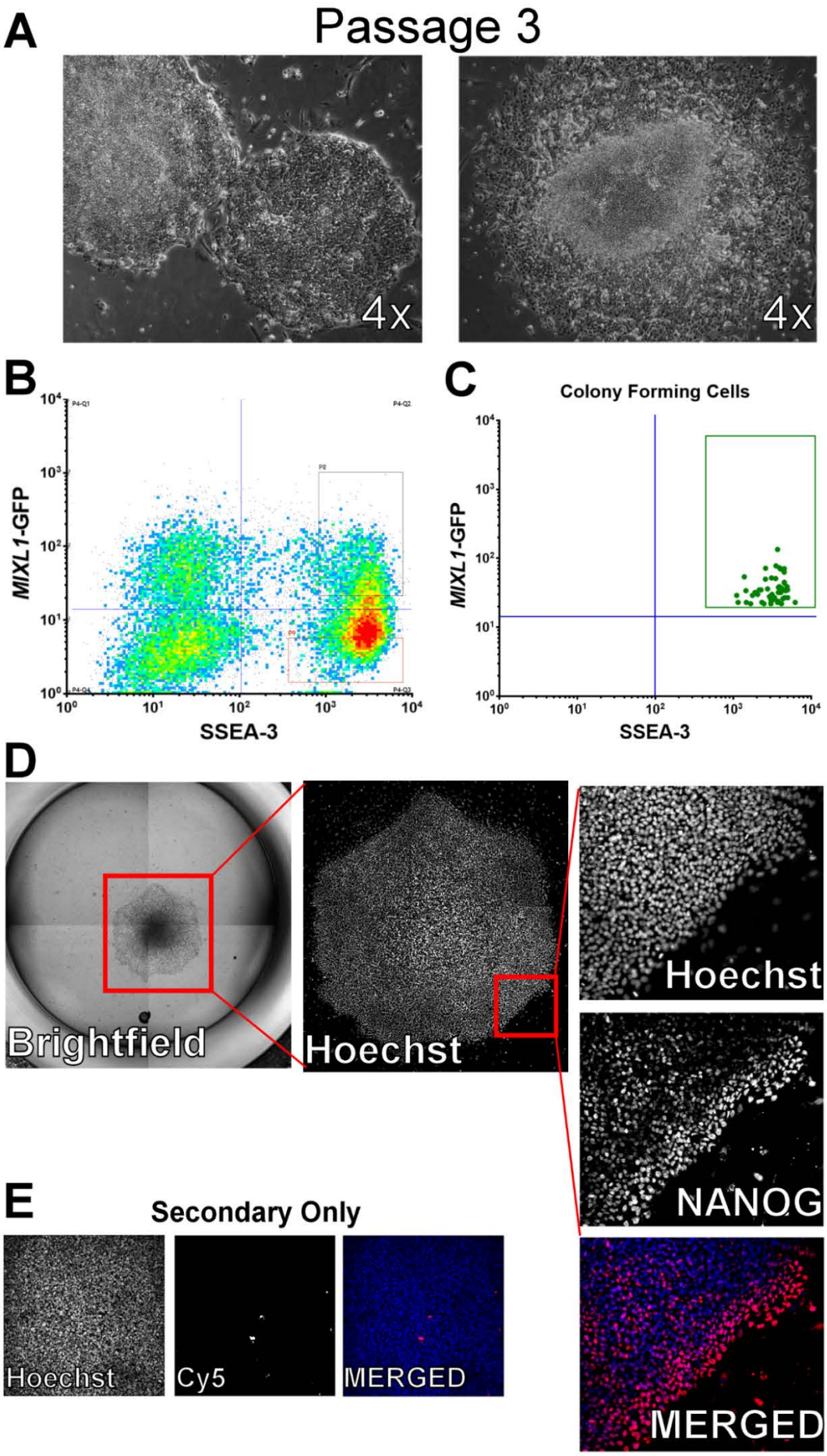

A
Passage 3
4x
4x
B
MIXL1-GFP
SSEA-3
C
Colony Forming Cells
MIXL1-GFP
SSEA-3
D
Brightfield
Hoechst
Hoechst
NANOG
MERGED
E
Secondary Only
Hoechst
Cy5
MERGED

Figure 4-18 NANOG Positive Colonies Generated from *MIXL1*(+)/SSEA-3(+) Single Cells after 3 Passages in 2xPrimo.

A) Phase contrast images at 4x magnification of representative colonies at passage 3 in 2xPrimo. **B)** Flow cytometry density plot of SSEA-3 versus *MIXL1*-GFP expression. **C)** Indexed positions of single cells which formed stem cell like colonies post sorting. **D)** Phase contrast and immunofluorescence analysis of a representative colony from single cell cloning. Shown is a 4x brightfield image, 4x Hoechst image, then 10x Hoechst, and NANOG expression. The merged image displays Hoechst in blue and NANOG in red. **E)** Hoechst and Cy5 (NANOG Channel) of secondary only stained colony, the merged image displays Hoechst in blue and Cy5 in red.

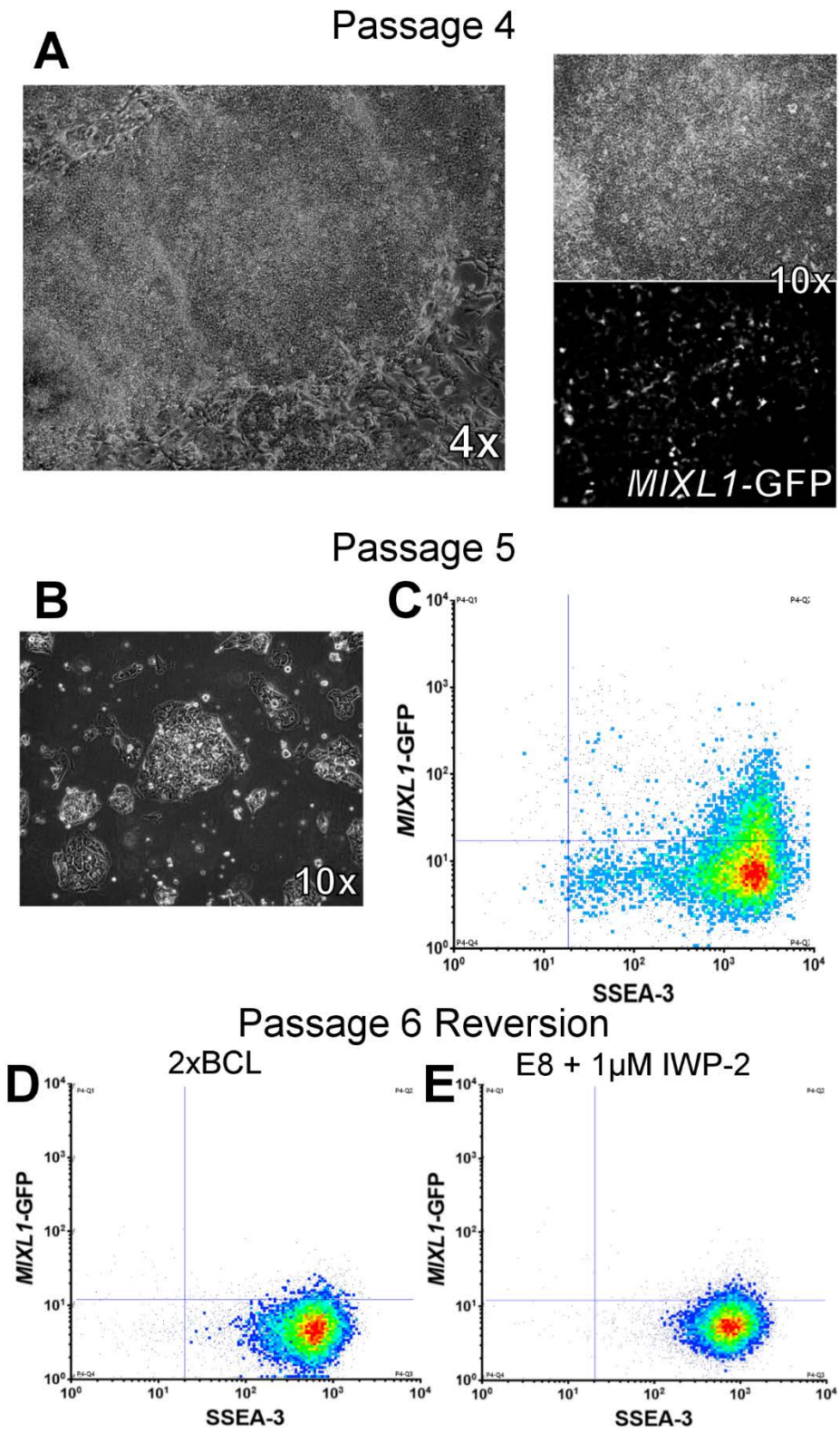


Figure 4-19 Colony Morphology and *MIXL1*-GFP/SSEA-3 Expression Throughout Passaging

A) Phase contrast and *MIXL1*-GFP fluorescent images at 4x and 10x magnification of representative colonies at passage 4 in 2xPrimo after removal of 2-mercaptoethanol. **B)** Phase contrast image of cells one day post passage 5. **C)** Flow cytometry density plot of SSEA-3 versus *MIXL1*-GFP expression at passage 5. **D-E)** Flow cytometry density plot of SSEA-3 versus *MIXL1*-GFP expression at passage 6 after 5 passages in 2xPrimo and reversion in presence of LPA **(D)** and IWP-2**(E).**

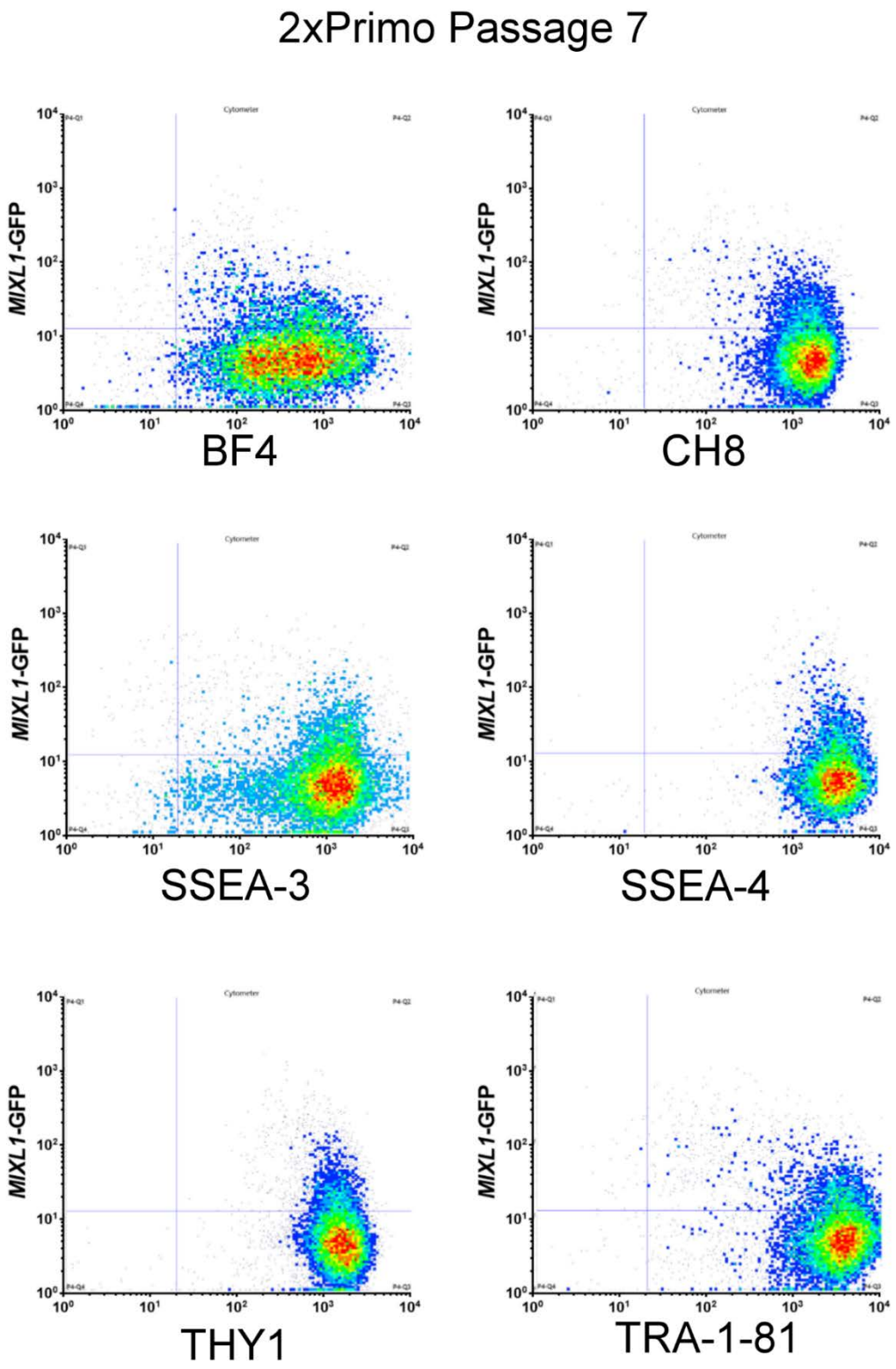


Figure 4-20 Pluripotency-associated Surface Marker Expression of Cells Growing in 2xPrimo

Flow cytometry density plots of *MIXL1*-GFP versus a given pluripotency-associated surface marker for HES3 *MIXL1*-GFP at passage 7 in 2xPrimo conditions. Markers analysed are BF4, CH8(CD9), SSEA-3, SSEA-4, THY1 and TRA-1-81. Cells show high expression of all markers analysed and a *MIXL1* positive population present.

4.2.12. Characteristics and reversion potential of lineage biased human ESC after 10 Passages in Primo

After the tenth passage in 2xPrimo medium, I assessed the cells within 2xPrimo conditions and the effect on the cells when split into different culture media. Cells were split into 1xPrimo, and 2xPrimo to assess the maintenance of the lineage biased population, *MIXL1*(**+**). Cells were also split into 2xBCL (E8 with LPA), E8 with and without 1µM IWP-2 to assess the cells ability to revert to a *MIXL1*(-) state. The expression of *MIXL1*-GFP expression in the cultures, except 1xPrimo, was monitored over 3 days post passage using time-lapse microscopy. In 2xPrimo conditions as colonies developed there was an increase in *MIXL1* positive cells in the colonies with a relatively low intensity (Figure 4-21A). Colonies monitored in 2xBCL and E8, both with and without 1µM IWP-2, showed little detectable *MIXL1*-GFP over the three days, indicating reversion to a *MIXL1*(-) state.

It is possible that the reversion, generation of *MIXL1*(-)/SSEA-3(**+**) cells population, arises from selection for *MIXL1*(-)/SSEA-3(**+**) cells in the Primo culture. In order to assess whether reversion was a selective process I also assessed the growth rates of these cells under the different conditions. Growth rates were calculated using time lapse microscopy of cells, assessing a 5x5 grid in the middle of wells at 10x magnification. CL-quant software (NIKON) was used to calculate the area covered by cells from phase contrast images approximately every 2 hours between 8-64 hours in culture. The growth rates were comparable between 2xPrimo, E8 and E8 **+** IWP-2, with similar increases in percentage area covered and minimal cell death observed (Figure 4-21B). 2x BCL also showed very minimal death but due to the excessively spread cell morphology exhibited by cells under just LPA, tracking the area covered was not suitable for these cultures. Hoechst staining of these cultures demonstrates high cell numbers present in this culture as well (Figure 4-21C).

Immunofluorescent analysis of NANOG expression demonstrated high expression in all conditions, while *MIXL1*-GFP expression was mainly confined to 2xPrimo conditions only (Figure 4-21C). Assessing the NANOG expression with *MIXL1*-GFP expression there were colonies within the 2xPrimo conditions which showed coexpression of *MIXL1*-GFP and NANOG, with ~45% of the culture identified as double positive (Figure 4-22A). The other conditions demonstrated a significant reduction in *MIXL1*-GFP expression, but a maintenance of high percentage of NANOG expression (Figure 4-22B).

The cultures were assessed with another pluripotency-associated marker, SOX2, and coupled this with YAP staining. Immunofluorescent analysis shows high percentage expression of SOX2

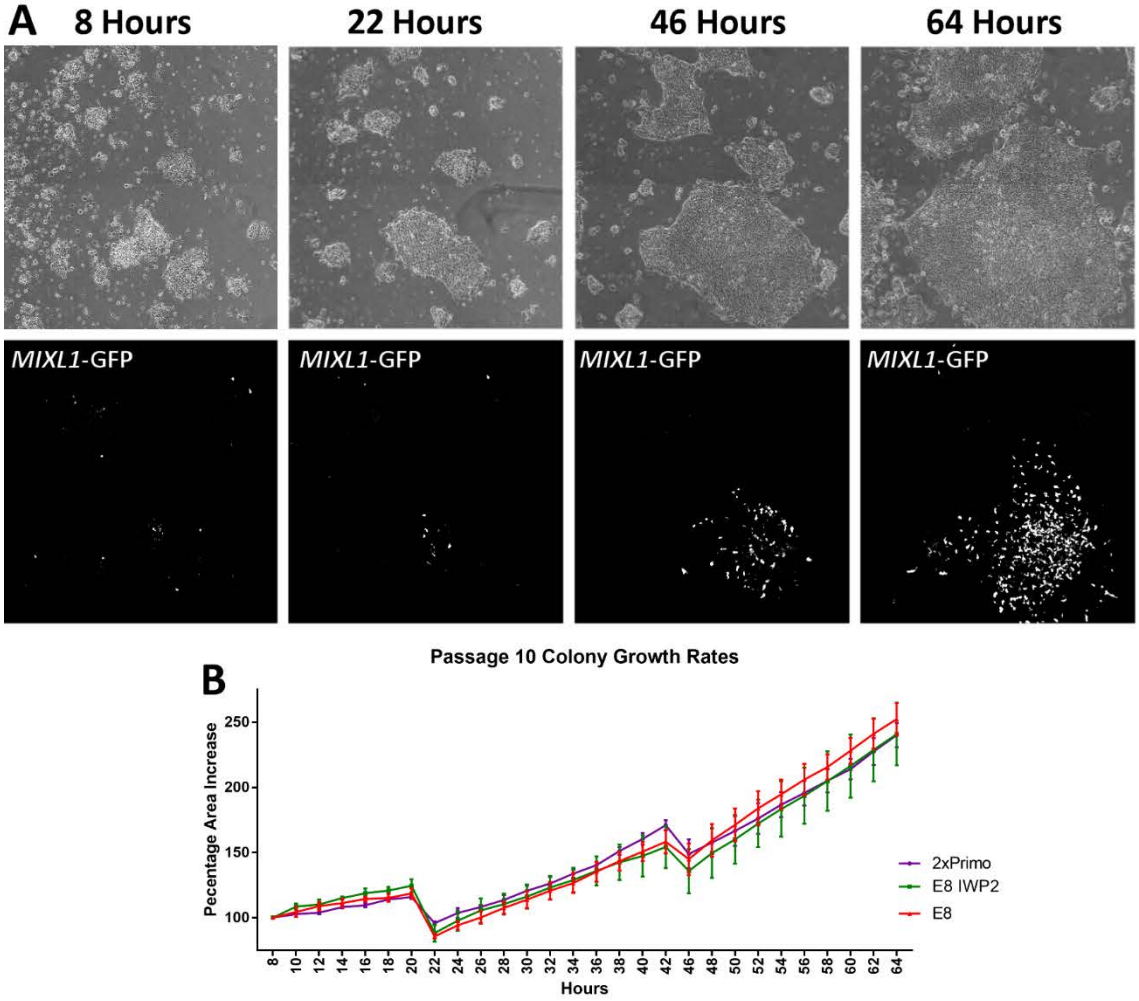

A
8 Hours
22 Hours
46 Hours
64 Hours
MIXL1-GFP
MIXL1-GFP
MIXL1-GFP
MIXL1-GFP
B
Passage 10 Colony Growth Rates
Pecentage Area Increase
250
200
150
100
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
46
48
50
52
54
56
58
60
62
64
Hours
2xPrimo
E8 IWP2
E8
C
2xPrimo
2xBCL
E8
E8 + IWP-2
2xBCL
Secondary Only
HOECHST
HOECHST
HOECHST
HOECHST
HOECHST
NANOG
NANOG
NANOG
NANOG
MIXL1-GFP
MIXL1-GFP
MIXL1-GFP
MIXL1-GFP
MIXL1-GFP
MERGED
MERGED
MERGED
MERGED
MERGED

Figure 4-21 Passage 10 Growth Analysis and NANOG Expression

A) Phase contrast and green fluorescent images of HES3 *MIXL1*-GFP growing in 2xPrimo conditions after the tenth passage in this condition. Cells were imaged every 2 hours between 8 and 64 hours post plating. **B)** The graph displays colony growth rates, calculated as percentage covered increase, for cells growing in 2xPrimo, E8 and E8 **+** 1µM IWP-2 post 9 passages in 2xPrimo. **C)** Immunofluorescence analysis of Hoechst, *MIXL1*-GFP, and NANOG expression of HES3 *MIXL1*-GFP cells in 2xPrimo, 2xBCL, E8 alone, E8 with 1µM IWP-2 added and 2xBCL (Secondary antibody only staining) for 3 days post to 9 passages in 2xPrimo (Bars are mean ± SD, n=three technical repeats). A merged image of all four channels is present below Hoechst (Blue), *MIXL1*-GFP (Green), and NANOG(red).

positive cells in all conditions. The majority of these SOX2 positive cells also demonstrated nuclear YAP localisation. As I had seen previously, nuclear YAP correlated very well with the expression of a pluripotency-associated marker (Figure 4-23)

I also assessed *MIXL1*-GFP and SSEA-3 expression in these cultures by flow cytometry. Cells passaged into 1xPrimo and 2xPrimo had a large proportion of *MIXL1*(**+**)/SSEA-3(**+**) ~40% and ~15%, respectively at 3 days post passage (Figure 4-24A). Cultures passaged into 2xBCL and E8 with IWP-2 had decreased levels of *MIXL1*(**+**)/SSEA-3(**+**) cells ~3% and ~8% respectively, the intensity of *MIXL1* was also much lower in these cultures compared to Primo cultures (Figure 4-24A). I further passaged the cells from 2xBCL and E8 with IWP-2 cultures into standard E8 medium to check if the *MIXL1* expression was lost after passage, from both condition *MIXL1* expression had decreased to almost 0% (Figure 4-24B).

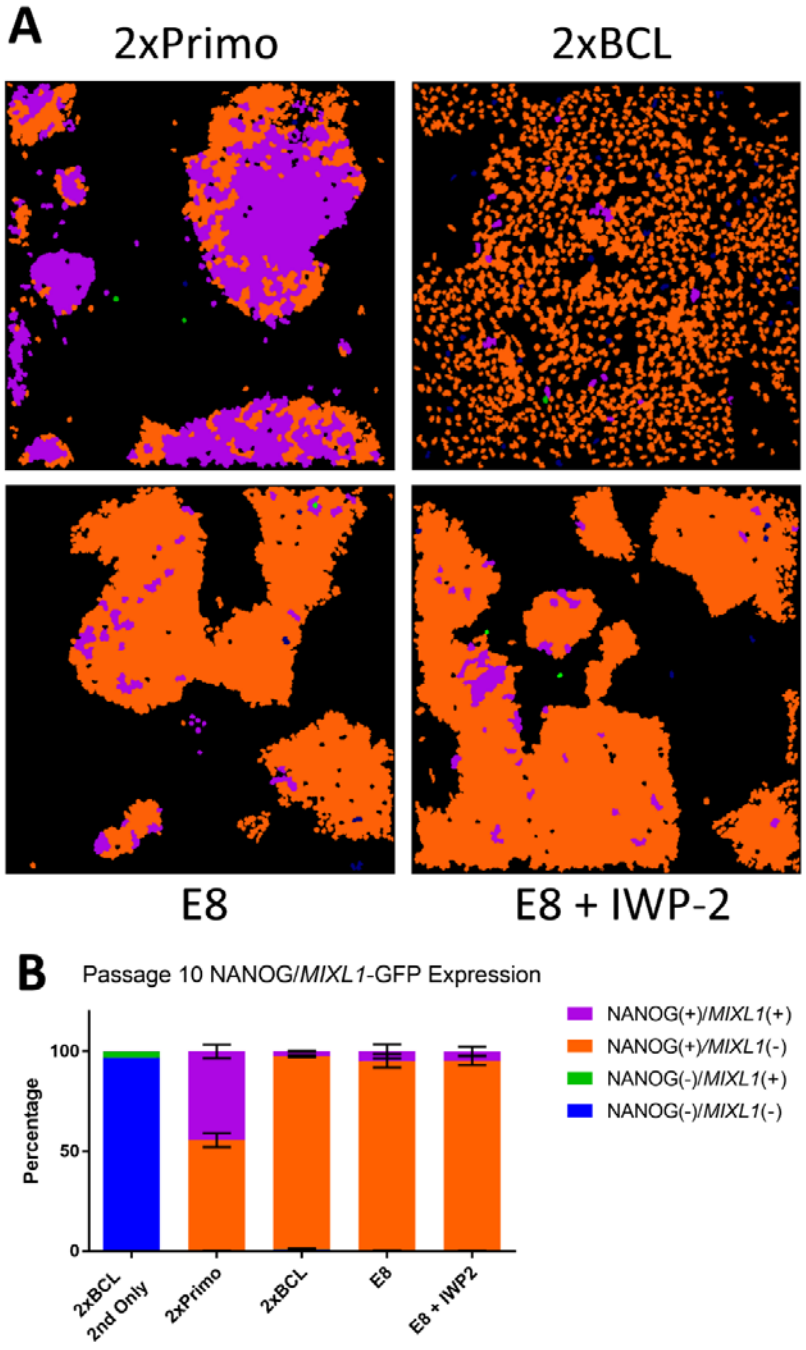


Figure 4-22 Passage 10 NANOG and *MIXL1*-GFP Expression Analysis

Cells maintained in 2xPrimo show a large proportion of NANOG(**+**)/*MIXL1*(**+**) cells which is greatly reduced when passaged into other conditions assessed. **A)** Examples of image analysis for HES3 *MIXL1*-GFP colonies immunofluorescence imaged for *MIXL1*-GFP, and NANOG expression grown in 2xPrimo, 2xBCL, E8 alone and E8 with 1µM IWP-2 added, for 3 days post to 9 passages in 2xPrimo. Colonies are coloured according to positive and negative status of cells for *MIXL1*-GFP and NANOG expression, NANOG(-)/*MIXL1*(-) (Blue), NANOG(-)/*MIXL1*(**+**) (Green), NANOG(**+**)/*MIXL1*(-) (Orange) and NANOG(**+**)/*MIXL1*(**+**) (Purple). **B)** Stacked percentage bar charts displaying cell profiler analysis of 3 wells for each condition (Bars are mean ± SD, n= three technical repeats).

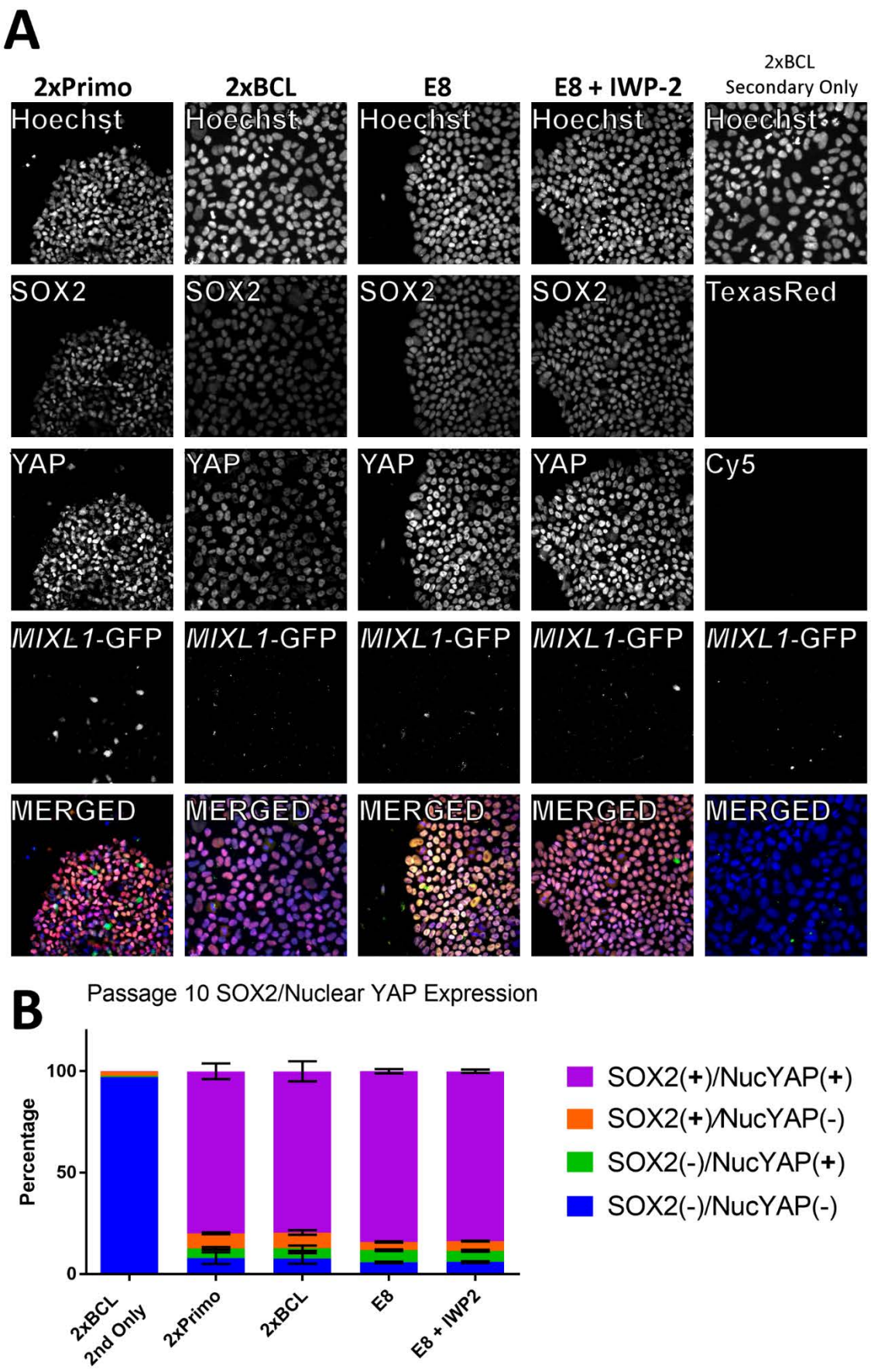

A
2xPrimo
2xBCL
E8
E8 + IWP-2
2xBCL
Secondary Only
Hoechst
SOX2
TexasRed
YAP
Cy5
MIXL1-GFP
MERGED
B
Passage 10 SOX2/Nuclear YAP Expression
Percentage
100
50
0
2xBCL
2nd Only
2xPrimo
2xBCL
E8
E8 + IWP2
SOX2(+)/NucYAP(+)
SOX2(+)/NucYAP(-)
SOX2(-)/NucYAP(+)
SOX2(-)/NucYAP(-)

Figure 4-23 Passage 10 Analysis of SOX2 Expression and YAP Localisation

Cells in all conditions at the tenth passage show high expression of SOX2 and Nuclear Yap **A)** Immunofluorescence analysis of Hoechst, *MIXL1*-GFP, SOX2 and YAP expression of HES3 *MIXL1*-GFP cells in 2xPrimo, 2xBCL, E8 alone, E8 with 1µM IWP-2 added and 2xBCL (Secondary antibody only staining) for 3 days post to 9 passages in 2xPrimo. A merged image of all four channels is present below Hoechst (Blue), *MIXL1*-GFP (Green), SOX2 (Yellow) and YAP (Red). **B)** Stacked bar chart displaying percentage of SOX2 and Nuclear YAP (NucYap) in the conditions tested, SOX2(-)/NucYAP(-) (Blue), SOX2 (-)/NucYAP (**+**) (Green), SOX2 (**+**)/NucYAP (-) (Orange) and SOX2 (**+**)/NucYAP (**+**) (Purple) (Bars are mean ± SD, n= three technical repeats).

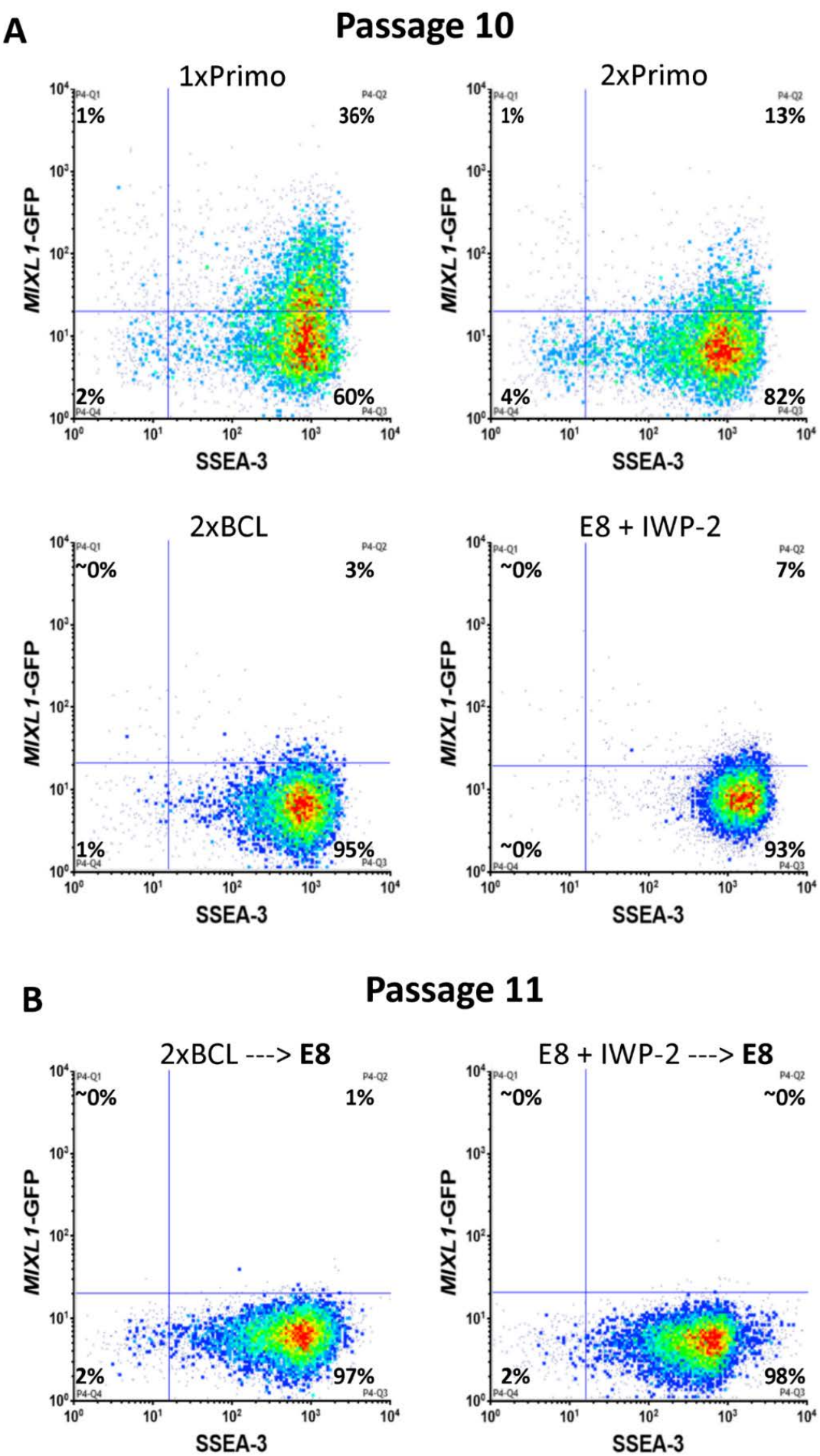

A
Passage 10
1xPrimo
2xPrimo
2xBCL
E8 + IWP-2
MIXL1-GFP
SSEA-3
1%
36%
2%
60%
1%
13%
4%
82%
~0%
3%
1%
95%
~0%
7%
~0%
93%
B
Passage 11
2xBCL ---> E8
E8 + IWP-2 ---> E8
~0%
1%
2%
97%
~0%
~0%
2%
98%

Figure 4-24 Passage 10 - 11 Flow Cytometric Analysis of *MIXL1*-GFP and SSEA-3

A-B) Flow cytometry density plots of SSEA-3 versus *MIXL1*-GFP expression for cells grown in various conditions. **A)** 1xPrimo, 2xPrimo, 2xBCL and E8 with 1µM IWP-2, for 3 days post to 9 passages in 2xPrimo. **B)** Passage 10 cells grown in 2xBCL and E8 with 1µM IWP-2 were further passaged into standard E8 medium alone and analysed after 4 days growth.

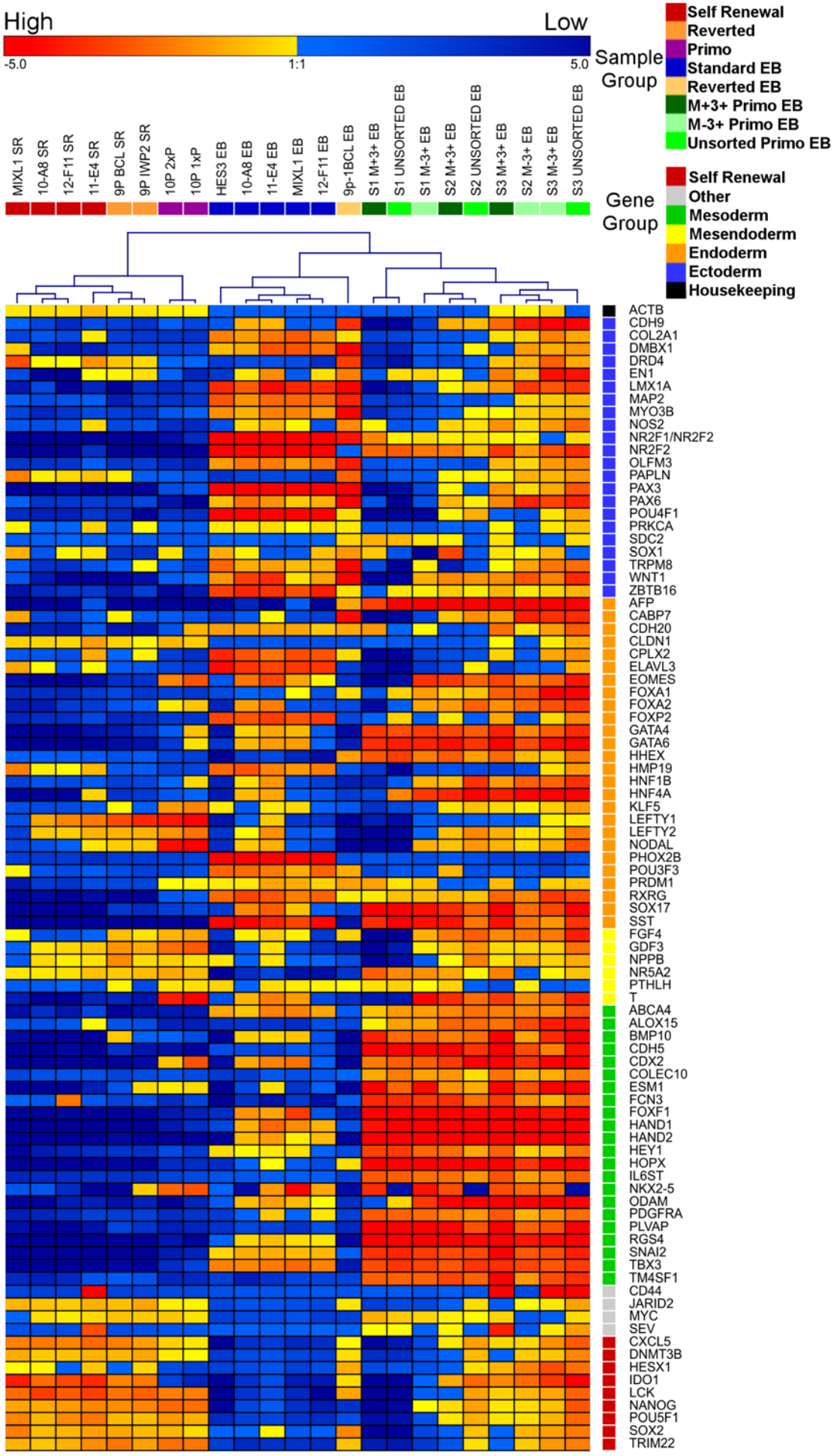

High
Low
-5.0
1:1
5.0
Sample Group
Self Renewal
Reverted
Primo
Standard EB
Reverted EB
M+3+ Primo EB
M-3+ Primo EB
Unsorted Primo EB
Gene Group
Self Renewal
Other
Mesoderm
Mesendoderm
Endoderm
Ectoderm
Housekeeping
MIXL1 SR
10-A8 SR
12-F11 SR
11-E4 SR
9P BCL SR
9P IWP2 SR
10P 2xP
10P 1xP
HES3 EB
10-A8 EB
11-E4 EB
MIXL1 EB
12-F11 EB
9p-1BCL EB
S1 M+3+ EB
S1 UNSORTED EB
S1 M-3+ EB
S2 M+3+ EB
S2 UNSORTED EB
S3 M+3+ EB
S2 M-3+ EB
S3 M-3+ EB
S3 UNSORTED EB
ACTB
CDH9
COL2A1
DMBX1
DRD4
EN1
LMX1A
MAP2
MYO3B
NOS2
NR2F1/NR2F2
NR2F2
OLFM3
PAPLN
PAX3
PAX6
POU4F1
PRKCA
SDC2
SOX1
TRPM8
WNT1
ZBTB16
AFP
CABP7
CDH20
CLDN1
CPLX2
ELAVL3
EOMES
FOXA1
FOXA2
FOXP2
GATA4
GATA6
HHEX
HMP19
HNF1B
HNF4A
KLF5
LEFTY1
LEFTY2
NODAL
PHOX2B
POU3F3
PRDM1
RXRG
SOX17
SST
FGF4
GDF3
NPPB
NR5A2
PTHLH
T
ABCA4
ALOX15
BMP10
CDH5
CDX2
COLEC10
ESM1
FCN3
FOXF1
HAND1
HAND2
HEY1
HOPX
IL6ST
NKX2-5
ODAM
PDGFRA
PLVAP
RGS4
SNAI2
TBX3
TM4SF1
CD44
JARID2
MYC
SEV
CXCL5
DNMT3B
HESX1
IDO1
LCK
NANOG
POU5F1
SOX2
TRIM22

Figure 4-25 Human PSC Scorecard Heatmap Analysis

Heatmap of samples assessed by the human PSC scorecard. Ct values for each gene are normalised to *ACTB.* Heatmap colouring is done after mean centring the genes across the samples. Hierarchical clustering was performed on the samples. Samples are colour coded. Self Renewal (SR) refers to samples grown in E8V conditions. Genes are ordered according to the grouping indicated in the gene group colour key. Three main clusters were identified by hierarchical clusters. Cluster one consisted of samples from self-renewal and Primo conditions. Cluster two contains EBs generated from self-renewal conditions (E8). Cluster 3 contains EBs generated from Primo conditions.

4.2.13. Gene Expression of Passage 10 Samples

Cells grown in 1xPrimo (10P 1xP), 2xPrimo (10P 2xP), 2xBCL (9P BCL SR) and E8 **+** IWP-2 (9P IWP2 SR) were harvested for gene expression analysis by human PSC scorecards (Figure 4-25). I compared this gene expression to data obtained from pluripotent and differentiated cells. For comparisons to the pluripotent state I used gene expression data for cells grown in self-renewal conditions (E8 and Vitronectin), this included the parental HES3 *MIXL1*-GFP line and three Primo clones, 10-A8, 11-E4 and 12-F11. For comparisons to differentiated states I utilised gene expression data from Embryoid Bodies generated from cells grown in self renewal (Standard EB) and Primo (Primo EB) conditions. Hierarchical clustering segregated the cells into 3 clear clusters (Figure 4-25).

The first cluster contains samples grown under standard self-renewal conditions, this includes three clonal lines derived from Primo *MIXL1*(**+**)/SSEA-3(**+**) fraction. The cells at passage ten from all conditions showed high expression of pluripotency-associated genes and clustered with parental HES3 *MIXL1* line grown under self-renewal conditions, E8V. The samples which have been growing in Primo for 9 passages and then reverted to *MIXL1*-GFP negative, in the presence of IWP-2 or BCL (LPA) form a sub-cluster with the standard self-renewal cells. Both 10 passage cells grown in Primo, 1x and 2x Primo, are within the main cluster but are separated into their own sub-cluster. Cells from both 1x and 2x Primo had elevated expression of differentiation associated markers such as *EOMES*, *FOXA2* and *T* whilst maintaining similar levels of *NANOG, POU5F1,* and *SOX2* to the other samples in this first cluster.

The second cluster contains EBs made from cells grown under standard self-renewal conditions (E8 and Vitronectin). EBs from clonal lines derived from Primo *MIXL1*(**+**)/SSEA-3(**+**) fraction and grown in E8 and vitronectin clusters with the EBs derived from parental HES3 and HES3 *MIXL1*-

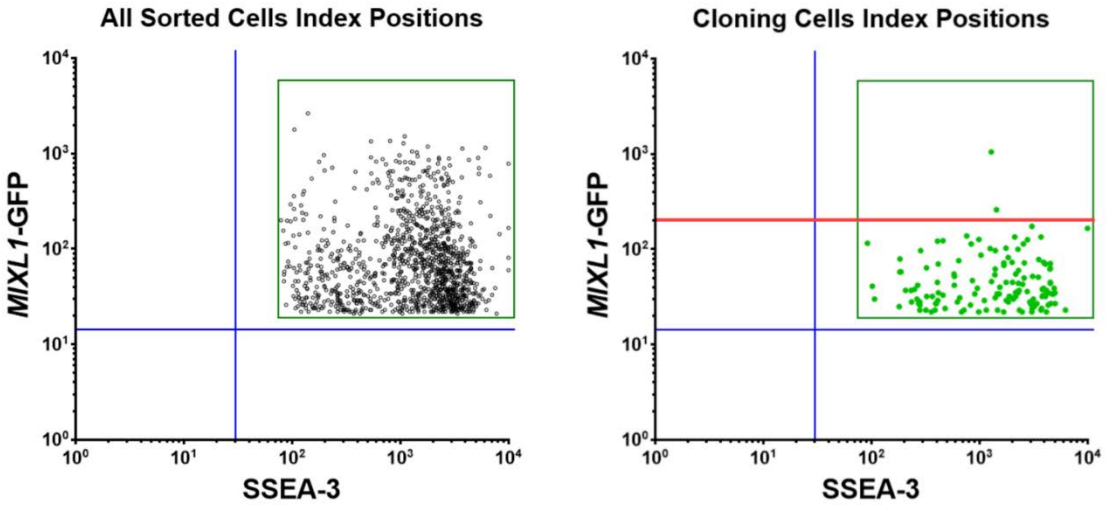
GFP line in self renewal conditions prior to EB formation. EBs generated from cells growing in Primo for 9 passages and then reverted to *MIXL1*-GFP negative, in the presence of LPA (BCL) also reside in this cluster albeit as an outgroup of this cluster.

The third cluster encompasses all the samples which came directly from 2xPrimo conditions, including both unsorted and sorted samples *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**). This cluster has some upregulation of ectoderm and endoderm associated genes but has a greater upregulation of genes associated with mesoderm. Downstream mesoderm markers such as *HAND1* and *HAND2* are strongly upregulated in these samples.

Analysis using the human PSC scorecards demonstrates a stark difference between the EBs generated from self-renewal conditions and that of EBs generated from Primo conditions, in respect to expression of mesoderm associated genes, whilst also demonstrating that both clonal lines and reverted samples exhibit similar EB gene expression patterns as EBs generated from self-renewal conditions.

4.2.14. *MIXL1* Intensity and loss of cloning ability

Unlike mouse PSC cells, human PSC have been hampered by low cloning ability. The use of ROCKi, (Y-27632) increased cloning efficiencies from ~ 1% to between 10-20% (Watanabe et al., 2007). Throughout the project, in different experiments, I have sorted a total of1152 single cells from the *MIXL1*(**+**)/SSEA-3(**+**) fraction obtained from cultures in various media conditions and from these obtained a total of 136 colonies. When I observed the index position data of these cells, I noticed a *MIXL1*-GFP intensity threshold for cloning samples. In all conditions there was a distinct drop in cloning efficiency when the intensity of *MIXL1*-GFP was over 10^2.3 on the log scale (Figure 4-26). In most conditions I saw no colonies formed beyond this point except for two cells from 2xPrimo conditions, and none of these cells went on to make our established clonal lines. When assessing all the indexed positions there were 177 cells sorted with an intensity above 10^2.3, of which only 2 cells went on to form colonies, equating to a cloning efficiency of 1.1% (Figure 9). Whereas cells with an intensity below 10^2.3, had a cloning efficiency of 13.7%, 134 out of 975, this is similar to normal cloning efficiencies using ROCKi. This finding is consistent with our notion that maintaining a *MIXL1*-GFP low expression is important for trapping cells in a biased state and validates the use of Primo culture medium for this purpose.



Condition	MIXL1 FACS intensity	Number of Cells	Number of Clones	Efficiency %
MEF/KOSR	All	288	46	16.0
	Below 10^2.3	264	46	**17.4**
	Above 10^2.3	24	0	0.0
3 days Primo	All	384	38	9.9
	Below 10^2.3	244	36	**14.8**
	Above 10^2.3	140	2	1.4
3 Passages Primo	All	480	52	10.8
	Below 10^2.3	467	52	**11.1**
	Above 10^2.3	13	0	0.0
All conditions	All	1152	136	11.8
	Below 10^2.3	975	134	**13.7**
	Above 10^2.3	177	2	1.1

Figure 4-26 *MIXL1* Intensity and Cloning Ability

A - B) Flow cytometry scatter plots **A)** Indexed positions of sorted *MIXL1*(**+**)/SSEA-3(**+**) cells from all single cell cloning experiments. **B)** Indexed positions of *MIXL1*(**+**)/SSEA-3(**+**) which went on to form stem cell like colonies. Red line indicates a cloning threshold of *MIXL1*-GFP intensity of 10^2.3.

4.2.15. Assessment of Primo medium on other cell lines.

Differences in derivation and culture techniques as well as genetic background can have an effect on human PSC cell lines and their growth and differentiation potential (Osafune et al., 2008). We assessed another human PSC line H9, carrying a reporter for the mesoderm associated gene *T*. As with the HES3 *MIXL1*-GFP line we were able to maintain good cell growth and colony morphology for H9 *T*-Venus grown in 2xPrimo medium. After three passages in the Primo medium I assessed the T-Venus and SSEA-3 expression of the cells (Figure 4-27). The

majority of the cells in this condition were double positive for *T*-Venus and SSEA-3. Cells were passaged further and maintained in 2xPrimo for 10 passages without any visible differentiation arising in the cultures. Cells were maintained by a colleague, Ivana Barbaric.

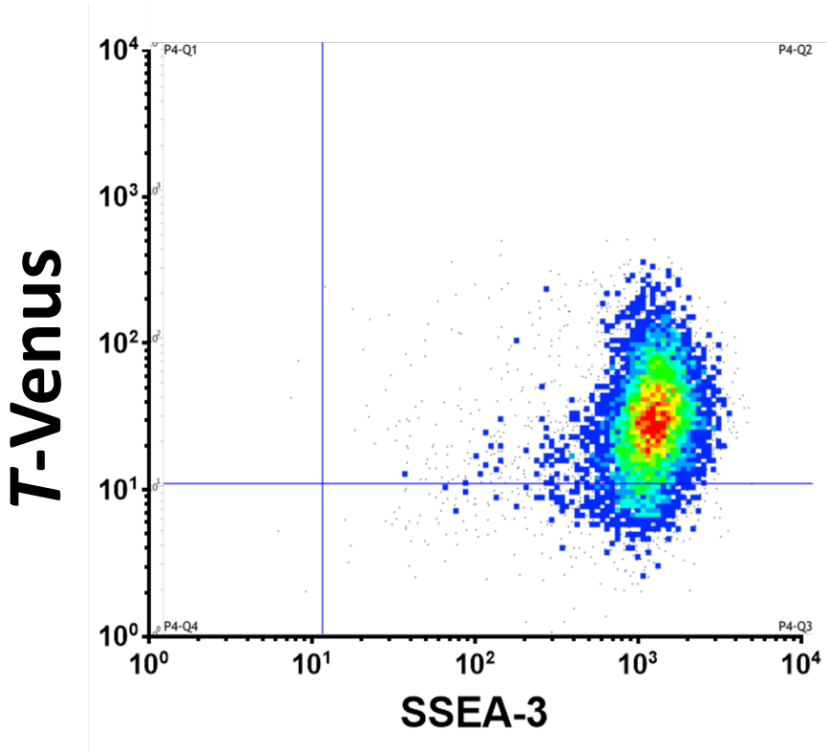


Figure 4-27 H9 T-Venus Reporter grown in 2xPrimo Conditions

Flow cytometry density plot of SSEA-3 versus *T*-Venus expression of H9 *T*-Venus reporter line grow in 2xPrimo after 3 passages.

4.2.16. Genetic stability of human ESC grown in Primo medium

Genetic variants can arise in standard human PSC culturing, common such changes involve the gaining of parts or whole chromosomes 1, 12, 17 and 20 (Amps et al., 2011; Baker et al., 2016; Taapken et al., 2011). These changes can have an effect on functional aspects of the cells including proliferation (Draper et al., 2004), cloning efficiency (Barbaric et al., 2014; Enver et al., 2005), and their ability to differentiate (Fazeli et al., 2011). The manipulation of human PSC growth medium can have unwanted impact on the genetic stability of cells. A genetic variant which resisted differentiation might be selected for in a lineage priming condition. With this in mind, cells from passage ten of 2xPrimo were analysed by metaphase spreads, for karyotypical changes. After ten passages in the Primo medium the cells had maintained a normal diploid karyotype, 46 XX, when 30 metaphases were analysed (Figure 4-28A). This maintenance of a normal diploid karyotype was also seen using the H9 *T*-Venus reporter line grown for 3 passages in Primo conditions by a colleague, Jonathon Carr (Figure 4-28B).

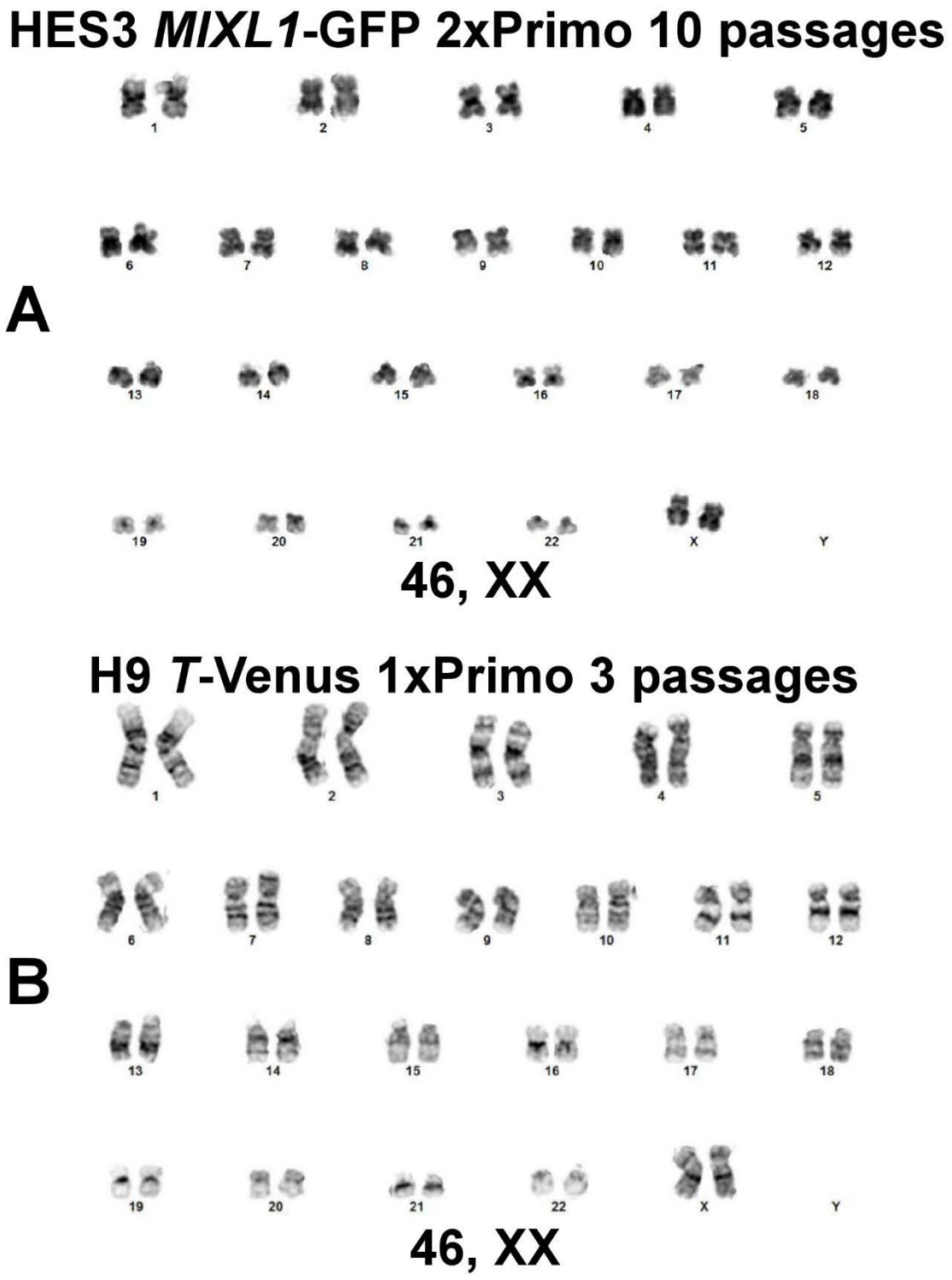


Figure 4-28 Lines Grown in Primo Conditions retain a normal karyotype

Representative images of G-banded metaphase spreads for **A)** HES3 *MIXL1*-GFP at 10 passages in 2xPrimo and **B)** H9 T-Venus at 3 passages in 1xPrimo. Both lines display a normal karyotype 46, XX.

4.3. Discussion

Functionally diverse substates of human PSC appear at different frequencies in standard culture conditions. Our previous work highlighted the presence of a particular substate identified by *MIXL1*-GFP reporter gene expression and the pluripotency-associated marker SSEA-3. This substate exhibits functional differences with respect to differentiation and its presence within a culture is sensitive to the micro-environment and matrix in which human PSC are grown. Previous work has mainly focused on substates present in standard human PSC culture conditions (Allison et al., 2018a; Gokhale et al., 2015; Hough et al., 2009; Hough et al., 2014). Here I focus on manipulating the micro-environment through the addition of signalling molecules and inhibitors to create and possibly maintain these substates. A critical observation that provided the starting point for this work was that the *MIXL1*(**+**)/SSEA-3(**+**) substates, seen when the cells were cultured on MEFs, were absent when the cells were cultured under the fully defined conditions of E8 and vitronectin.

I focused on one of the main signalling pathways involved in human PSC differentiation, WNT signalling. I also explored the pathways involved in maintaining pluripotency. While the relationship of FGF and TGFβ to maintaining pluripotency has been well investigated (Chen et al., 2011; Eiselleova et al., 2009; James et al., 2005; Thomson et al., 1998; Valdimarsdottir and Mummery, 2005; Vallier et al., 2005; Yu et al., 2011), the action of LPA has been less so (Blauwkamp et al., 2012; Garcia-Gonzalo and Izpisúa Belmonte, 2008; Qin et al., 2016). The dynamics of these signalling pathways is complexed and overlapping so manipulation is a delicate process. (Figure 4-29).

The GSK3β inhibitor, CHIRON, has been shown to be an effective mimic of WNT signalling through promotion of beta-catenin's translocation to the nucleus (Blauwkamp et al., 2012). In standard human PSC culture, CHIRON addition causes differentiation, it has been shown that LPA can be used to attenuate these effects . Qin et al, 2016 indicated that this may be as a result of LPA's promotion of YAP expression which in turn leads to beta catenin degradation.

Our work has shown that KOSR can be used as an effective modulator of CHIRON-induced *MIXL1* expression. Further to this I showed how LPA, a key component of KOSR, can also attenuate the *MIXL1* expression seen when CHIRON is added. 15% KOSR has a predicted LPA content of approximately 4.8µM (Garcia-Gonzalo and Izpisúa Belmonte, 2008). Both 15% KOSR and 4.8µM LPA resulted in effectively no *MIXL1* expression in the presence of CHIRON.

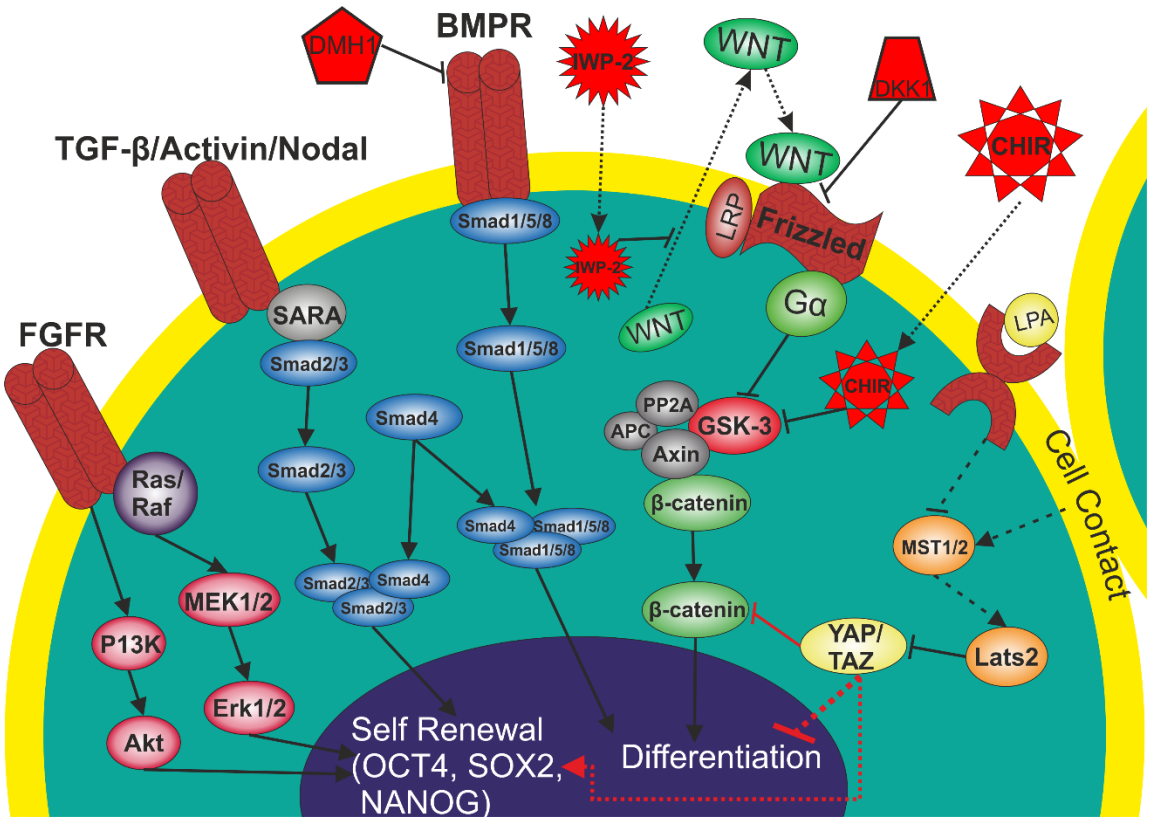


Figure 4-29 Human PSC Self-Renewal and Differentiation Pathways.

A schematic diagram detailing common signalling pathways involved in human PSC self-renewal and the interactions between different components of the pathway. Arrows indicate promotive effects capped lines indicate inhibitory effects. Solid lines denotes recognised interactions, dotted lines indicate possible interactions, red lines indicate proposed effects.

Low levels of LPA were also effective at controlling the intensity of *MIXL1*-GFP expression. At 0.48µM LPA and 3µM CHIRON I did see *MIXL1*-GFP expression increased but it was mainly confined to low expression. Without LPA I observed more *MIXL1*-GFP expression and a large proportion at a higher intensity level. I also saw the seemingly committed, *MIXL1*(High)/SSEA-3(-), fraction decrease from ~15% to less than 1% with the addition of LPA. Maintaining a low level of *MIXL1* expression may be key to trapping cells in a lineage biased substate, as highlighted by the strong SSEA-3 expression these cells retained.

I investigated the relationship between increasing concentrations of LPA and ability to both prevent *MIXL1*-GFP and promote NANOG expression in the presence of GSK3β inhibition. I observed a change in SSEA-3, *MIXL1*-GFP and cell size, indicated by forward versus side scatter analysis, under GSK3β inhibition. All of the observed changes could be alleviated through increasing LPA addition; the same was true for NANOG expression levels. Assessing the possible mechanism for this effect, I highlighted an increase in YAP nuclear localization. When YAP was

nuclear cells were almost always NANOG positive, implying that maintaining nuclear YAP might help to safeguard pluripotency. YAP over-expression has previously been reported to combat the effects of GSK3β inhibition in human PSC (Qin et al., 2016).

While this LPA, CHIRON balancing system gave us our first trapping medium formulation, further optimization was required in order to maintain the stem cell state. This first formulation, gave good induction of the *MIXL1*(**+**)/SSEA-3(**+**) substate and could easily revert back to *MIXL1*(-)/SSEA-3(**+**) fraction, in the presence of just LPA. However, post passage in trapping medium started to demonstrate signs of early differentiation, with the loss of pluripotency-associated surface markers.

With any manipulation system control is key. I found that I need tighter control over the activity of WNT signalling (Figure 4-8). I implemented a similar approach to Baseline Activation first proposed by Hackland et al, 2017. The authors proposed using DKK1 to block endogenous WNT signalling. DKK1 has been used to attenuate WNT signalling during human PSC differentiation (Paige et al., 2010). Rather than a recombinant DKK1, I utilised IWP-2 to inhibit WNT secretion. The idea of using a BLA approach was thought of as early differentiation would lead to the secretion of WNT ligands, as seen by the expression on WNT3 in our single cell transcriptomics, presented later in chapter 5. The addition of the WNT secretion inhibitor, IWP-2, helped to stabilise cultures in trapping medium and maintain the expression of SSEA-3, BF4 and CD9.

I incorporated 1μM IWP-2 into our optimized "Primo" conditions and saw improved expression of surface and intracellular pluripotency-associated markers in passaged and reverted cultures. The system also demonstrated high reversion efficiency when cells were transitioned back into just E8 alone without LPA or IWP-2 addition. I also demonstrated the reversion at the single cell level by creating clonal lines from Primo conditions. Much like the clones derived from MEF/KOSR *MIXL1*(**+**)/SSEA-3(**+**) cells, the clonal lines exhibit normal cell human PSC growth and surface marker expression. I also demonstrated trilineage potential of these lines under "Neutral" embryoid body formation.

Now that I had a system which could more robustly generate *MIXL1*(**+**)/SSEA-3(**+**) which demonstrated self-renewal abilities, I assessed their differentiation potential under "Neutral" conditions. In contrast to Primo clonal lines grown in E8, self renewal medium, cells grown in Primo medium generated EBs with a significantly stronger mesoderm signature, by transcriptional analysis than any other lineage. Unsorted and both *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**) cells exhibited this increased mesoderm signature. This development of "Primo" medium to trap cells in a mesoderm biased state is integral to the core objectives of the project.

Initially in both 1x and 2x Primo it became particularly difficult to expand the cells beyond 3 passages. Cells were lost to both death and differentiation. This result was similar to the findings of Blaukwamp et al, 2012 which found cells in the presence of LPA difficult to maintain. Cells that survived passaging appeared in large tightly compact colonies as smaller colonies tended to die. The large colonies favoured differentiation as the LPA was unable to act on the surface receptors of cells inside the colonies. Much like Chen et al, 2011 found, the cell death seemed to be a result of the presence of 2-mercaptoethanol. 2-mercaptoethanol is an antioxidant found in many culture mediums including N2B27 which was used by Blaukwamp et al, 2012. However once 2-mercaptoethanol was removed from Primo medium I observed better growth and proliferation of cells with less differentiation. 2-mercaptoethanol might have been the cause of the death of smaller colonies thus selecting for larger colonies that are more prone to differentiate.

After the removal of 2-mercaptoethanol from Primo, cells could be passaged beyond passage three. For continued passage 2X Primo was used. 2X Primo exhibited a lower yield of *MIXL1*(**+**)/SSEA-3(**+**) cells but cells maintained higher SSEA-3 expression levels than 1X Primo, the *MIXL1*-GFP intensity also remained low. There is some variability with the abundance of *MIXL1*(**+**)/SSEA-3(**+**) this can be affected by the cell density in the cultures. At passage 3 I demonstrated cells from *MIXL1*(**+**)/SSEA-3(**+**) positive cells could still produce NANOG positive colonies. At passage 5 I demonstrated cultures could revert to the *MIXL1*(-)/SSEA-3(**+**) state by culturing in 2X BCL (E8 **+** LPA) or E8 **+** 1μM IWP2. At passage 7 cultures were screened for 6 different pluripotency-associated surface antigens, maintaining high expression and a continued *MIXL1* (**+**) population. At passage ten, I stained for SSEA-3, NANOG, SOX2 and Nuclear YAP, reverted cultures to *MIXL1*(-)/SSEA-3(**+**), had a sample karyotyped and took samples for gene expression analysis. All findings were consistent with the ability to maintain a biased state throughout long term culture of these cells.

5. Chapter 5: Transcriptional Analysis of endogenous and induced substates

5.1. Introduction

5.1.1. The Pluripotency Gene Regulatory Network

The maintenance of pluripotency is controlled by a tightly regulated network of transcription factors. The maintained expression of these genes and their subsequent protein translation is crucial for sustaining pluripotency. Mouse studies revealed key components of this regulatory network. Oct4, Sox2 and Nanog are master regulators of this pluripotency network and work with other co-factors in the maintenance of this network. While these factors are crucial for the maintenance of pluripotency they also have roles in the early differentiation of PSC. In vitro assessment of these genes highlights their role in not only the maintenance of self-renewal but also in the competency to generate all three germs layers.

The *Pou5F1* gene encodes for the DNA binding protein Oct4. Oct4's importance to development was first highlighted as its protein expression is restricted to the ICM of the developing mice embryos (Rosner et al., 1990). This expression was also demonstrated is mouse ESC in culture. Further to this, Oct4 null mouse ESC present no formation of the three gem layers with all cells differentiating into trophectoderm (Nichols et al., 1998).

Oct4 governs pluripotency by binding to specific promoter and enhancer sites within the genome and activating gene expression. Many of these targets are also targeted by the Sox2, a member of the High Mobility Group (HMG) of proteins (Sharov et al., 2008) . Knockout studies revealed the requirement for Sox2 in the development of the ICM (Avilion et al., 2003).

Unravelling the pluripotency network and identifying the core genes, spurred the notion of inducing pluripotency. It was demonstrated that by exogenously expressing a handful of pluripotency markers in somatic cells they could be reprogrammed into a pluripotent state, akin to the cells isolated from the ICM (Takahashi and Yamanaka, 2006; Watanabe et al., 2007). Oct4, Sox2, c-Myc and Klf4 were overexpressed in the first work on induced pluripotent stem (iPS) cells. Further studies have revealed some flexibility in the genes required for induced pluripotency with other pluripotency-associated genes replacing these factors. Previously, it has been shown that overexpression of Oct4 alone could induce pluripotency in neural stem cells (Kim et al., 2009).

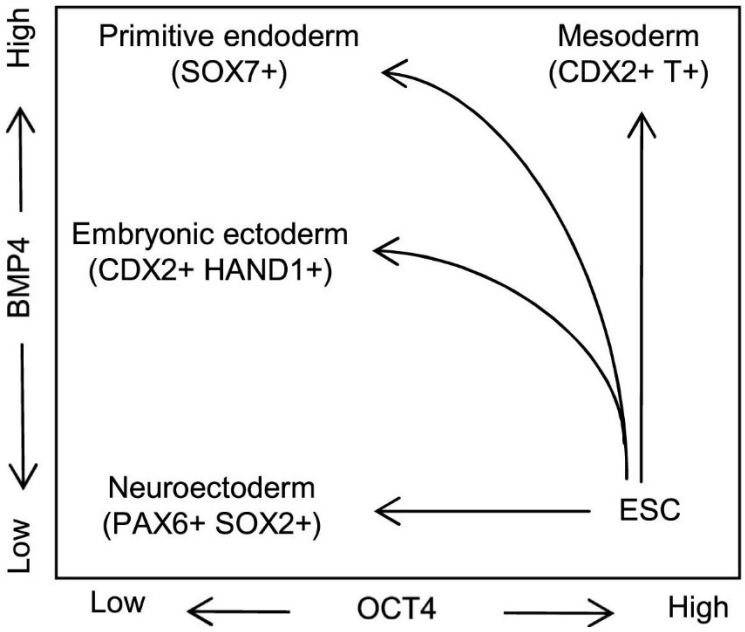


Figure 5-1 OCT4's Modulation of Differentiation of human PSC in vitro

Extracted model from Wang et al (2014). The proposed model displays how levels of OCT4 expression and BMP4 signalling can alter cell fate towards ectodermal, endodermal or mesodermal lineages in differentiating human PSC.

Expression of certain genes within the pluripotency gene network has a crucial impact on cell fate decisions in further differentiation. OCT4 has been identified as a critical factor in the formation of mesoderm lineages (Wang et al., 2012). Down-regulation of *OCT4* has been shown to promote ectoderm differentiation both in the presence and absence of BMP signalling. Further to this, the expression level of *OCT4* was shown to impact on the formation of mesoderm or endoderm in the presence of BMP signalling (Wang et al., 2012).

Low levels of SOX2 expression have been able to still permit self renewal but facilitates the formation of a primitive streak like population during differentiation (Wang et al., 2012). Complete knock-out studies with *SOX2* leads to trophectoderm (trophoblastic) differentiation and its expression has a vital function in the formation of ectoderm and neural images (Zhang and Cui, 2014). The regulation of *SOX2* expression is very important to maintaining pluripotency as both low or high levels of SOX2 can lead to differentiation. SOX2 works cooperatively with pluripotency factors in establishing the pluripotency network but switches to cooperate with lineage specific factors in differentiation (Figure 5-2). SOX2 expression is continued in further differentiated progeny of the neural lineages (Graham et al., 2003). The effect of *SOX2* expression appears to be very context dependent and is further complicated with its noted

balanced coexpression with *T* (BRACHYURY) in Neuro-Mesodermal progenitors (NMPs) (Koch et al., 2017).

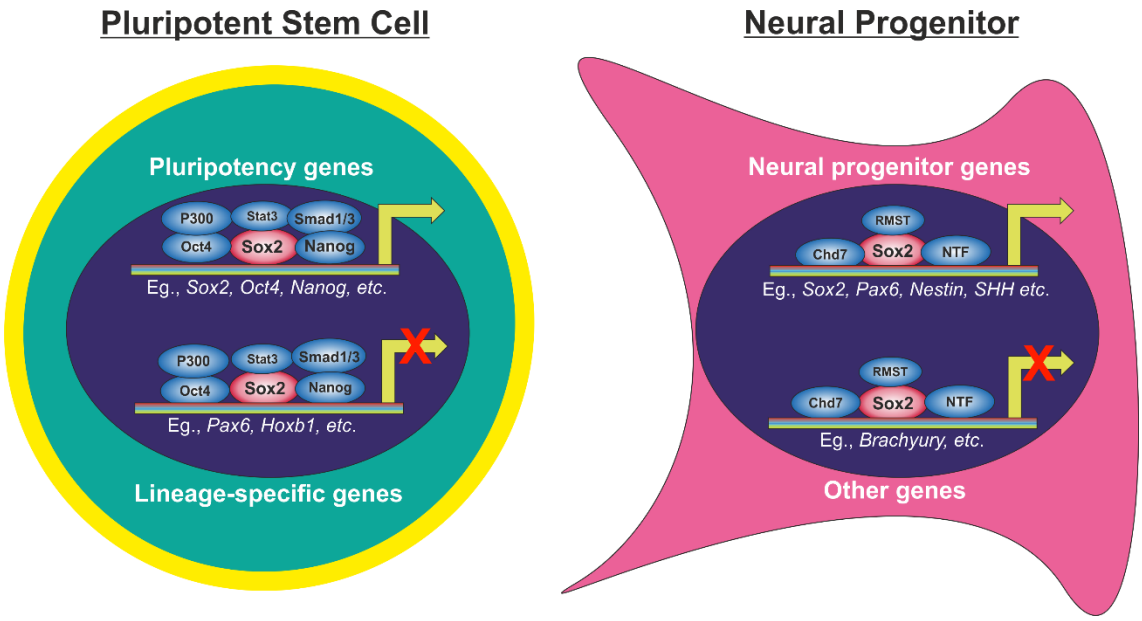


Figure 5-2 Sox2 and its Dynamic Cooperative Relationships

Adapted from (Zhang and Cui, 2014). In Pluripotent cells, Sox2 works cooperatively with pluripotency markers such as Oct4 and Nanog to both promote pluripotency gene transcription and repress lineage-specific genes. In Neural progenitors, Sox2 works cooperatively with neural progenitor markers such as Pax6 and Nestin to both promote neural progenitor gene transcription and repress other lineage-specific genes.

CDX2 is a homeobox protein that is expressed upon differentiation of human PSC towards mesoderm and shares a negative relationship with NANOG in the patterning of early mesodermal lineages (Mendjan et al., 2014). In human PSC differentiation towards anterior primitive streak, NANOG was shown to colocalise with the early mesodermal protein EOMES and exhibited co-expressing with *MESP1.* However, when CDX2 expression was increased during posterior primitive streak formation, it coincided with repression of *NANOG* but not *SOX2* expression. Overall the authors found that NANOG was required for cardiac mesoderm formation but inhibited the formation of other somatic mesoderm types, CDX2 had the opposite effect in both lineages (Figure 5-3) (Mendjan et al., 2014).

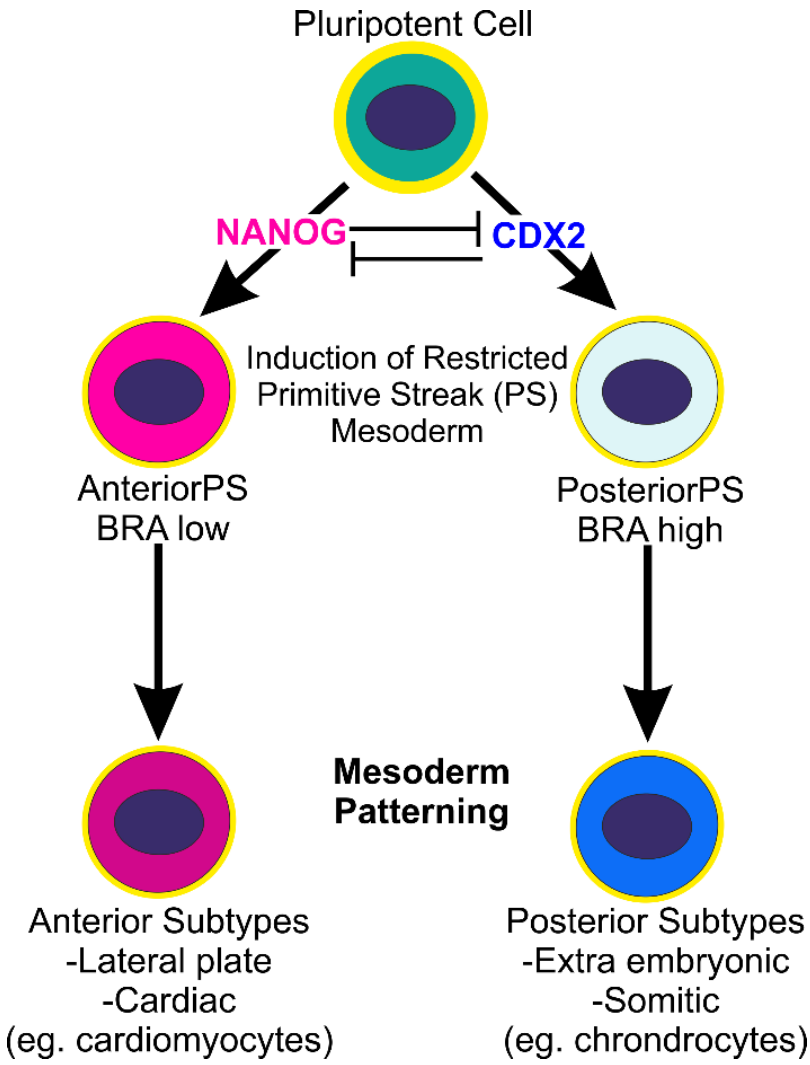


Figure 5-3 NANOG and CDX2 Pattern Anterior and Posterior Primitive Streak

Adapted from graphical abstract of Mendjan et al, 2014. Pluripotent cells differentiating towards primitive streak lineages are patterned by the negative relationship of NANOG and CDX2. NANOG expression guides cells towards anterior PS, marked by BRACHYURY (BRA) low expression. CDX2 expression guides cells towards posterior PS, marked by BRACHYURY (BRA) high expression.

Therefore, the maintenance of pluripotency, by sustaining this gene regulatory network is not only a requirement for promoting self-renewal but actually is important for making sure all differentiation avenues remain possible. The downregulation of this network is crucial to the process of differentiation, but equally important in the specification of particular lineages is a coordinated downregulation. This allows for certain elements of the pluripotency network to actively contribute to lineage specification (Wang et al., 2012).

5.1.2. Evolution of transcriptomic techniques

Gene expression analysis has always been a crucial tool for cellular biologists. Previously studies were hampered by the scope of their technologies. The field of transcriptomics has progressed from looking at single genes by PCR, through 1000s using microarrays, to the more recent entire transcriptome with RNA sequencing (RNA-seq). Human PSC have been analysed by some of these methods previously. For example, Enver et al (2005), utilised microarrays and more recently Yan et al (2013) utilised single cell RNA-seq. While RNA-sequencing offers large coverage of the transcriptome, its sensitivity can be an issue in relation to genes with relatively small transcript numbers. Therefore, targeted techniques such as qPCR and microarrays can still offer powerful analysis of specific gene expression fluctuations.

The evolution of transcriptomic techniques has emerged as two inter-related topics. The technology assessing gene expression and the application of the technology to smaller and smaller populations. This has allowed for robust transcriptomic techniques to be performed all the way down to the single cell level. Single cells can be isolated from clinical samples, tissues or cell culture and analysed by targeted qPCR or whole transcriptome RNA-seq (Figure 5-4).

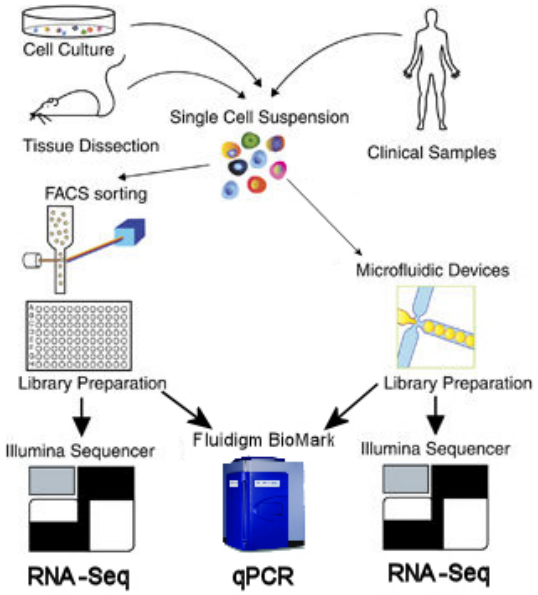


Figure 5-4 Single Cell Transcriptomic Techniques

Adapted from (Proserpio and Lönnberg, 2015). Schematic diagram of single cell transcriptomic techniques. Single cell suspensions can be collected from a variety of sources. Single cells can be isolated by fluorescent activated cell sorting (FACS) or by microfluidic devices. cDNA Libraries are prepared through reverse transcription and then analysed by qPCR (Fluidigm BioMark HD) or Rna-sequencing (Illumina Sequencer)

The use of RNA-sequencing has rapidly increase since its introduction. In 2008, only 2 publications used the technique but by 2015 over 2500 publications used this technique. In terms of its use in publications the technique has even over taken the once popular technique of microarrays (Lowe et al., 2017). The feasibility of RNA-seq experiments has improved over the years (Wang et al., 2009) as highlighted by Yan et al (2013) who were able to look at single cells related to different stages of embryo development. Those authors analysed the stages of embryo development by RNA-seq and were able to cluster samples according to their developmental stage. This shows that RNA-seq can be used to provide a profile for cells in a particular state and that other samples can then be compared to this profile. In that study, the authors also used the power of RNA-seq to address whether human PSC are similar to that of the human epiblast. They found 1,498 genes to be differentially expressed between human PSC and cells from the human epiblast and their gene expression profiles to be vastly different (Yan et al., 2013).

Expanding on the work of Yan et al, 2013, the laboratories of Sandberg and Lanner transcriptionally profiled 1,529 single cells from 88 human embryos (Petropoulos et al., 2016). This study disentangled some of the early gene expression changes that underpin key developmental moments in embryo maturation. Particularly, they were able to track the segregation of the inner cell mass and trophoectoderm and map the timing of this event to E5 stage of embryonic development. Similarly, the transition of the ICM into Epiblast and primitive endoderm was examined. They found that within single cells co-expression of lineage markers precedes the formation of matured lineages, with cells expressing genes correlating to epiblast, trophectoderm and primitive endoderm lineages (Petropoulos et al., 2016).

5.1.1. Analysis of stem cell heterogeneity by transcriptomic approaches.

The heterogeneous population that is present in human PSC in culture has caused issues, controversy and spurred research in the field. Classification of substates within these heterogeneous population has been difficult. A common segregation of substates has been to look at antigen expression. Certain antigens have been related to pluripotency or the stem cell state such as SSEA-3, TRA-1-60. TRA-1-81 (Andrews et al., 1984a; Kannagi et al., 1983; Wright et al., 2011). These antigens tend to be utilised as positive and negative discriminants of stem cell populations from within a culture. Their expression and intensity can vary between cells in culture and relate to functional differences (Tonge et al., 2011). Enver et al 2005 used SSEA-3 expression as a discriminant between stem cell populations and differentiated cells. Enver et al,

did find however that culture adaptation, the process in which cells accrue genetic mutations in culture, could perturb the relationship of SSEA-3 to the stem cell state. This highlights the issue with using a single discriminant to segregate stem cell substates.

Work carried out by our lab and published in Gokhale et al, 2015, used a single cell transcriptomic approach to analyse the differences in gene expression between SSEA-3(**+**) and SSEA-3(-) cells. The paper investigated the gene expression patterns of karyotypically Normal human PSC and karyotypically Abnormal ('Culture Adapted') human PSC. The results showed primarily that culture adaptation perturbed the dynamics of SSEA-3 expression and the stem cell state, with adapted SSEA-3(-) gene expression mapping to the stem cell compartment whereas Normal SSEA-3(-) cells show higher expression of differentiation markers. Another interesting finding when comparing expression patterns of the normal SSEA-3(**+**) is that a substate of these cells, whilst still expressing stem cell genes such as *POU5F1 (OCT4)* and *NANOG,* showed increased expression of genes associated with endoderm differentiation. These cells possibly represent a lineage-primed substate of human PSC.

Hough et al, 2014 used a combination of three pluripotency-associated surface markers, GCTM2, CD9 and EPCAM to isolate substates of human PSC. After applying single cell transcriptomic techniques (single cell qPCR), the segregated populations demonstrated a continuum of diverse transcriptional states. The cells which had high expression of all three surface markers, displayed a pristine pluripotent gene expression signature, whereas mid level and low expression of these surface markers displayed expression of lineage specific markers for mesendoderm and ectoderm, respectively. The substates also exhibited functional differences with respect to self renewal and differentiation, highlighting how transcriptional changes can have functional consequences to cells still within the pluripotent stem cell compartment.

Human PSC reporter lines were developed mainly with tracking differentiation in mind. The expression of some of these lineage specific markers may occur in the stem cell compartment. We can identify different sub-populations of stem cells based on reporter expression. Reporter lines for mesendoderm markers *MIXL1* (Davis et al., 2008) and *GATA6* (Allison et al., 2018a) can offer another level for isolating substates. A powerful combination of immunocytology, flow cytometry and genetic engineering has allowed us to identify and monitor sub-states in the PSC compartment under different conditions that support self renewal. I looked for fluorescence from a mesendodermal gene reporter such as *MIXL1* coupled with antibody staining for a stem cell marker such as SSEA-3. Isolation of these substates and the application of transcriptomic analysis techniques allowed us to identify changes in the gene expression patterns towards a particular lineage and the biasing that occurs within the stem cell compartment.

5.1.2. Transcriptomic approaches to model dynamic differentiation processes.

Accurately tracking genes throughout the dynamic process of differentiation can be difficult, while fluorescent reporter lines can offer transcriptional feedback of a gene throughout this process this represents less than 0.01% of the entire transcriptome. Of course, obtaining transcriptome data by qPCR or RNA-seq requires the cells to be destroyed and you cannot assess what the cell might have become if it had continued to differentiate. In order to resolve this dilemma cells can be taken at given timepoints along the trajectory and the gene expression changes are somewhat inferred between the timepoints.

The differentiation of human PSC towards a particular cell type can be a long guided process, in particular the differentiation towards the otic lineage takes 18 days and involves many steps of specification towards the final cell type (Ealy et al., 2016). From the pluripotent state, cells differentiate towards non-neural ectoderm, then pre-placodal ectoderm and finally transitioning towards the posterior otic lineage. During this process cells demonstrate asynchronous differentiation, and at a given timepoint cells can be in different phases of the differentiation. The process is also hampered from poor efficiency, with cells being lost to other lineages. This makes following gene expression changes very difficult as bulk approaches cannot resolve these differences. Ealy, et al, in 2016 demonstrated that single cell qPCR could delineate the trajectory of human PSC differentiating towards the human early otic lineage. By assessing single cells from timepoints throughout the differentiation, they were able to reconstruct the gene expression of 90 genes that change during these transitions. The analysis highlighted crucial branch points in the development in otic progenitors which in future could be focused on to improve the efficiency of future differentiations (Ealy et al., 2016).

The time course method is useful and has been employed in many studies investigating the process of differentiation. However, what if the differentiation cannot be split into clean time points? In many culture systems heterogenous populations can exist, with sporadic differentiation occurring seemingly randomly throughout. Separation by surface markers and reporter expression cannot resolve where the cells sit in a continuum of differentiation and the timing of these changes cannot be determined and might not be uniform for all cells. Nevertheless, because we know the final outcome of these changes is the generation of differentiated progeny, I can begin to model the process using pseudotime. If I make the assumption that we are observing a dynamic process of differentiation, this must have a start and a finish point. Pseudotime therefore tries to arrange cells in order from start to finish based on the points assessed. This ordering works through assessing gene expression changes between cells and arranging them according to expression patterns. This method allows for the

identification of discrete populations of cells and analyse where their trajectory branched from another population.

The rapid expansion of transcriptomic techniques has been matched by the computational approaches to its analysis. Monocle and its successor Monocle2, is a piece of analysis software that utilises transcriptomic data to model single-cell kinetics by pseudotemporal ordering of cells (Trapnell et al., 2014). In their original publication, Trapnell et al, 2014, demonstrated that their software could identify regulators of cell fate decisions during myoblast differentiation. The software was also able to model the branch point which segregates the small population of interstitial mesenchymal cells, from the larger differentiating myoblasts (Trapnell et al., 2014). This approach has already been used to investigate the key genes in the process of definitive endoderm specification from mesendoderm, highlighting *KLF8*'s importance (Chu et al., 2016).

Our knowledge of the pluripotency network and the factors associated with differentiation combined with the advancements in transcriptomic analysis allows for detailed analysis of the substates that we identified in chapters 4 and 5. Firstly, we can see the gene expression changes seen between the fractions we identified in standard human PSC culture, possibly allowing for the modelling of early differentiation. Further to this, we can see how our endogenous fractions from standard conditions relate to the newly generated fraction in our defined Primo system. The analysis can highlight important gene expression and also go some way to identifying the position of our cells on the continuum of pluripotency.

5.2. Results

5.2.1. RNA sequencing of Bulk sorted populations based on *MIXL1*-GFP and SSEA-3 Expression.

Using fluorescence-activated cell sorting (FACS) we have extracted different populations of cells based on expression of the fluorescent reporter *MIXL1*-GFP and the pluripotency-associated surface antigen SSEA-3. Once the sub-populations were collected we wanted to look for intrinsic changes in gene expression, thus the transcriptome of these populations was obtained by RNA sequencing. The hypothesis behind this is that populations with expression of a mesodermal marker should have different transcriptome profiles from the populations without marker expression beyond just the expression of the marker itself. The profile of the marker expressing cells might reveal an upregulation of further mesodermal markers. The transcriptome highlights important gene expression differences which could be under the control of cell signalling.

Cells from three substates, *MIXL1*(-)/SSEA-3(**+**), *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(-) were isolated by FACS from cultures growing under the MEF/KOSR conditions. 10,000 cells were taken from three technical replicates from two biological samples for each fraction. RNA-seq data was collected for each of these samples and analysed together.

5.2.2. Gene Grouping

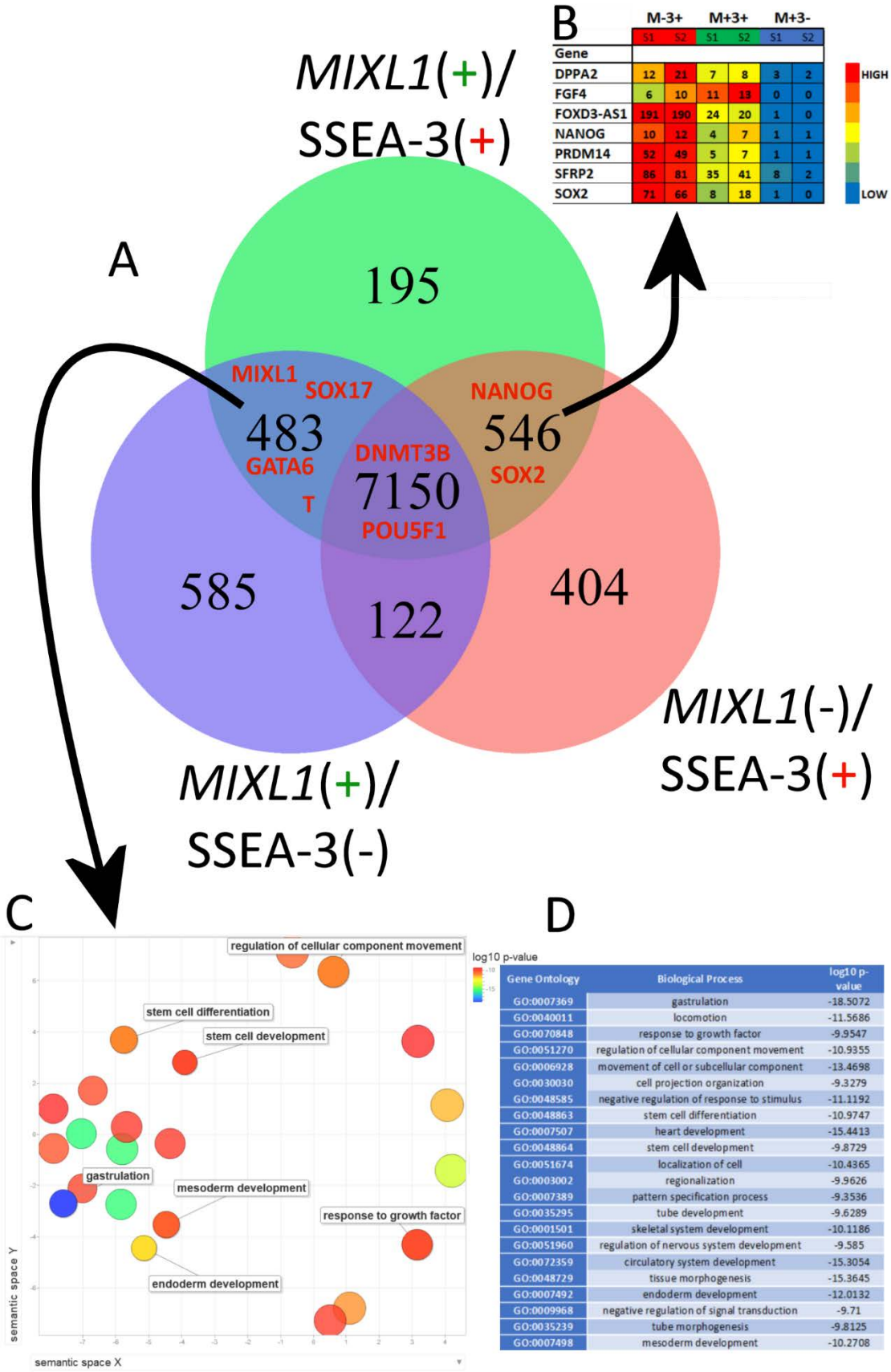


Figure 5-5 Expression Grouping.

A)This Venn diagram was produced using VennPlex program. Each circle corresponds to the total number of genes identified with a FPKM >5 for a given population based on *MIXL1* and SSEA-3 expression. Overlaps in the circles relate to genes shared between the cell populations. The numbers for each overlap shows the number of genes which are found in the group. Overlapping area size is not proportional to the number of genes identified in the group. The group position of a few key genes related to pluripotency and differentiation are highlighted in red writing. **B)** The table above highlights some pluripotency related genes that sit in the SSEA-3**+** fractions only. The table displays the FPKM values for the genes in each fraction and is colour coded based on expression, High expression (Red), Low expression (Blue). **C)** Go enrichment analysis of genes found in *MIXL1*(**+**) fractions. Biological Processes are coloured by their Log 10 p-values, as calculated by ToppGene (Chen et al., 2009b), higher p-values (Red), lower p-values (Blue). Their positions in sematic space are arranged by ReviGO(Supek et al., 2011), calculated by the correlation between biological processes. Some key biological processes are highlighted. **D)** Table of key biological processes identified for *MIXL1*(**+**) fractions and their associated Log10 p-values. These values are used to generate graph **C)**.

The processed RNA sequencing data was normalised to produce values of Fragments Per Kilobase per Million Reads (FPKM), this approach normalises the expression data to the size of the gene which the sequencing matches and the overall number of reads sequenced. I firstly averaged the FPKM values for each gene from the two sample repeats and applied a >5 FPKM cutoff value to the list. The data was then plotted into a venn diagram using Vennplex software (Cai et al., 2013). This allowed us to see the distribution of gene expression and probe certain groups for unique gene expression. Core pluripotency-associated genes such as *POU5F1* and *DNMT3B* were found present in all populations. Pluripotency-associated genes such as *NANOG* and *SOX2* however showed expression only within the SSEA-3(**+**) fractions. Further examination of the genes present in this group revealed other genes associated with pluripotency being present (Figure 5-5B), heatmap analysis does reveal a downregulation of these genes within the *MIXL1*(**+**)/SSEA-3(**+**) fraction but still present above 5 FPKM, whereas FPKM values in *MIXL1*(-)/SSEA-3(**+**) fractions are much lower.

In the *MIXL1*(**+**)/SSEA-3(**+**), *MIXL1*(**+**)/SSEA-3(-) overlapping group I identified the expression of mesoendodermal markers such as *EOMES, FGF17, FOXA2, GATA4, GATA6, GSC, HAND1, LHX1, SOX17, T* and *WNT3*, highlighting the co-expression of these mesoendodermal associated genes with *MIXL1*. Gene Ontology (GO) enrichment analysis reveals the key biological processes shared by the genes present in this group. ToppGene (Chen et al., 2009b) was used to perform a statistical over representation assessment of the genes in the list. This assesses whether the genes in the list pertain to a particular GO term or biological process, and the p-values for each process indicate the probability that this co-expression of genes could be random. This list is

then further refined using reviGO (Supek et al., 2011) software to remove overlapping GO terms and highlight key terms based on their p-values. The resulting list (Figure 5-5D) is then plotted in a semantic space, where correlated GO terms cluster closer together (Figure 5-5C). GO enrichment analysis highlighted some key biological processes which pertained to early differentiation and development including, gastrulation, mesoderm development and endoderm development were highly enriched.

Direct comparison of the *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**) fractions revealed the gene expression changes. The volcano plot demonstrates the variation in gene expression between the fractions, with the X axis displaying the LOG2 fold change of a gene between the fractions and the Y axis displaying the statistical significance of the difference as a LOG10 p-value (Figure 5-6A). The midpoint line indicates where genes are equally expressed in both fractions. Genes such as *POU5F1* and *NANOG* reside close to the mid line but further towards the *MIXL1*(-)/SSEA-3(**+**) fraction and further still is *SOX2*.

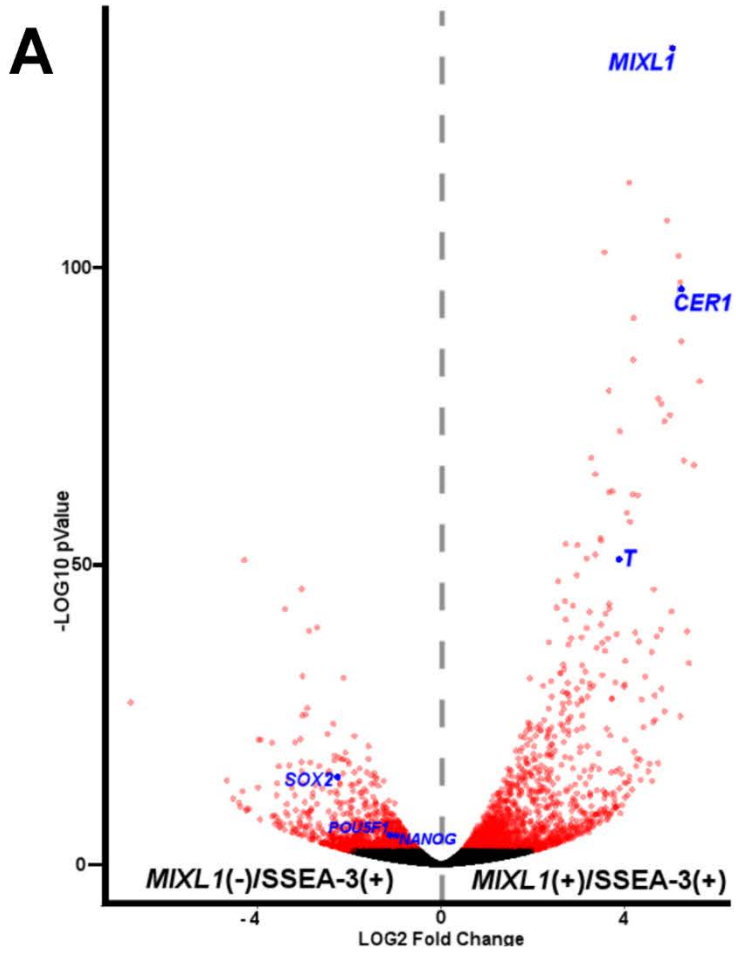
The *MIXL1*(**+**)/SSEA-3(**+**) has a larger spread of genes being upregulated within the fraction compared to *MIXL1*(-)/SSEA-3(**+**). Unsurprisingly, *MIXL1* demonstrated the larger differences between the fractions, which goes some way to validating that our sorting technique accurately separates these populations. *CER1* and *T* feature within the list of genes which show dramatic upregulation within the *MIXL1*(**+**)/SSEA-3(**+**) fraction. *CER1* encodes the protein Cerberus (CER1), a secreted factor in normal embryonic development which can inhibit BMP, WNT and Nodal signalling (Torres-Padilla et al., 2007).

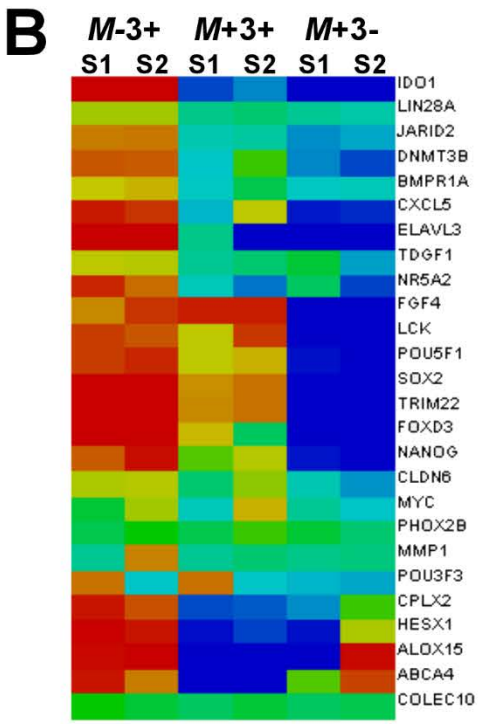
Heatmap analysis demonstrates the progressive nature of gene expression changes as cells move from *MIXL1*(-)/SSEA-3(**+**) to *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(-). Genes expressed highly in *MIXL1*(-)/SSEA-3(**+**) included pluripotency-associated genes, such as *POU5F1* and *NANOG* display a progressive down regulation towards *MIXL1*(**+**)/SSEA-3(-) (Figure 5-6B). Conversely, a collection of genes associated with differentiation show relatively low expression in *MIXL1*(-)/SSEA-3(**+**) and progressive upregulation towards *MIXL1*(**+**)/SSEA-3(-). Genes showing this pattern of expression included *EOMES*, *SOX17*, *HAND1* and other markers associated with mesendodermal lineages (Figure 5-6C).

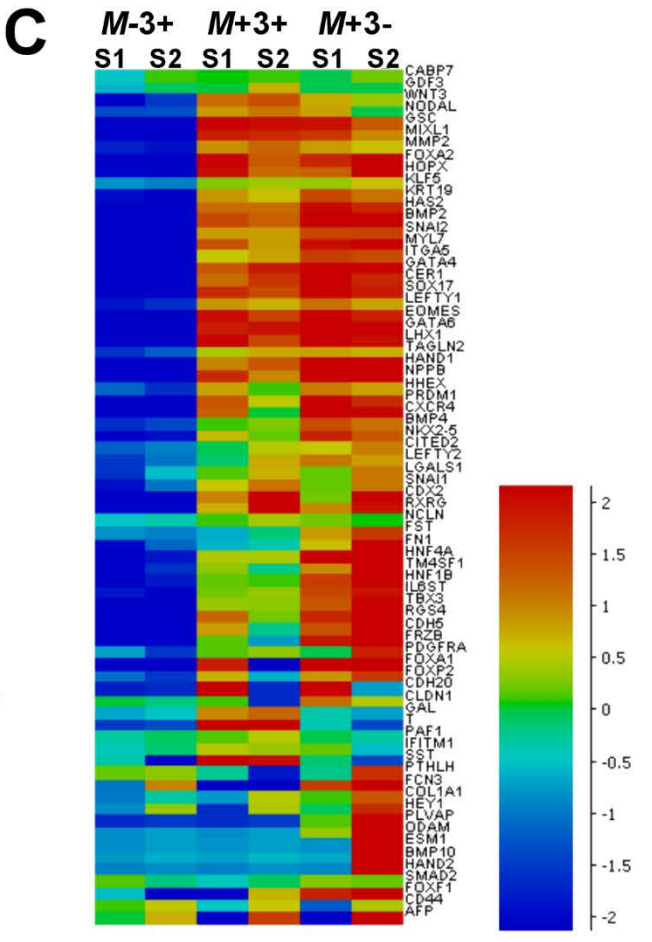
Figure 5-6 Gene Expression Changes Between Fractions

(Figure on next page)

A) Volcano plot displaying differential gene expression between *MIXL1*(-)/SSEA-3(**+**) (left of the plot) and *MIXL1*(**+**)/SSEA-3(**+**) (right of the plot). *MIXL1* demonstrated the highest differential expression between the two fractions. **B, C)** Heatmaps analysis of a selection of genes that show a downregulation (B) or upregulation **(C)** between *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(-).


A
MIXL1
CER1
T
SOX2
POU5F1
NANOG
100
50
0
-LOG10 pValue
MIXL1(-)/SSEA-3(+)
MIXL1(+)/SSEA-3(+)
-4
0
4
LOG2 Fold Change


B
M-3+
M+3+
M+3-
S1 S2 S1 S2 S1 S2
IDO1
LIN28A
JARID2
DNMT3B
BMPR1A
CXCL5
ELAVL3
TDGF1
NR5A2
FGF4
LCK
POU5F1
SOX2
TRIM22
FOXD3
NANOG
CLDN6
MYC
PHOX2B
MMP1
POU3F3
CPLX2
HESX1
ALOX15
ABCA4
COLEC10


C
M-3+
M+3+
M+3-
S1 S2 S1 S2 S1 S2
CABP7
GDF3
WNT3
NODAL
GSC
MIXL1
MMP2
FOXA2
HOPX
KLF5
KRT19
HAS2
BMP2
SNAI2
MYL7
ITGA5
GATA4
CER1
SOX17
LEFTY1
EOMES
GATA6
LHX1
TAGLN2
HAND1
NPPB
HHEX
PRDM1
CXCR4
BMP4
NKX2-5
CITED2
LEFTY2
LGALS1
SNAI1
CDX2
RXRG
NCLN
FST
FN1
HNF4A
TM4SF1
HNF1B
IL6ST
TBX3
RGS4
CDH5
FRZB
PDGFRA
FOXA1
FOXP2
CDH20
CLDN1
GAL
T
PAF1
IFITM1
SST
PTHLH
FCN3
COL1A1
HEY1
PLVAP
ODAM
ESM1
BMP10
HAND2
SMAD2
FOXF1
CD44
AFP
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2

5.2.3. Development of qPCR profile

Our bulk transcriptome analysis highlighted that a pluripotency network might still be active within the *MIXL1*(**+**)/SSEA-3(**+**) fraction, albeit at a lower expression level. To investigate whether this lower representation of pluripotency markers was homogenous across the population or rather that the population contained cells that exhibited high and low expression of pluripotency markers, I developed a qPCR profile to interrogate the fractions at a single cell level. I also wanted to further investigate this heterogeneity, in relation to early differentiation associated genes I saw upregulated in this fraction by bulk RNA-Seq.

The RNA-seq data collected was utilised to create a list of genes which could probe the heterogeneity within these cells. A list of genes was created by combining gene lists from these sources

1. The 200 differentially regulated genes from the RNA-Seq data.
2. 180 genes related to Pluripotency and the three germ layers, endoderm, ectoderm and mesoderm.
3. The CD antigen family.
4. Taqman Pluripotency Array (Thermofisher, #4385344) gene list.
5. The gene list used in a similar single cell study, Hough et al 2014.

These lists were used as part of the initial screen of over 500 selected genes. A FPKM cutoff point >1 across all fractions was applied initially to remove non or low expressing genes. The remaining genes were then clustered into 10 groups using K-clustering. The *MIXL1*, *POUF51* and *CER1* groups were taken forward in full. The others were screened by expression and function with approximately 8 genes from each cluster carried forward to the 96 gene list. To further reduce the list from 96 to 48 genes I performed complete hierarchical clustering on the remaining genes (Figure 5-7). I then removed genes which mapped closely to others, removal was based on either their expression level or the genes function. The final gene list and primers are represented in Table 5.1, a more detailed table is present in appendix Table 9.1.

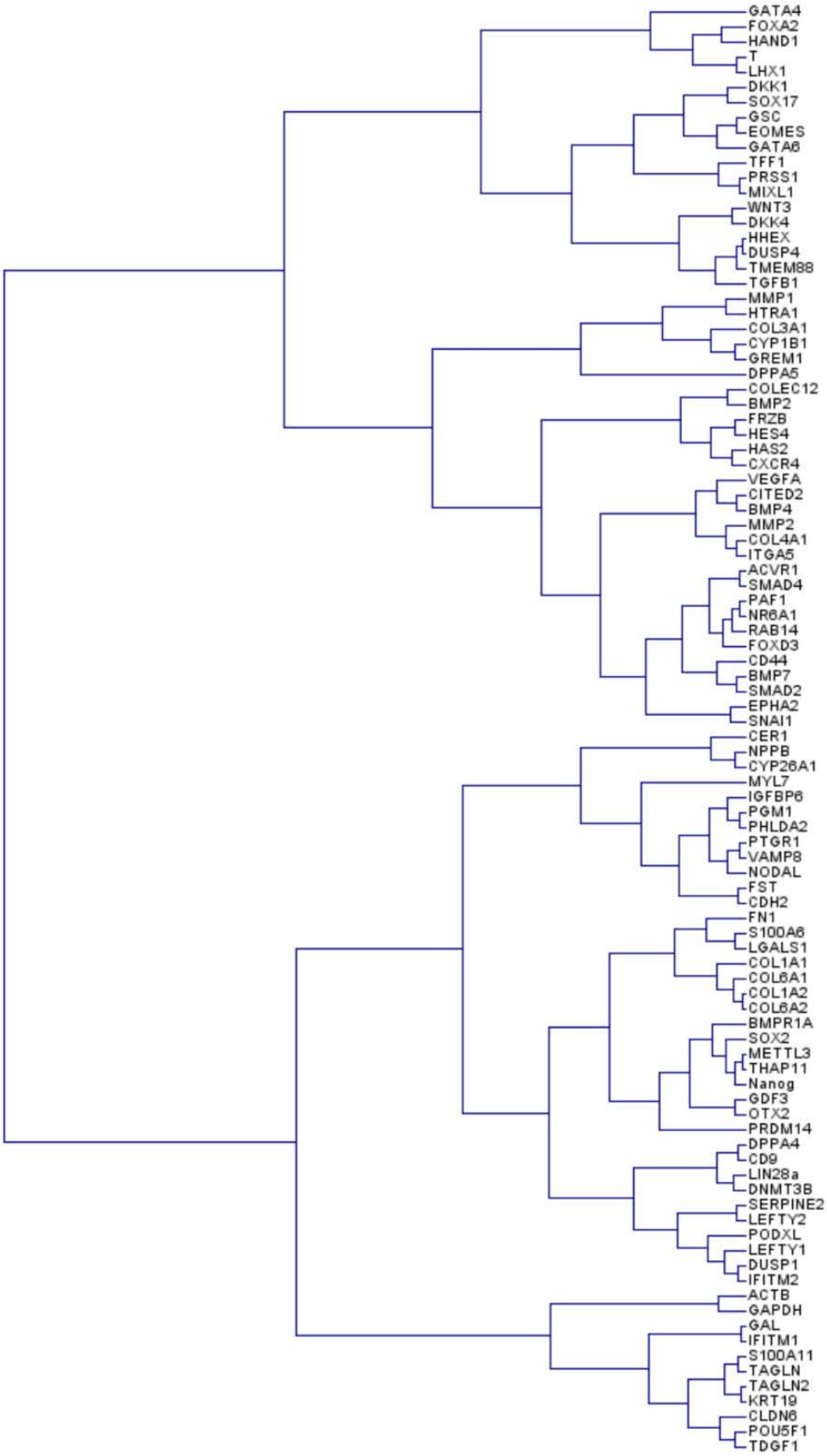


Figure 5-7 Cluster analysis for single cell qPCR gene list.

Hierarchical Clustering based on FPKM values from RNA-seq data of the 96 genes remaining after screening and K-means Clustering.

Table 5.1 Single Cell qPCR Assay ID List

Assay ID	Gene Symbol	Gene Name
Hs01060665_g1	*ACTB*	actin; beta
Hs00154192_m1	*BMP2*	Bone morphogenetic protein 2
Hs03676628_s1	*BMP4*	bone morphogenetic protein 4
Hs01034913_g1	*BMPR1A*	bone morphogenetic protein receptor; type IA
Hs00193796_m1	*CER1*	cerberus 1; DAN family BMP antagonist
Hs01897804_s1	*CITED2*	Cbp/p300-interacting transactivator; with Glu/Asp-rich carboxy-terminal domain; 2
Hs00607528_s1	*CLDN6*	claudin 6
Hs00164004_m1	*COL1A1*	collagen; type I; alpha 1
Hs00607978_s1	*CXCR4*	chemokine (C-X-C motif) receptor 4
Hs00171876_m1	*DNMT3B*	DNA (cytosine-5-)-methyltransferase 3 beta
Hs00172872_m1	*EOMES*	eomesodermin
Hs01549976_m1	*FN1*	fibronectin 1
Hs00232764_m1	*FOXA2*	forkhead box A2
Hs00255287_s1	*FOXD3*	forkhead box D3
Hs00173503_m1	*FRZB*	frizzled-related protein
Hs00246256_m1	*FST*	follistatin
Hs00544355_m1	*GAL*	galanin/GMAP prepropeptide
Hs00171403_m1	*GATA4*	GATA binding protein 4
Hs00232018_m1	*GATA6*	GATA binding protein 6
Hs00906630_g1	*GSC*	goosecoid homeobox
Hs00193435_m1	*HAS2*	hyaluronan synthase 2
Hs00242160_m1	*HHEX*	hematopoietically expressed homeobox
Hs00705137_s1	*IFITM1*	Interferon-induced transmembrane protein 1
Hs01547673_m1	*ITGA5*	integrin; alpha 5 (fibronectin receptor; alpha polypeptide)
Hs00761767_s1	*KRT19*	keratin 19
Hs00764128_s1	*LEFTY1*	left-right determination factor 1
Hs00745761_s1	*LEFTY2*	left-right determination factor 2
Hs00355202_m1	*LGALS1*	lectin; galactoside-binding; soluble; 1
Hs00232144_m1	*LHX1*	LIM homeobox 1
Hs00702808_s1	*LIN28A*	lin-28 homolog A (C. elegans)
Hs00430824_g1	*MIXL1*	Mix paired-like homeobox

Hs00899658_m1	*MMP1*	matrix metallopeptidase 1 (interstitial collagenase)
Hs01548727_m1	*MMP2*	matrix metallopeptidase 2 (gelatinase A; 72kDa gelatinase; 72kDa type IV collagenase)
Hs01085598_g1	*MYL7*	myosin; light chain 7; regulatory
Hs04399610_g1	*NANOG*	*NANOG* homeobox
Hs00378379_m1	*NCLN*	nicalin
Hs00415443_m1	*NODAL*	nodal growth differentiation factor
Hs00219496_m1	*PAF1*	Paf1; RNA polymerase II associated factor; homolog (S. cerevisiae)
Hs04260367_gH	*POU5F1*	POU class 5 homeobox 1
Hs01375212_g1	*RPS18*	ribosomal protein S18
Hs00183425_m1	*SMAD2*	SMAD family member 2
Hs00195591_m1	*SNAI1*	snail family zinc finger 1
Hs00751752_s1	*SOX17*	SRY (sex determining region Y)-box 17
Hs01053049_s1	SOX2	SRY-box2
Hs00610080_m1	T	T; brachyury homolog (mouse)
Hs00761239_s1	TAGLN2	transgelin 2
Hs02339499_g1	TDGF1	teratocarcinoma-derived growth factor 1
Hs00902257_m1	WNT3	wingless-type MMTV integration site family; member 3

5.2.4. Single Cell qPCR

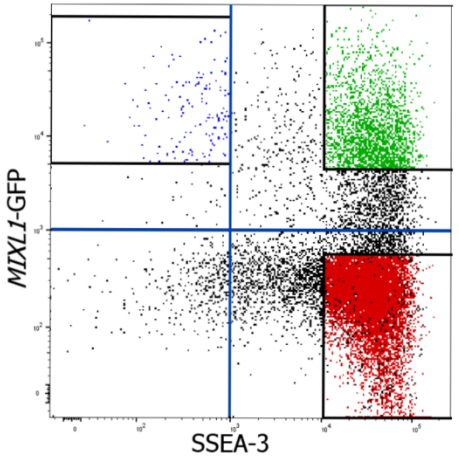
5.2.4.1. Gene Expression Analysis

To further investigate the heterogeneity that might exist within *MIXL1*(**+**)/SSEA-3(**+**) fraction of human PSC we performed single cell transcriptomic analysis. Single cells were sorted from the three fractions of interest and single-cell RT-qPCR data obtained on a Fluidigm platform (Figure 5-8A). Global Z-scores were calculated using Cycle threshold (Ct) values for all genes in all cells whilst assuming an amplification efficiency of 100%, as stated by the manufacturers of the gene assays. Hierarchical clustering of the single cell data shows good discrimination of individual cells in the distinct subsets, with *MIXL1*(-)/SSEA-3(**+)** away from *MIXL1*(**+**)/SSEA-3(**+**) and the early differentiated fraction *MIXL1*(**+**)/SSEA-3(-) cells. Sub-clustering then broadly delineates the 3 individual subsets (Figure 5-8B). The heatmap shows some *MIXL1*(**+**)/SSEA-3(**+**) cells cluster close to the *MIXL1*(-)/SSEA-3(**+**) population but a strong overlap of *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(-)/SSEA-3(**+**). This could be reflective of the qPCR panel weighted towards differentiated genes with only 9 out of 48 genes associated with pluripotency. As the clustering is determined by cells similarity in gene expression having a larger number of differentiation genes slightly skews this clustering as cells expressing differentiation markers will cluster closer with pluripotency-associated factors now hold less weight in the overall gene expression.

The clustering of genes also highlights some important correlations, with pluripotency factors *POU5F1*, *SOX2*, *NANOG* and *TDGF1* clustering together while *MIXL1* clusters closest with differentiation markers such as *GATA4*, *WNT3* and *LHX*. Observing the patterns of differentiation, I can see a gradual decrease in pluripotency-associated markers such as *POU5F1* and *NANOG* but a sharper decline in *SOX2*. I also observe an increase of differentiation associated markers such as *T*, *FOXA2* and particularly a high increase in *CER1* expression. This mirrors our previous finding from the bulk RNA-sequencing that *CER1* is highly upregulated in these fractions.

I also examined the individual cells utilising a similarity matrix approach (Figure 5-9). The gene expression for each cell was compared to all the other cells and a similarity score was generated depending on the gene expression profiles. There was a strong correlation between cells within the *MIXL1*(-)/SSEA-3(**+**) fraction, again possibly reflective of the gene selection, with most of these cells not expressing any differentiation markers. There is a larger disparity between the *MIXL1*(**+**) fractions compared to *MIXL1*(-) fraction displaying a lower similarity. However, within the *MIXL1*(**+**)/SSEA-3(**+**) fraction some cells show a higher similarity to the *MIXL1*(-)/SSEA-3(**+**) fraction, whereas few cells in the *MIXL1*(**+**)/SSEA-3(-) fraction have this level of similarity. The similarity matrix also highlights the further heterogeneity within the *MIXL1*(**+**) fractions.

A



Sample Groups

MIXL1(-)/SSEA3(+)
- Sample 1
- Sample 2

MIXL1(+)/SSEA3(+)
- Sample 1
- Sample 2

MIXL1(+)/SSEA3(-)
- Sample 1
- Sample 2

Gene Groups

- Pluripotency
- Gastrulation
- Mesendoderm
- Mesoderm
- WNT Signalling
- BMP Signalling
- Endoderm
- Neural
- Differentiation Factor

B

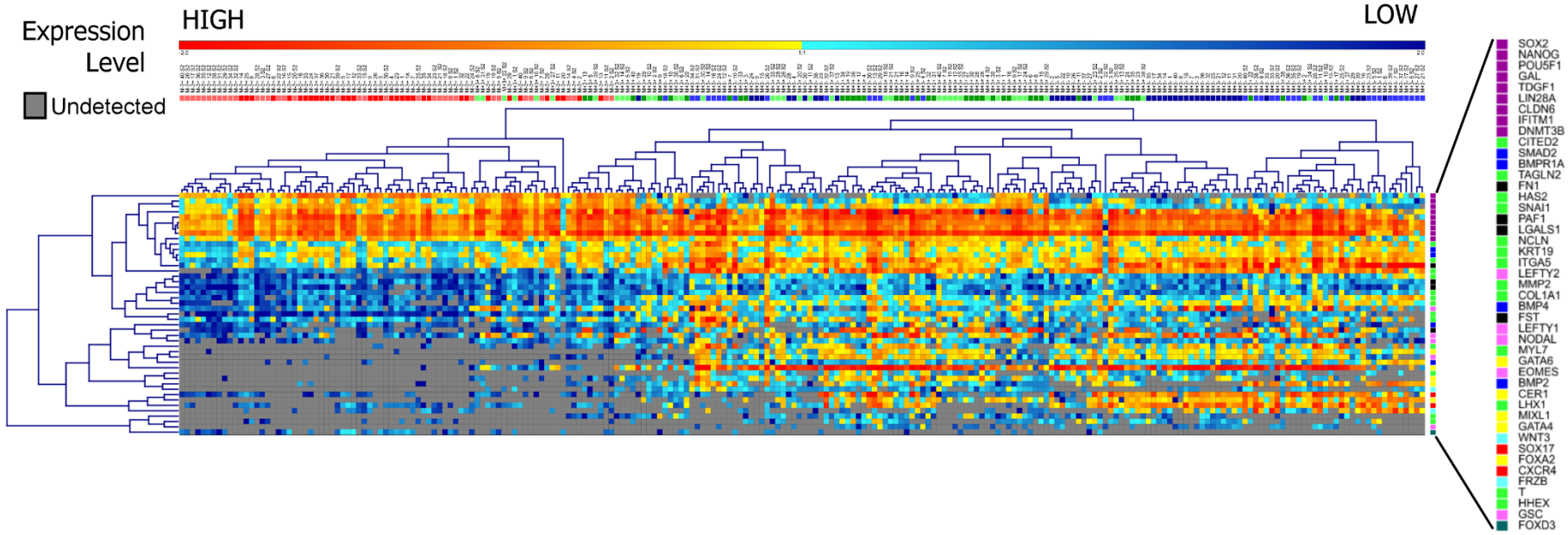


Figure 5-8 Single Cell qPCR Heatmap

A) Flow cytometry scatter plot of HES3-*MIXL1* stained for SSEA-3 from MEF/KOSR conditions. Coloured populations indicate fractions sorted by FACS for single cell analysis, *MIXL1*(-)/SSEA-3(**+**) (Red) *MIXL1*(**+**)/SSEA-3(**+**) (Green), *MIXL1*(**+**)/SSEA-3(-) (Blue). 80 cells were sorted from each of the 3 fractions based on *MIXL1*-GFP and SSEA-3 expression. **B)**Two way Hierarchical clustering analysis of 232 single cells analysed across 48 genes using a Fluidigm BioMark system. The heatmap visualises the individual gene expression after normalisation across genes and samples. Grey coloured genes indicates undetected levels.

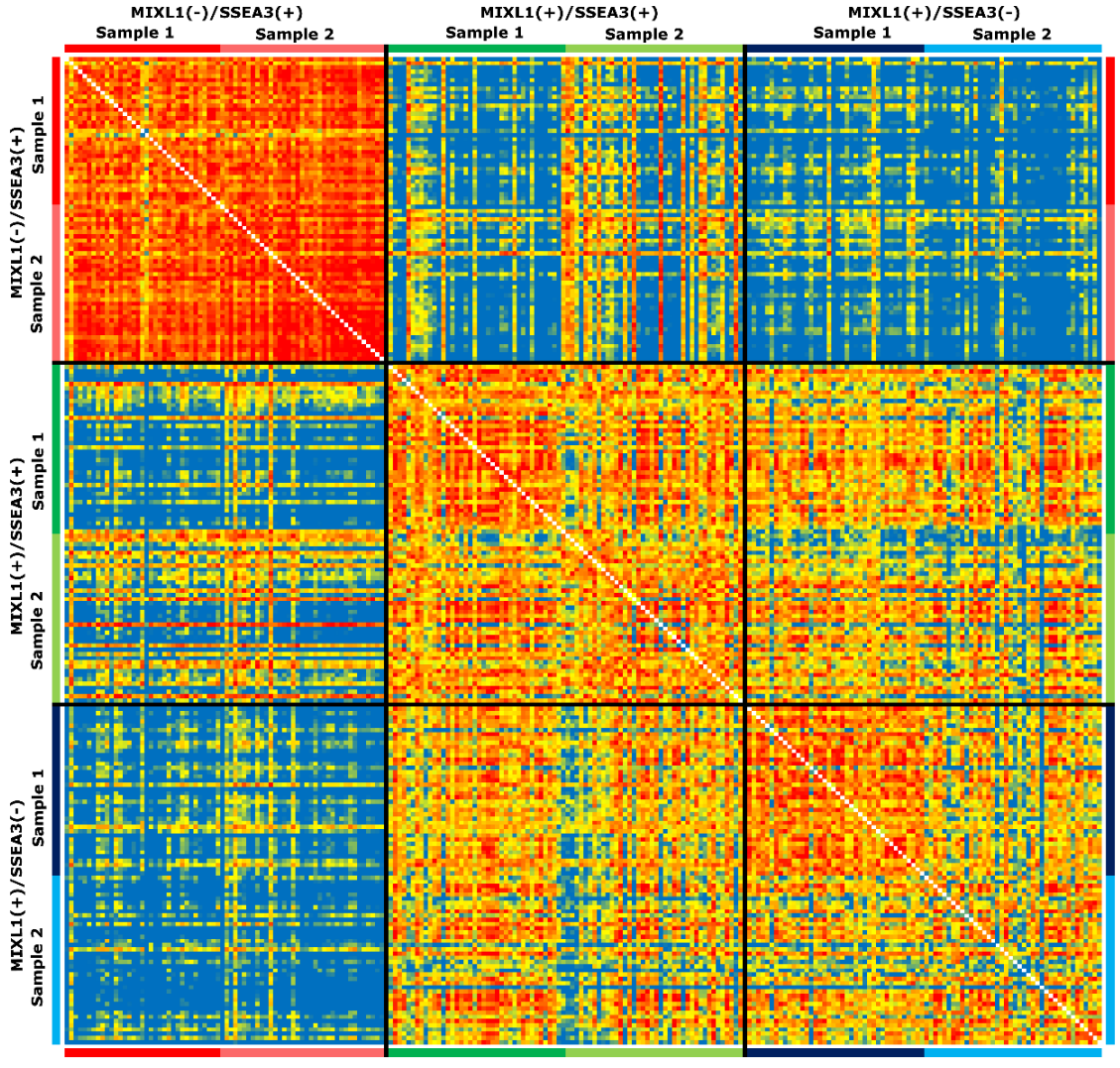


Figure 5-9 Single Cell qPCR similarity matrix

A similarity matrix of individual single cells from three fractions *MIXL1*(-)/SSEA-3(**+**)(Red), *MIXL1*(**+**)/SSEA-3(**+**)(Green) and *MIXL1*(**+**)/SSEA-3(-)(Blue). Single cell qPCR data was compared using Genesis software and a similarity score was generated for each cell compared to every other cell. The data present is then coloured with lower similarity (blue), mid level similarity (yellow) and higher similarity (red). White blocks indicate when cells are compared to themselves.

5.2.4.2. Cell Trajectory Analysis

While the classical heatmap and clustering analysis revealed some overlap between *MIXL1*(**+**)/SSEA-3(**+**) and *MIXL1*(-)/SSEA-3(**+**) fractions their actual relationship is harder to resolve using these methods. Our hypothesis is that cells are transitioning through the dynamic process of differentiation and that we begin with pluripotent *MIXL1*(-)/SSEA-3(**+**) cells through lineage biasing *MIXL1*(**+**)/SSEA-3(**+**) and eventually towards differentiation *MIXL1*(**+**)/SSEA-3(-). If we interpret the data as a dynamic process, it allows us to model cell trajectories with Monocle2 (Qiu et al., 2017a; Qiu et al., 2017b; Trapnell et al., 2014) and assess where cells fall in this continuum from pluripotency to differentiation.

With the hypothesis that we are looking at a dynamic process of differentiation then Monocle2 can begin to assess where cells might fit in this continuum. Thus, based on the changes in gene expression Monocle2 orders the cells in "pseudotime" as a representation of where they are situated in this process from undifferentiated to differentiated. Once ordered I firstly assessed the cells using a t-Distributed Stochastic Neighbour Embedding (t-SNE) approach within Monocle2. t-SNE is a multi-dimensional reduction approach to resolve the similarities within large data sets (Maaten and Hinton, 2008). The t-SNE uses both the qPCR data and pseudotime projection for each cell and then reduces the dimensions to yield a 2D graph.

Observing the t-SNE plot we see, *MIXL1*(-)/SSEA-3(**+**) cells in red, appearing in the bottom right corner of the plot, they are distinctly separated from the *MIXL1*(**+**)/SSEA-3(-) cells which span the upper part of the plot. *MIXL1*(**+**)/SSEA-3(**+**) cells however seem to span the space between these two fractions, with some cells closer to *MIXL1*(-)/SSEA-3(**+**) group and some within the *MIXL1*(**+**)/SSEA-3(-) group (Figure 5-10A). Since I have three fractions that I am assessing if they were distinctly different you would expect them to cluster into 3 separate cluster, but when I cluster into three groups this is not what was seen. Interestingly, *MIXL1*(-)/SSEA-3(**+**) cells span clusters 1 and 2 towards the bottom of the t-SNE plot and *MIXL1*(**+**)/SSEA-3(-) cells are restricted to cluster 3 alone (Figure 5-10B). However as seen in our standard t-SNE plot *MIXL1*(**+**)/SSEA-3(**+**) cells overlap with both *MIXL1*(**+**)/SSEA-3(-) cells and *MIXL1*(-)/SSEA-3(**+**) cells, with cells within cluster 2 and cluster 3.

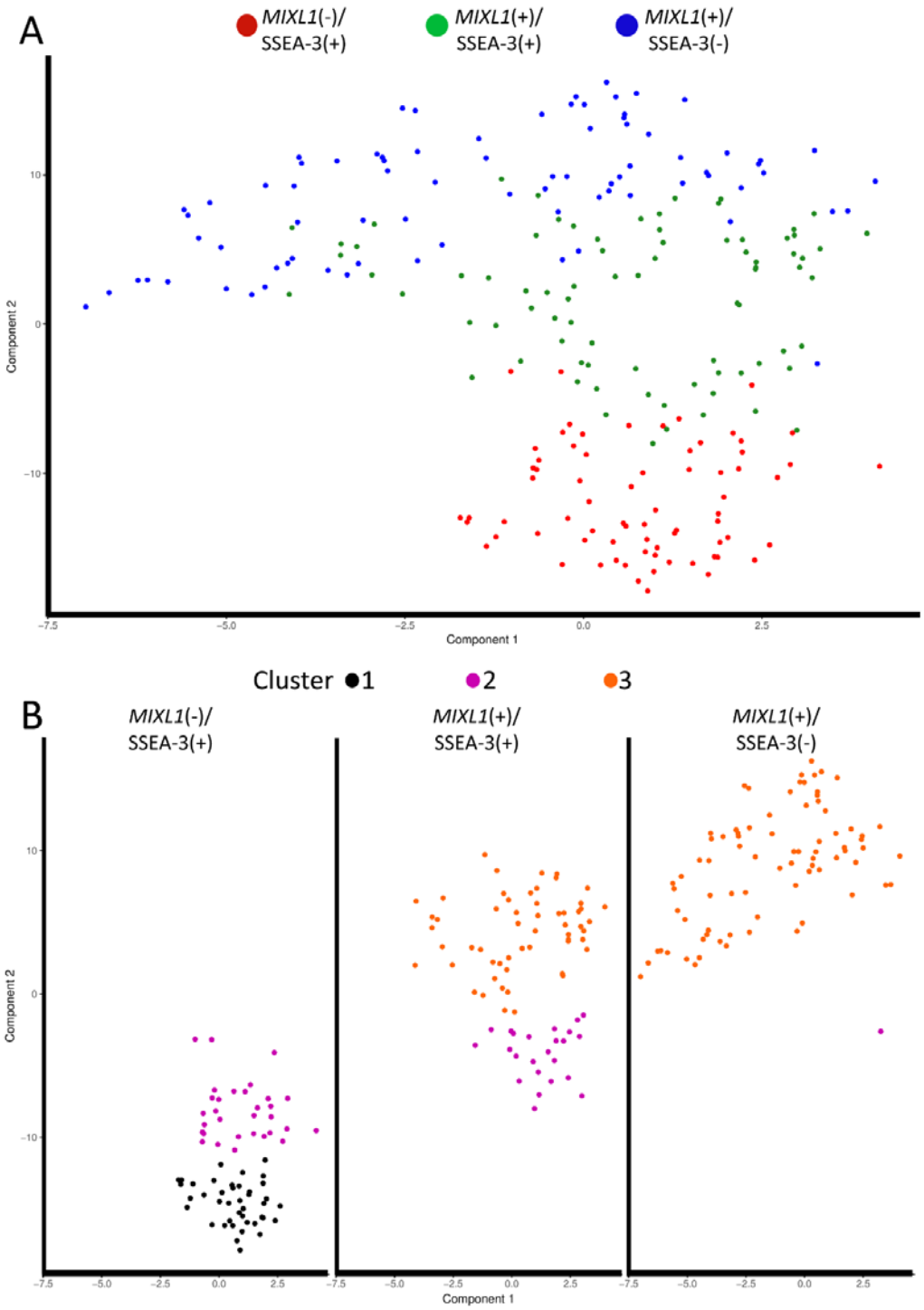


Figure 5-10 Dimension reduction clustering of single cells

The Single cell qPCR data generated from the fractions is multidimensional, having many aspects that can be analysed. In order to get a clearer idea of the cells relationship to each other I have reduced the dimensions and plotted the data in a two dimensional space. t-Distributed Stochastic Neighbour Embedding (t-SNE) plots of the expression of 45 genes in individual cells measured by Fluidigm Biomark qPCR. **A)** Displays all fractions in the same dimensional space *MIXL1*(-)/SSEA-3(**+**) cells in red, *MIXL1*(**+**)/SSEA-3(**+**) cells in green and *MIXL1*(-)/SSEA-3(**+**) cells in blue. **B)** Displays fractions separately and cells are coloured according to the cluster they correspond to after cluster analysis by Monocle2, cluster 1(black), 2(purple) and 3(orange).

5.2.4.3. Minimal spanning trees.

After assessment by t-SNE we can see a continuum from undifferentiated to differentiated with *MIXL1*(**+**)/SSEA-3(**+**) occupying a mid-point in this process. To assess further the dynamics of this continuum and model a cell trajectory based on these fractions, I used a minimal spanning tree approach, which arranges cells in pseudotime based on gene expression and attempts to model the process of differentiation and predict any branch points in this process where two cell types might be emerging (Figure 5-11).

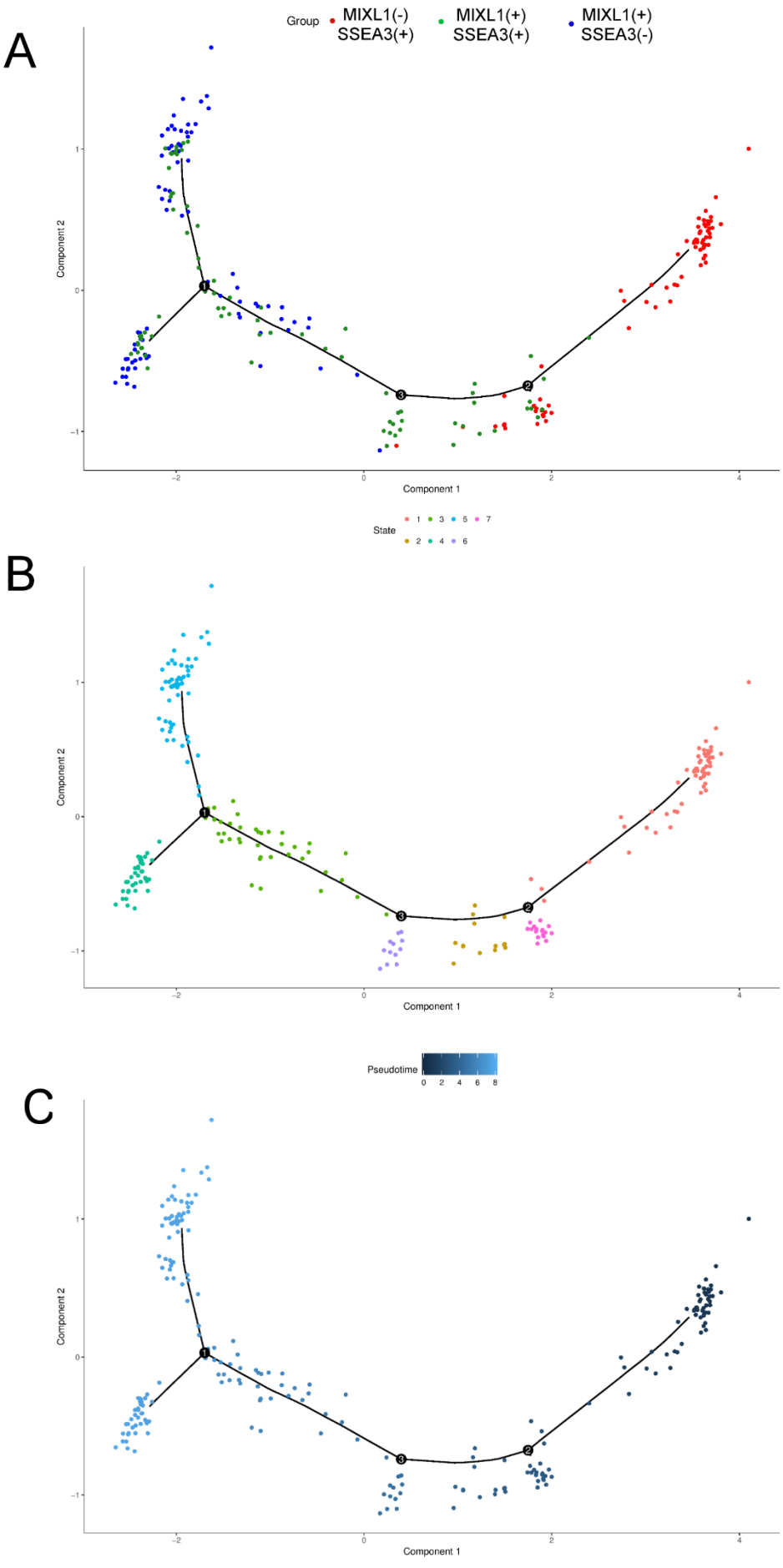
Again, as with the t-SNE approach we can see a clear separation between the *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(-) cells. The majority of the *MIXL1*(-)/SSEA-3(**+**) cells occupy the start of the tree, on the far right of the plot. As we move down the trajectory we have some *MIXL1*(-)/SSEA-3(**+**) cells occupying the space and eventually there is a collection of *MIXL1*(**+**)/SSEA-3(**+**) cells. As we move towards the ends of the trajectory tree we begin to see *MIXL1*(**+**)/SSEA-3(-) cells. The *MIXL1*(**+**)/SSEA-3(-) differ enough in gene expression to show a distinctive branching point at the end of the trajectory. The majority of *MIXL1*(**+**)/SSEA-3(-) cells occupy these end branch points but we do also see some of the *MIXL1*(**+**)/SSEA-3(**+**) cells also present towards the end of these trajectories (Figure 5-11A).

When I colour the cells by state we can observe a clear separation between the *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(-) groups with the *MIXL1*(**+**)/SSEA-3(**+**) overlapping both populations (Figure 5-11B). Monocle identified 7 cell states within the population based on their gene expression patterns, plotted along the trajectory. *MIXL1*(-)/SSEA-3(**+**) cells were found in states 1, 2 and 7. *MIXL1*(**+**)/SSEA-3(**+**) cells are also found in states 2 and 7 with a few cells also appearing in state 1. State 6 contains mainly cells from *MIXL1*(**+**)/SSEA-3(**+**) and marks the last state on the trajectory before the appearance of a large proportion of *MIXL1*(**+**)/SSEA-3(-) cells. The remaining groups are made up of mainly *MIXL1*(**+**)/SSEA-3(-) but also contains *MIXL1*(**+**)/SSEA-3(**+**) cells. I also coloured the cells with their estimated pseudotime, and see the progression along the trajectory through pseudotime (Figure 5-11C).

Figure 5-11 Cell Trajectories in Pseudotime

(Figure on next page)

A-C) Minimal spanning tree cell trajectories. Each dot represents and individual cells expression profile, plotted in a two-dimensional independent component space according to Monocle2s pseudotime ordering. The solid black line represents the path of the trajectory and branch points are labelled 1, 2 and 3. **A)** Cells are coloured according to the fraction which they are derived from *MIXL1*(-)/SSEA-3(**+**) (Red), *MIXL1*(**+**)/SSEA-3(**+**) (Green) and *MIXL1*(**+**)/SSEA-3(-) (Blue). **B)** Cells are coloured according to the state predicted by Monocle2, 7 states were predicted. **C)** Cells are coloured by their position in pseudotime.


A
Group
MIXL1(-) SSEA3(+)
MIXL1(+) SSEA3(+)
MIXL1(+) SSEA3(-)
Component 2
Component 1
State
1 3 5 7
2 4 6
B
Component 2
Component 1
Pseudotime
0 2 4 6 8
C
Component 2
Component 1

5.2.4.4. Gene expression through pseudotime

After ordering the cells in pseudotime I tracked the gene expression changes through pseudotime and visualised the changes by both heatmap (Figure 5-12) and branching gene expression plots (Figure 5-13), because the data comes from Ct values, the expression colouring is inverted with minus numbers being higher expression and positive numbers being lower expression. The heatmap clustered genes by their pattern of expression through pseudotime. Genes associated with pluripotency show a small upregulation followed by a down regulation near the end of pseudotime (Figure 5-12). Of these genes *SOX2* showed the greatest downregulation towards the end of pseudotime but some cells even in the *MIXL1*(-)/SSEA-3(**+**) fraction maintain *SOX2* expression, this is highlighted by the branch formation in the gene expression plot (Figure 5-13A)

Some genes associated with gastrulation such as *MIXL1*, *CER1, EOMES,* etc, exhibit low level expression at the start of pseudotime with increasing expression through pseudotime (Figure 5-13B). Some genes exhibited particularly high expression near the end of pseudotime such as *CXCR4, FRZB, HHEX,* and *SOX17*. Other genes associated with differentiation exhibit very low expression at the beginning of pseudotime and have a divergence of two populations both positive and negative for the marker at the end of pseudotime such as *T* and *FOXA2*(Figure 5-13B). The branching gene expression plots for all the genes are depicted in Figure 5-14

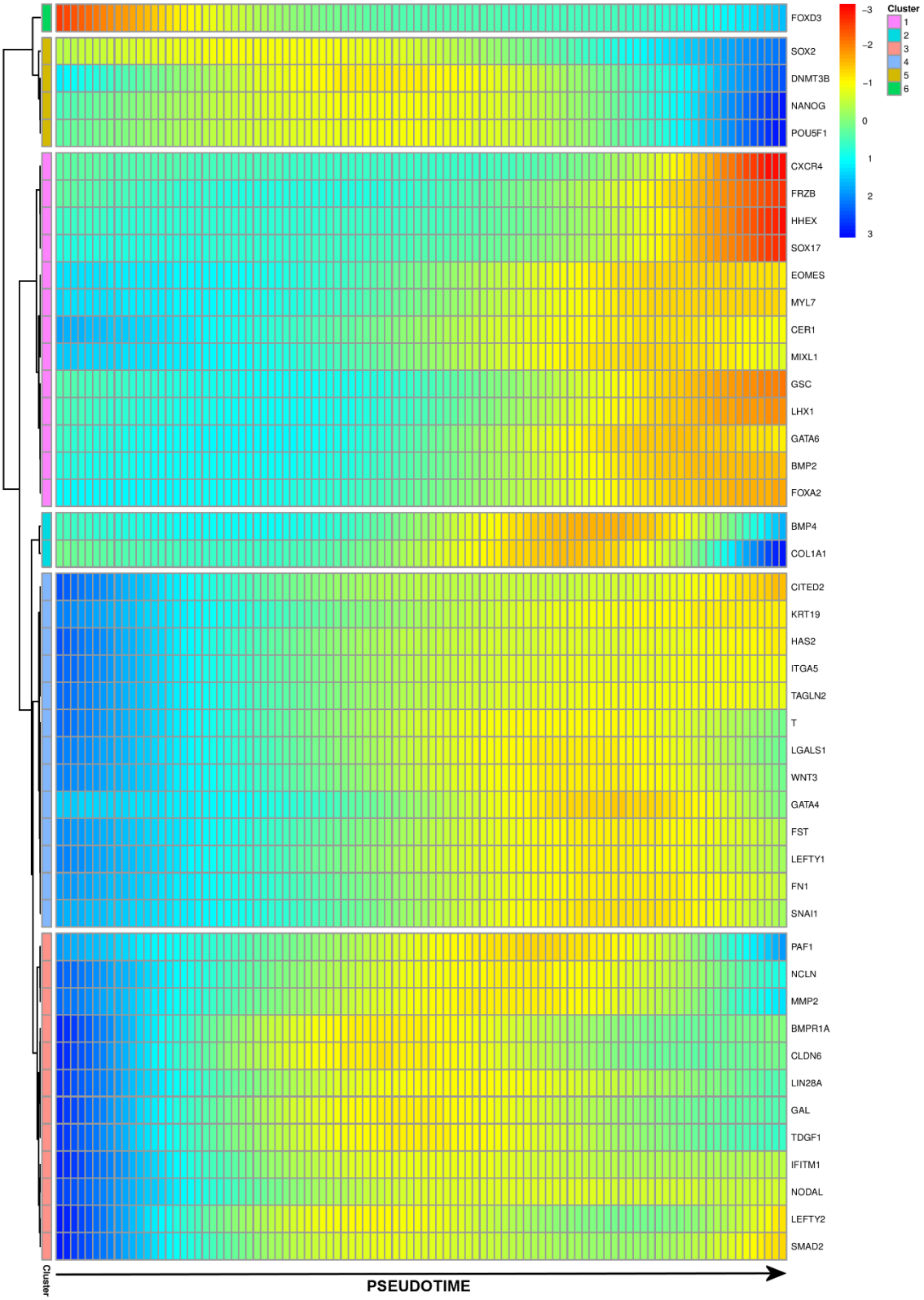
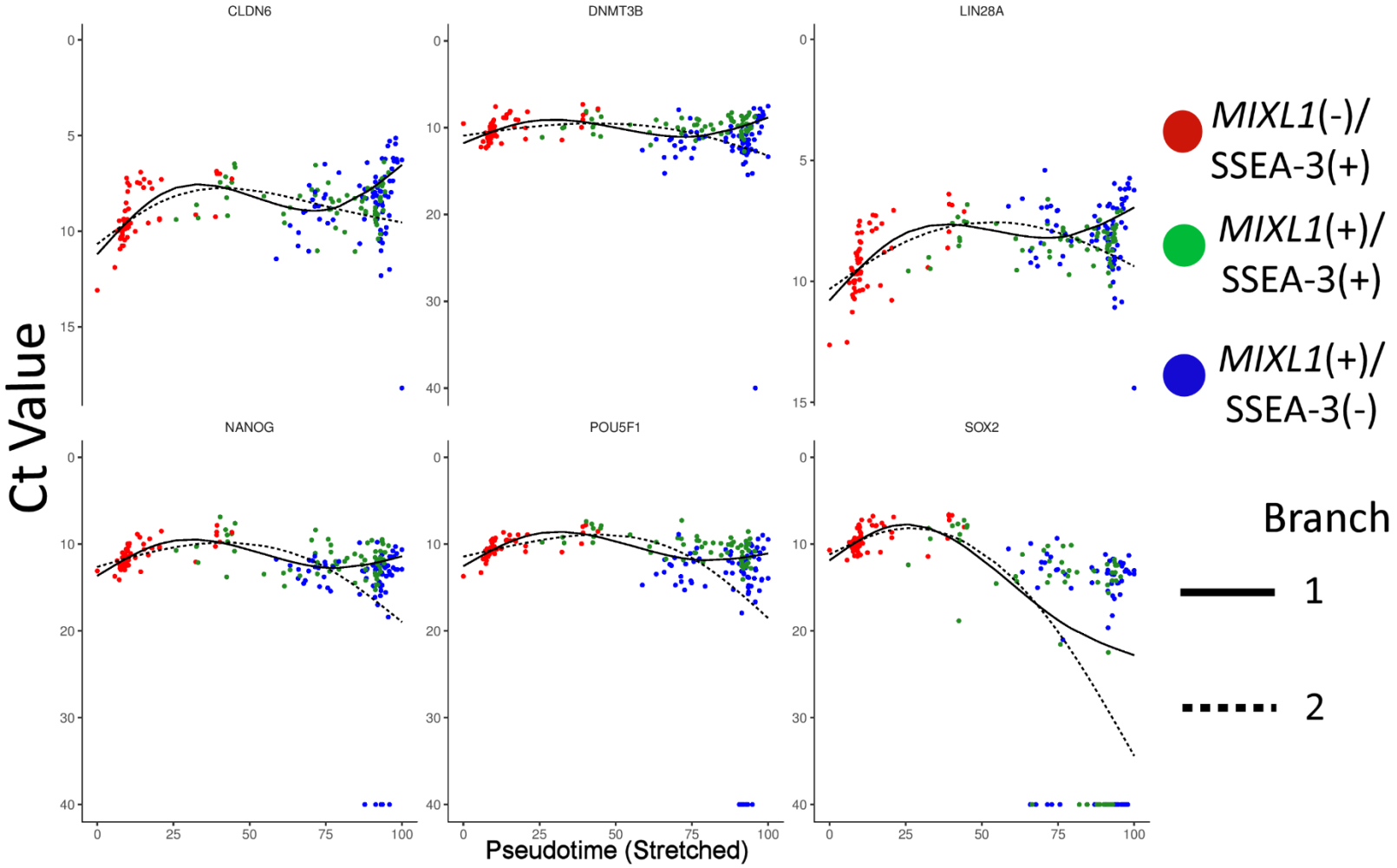


Figure 5-12 Pseudotime Gene Expression Heatmap

Heatmap representation of gene expression throughout pseudotime based on Monocle2 ordering of single cells assessed by qPCR. Genes are clusters into 6 clusters based on their expression profiles throughout pseudotime.

A


CLDN6
DNMT3B
LIN28A
NANOG
POU5F1
SOX2
Ct Value
Pseudotime (Stretched)
MIXL1(-)/ SSEA-3(+)
MIXL1(+)/ SSEA-3(+)
MIXL1(+)/ SSEA-3(-)
Branch
1
2

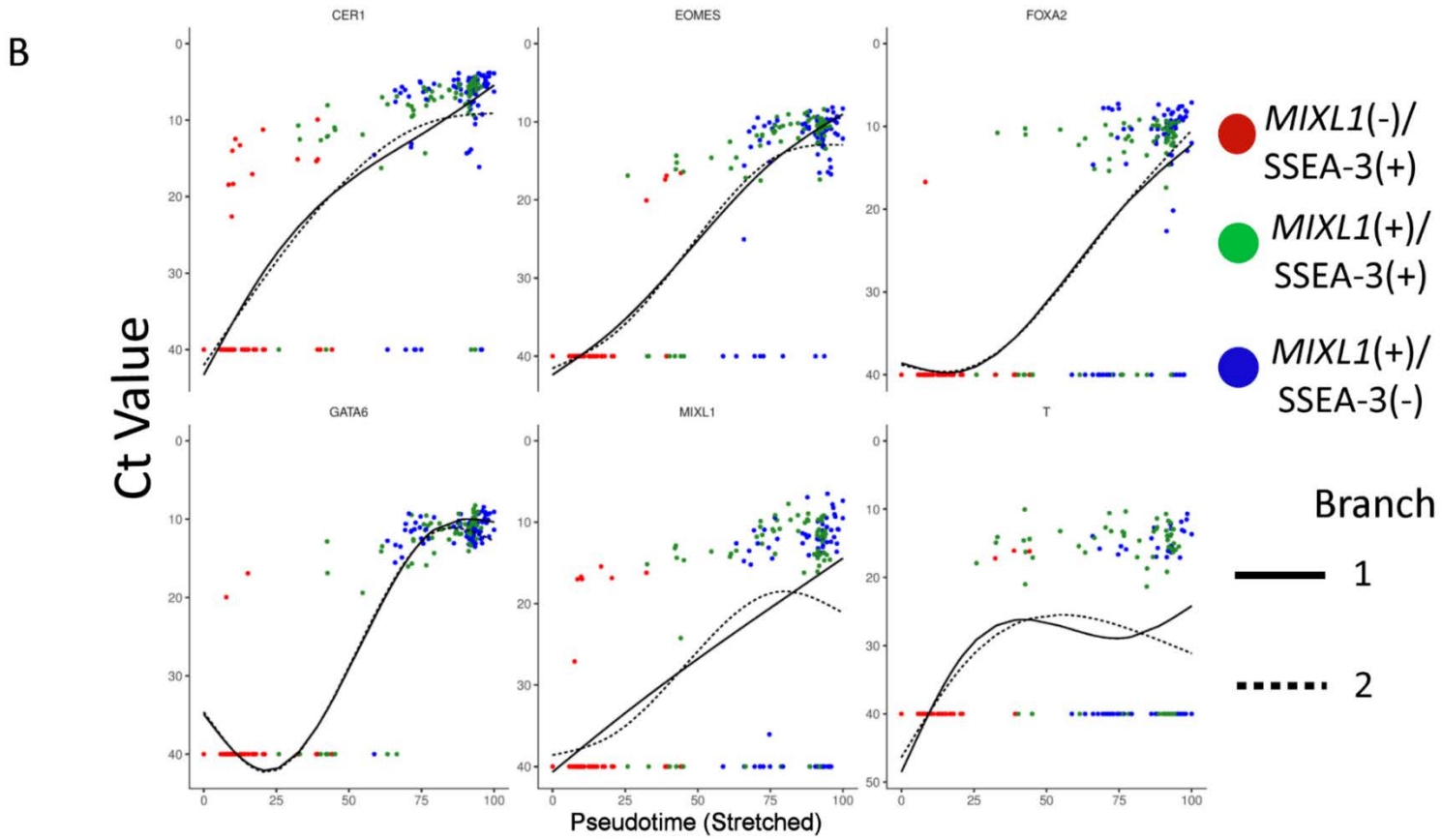
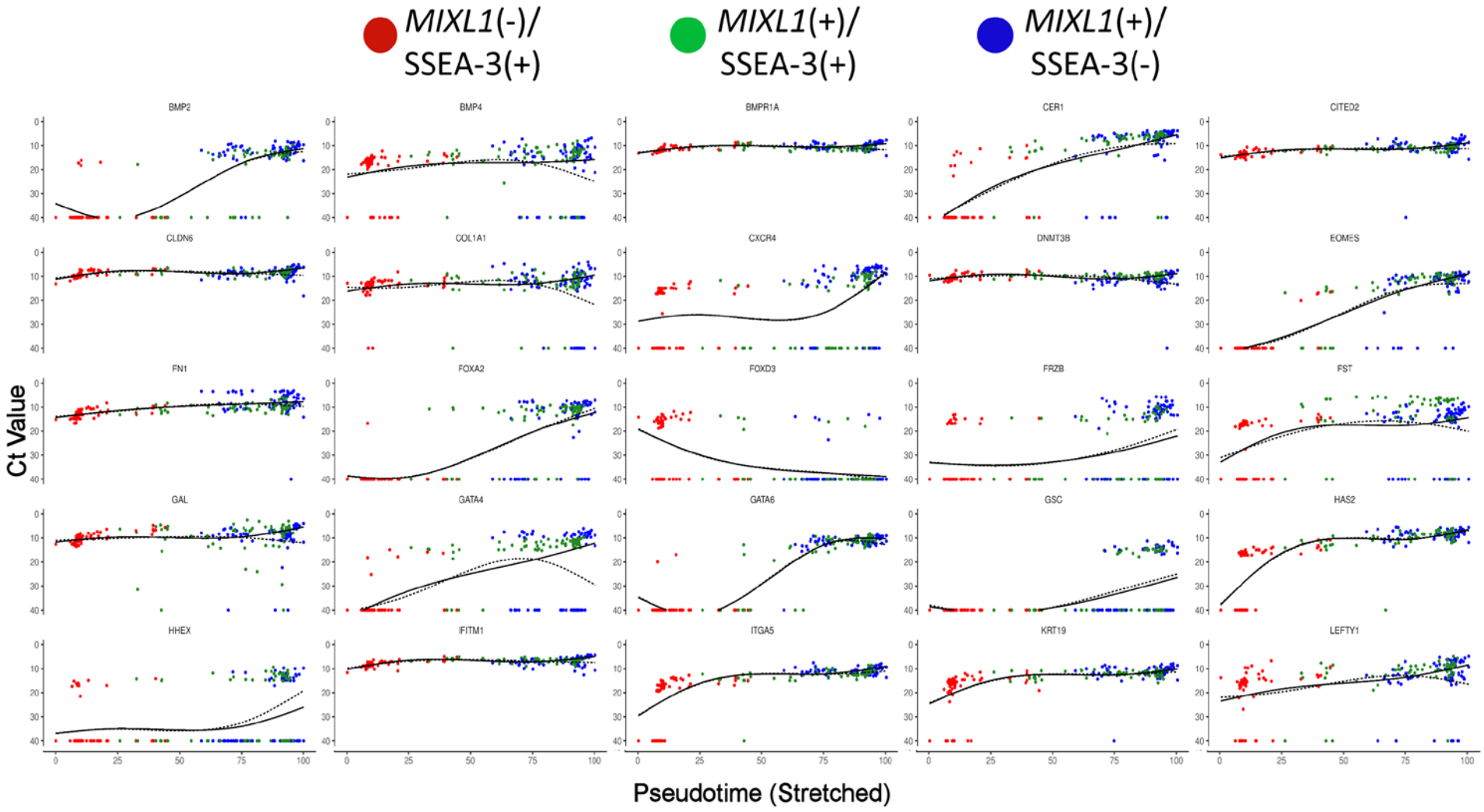


Figure 5-13 Gene Expression Profile through Pseudotime

Branching gene expression profiles of key genes from single cell qPCR analysis. Solid and dashed black lines represent the gene distribution of branching populations. Each dot represents the Ct value of the gene from a given cell from *MIXL1*(-)/SSEA-3(**+**), *MIXL1*(**+**)/SSEA-3(**+**), and *MIXL1*(**+**)/SSEA-3(-) from MEF/KOSR conditions. Cells are ordered left to right by their predicted position in pseudotime. **A)** Represents key genes related to pluripotency. **B)** Represents key genes related to differentiation.


MIXL1(-)/ SSEA-3(+)
MIXL1(+)/ SSEA-3(+)
MIXL1(+)/ SSEA-3(-)
BMP2
BMP4
BMPR1A
CER1
CITED2
CLDN6
COL1A1
CXCR4
DNMT3B
EOMES
FN1
FOXA2
FOXD3
FRZB
FST
GAL
GATA4
GATA6
GSC
HAS2
HHEX
IFITM1
ITGA5
KRT19
LEFTY1
Ct Value
Pseudotime (Stretched)

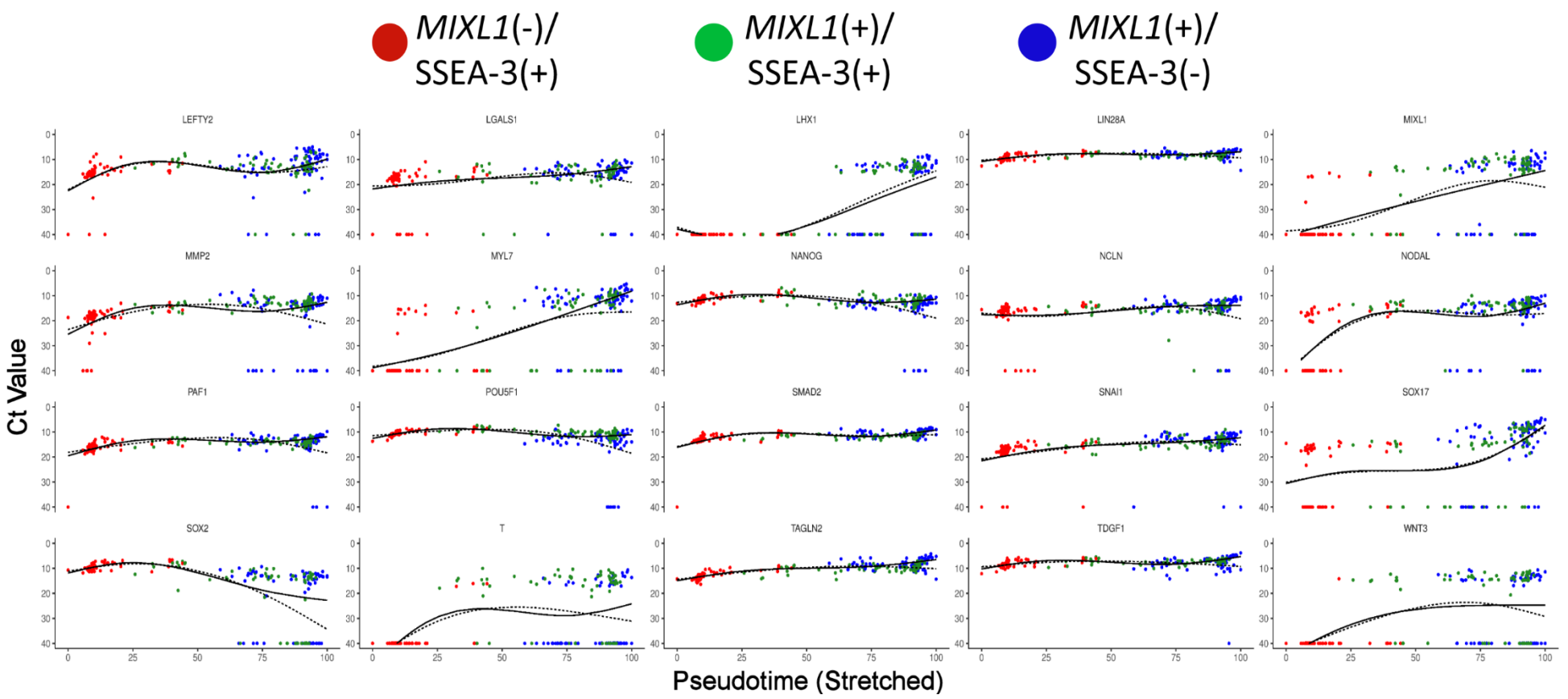


Figure 5-14 Gene Branched Trajectories

Branching gene expression profiles of all genes from single cell qPCR analysis. Solid and dashed black lines represent the gene distribution of branching populations. Each dot represents the Ct value of the gene from a given cell from *MIXL1*(-)/SSEA-3(**+**), *MIXL1*(**+**)/SSEA-3(**+**), and *MIXL1*(**+**)/SSEA-3(-) from MEF/KOSR conditions. Cells are ordered left to right by their predicted position in pseudotime.

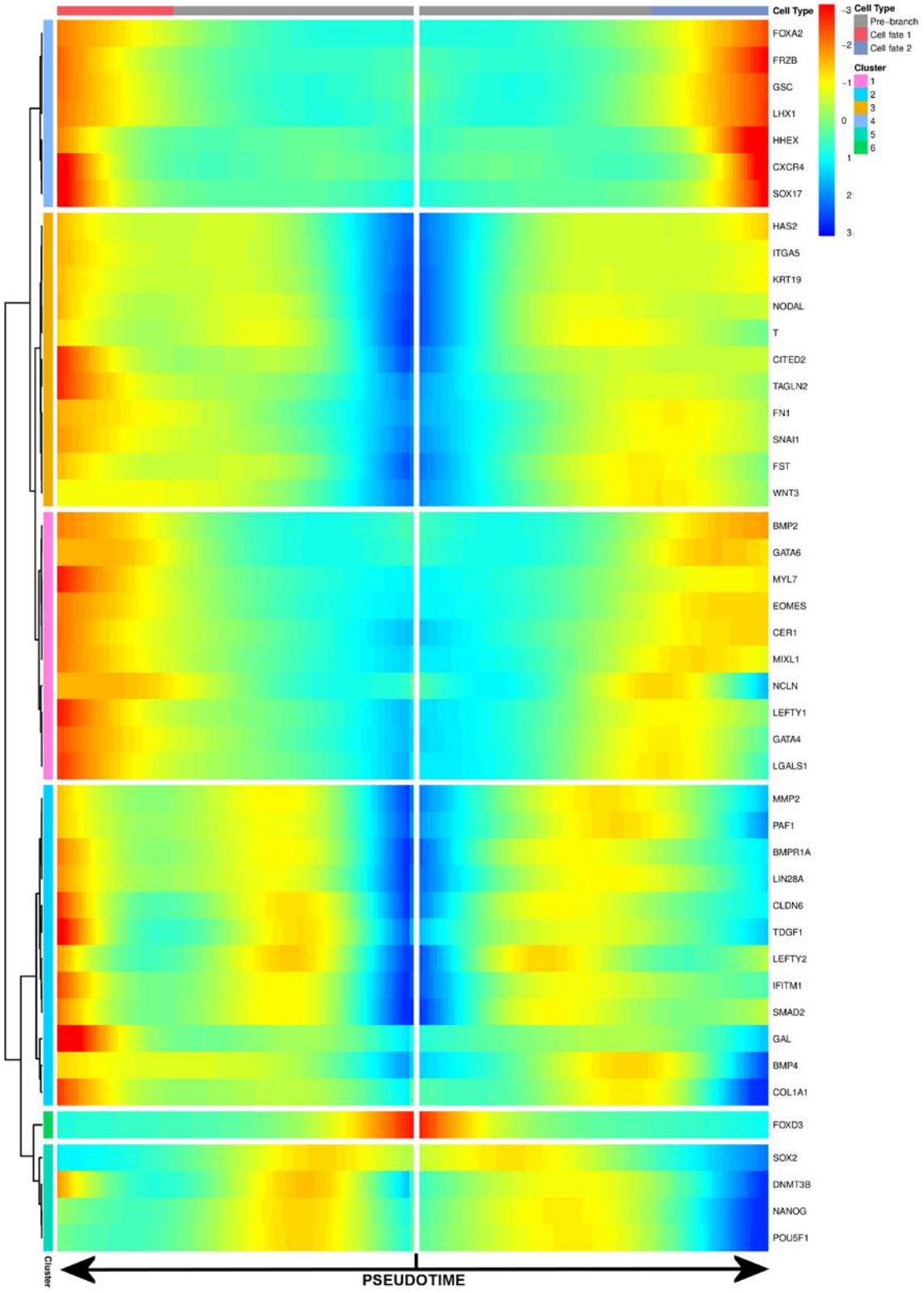


Figure 5-15 Branch Point Analysis

Heatmap representation of gene expression throughout pseudotime of two branching cell populations based on Monocle2 ordering of single cells assessed by qPCR. Genes are clusters into 6 clusters based on their expression profiles throughout pseudotime. Pseudotime begins in the centre and gene expression changes are displayed left and right towards two different cell types, or end branch points.

5.2.4.5. Branch point analysis

Monocle2 has identified 3 branch points but the most pronounced is branch point 1 (Figure 5-11) which segregates mainly the *MIXL1*(**+**)/SSEA-3(-) populations, but also cells from *MIXL1*(**+**)/SSEA-3(**+**) fraction. I assessed the differences between the two branches by heatmap analysis (Figure 5-15). The centre of the heatmap represents the gene expression at the start of pseudotime, in theory representing the undifferentiated state. The heatmap then displays the gene expression changes for the separate branches of the differentiation and clusters the genes based on expression pattern. At the top of the heatmap in the first cluster both branches have high expression of differentiation markers such as S*OX17*, *GSC*, *FOXA2*, with slightly higher expression in the right-hand branch. Further down the heatmap in the third cluster the left branch has higher expression of genes such as *GATA6* and *EOMES*. The bottom three clusters contain genes associated with pluripotency including *LIN28*, *NANOG*, *POU5F1* and *SOX2.* These clusters show a larger down regulation of these genes in the right-hand branch. The difference between these two branches appears as though the right-hand branch might be further differentiated than the left-hand branch. The right-hand branch is exhibiting high expression of differentiation genes such as *SOX17* and strong down regulation of pluripotency-associated genes, there is also a slight downregulation of the genes expressed during early mesendoderm differentiation such as *CER1*, *MIXL1* and *GATA6* compared to the left branch. The left branch does not exhibit as strong down regulation of pluripotency genes and has higher expression of early mesendoderm differentiation genes, and thus might relate to a slightly earlier stage in differentiation.

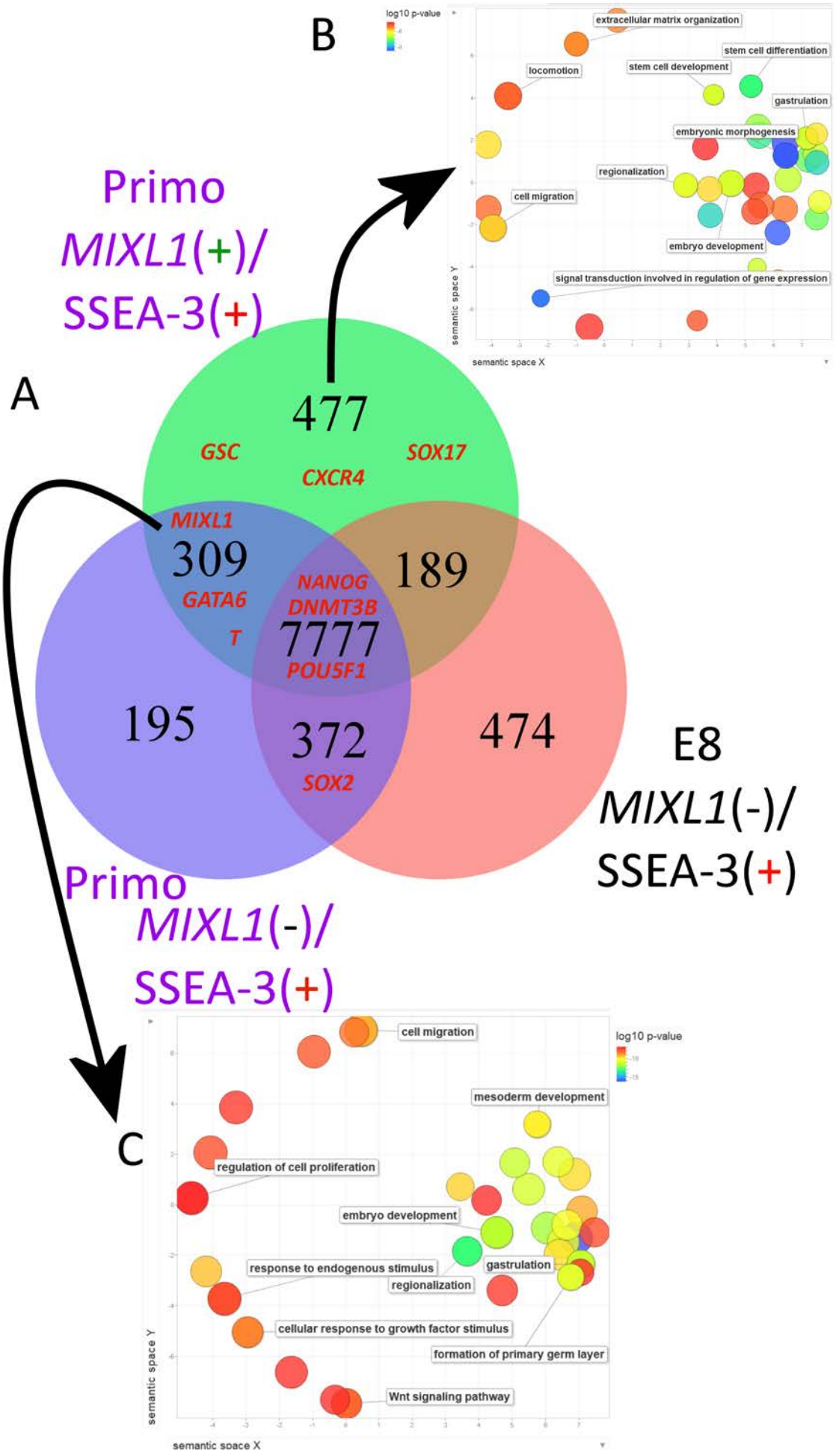

A
B
C
Primo
MIXL1(+)/
SSEA-3(+)
477
GSC
CXCR4
SOX17
MIXL1
309
GATA6
T
189
NANOG
DNMT3B
7777
POU5F1
195
372
SOX2
474
E8
MIXL1(-)/
SSEA-3(+)
Primo
MIXL1(-)/
SSEA-3(+)
log10 p-value
extracellular matrix organization
stem cell differentiation
locomotion
stem cell development
gastrulation
embryonic morphogenesis
regionalization
cell migration
embryo development
signal transduction involved in regulation of gene expression
semantic space Y
semantic space X
log10 p-value
cell migration
mesoderm development
regulation of cell proliferation
embryo development
response to endogenous stimulus
gastrulation
regionalization
cellular response to growth factor stimulus
formation of primary germ layer
Wnt signaling pathway
semantic space Y
semantic space X

Figure 5-16 Primo Fractions Gene Expression Grouping

A) This Venn diagram was produced using VennPlex program. Each circle corresponds to the total number of genes identified with a FPKM >5 for a given population based on *MIXL1* and SSEA-3 expression. Two fractions from Primo conditions, *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**) , were compared to *MIXL1*(-)/SSEA-3(**+**) fraction from E8 conditions. Overlaps in the circles relate to genes shared between the cell populations. The numbers for each overlap shows the number of genes which are found in the group. Overlapping area size is not proportional to the number of genes identified in the group. The group position of a few key genes related to pluripotency and differentiation are highlighted in red writing. **B-C)** Go enrichment analysis of genes found in *MIXL1*(**+**)/SSEA-3(**+**) fraction only **B)** and both Primo fractions **C)**. Biological Processes are coloured by their Log 10 p-values, as calculated by ToppGene, higher p-values (Red), lower p-values (Blue). Their positions in sematic space are arranged by ReviGO, calculated by the correlation between biological processes. Some key biological processes are highlighted.

5.2.5. Transcriptional analysis of human PSC sub-states generated using Primo medium.

Our work in chapter four indicated that the *MIXL1*(**+**)/SSEA-3(**+**) fractions generated in our Primo medium exhibited similar functional properties to that of *MIXL1*(**+**)/SSEA-3(**+**) fraction found in MEF/KOSR conditions. I wanted to assess whether the transcriptional similarities went beyond just the *MIXL1* expression that we identify with the GFP reporter. As with the fraction from MEF/KOSR I sought to analyse the fractions by bulk RNA-sequencing first (Figure 5-16).

Under Primo conditions there was less *MIXL1*(**+**)/SSEA-3(-) cells than in MEF/KOSR conditions and therefore I elected to focus only on SSEA-3(**+**) fractions. The expression of *MIXL1*(**+**)/SSEA-3(**+**) is variable in Primo conditions, between 20-60% of the population. However, a double positive fraction is always present. I sought to compare our generated fractions to the *MIXL1*(-)/SSEA-3(**+**) cells in E8 medium to assess any changes in gene expression from the starting self-renewal conditions (Figure 5-16A). Comparing these fractions there were stark differences between the *MIXL1*(-)/SSEA-3(**+**) fractions from Primo and E8. The *MIXL1*(-)/SSEA-3(**+**) fraction from the Primo medium shares overlapping expression of 309 genes with the *MIXL1*(**+**)/SSEA-3(**+**) fraction, that are non or lowly expressed in *MIXL1*(-)/SSEA-3(**+**) cells from E8 conditions. Genes in this overlapping section included genes such as *MIXL1, GATA6* and *T.* GO enrichment analysis of these genes correlates with biological processes such as gastrulation and mesoderm development (Figure 5-16C).

MIXL1(**+**)/SSEA-3(**+**) cells from the Primo condition have 477 genes uniquely expressed from the other fractions. Genes in this overlapping section included genes such as *GSC, CXCR4* and *SOX17.* Similar to the collection of genes from the *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**) overlapping fractions, GO enrichment analysis of these genes correlates with biological

processes such as gastrulation and stem cell development (Figure 5-16B). Pluripotency-associated genes such as *NANOG, DNMT3B* and *POU5F1* showed expression in all fractions, whereas *SOX2* in particular was confined to the *MIXL1*(-)/SSEA-3(**+**) fractions. GO enrichment of the remaining groups revealed no significant enrichment of GO terms.

5.2.6. Differentiation Time Course.

Figure 5-17 details a seventy-two hour time course of differentiation of HES *MIXL1* cells in E8 with 3μM CHIRON added to the medium. Cells were collected and stained with anti-SSEA-3 antibody at time points 6, 12, 18, 24, 48, and 72 hours. We can see the changes in both *MIXL1*-GFP and SSEA-3 expression over the time course. *MIXL1*-GFP becomes detectable in a small proportion of cells at 12 hours. This expression continues to increase in both intensity and proportion up to 48 hours into the differentiation. The expression of SSEA-3 however does not exhibit any discernible change until 72 hours, where a large proportion of cells are now SSEA-3 negative and *MIXL1*-GFP positive. For the purpose of our transcriptomic analysis I elected to take just the emerging population at each time point, by which I mean the population of cells that based on *MIXL1*/SSEA-3 expression that was not present in the previous time point, (Figure 5-17A) in order to track the progression through differentiation.

RNA-sequencing data reveals the changes in gene transcription between the different time points, heatmaps analysis details these changes (Figure 5-18). After analysing the time points, over 200 genes exhibited a standard deviation above 1.5 throughout the time course. This was further separated into genes that were both highly downregulated and highly upregulated. Genes associated with pluripotency show downregulation over the time course, the most sensitive being *NANOG* and *SOX2* whereas *POU5F1* maintains higher expression levels up until 72 hours (Figure 5-18A). Genes associated with gastrulation and early differentiation show increasing upregulation (Figure 5-18B). Genes such as *MIXL1*, and *T* show no expression at the 0 hours but by 6 hours are already detectable by RNA-seq Both *MIXL1* and *T* continue to increase and reach their peak expression level at 24 hours (Figure 5-18C). Some genes are expressed abundantly only after 72 hours such as *HAND1*.

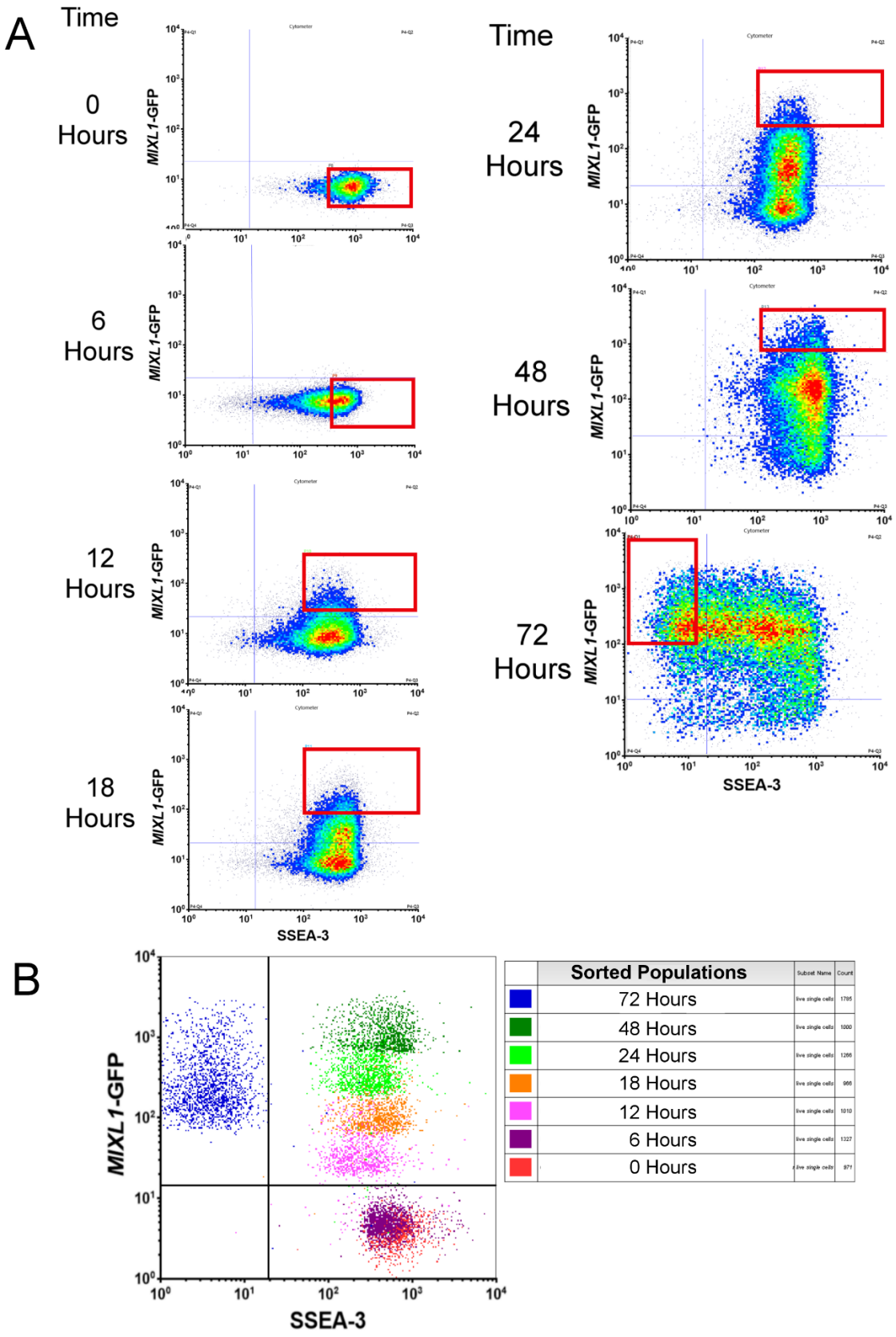


Figure 5-17 Differentiation Time Course

A) Flow cytometry density plots of HES3 *MIXL1*-GFP cells stained for SSEA-3 at indicated time points after induction of differentiation in E8 containing 3µM CHIRON. Red boxes indicate the sorting gates for each timepoint. **B)**The reanalysis of the emerging population sorted from each time point, 0 to 72 hours of differentiation.

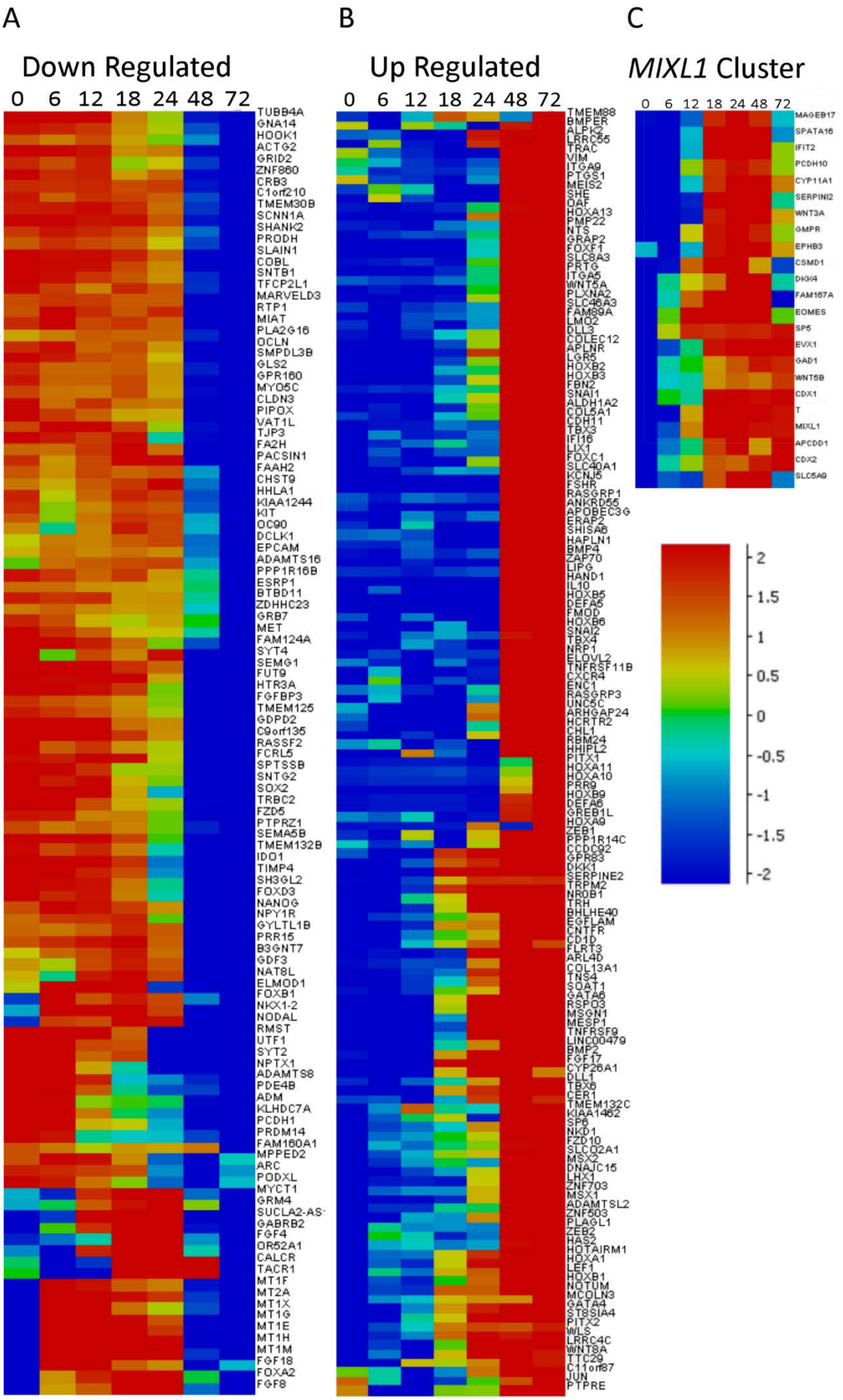

A
Down Regulated
0 6 12 18 24 48 72
TUBB4A
GNA14
HOOK1
ACTG2
GRID2
ZNF860
CRB3
C1orf210
TMEM30B
SCNN1A
SHANK2
PRODH
SLAIN1
COBL
SNTB1
TFCP2L1
MARVELD3
RTP1
MIAT
PLA2G16
OCLN
SMPDL3B
GLS2
GPR160
MYO5C
CLDN3
PIPOX
VAT1L
TJP3
FA2H
PACSIN1
FAAH2
CHST9
HHLA1
KIAA1244
KIT
OC90
DCLK1
EPCAM
ADAMTS16
PPP1R16B
ESRP1
BTBD11
ZDHHC23
GRB7
MET
FAM124A
SYT4
SEMG1
FUT9
HTR3A
FGFBP3
TMEM125
GDPD2
C9orf135
RASSF2
FCRL5
SPTSSB
SNTG2
SOX2
TRBC2
FZD5
PTPRZ1
SEMA5B
TMEM132B
IDO1
TIMP4
SH3GL2
FOXD3
NANOG
NPY1R
GYLTL1B
PRR15
B3GNT7
GDF3
NAT8L
ELMOD1
FOXB1
NKX1-2
NODAL
RMST
UTF1
SYT2
NPTX1
ADAMTS8
PDE4B
ADM
KLHDC7A
PCDH1
PRDM14
FAM160A1
MPPED2
ARC
PODXL
MYCT1
GRM4
SUCLA2-AS
GABRB2
FGF4
OR52A1
CALCR
TACR1
MT1F
MT2A
MT1X
MT1G
MT1E
MT1H
MT1M
FGF18
FOXA2
FGF8
B
Up Regulated
0 6 12 18 24 48 72
TMEM88
BMPER
ALPK2
LRRC55
TRAC
VIM
ITGA9
PTGS1
MEIS2
SHE
OAF
HOXA13
PMP22
NTS
GRAP2
FOXF1
SLC8A3
PRTG
ITGA5
WNT5A
PLXNA2
SLC46A3
FAM89A
LMO2
DLL3
COLEC12
APLNR
LGR5
HOXB2
HOXB3
FBN2
SNAI1
ALDH1A2
COL5A1
CDH11
TBX3
IFI16
LIX1
FOXC1
SLC40A1
KCNJ5
FSHR
RASGRP1
ANKRD55
APOBEC3G
ERAP2
SHISA6
HAPLN1
BMP4
ZAP70
LIPG
HAND1
IL10
HOXB5
DEFA5
FMOD
HOXB6
SNAI2
TBX4
NRP1
ELOVL2
TNFRSF11B
CXCR4
ENC1
RASGRP3
UNC5C
ARHGAP24
HCRTR2
CHL1
RBM24
HHIPL2
PITX1
HOXA11
HOXA10
PRR9
HOXB9
DEFA6
GREB1L
HOXA9
ZEB1
PPP1R14C
CCDC92
GPR83
DKK1
SERPINE2
TRPM2
NR0B1
TRH
BHLHE40
EGFLAM
CNTFR
CD1D
FLRT3
ARL4D
COL13A1
TNS4
SOAT1
GATA6
RSPO3
MSGN1
MESP1
TNFRSF9
LINC00479
BMP2
FGF17
CYP26A1
DLL1
TBX6
CER1
TMEM132C
KIAA1462
SP6
NKD1
FZD10
SLCO2A1
MSX2
DNAJC15
LHX1
ZNF703
MSX1
ADAMTSL2
ZNF503
PLAGL1
ZEB2
HAS2
HOTAIRM1
HOXA1
LEF1
HOXB1
NOTUM
MCOLN3
GATA4
ST8SIA4
PITX2
WLS
LRRC4C
WNT8A
TTC29
C11orf87
JUN
PTPRE
C
MIXL1 Cluster
0 6 12 18 24 48 72
MAGEB17
SPATA16
IFIT2
PCDH10
CYP11A1
SERPINI2
WNT3A
GMPR
EPHB3
CSMD1
DKK4
FAM167A
EOMES
SP5
EVX1
GAD1
WNT5B
CDX1
T
MIXL1
APCDD1
CDX2
SLC5A9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2

Figure 5-18 Gene Expression Changes During Differentiation

RNA-sequencing data from each time course was compared using SeqMonk and genes that displayed a standard deviation above 1.5 were selected. **A)** Displays a heatmap of the genes which are highly downregulated over the time course, this cluster includes genes related to pluripotency such as *SOX2, NANOG* and *POU5F1.* **B)** Displays a heatmap of the genes that highly upregulated over the time course, this cluster includes genes related to differentiation such as *CER1, GATA6* and *HAND1.* **C)** Displays a heatmap of genes which show a similar expression change to that of *MIXL1*, this includes genes such as *CDX2, EOMES* and *T*.

5.2.6.1. Gene Clouds of Differentiation Time Points

While traditional heatmaps offer a visualisation of gene expression change per gene, gene clouds offer us a visual "snapshot" of the transcriptomic landscape of a given state. To generate word clouds firstly the RNA-seq data from the time course was compared and genes with a standard deviation above 1.5 were selected for analysis. For the purposes of visualising this data as a gene cloud, the FPKM (Fragments Per Kilobase per Million reads) values for each gene was normalised to GPI (Glucose-6-Phosphate Isomerase), which had an FPKM value between 80 – 130 FPKM. The top 50 expressing genes were then taken forward to generate a gene cloud. The top 50 genes shows the major genes involved in the current gene regulatory network of the assessed state. The size of the gene in the cloud is correlated to their relative expression after normalisation to *GPI* (Figure 5-19).

At 0 hours, unsurprisingly, the gene expression is dominated by pluripotency-associated genes such as *POU5F1* and *CLDN6* represented with purple text, and very little expression of genes related to differentiation (Figure 5 19A). The pluripotency-associated genes show a progressive downregulation throughout the differentiation time course and by 48 hours do not form a major contribution to the gene clouds.

As expected, genes associated with differentiation show an inverse pattern of expression compared to pluripotency-associated genes. Gastrulation associated genes such as *NODAL* and *MIXL1* are among the first to contribute to the earliest changes in gene expression, visible in 6 and 12 hour gene clouds (Figure 5-19A). Genes associated with differentiation begin to form a more substantial part of the top 50 genes after 24 hours of differentiation, with genes such as *DKK1, MIXL1* and *T*(*BRACHYURY*) exhibiting similar expression levels as remaining pluripotency genes (Figure 5 16B). *DKK1* along with *KRT19* form major contributions to gene clouds at 48 and 72 hours. *HAND1,* a gene associated with cardiac mesoderm (McFadden et al., 2005) does not appear in the gene cloud until 72 hours (Figure 5-19B).

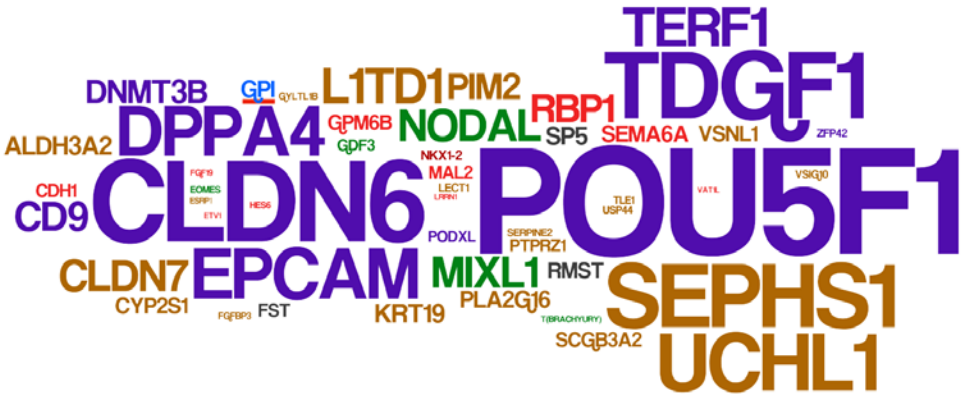
A

Hours

0

6

12



18

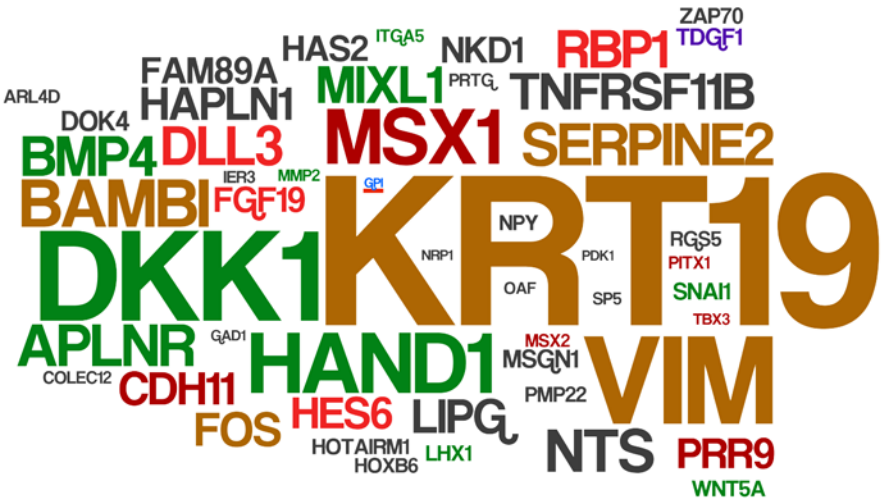
B

Hours

24

48

72



C

D

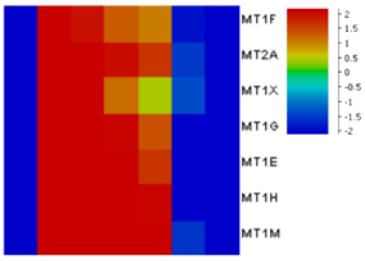


Figure 5-19 Gene Clouds of Differentiation Time Points

Gene expression changes visualised as Gene clouds. RNA-sequencing data from each time course was compared using SeqMonk and genes that displayed a standard deviation above 1.5 were selected. Gene expression for each time point was normalised to *GPI* (Glucose-6-Phosphate Isomerase), which expression ranged between 80 to 130 FPKM. *GPI* is present in each word cloud in blue underlined in red. After screening the top 50 expressing genes were chosen to make gene clouds. The size of the gene name in each gene cloud is relative expression of the genes calculated by the FPKM value after normalisation to *GPI*. Genes are coloured according to related biological process pluripotency (purple), gastrulation (green), regulation of cell differentiation (red), other biological processes (bronze and grey). **A)** Gene clouds for 0, 6, 12 and 18 hours of differentiation. **B)** Gene clouds for 24, 48 and 72 hours of differentiation. **C)** Displays a gene cloud of the metallothionein gene group at 12 hours, *GPI* is circled in red. **D)** Heatmap analysis of metallothionein gene group.

A collection of genes encoding metallothioneins showed abundant upregulation over the first 24 hours these were removed from the main gene clouds as their expression dwarfed the other genes in the clouds. Figure 5-19C displays a gene cloud using only the metallotionein genes and the housekeeping gene *GPI* for 12 hour time point. At the 12 hour time point *GPI* has an expression value of 80 FPKMs and *POU5F1* at 512, but *MTX1* has an expression value of over 2500 FPKM. This is reflected in the size difference between *GPI* and *MTX1* in the gene cloud. The metallothionein expression persisted abundantly over the first 24 hours and was all but absent in the 48 and 72 hour time point, heatmap analysis highlights this (Figure 5 19D).

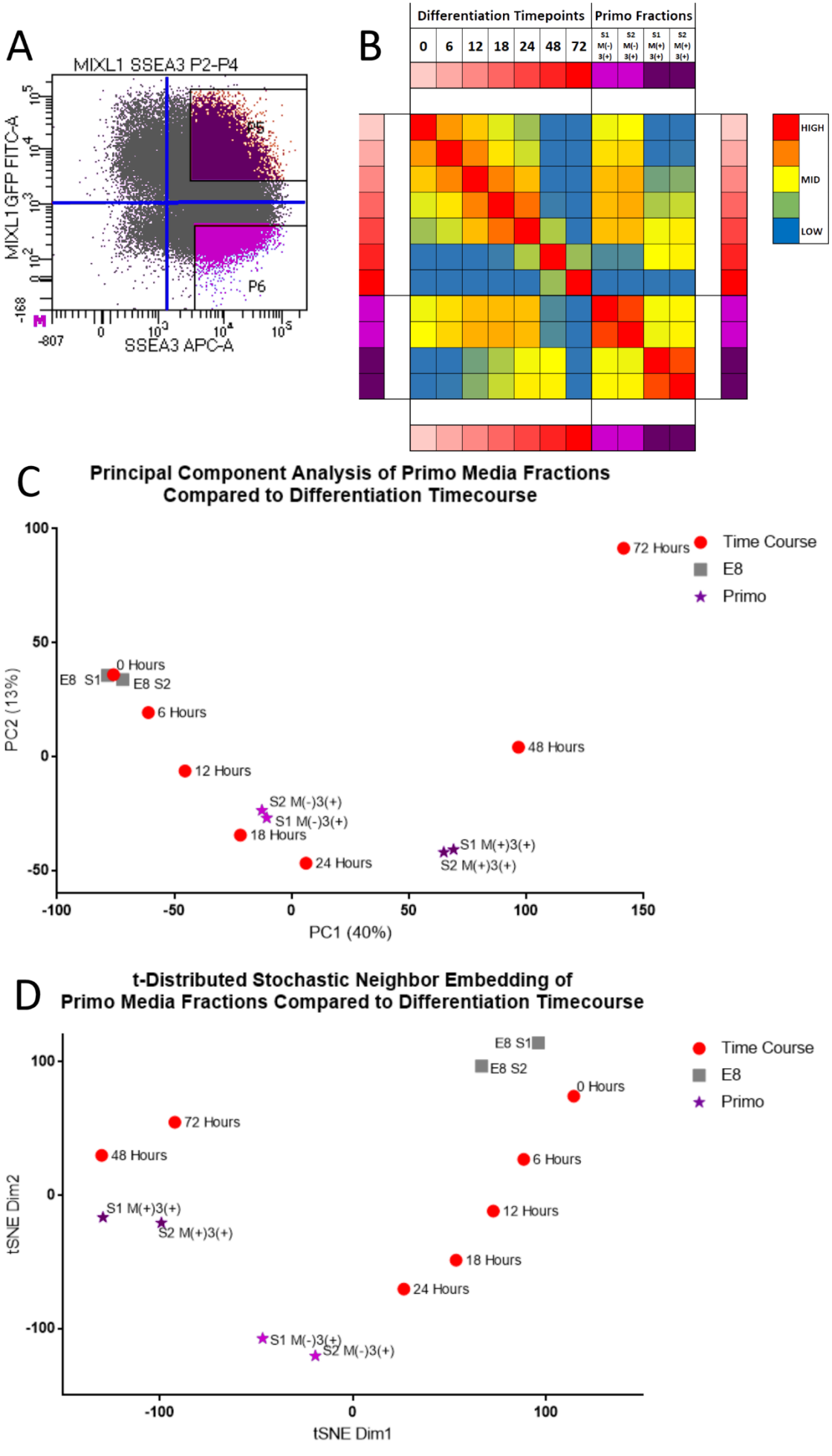

A
MIXL1 SSEA3 P2-P4
MIXL1 GFP FITC-A
SSEA3 APC-A
P5
P6
B
Differentiation Timepoints
Primo Fractions
0
6
12
18
24
48
72
S1 M(-) 3(+)
S2 M(-) 3(+)
S1 M(+) 3(+)
S2 M(+) 3(+)
HIGH
MID
LOW
C
Principal Component Analysis of Primo Media Fractions Compared to Differentiation Timecourse
PC2 (13%)
PC1 (40%)
72 Hours
Time Course
E8
Primo
0 Hours
E8 S1
E8 S2
6 Hours
48 Hours
12 Hours
S2 M(-)3(+)
S1 M(-)3(+)
18 Hours
S1 M(+)3(+)
S2 M(+)3(+)
24 Hours
D
t-Distributed Stochastic Neighbor Embedding of Primo Media Fractions Compared to Differentiation Timecourse
tSNE Dim2
tSNE Dim1
E8 S1
E8 S2
0 Hours
72 Hours
48 Hours
6 Hours
S1 M(+)3(+)
S2 M(+)3(+)
12 Hours
18 Hours
24 Hours
S1 M(-)3(+)
S2 M(-)3(+)
Time Course
E8
Primo

Figure 5-20 Comparing Primo Fraction to Differentiation Time course

A) Flow cytometry scatter plot of HES3 *MIXL1*-GFP cells stained for SSEA-3 grown in Primo conditions. Two fractions were sorted for RNA sequencing analysis *MIXL1*(**+**)/SSEA-3(**+**) (dark purple) and *MIXL1*(-)/SSEA-3(**+**) (light purple). **B)** RNA-sequencing similarity matrix of Primo fractions (purple) against differentiation time points(red) **C)** Principal component analysis of *MIXL1*(**+**)/SSEA-3(**+**) (dark purple) and *MIXL1*(-)/SSEA-3(**+**) (light purple) cells grown in Primo compared to the differentiation time course. **D)** t-Distributed Stochastic Neighbour Embedding (t-SNE) plot of *MIXL1*(**+**)/SSEA-3(**+**) (purple) and *MIXL1*(-)/SSEA-3(**+**) cells grown in Primo compared to the differentiation time course.

5.2.7. Comparing Primo Fraction to Differentiation Time course

I sorted *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**) from cells growing in 1xPrimo conditions for 3 days and performed bulk RNA-sequencing on both fractions (Figure 5-20 A). The transcriptome of these fractions was compared to the time course of differentiation. The *MIXL1*(-)/SSEA-3(**+**) fraction, showed strong correlation to the emerging populations at time points 0-24 hours (Figure 5-20 B). The *MIXL1*(**+**)/SSEA-3(**+**) fraction exhibited less correlation to earlier time points and shifted towards 48 hour time point. Neither fraction correlated with the *MIXL1*(**+**)/SSEA-3(-) fraction present at 72 hours (Figure 5-20 B).

I used both Principal component analysis and t-SNE analysis to compare *MIXL1*(**+**) and *MIXL1*(-), SSEA-3(**+**) cells from the Primo conditions. Both by PCA and t-SNE showed similar mapping. The *MIXL1*(-)/SSEA-3(**+**) fractions obtained from E8V conditions correlate well with our zero hour time point which also was grown in E8V conditions. The *MIXL1*(**+**)/SSEA-3(**+**) fraction maps between 24 and 48 hours even though they have been in the presence of 3µM CHIRON in Primo for 72 hours (Figure 5-20 C, D).

Interestingly *MIXL1*(-)/SSEA-3(**+**) fraction from Primo conditions did not correlate with the zero hour or *MIXL1*(-)/SSEA-3(**+**) fractions from E8V but rather it maps between 12 and 18 hours, closer to 18 hours by PCA analysis and between 24-48 hours by t-SNE (Figure 5-20 C, D). This fraction while having expression of pluripotency-associated genes such as *SOX2*, *NANOG*, etc. also has expression of differentiation genes such as *MIXL1*, *T* (*BRACHYURY)*, *GATA6*, etc. (Figure 5-16). These genes segregate this fraction from *MIXL1*(-)/SSEA-3(**+**) cells from E8V conditions. In the Primo conditions both *MIXL1*(**+**) and *MIXL1*(-) fractions exhibit expression markers of early differentiation, including *MIXL1*.

5.2.8. Gene Clouds of Primo Fractions

As I did with the time course data I created gene clouds to display to the top 50 expressing genes of the state. Both PCA and tSNE analysis of Primo *MIXL1*(-)/SSEA-3(**+**) fraction revealed a correlation with early differentiation. When we observe the generated gene cloud of the state we can see the presence of a pluripotency network with high expression of genes such as *CLDN6* and *POU5F1.* We also see coexpression of differentiation markers such as *CER1, MIXL1* and *NODAL* (Figure 5-21A). This gene cloud displayed similar gene expression patterns as 12 and 18 hour differentiation time points, complementing the findings of the PCA analysis (Figure 5 16A).

The gene cloud for Primo *MIXL1*(**+**)/SSEA-3(**+**) fraction (Figure 5 16B) has similar expression patterns to the MEF/KOSR *MIXL1*(**+**)/SSEA-3(**+**). As with the fraction from MEF/KOSR, *CER1* and *KRT19* show abundant expression. *CER1,* whilst present did not feature predominantly in any of the differentiation time points analysed. Genes associated with pluripotency such as *CLDN6* and *POU5F1* are still present but have an apparent downregulation, forming a smaller part of the gene expression. *DKK1* also forms a major part of the gene cloud which is similar to what is seen at 24 and 48 hours differentiation time points. The remaining presence of some pluripotency-associated genes and the lack of expression of further differentiation genes such as *HAND1,* separates this fraction from the 72 hour differentiation time point (Figure 5 16B).

Figure 5-21 Gene Expression of Primo Fractions

(Figure on next page)

The same gene list generated in the analysis of the timecourse was used and screened for the top 50 expressing genes. Gene clouds from the *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(**+**) fractions from Primo conditions are displayed. The size of the gene name in each gene cloud is relative expression of the genes calculated by the FPKM value after normalisation to *GPI*. *GPI* is present in each word cloud in blue underlined in red. Genes are coloured according to related biological process pluripotency (purple), gastrulation (green), regulation of cell differentiation (red), other biological processes (bronze and grey).

MIXL1(-)/
SSEA-3(+)

MIXL1(+)/
SSEA-3(+)

5.2.9. Bulk RNA-sequencing Assessment of fractions isolated from MEF/KOSR and Primo conditions.

The cells generated in Primo displayed upregulation of genes associated early differentiation and the fractions correlated to early differentiation compared a differentiation time-course. Further to this, I assessed the similarity of our generated Primo fractions to the fractions which we identified in MEF/KOSR conditions. I also included the *MIXL1*(-)/SSEA-3(**+**) fraction from E8 in the comparison as this represents the population from which cells in Primo medium are generated. Hierarchical clustering of the fractions revealed separation into two main clusters based on *MIXL1*-GFP expression (Figure 5-22A). The *MIXL1*(-)/SSEA-3(**+**) fractions from all conditions clustered together within the main clusters, sub-clusters segregated the Primo fraction from the others in standard media. Despite the upregulation of differentiation associated genes in the Primo *MIXL1*(-)/SSEA-3(**+**) fraction it still clustered closer to the more pristine populations present in standard media.

The second main cluster encompasses all the *MIXL1*-GFP (**+**) fractions from all conditions. Sub-clusters revealed a closer correlation with the *MIXL1*(**+**)/SSEA-3(**+**) from Primo and MEF/KOSR conditions than that of *MIXL1*(**+**)/SSEA-3(-). Visualising the data using a t-SNE plot displayed a similar correlation between the fractions (Figure 5-22B). If we think of the t-SNE plot as a continuum from *MIXL1*(-)/SSEA-3(**+**) in E8 and MEF/KOSR as pristine pluripotent cells to *MIXL1*(**+**)/SSEA-3(-) as our differentiated fraction, our Primo fractions and MEF/KOSR *MIXL1*(**+**)/SSEA-3(**+**) spanning the space between.

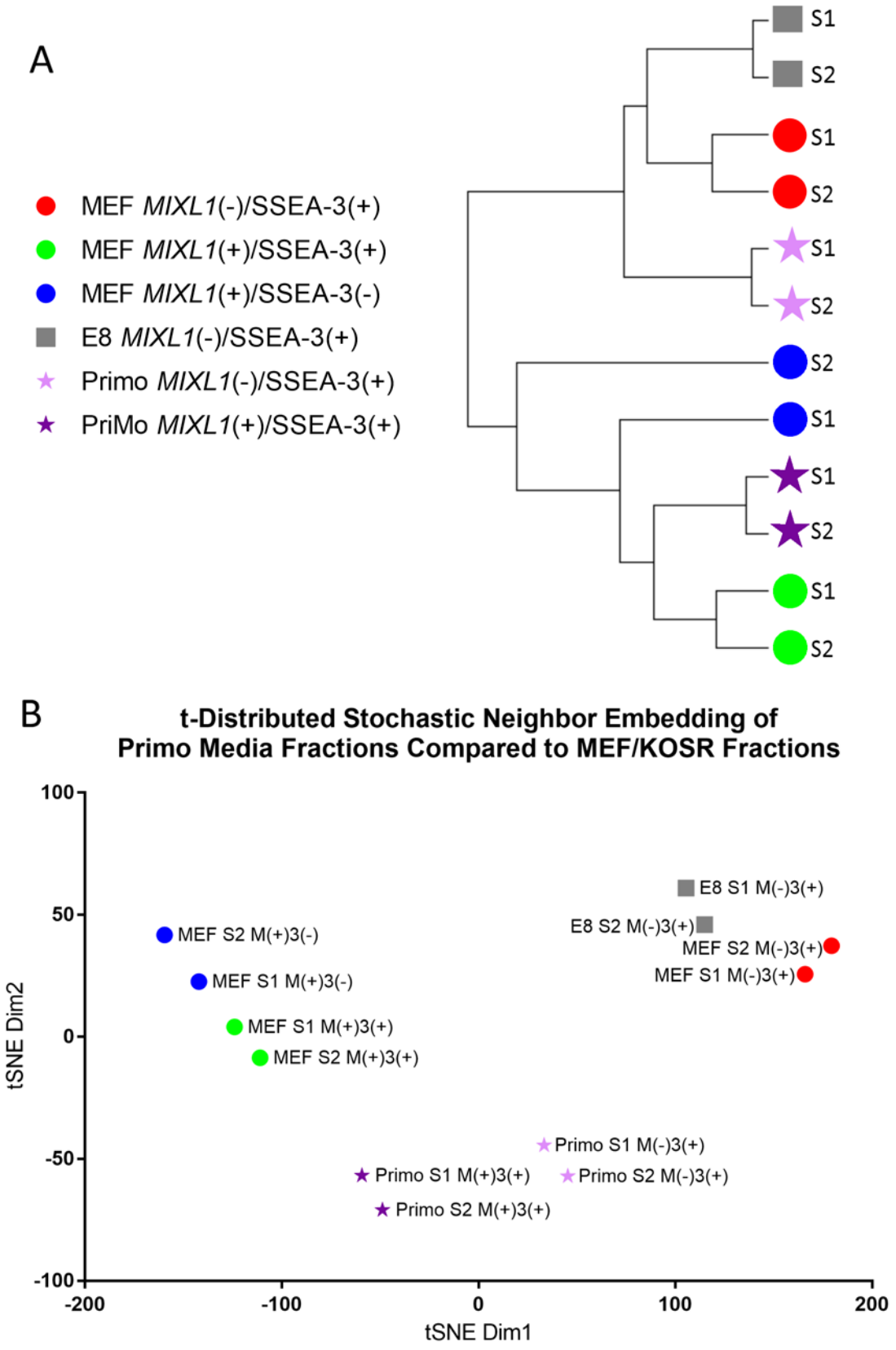


Figure 5-22 Bulk RNA-seq Comparison of Fractions from Different Culture Systems

Cells grown in E8, MEF/KOSR and Primo conditions, sorted based on *MIXL1*-GFP and SSEA-3 expression, assessed by bulk RNA-sequencing. **A)** Hierarchical cluster analysis and **B)** tSNE analysis of RNA-seq data.

5.2.9.1. Single Cell qPCR analysis

As we had done with our *MIXL1* fractions from MEF/KOSR conditions, we utilised single cell qPCR analysis to assess the both the heterogeneity of our *MIXL1*(**+**)/SSEA-3(**+**) fraction generated in Primo medium and their similarities to fractions seen from MEF/KOSR conditions. In order to get a signature of the state I took the average gene expression of the cells and compared that to the signature of fractions seen in MEF/KOSR conditions. The genes are ordered on the graphs from high to low according to the average expression for the *MIXL1*(-)/SSEA-3(**+**) fractions (Figure 5-23A). We can then see how gene expression deviates in the other fractions analysed. When we look at the MEF/KOSR *MIXL1*(**+**)/SSEA-3(**+**) fraction signature we can see an upregulation of genes associated with differentiation including *CER1, EOMES, MIXL1* and *T.* This upregulation is further pronounced in the MEF/KOSR *MIXL1*(**+**)/SSEA-3(-) fraction. Genes associated with pluripotency including *DMNT3B, POU5F1* and *NANOG* are maintained at a similar level in the *MIXL1*(**+**)/SSEA-3(**+**) fraction but begin to show downregulation in the *MIXL1*(+)/SSEA-3(**-**) fraction. *SOX2* however does display some downregulation in the *MIXL1*(**+**)/SSEA-3(**+**) but again this is further pronounced in the *MIXL1*(+)/SSEA-3(**-**) fractions (Figure 5-23A).

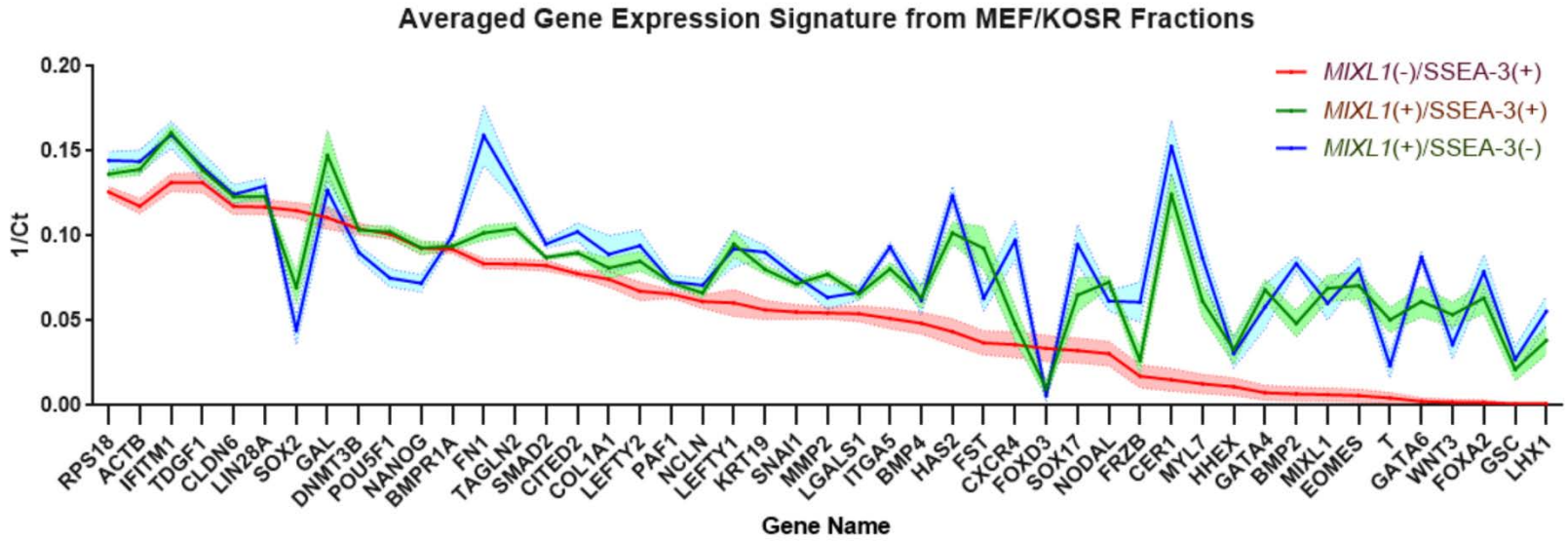
The signatures between both *MIXL1*(**+**)/SSEA-3(**+**) cells from MEF/KOSR and Primo conditions showed a near identical signal, including the maintenance of pluripotency marker expression) (Figure 5-23B). Some of the genes associated with mesendoderm display higher upregulation in Primo conditions including *GATA6, EOMES, MIXL1* and *T*. Genes related to pluripotency display similar maintenance in the Primo fraction, but we also see a similar dip *SOX2* expression. When looking at the individual cell gene distribution revealed similar distribution in both fractions, and the dip in average *SOX2* expression is in part due to the presence on *SOX2* negative cells in these population. As in the Bulk Rna-seq analysis *CER1* was one of the most heavily upregulated genes in all *MIXL1*(**+**) fractions, however a collection of *MIXL1*(-) cells were expressing *CER1* as well.

Figure 5-23 Averaged qPCR Signature Comparison

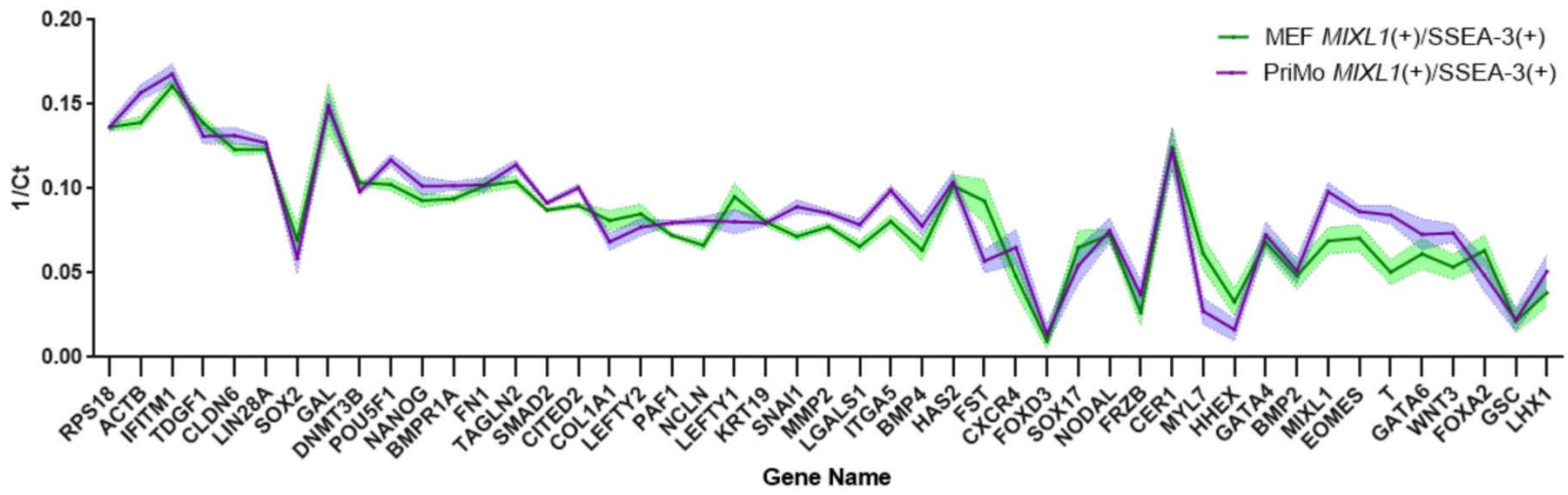
(Figure on next page)

The average 1/Ct values for 47 genes from single cell qPCR analysis. Genes were ordered from highest to lowest expression based on the *MIXL1*(-)/SSEA-3(**+**) fraction. A solid line connects the mean expression points to give a state "signature" with surrounding shaded area represents the 95% confidence interval of the data. **A)** Displays the state signatures of *MIXL1*(-)/SSEA-3(**+**)(red), *MIXL1*(**+**)/SSEA-3(**+**) (green) and *MIXL1*(**+**)/SSEA-3(-) (blue) grown in MEF/KOSR conditions. **B)** Displays the state signatures of *MIXL1*(**+**)/SSEA-3(**+**) cells grown in MEF/KOSR (green) and Primo (purple) conditions.

A


Averaged Gene Expression Signature from MEF/KOSR Fractions
MIXL1(-)/SSEA-3(+)
MIXL1(+)/SSEA-3(+)
MIXL1(+)/SSEA-3(-)
1/Ct
0.20
0.15
0.10
0.05
0.00
RPS18 ACTB IFITM1 TDGF1 CLDN6 LIN28A SOX2 GAL DNMT3B POU5F1 NANOG BMPR1A FN1 TAGLN2 SMAD2 CITED2 COL1A1 LEFTY2 PAF1 NCLN LEFTY1 KRT19 SNAI1 MMP2 LGALS1 ITGA5 BMP4 HAS2 FST CXCR4 FOXD3 SOX17 NODAL FRZB CER1 MYL7 HHEX GATA4 BMP2 MIXL1 EOMES T GATA6 WNT3 FOXA2 GSC LHX1
Gene Name

B


Averaged Gene Expression Signature from MIXL1(+)/SSEA-3(+) Fractions
MEF MIXL1(+)/SSEA-3(+)
PriMo MIXL1(+)/SSEA-3(+)
1/Ct
0.20
0.15
0.10
0.05
0.00
RPS18 ACTB IFITM1 TDGF1 CLDN6 LIN28A SOX2 GAL DNMT3B POU5F1 NANOG BMPR1A FN1 TAGLN2 SMAD2 CITED2 COL1A1 LEFTY2 PAF1 NCLN LEFTY1 KRT19 SNAI1 MMP2 LGALS1 ITGA5 BMP4 HAS2 FST CXCR4 FOXD3 SOX17 NODAL FRZB CER1 MYL7 HHEX GATA4 BMP2 MIXL1 EOMES T GATA6 WNT3 FOXA2 GSC LHX1
Gene Name

5.2.10. Gene expression Plots.

Single cell expression allowed us to look at the correlation between genes based on their expression patterns within cells. I utilised the SCExV (Lang et al., 2015) single cell qPCR analysis software to assess gene relationships Assessment by PCA, revealed the relationship between genes (Figure 5-24). A major cluster of genes is near the 0 point on both axis and contains genes associated with pluripotency such as *POU5F1* and *NANOG.* To the left of this cluster and clearly separated from the other pluripotency-associated genes is *SOX2*. *SOX2* showed a dip in the average expression level in the *MIXL1*(**+**)/SSEA-3(**+**) fractions from both Primo and MEF/KOSR conditions which was further compounded in the *MIXL1*(**+**)/SSEA-3(-) fraction (Figure 5-23). Other pluripotency markers showed less downregulation between fractions which speaks to the separation there was by PCA.

To the right of the main cluster there are genes associated with differentiation. *MIXL1* is situated on the far right of the plot and is clustered close to *CER1, GATA4, WNT3* and *T. WNT3* and *T* demonstrate a very strong relationship, this is also seen *in vivo* during the earliest event of primitive streak formation (Rivera-Perez and Magnuson, 2005). Another set of genes showing a tight correlation, is that of *CXCR4* and *SOX17*, these genes have demonstrated synchronised expression during the formation of endoderm from human PSC previously (Ghosheh et al., 2016).

The distribution of gene expression per fraction from each cell was also assessed (Figure 5 22 displayed over the following six pages). Two trends are apparent within the data. Firstly, changes in expression from MEF/KOSR *MIXL1*(-)/SSEA-3(**+**) (red group in figure) to the other fractions in MEF/KOSR tended to either increase or decrease progressively from *MIXL1*(**+**)/SSEA-3(**+**) (green group in figure) to *MIXL1*(**+**)/SSEA-3(-) (blue group in figure). This included the downregulation of pluripotency-associated genes (Figure 5 22A), in particular *SOX2* as well as the upregulation of differentiation associated genes, including *CER1, MIXL1, T* and many others (Figure 5 22B). The second visible trend is the similarity in distributions between the *MIXL1*(**+**)/SSEA-3(**+**) from both MEF/KOSR and Primo conditions (purple group in figure) (Figure 5 22).

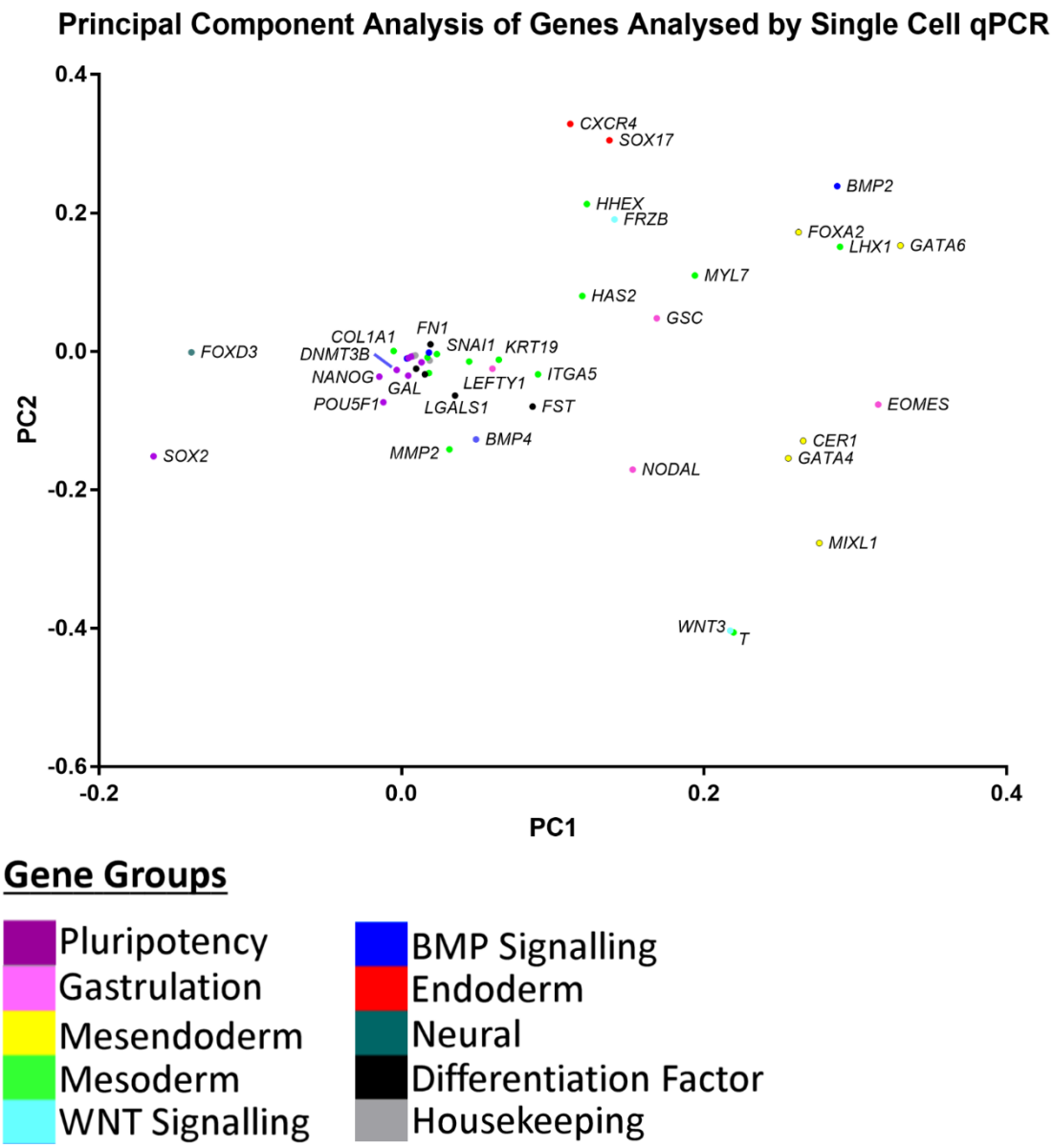


Figure 5-24 Gene Expression Correlations

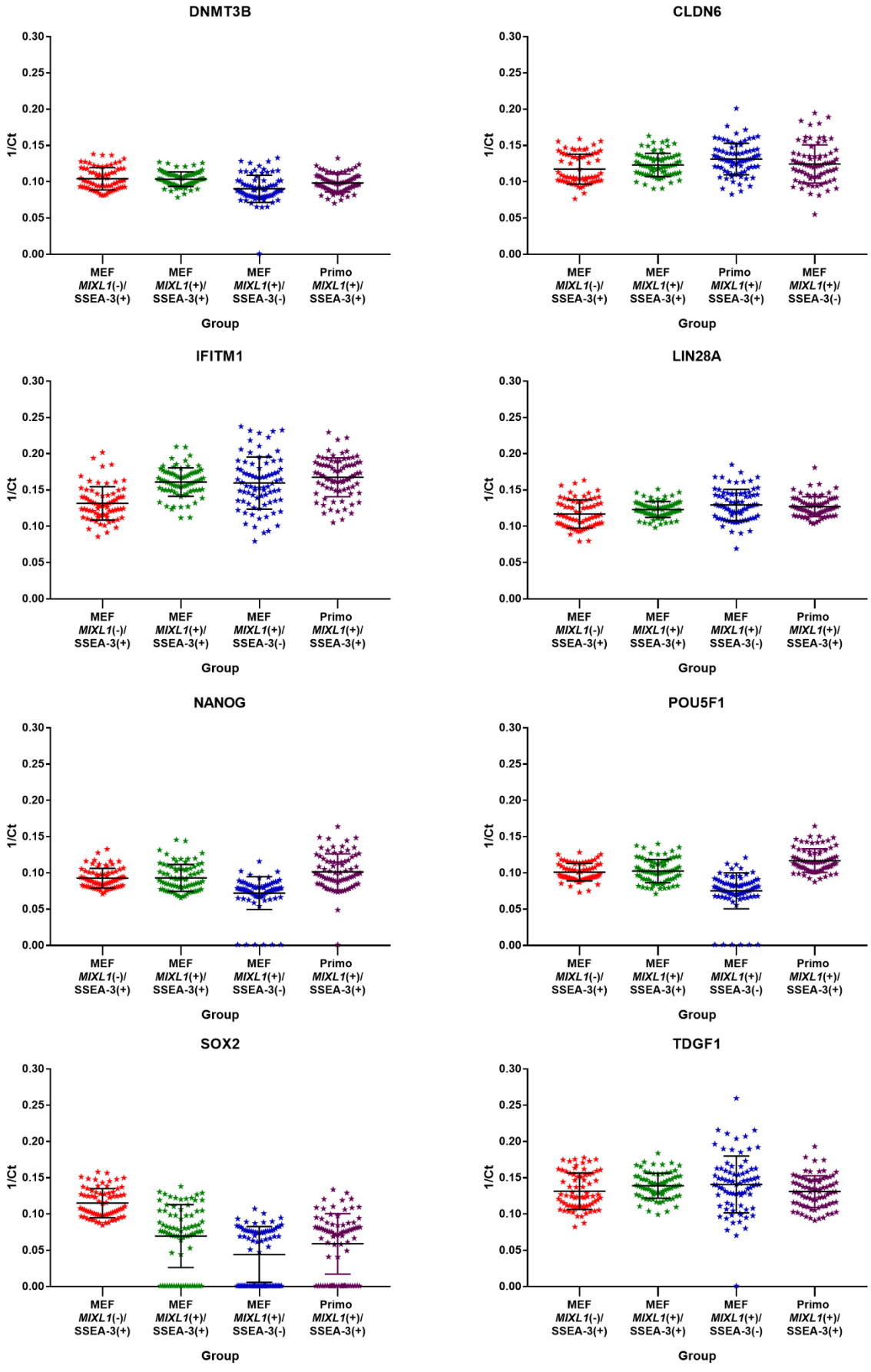
Principal Component Analysis of genes assessed by single cell qPCR analysis from cells populations defined by *MIXL1* and SSEA-3 expression grown in MEF/KOSR and Primo conditions. Genes are colour coded according to their associated state or lineage.

Figure 5-25 Single Cell Gene Expression Plots

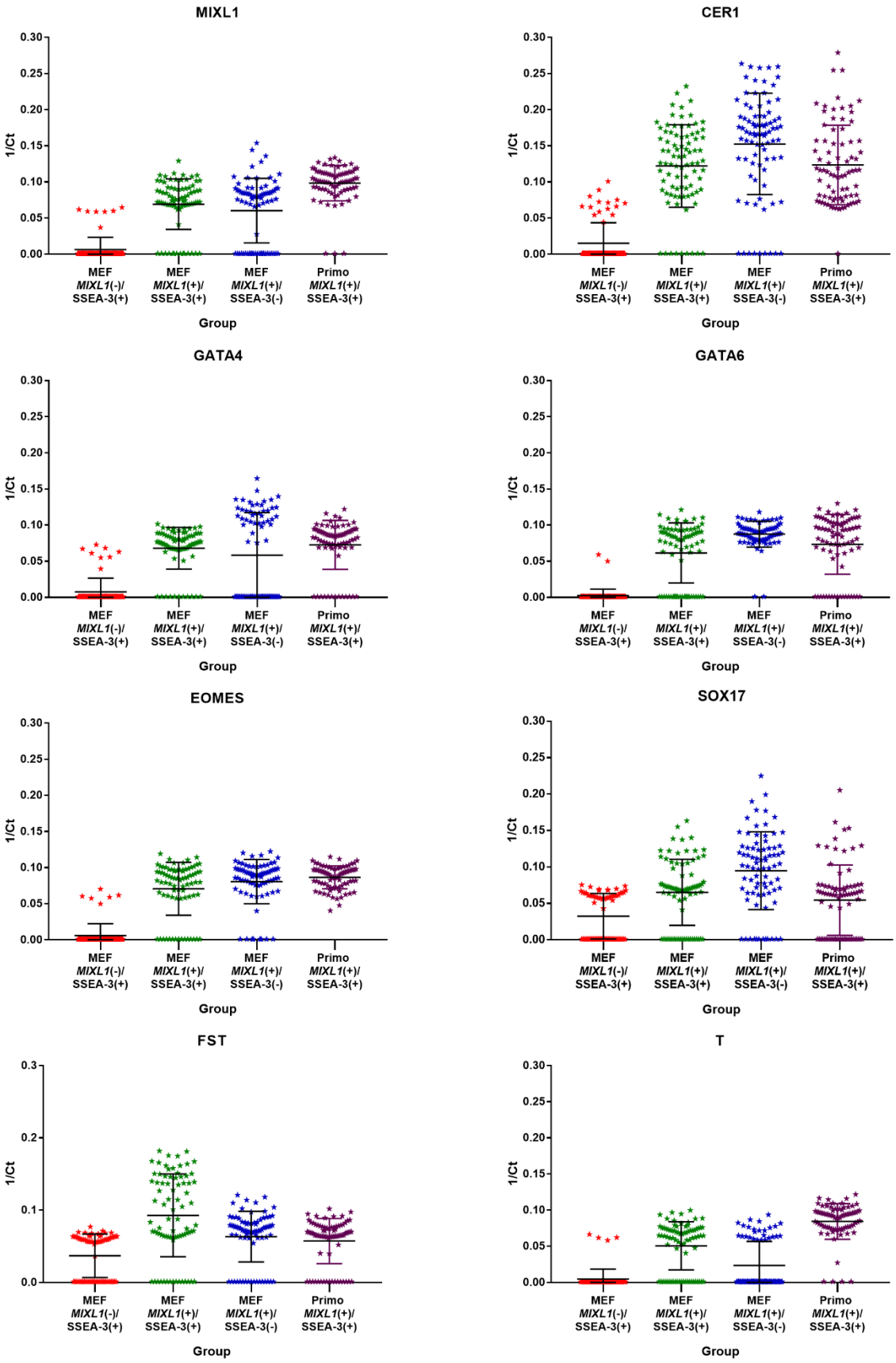
(Figure spans the next six pages)

1/Ct values for each single cell for a given gene. Mean and standard deviation are displayed on top of data sets as black bars. Cells are split into their respective sorted fractions MEF/KOSR conditions *MIXL1*(-)/SSEA-3(**+**) cells in red, *MIXL1*(**+**)/SSEA-3(**+**) cells in green, *MIXL1*(-)/SSEA-3(**+**) cells in blue and Primo conditions *MIXL1*(**+**)/SSEA-3(**+**) cells in purple. **A)** Contains a collection of plots from genes associated with pluripotency. **B)** Contains a collection of plots from key genes associated with mesendoderm differentiation. **C, D,E, F)** Contains plots from the remaining genes assessed by single cell qPCR.

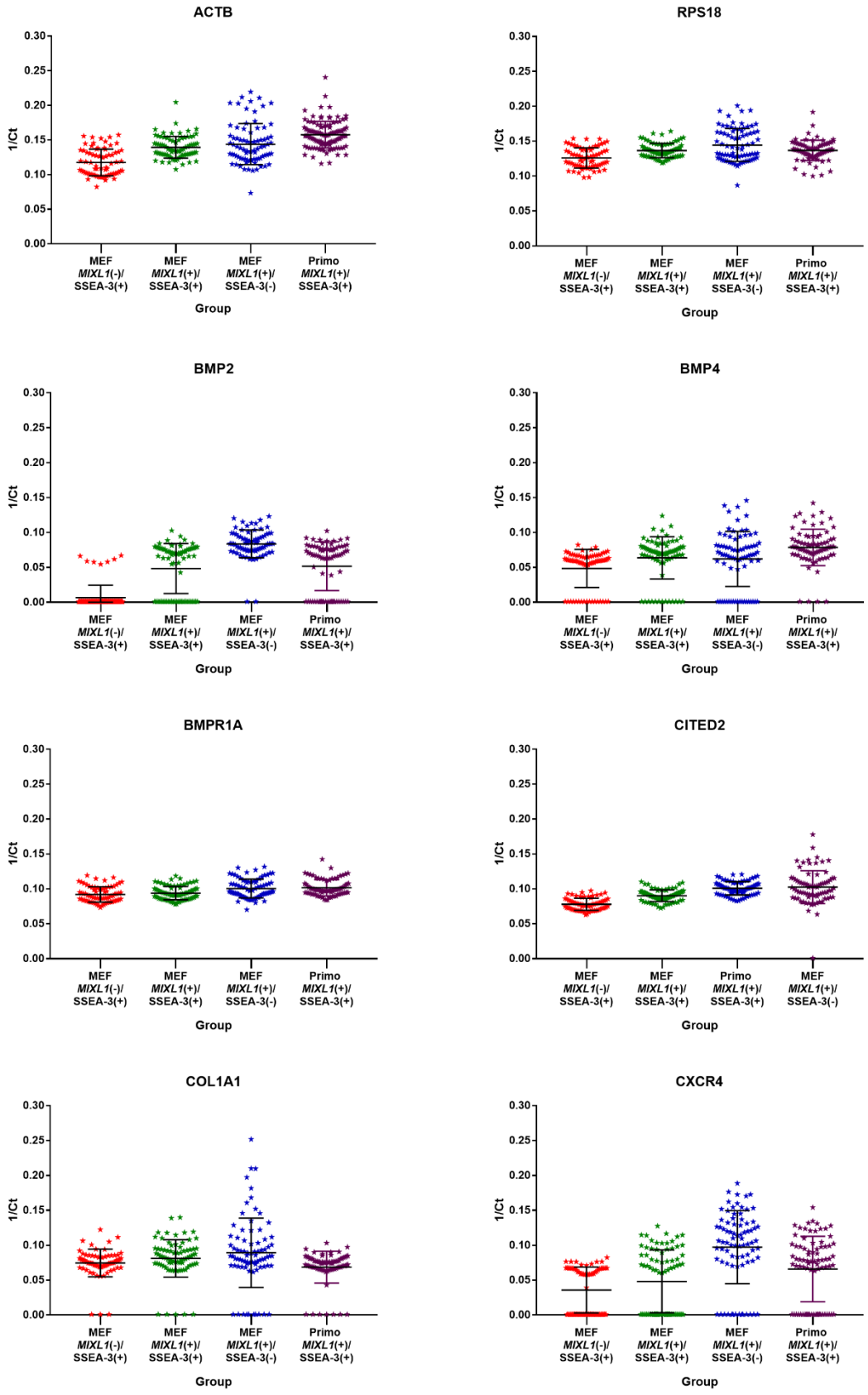
A)


DNMT3B
CLDN6
IFITM1
LIN28A
NANOG
POU5F1
SOX2
TDGF1
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group

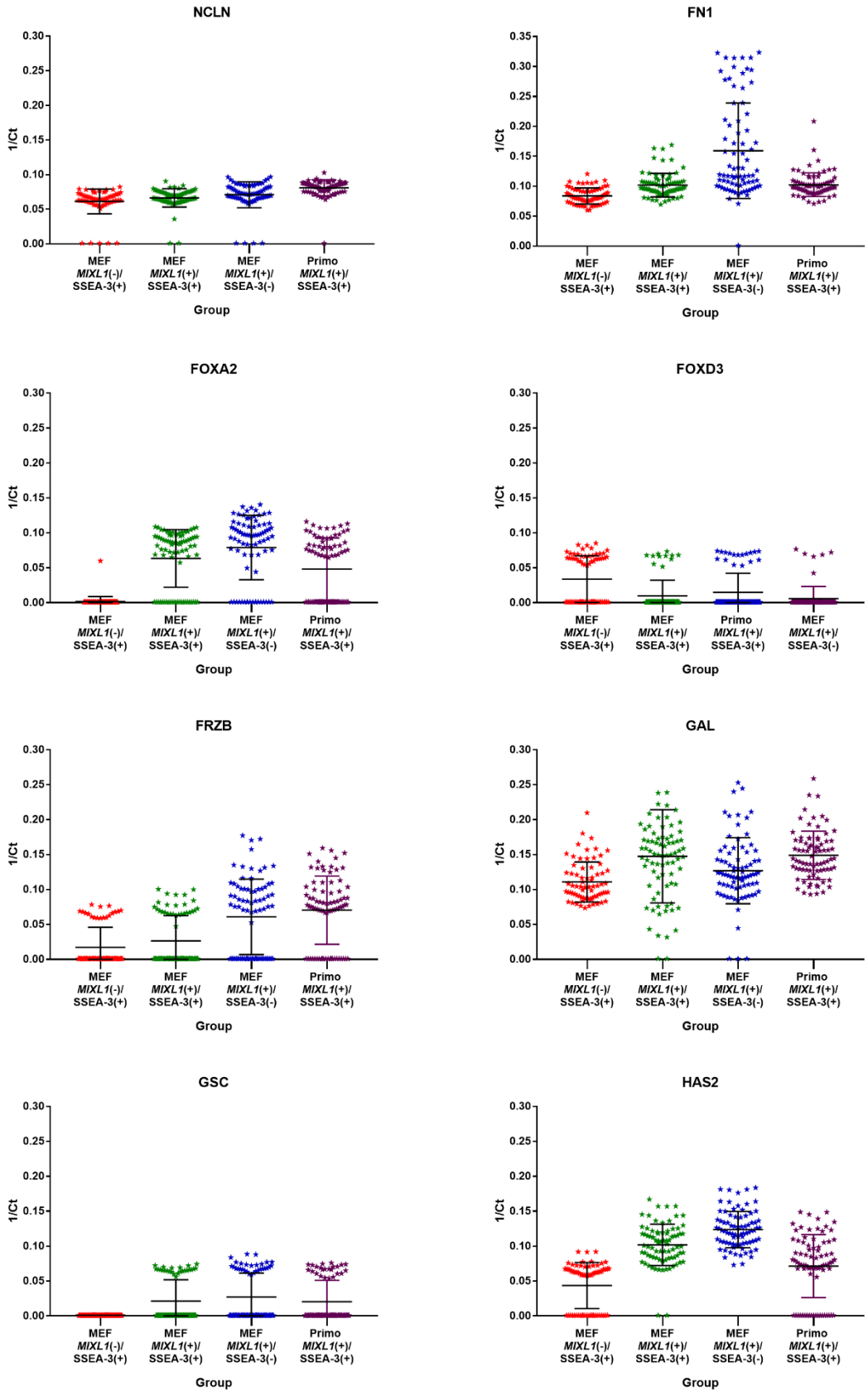
B)


MIXL1
CER1
GATA4
GATA6
EOMES
SOX17
FST
T
1/Ct
Group
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)

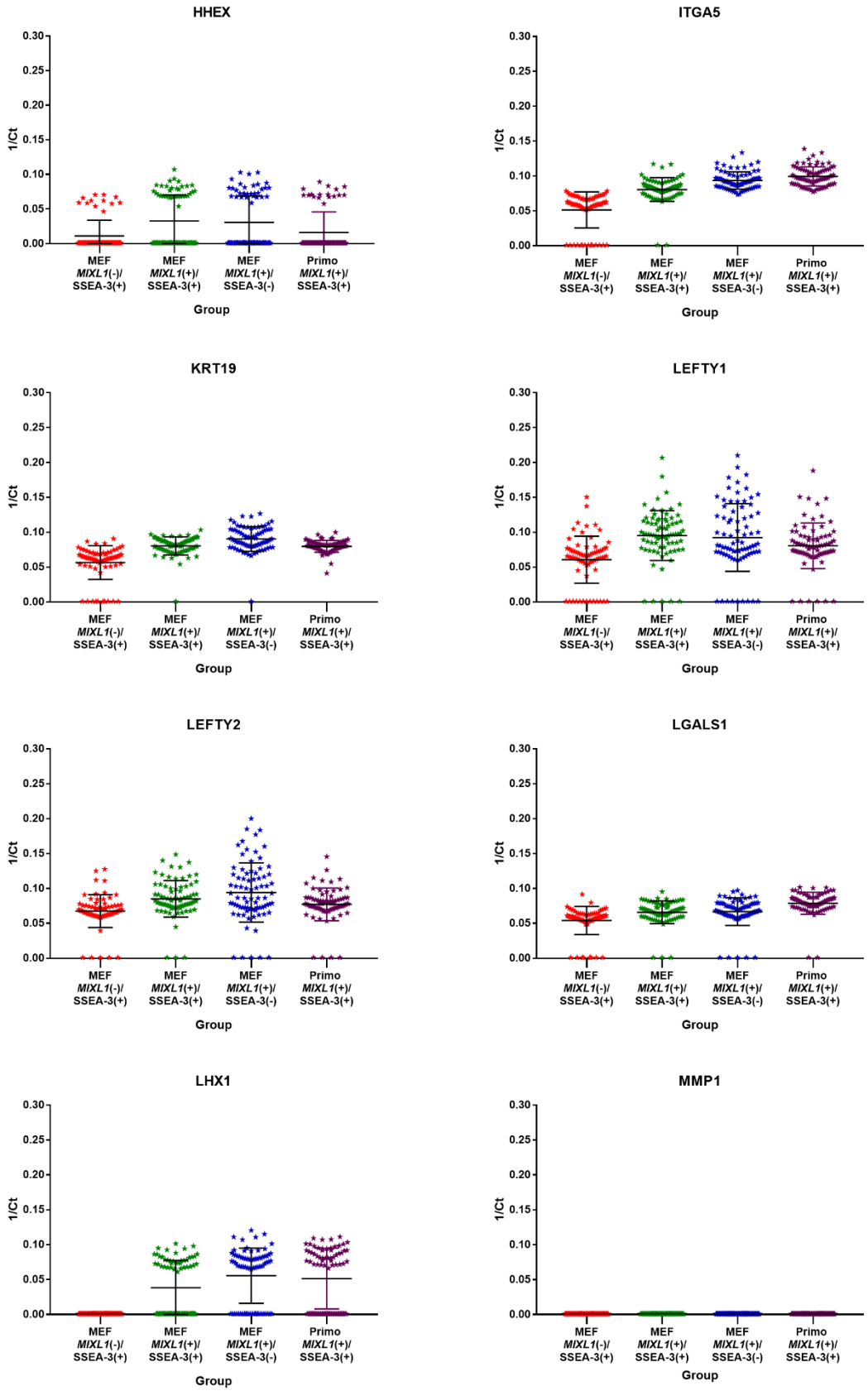
C)


ACTB
RPS18
BMP2
BMP4
BMPR1A
CITED2
COL1A1
CXCR4
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Group

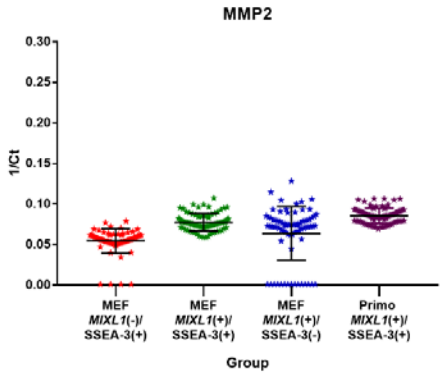
D)

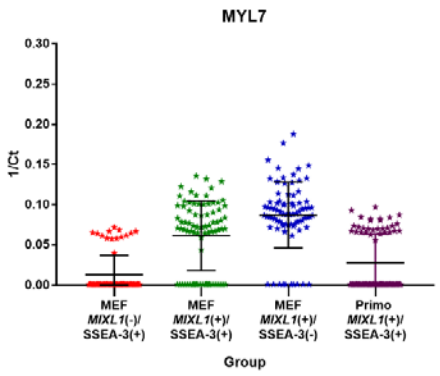

NCLN
FN1
FOXA2
FOXD3
FRZB
GAL
GSC
HAS2
1/Ct
Group
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)

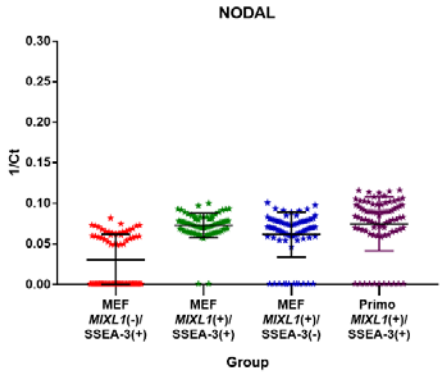
E)

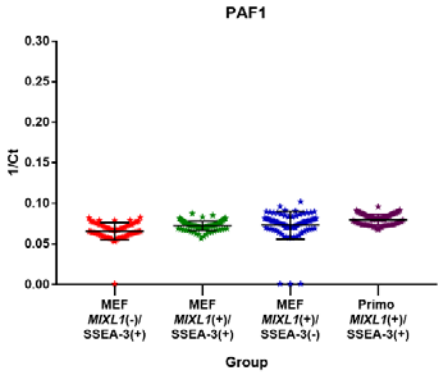

HHEX
ITGA5
KRT19
LEFTY1
LEFTY2
LGALS1
LHX1
MMP1
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group

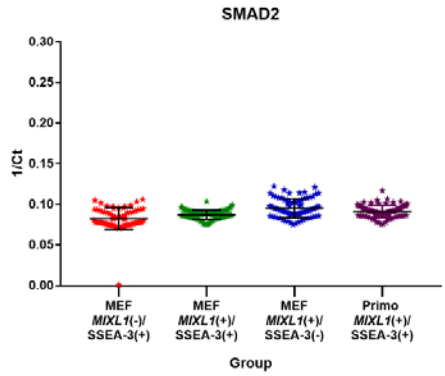
F)

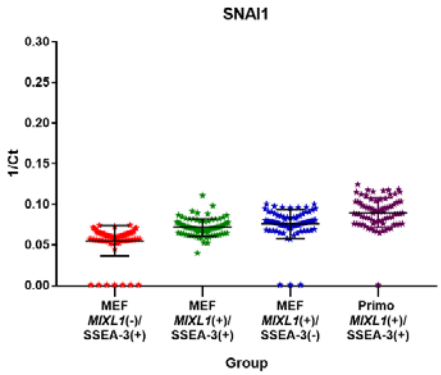

MMP2
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group

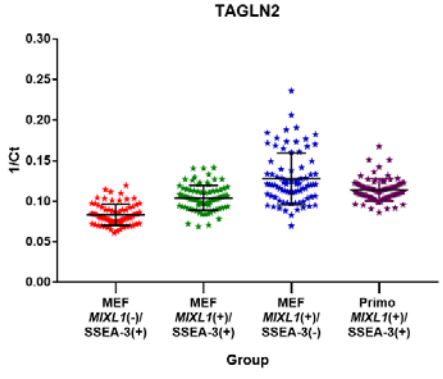

MYL7
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group

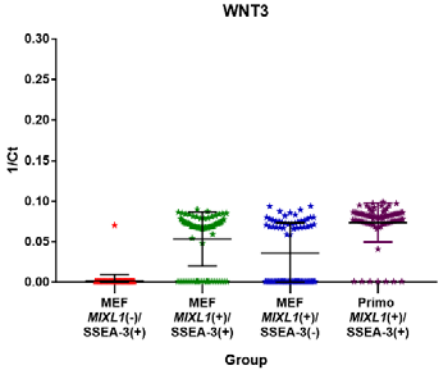

NODAL
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group


PAF1
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group


SMAD2
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group


SNAI1
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group


TAGLN2
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group


WNT3
1/Ct
0.30
0.25
0.20
0.15
0.10
0.05
0.00
MEF MIXL1(-)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(+)
MEF MIXL1(+)/ SSEA-3(-)
Primo MIXL1(+)/ SSEA-3(+)
Group

5.2.11. Trajectory Comparisons between MEF/KOSR and Primo Fractions

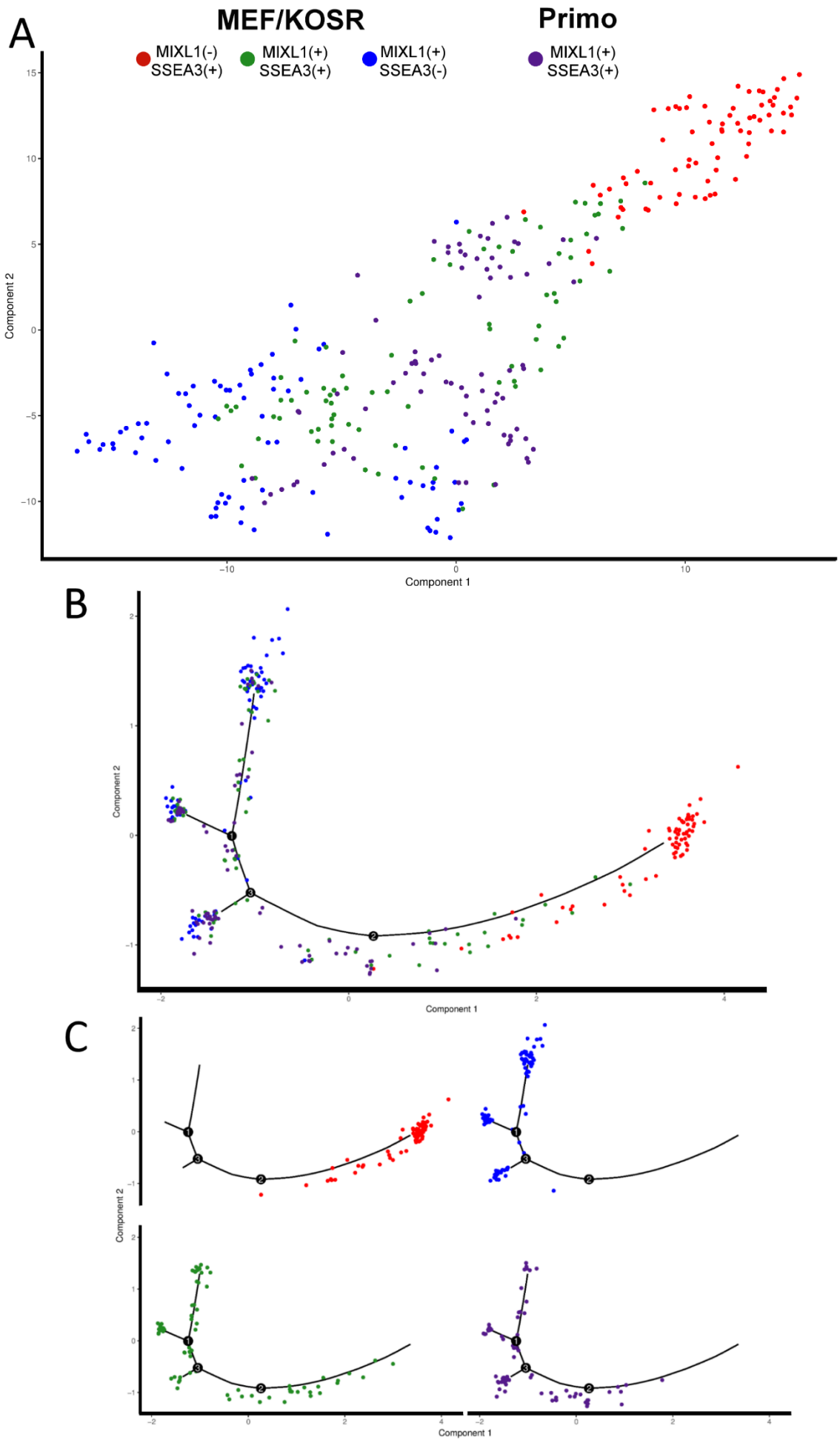
Again, using Monocle2, I incorporated the single cell data acquired from *MIXL1*(**+**)/SSEA-3(**+**) from Primo conditions to see where the cells would map in our original trajectory. Observing the data in the t-SNE plot we can still see a clear separation of the *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(-) from MEF KOSR conditions at the top right and bottom left of the plot, respectively. Again, spanning the space in the middle are cells from the *MIXL1*(**+**)/SSEA-3(**+**) fraction from MEF/KOSR in green. With the Primo fraction also incorporated we can also see the cells occupying a similar distribution of that double positive fraction form MEF/KOSR conditions spanning the middle of the apparent continuum.

With the data observed as a minimal spanning tree there was a slightly different trajectory than previously. With more cells/data points now representing lineage primed and differentiated cells compared to pristine *MIXL1*(-) fractions we can see the tree shape and branch structure change towards the end points. Again, both *MIXL1*(**+**)/SSEA-3(**+**) fractions from MEF/KOSR and Primo show a strong overlap in their positions in the trajectories. However, as seen before with the *MIXL1*(**+**)/SSEA-3(**+**) fraction from MEF/KOSR many of the cells from Primo occupy positions at branch end points as well, overlapping with cells from *MIXL1*(**+**)/SSEA-3(-) cells from MEF/KOSR.

Figure 5-26 Cell Trajectory Comparison

(Figure on next page)

A) t-Distributed Stochastic Neighbour Embedding (t-SNE) plots of the expression of 45 genes in individual cells measured by Fluidigm Biomark qPCR. Displays all fractions in the same dimensional space from MEF/KOSR conditions *MIXL1*(-)/SSEA-3(**+**) cells in red, *MIXL1*(**+**)/SSEA-3(**+**) cells in green, *MIXL1*(-)/SSEA-3(**+**) cells in blue and Primo conditions *MIXL1*(**+**)/SSEA-3(**+**) cells in purple. **B)** Minimal spanning tree cell trajectories. Each dot represents and individual cells expression profile, plotted in a two-dimensional independent component space according to Monocle2s pseudotime ordering. The solid black line represents the path of the trajectory and branch points are labelled 1, 2 and 3. **C)** Displays fractions separately on the same minimal spanning tree as in **B)**.


A
MEF/KOSR
Primo
MIXL1(-) SSEA3(+)
MIXL1(+) SSEA3(+)
MIXL1(+) SSEA3(-)
MIXL1(+) SSEA3(+)
Component 2
Component 1
B
Component 2
Component 1
C
Component 2
Component 1

5.3. Discussion

We have transcriptionally characterised subsets of human PSC based on reporter gene, *MIXL1*-GFP and surface antigen expression, SSEA-3. The transcriptome of these SSEA-3(**+**) subsets indicates the presence of pluripotency network. However, lineage specific genes exhibit transcriptional difference between subsets. Some subsets display an upregulation of lineage specific genes, this may equate to a lineage bias in a particular subset of human PSC.

The analysis of RNA-Seq data for *MIXL1* verified that the separation of populations based on *MIXL1*-GFP expression was accurate. With this knowledge I proceeded to probe the data and assessing the substates that exist in MEF/KOSR conditions. Expression of *MIXL1*-GFP coincided with a change in the transcriptomic profile of the population, with samples clustering together based on their *MIXL1* expression. There was a large upregulation of genes associated with gastrulation, the primitive streak and mesendoderm. This in itself is not surprising since *MIXL1* is a transcription factor that is crucial for the formation of the primitive streak (Hart et al., 2002). Genes associated with pluripotency show relatively small down regulation in the *MIXL1*(**+**)/SSEA-3(**+**) substates when compared to the loss in the *MIXL1*(**+**)/SSEA-3(-) substate. The functional data represented in chapter 3 demonstrated that *MIXL1*(**+**)/SSEA-3(**+**) substate can form self-renewing stem cell colonies and bulk RNA-seq analysis revealed expression of pluripotency-associated genes in this fraction.

I have been able to further analyse the gene expression of substates and highlight areas of further investigation. One such area is that of *CER1* which showed increased upregulation in the *MIXL1*(**+**) groups. *CER1* encodes for the protein Cerberus, which is expressed during embryonic development (Bouwmeester et al., 1996). Cerberus is secreted during gastrulation and is an inhibitor of TGFβ, Wnt, Nodal and BMP(Piccolo et al., 1999). The expression of *CER1* has been studied in human embryonic stem cells previously, with respect to its impact on differentiation (Katoh and Katoh, 2006; Kempf et al., 2016).

Tracking Cerberus expression is difficult since it is a secreted protein, which makes antibody staining an inefficient technique. A fluorescent reporter targeted to *CER1* has been generated within the PluriMes project and will be useful in the future. The secretion of Cerberus by human PSC could perturb the states which exists in the heterogeneous population. Cerberus may be a key factor in the progression or prevention of differentiation human PSC colonies, highlighted by Kempf et al., 2016. The fact that SSEA-3(**+**) cells seem to be at least expressing the *CER1* gene, could answer the question of why stem cell differentiation protocols can be inefficient. Many differentiation protocols use TGFβ, Wnt, Nodal and BMP signalling for differentiation. If a

substate of the cells are secreting Cerberus this can impact upon the desired signalling towards a certain lineage.

When present, *MIXL1*(**+**) cells within stem cell colonies often appeared as small cell clusters. It might be that the secretion of Cerberus by a small substate of cells leads to a change in the microenvironment and the upregulation of *MIXL1* in neighbouring cells. Since Cerberus is a secreted protein, it can also be supplemented into media as a recombinant protein. The addition of Cerberus may highlight important signalling networks in the substate creation.

The transcriptomic profile of the *MIXL1*(**+**) states also allows us to identify possible surface antigens in future for screening. Reporter lines are beneficial for research purposes but for translational medicine they are less applicable. Identifying a surface antigen that would allow us to select for particular substates would be helpful for future investigation and therapeutic development.

While bulk RNA-sequencing provides considerable in-depth information about our particular substates it cannot enquire as to whether heterogeneity exists between single cells within the population. I was able to utilise single cell transcriptomics to investigate the heterogeneity within substates. We segregated our human PSC grown on feeders into three fractions *MIXL1(***+**)/SSEA-3(**+**), *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(-). These fractions represent three different stages between pluripotency and committed differentiation. *MIXL1(***+**)/SSEA-3(**+**)(Bias), *MIXL1*(-)/SSEA-3(**+**) (Pluripotent) and *MIXL1*(**+**)/SSEA-3(-) (Differentiated).

Classic hierarchical clustering and dimensionality plots gave us some intriguing results although the high degree of overlap between the populations makes it hard to distinguish them. While these methods provide helpful insights, analysing the data using Monocle2 highlights differences in the transcriptional regulation of the subpopulations that are difficult to see by dimensionality reduction alone. This allows the representation of the dynamic evolution of gene expression through commitment and differentiation, highlighting the importance of individual genes and timing of gene expression changes during this process, and more globally, in indicating the relative contribution of gene coordination and programme conflict during priming and commitment to differentiation.

When we observe the gene expression of pluripotency-associated genes, with the exception of *SOX2*, there is still a high correlation between cells the pluripotent *MIXL1*(-)/SSEA-3(**+**) and early differentiated *MIXL1*(**+**)/SSEA-3(**+**) states. This implies that the loss of parts of the pluripotency networks are more gradual than others. This is unsurprising as these genes can still have an impact on the differentiation of these cells (Wang et al., 2012). Some groups have suggested a more stochastic model of differentiation, that cells may enter lineage commitment with variable

gene expression programmes as a result of diverse transition events effecting the decision (Pina et al., 2012; Teles et al., 2013). A stochastic model may better explain the nature of these substates in standard culture and the further heterogeneity that exists. Some genes do clearly follow a progressive pattern though, with genes such as *GATA6* appearing further in pseudotime, of course the initial changes could be stochastic in nature leading to a more scripted programme.

Another observation is the clustering of differentiation markers including *T, GATA4, GATA6*, and *EOMES*. Examination of this cluster of the network may help to determine biasing versus commitment. While our Monocle2 analysis indicates *GATA6* expression is further in pseudotime, work conducted in our laboratory with a *GATA6* reporter in human PSC demonstrates substates expressing *GATA6* still contain a stem cell population (Allison et al, 2018 Unpublished). So *GATA6* expression alone is not an indicator of commitment. We have also separately assessed *T* (Brachyury) reporter line from Roger Pedersen's Laboratory (Mendjan et al., 2014), and found the same to be true of *T* positive substates. Matching single cell analysis on subsets expressing *T, GATA6* and in future *CER1* may help to disentangle the gene regulatory network governing commitment.

Our newly generated fractions from Primo conditions exhibited similar expression of both *MIXL1*-GFP and SSEA-3 to those identified in MEF/KOSR conditions. After analysing the transcriptome of the *MIXL1*(**+**)/SSEA-3(**+**) fractions from both conditions we can see similar gene expression patterns. Using bulk analysis, there is similar transcriptomic landscapes depicted in our gene cloud analysis, with coexpression of pluripotency-associated and mesoderm associated genes. Further to this, single cell analysis revealed similar heterogeneity and gene distribution in these fractions, occupying overlapping positions in both tSNE and minimal spanning tree analysis. Minor differences are apparent but transcriptomically these *MIXL1*(**+**)/SSEA-3(**+**) fractions from these two conditions appear very similar.

When comparing the fractions in 1xPrimo conditions to the differentiation time course we get an idea of where our fractions relate to on a differentiation trajectory. Cells were analysed after 3 days in Primo medium and exhibited a large proportion of *MIXL1*(**+**)/SSEA-3(**+**) cells. When I compare this fraction to the emerging populations during differentiation, we can see that by PCA and tSNE analysis these cells correlate between the emerging populations at 24-48 hours of differentiation.

Interestingly the *MIXL1*(-)/SSEA-3(**+**) fraction from Primo conditions equate to between 12-18 hours of differentiation. The original prediction was that this fraction would correlate closer to 0 hours as all cells are *MIXL1*(-)/SSEA-3(**+**) at the 0 hour timepoint but this was not the case. When I looked at the gene expression of this fraction I saw a significant overlap in expression

with the *MIXL1*(**+**)/SSEA-3(**+**) fraction, the genes in this overlap exhibited an association with gastrulation and early development by GO enrichment analysis. *MIXL1* is amongst this group of genes, even though the cells are negative for *MIXL1*-GFP by flow cytometry. Examining the expression level, *MIXL1* has an average FPKM value of ~65 from the *MIXL1*-GFP(-) fraction compared to ~312 in the *MIXL1*-GFP(**+**) fraction.

These data help in validating the differentiation bias that I saw in our neutral EB assay, in 4.2.10, but also highlights the possible drawbacks of using this *MIXL1* reporter. In terms of our differentiation assay, the *MIXL1*(-)/SSEA-3(**+**) fraction already has upregulation of genes associated with gastrulation and mesoderm development prior to the start of differentiation. From this viewpoint it is not surprising that this fraction generates EBs with a particularly high mesodermal gene expression signature.

HES3 *MIXL1*-GFP was generated using a variant of GFP named, enhanced GFP (eGFP) (Davis et al., 2008) which offers better sensitivity by being 35 times brighter than wild-type GFP (Zhang et al., 1996). eGFP also exhibits a significant increase in maturation time when compared to wild type GFP (Iizuka et al., 2011). This allows for sensitive tracking of *MIXL1* expression throughout differentiation as designed (Davis et al., 2008). However, our application is slightly different to assessing differentiation. Our work with both cultures in MEF/KOSR and Primo conditions reveals the importance of *MIXL1*-GFP low expression, and our differentiation timecourse reveals how high the *MIXL1* intensity can increase during this process. Our low level of *MIXL1* expression in some cells may not be enough to bring the cells above a detectable level by flow cytometry.

Our initial idea, was to use *MIXL1*-GFP expression as a surrogate for the transcriptional changes associated with early differentiation and this ideal held true for bulk RNA-seq analysis from MEF/KOSR conditions. However, in our bulk RNA-seq analysis of the *MIXL1* negative fraction from Primo medium we detected *MIXL1* expression and coexpression of other differentiation genes. Our single cell analysis from MEF/KOSR did highlight that seven out of seventy-two *MIXL1*(-)/SSEA-3(**+**) cells had low level expression of *MIXL1* by qPCR analysis. While this could be contaminating *MIXL1*(**+**)/SSEA-3(**+**) cells during the sorting process, it may also be a consequence of the detection limit of *MIXL1*-GFP by flow cytometry.

Our work with the H9 *T*-Vensus reporter line displayed a more homogenous population when grown in Primo condition. *MIXL1* and *T* share a similar expression pattern and timing during mesoderm differentiation (Loh et al., 2016). Therefore, we would expect similar expression of these markers in lineage priming medium. However, there is a disparity between *MIXL1*-eGFP and *T*-Venus population size in culture. Brachyury has been shown to repress the expression of Mixl1 in differentiation mouse PSC (Pereira et al., 2011) and whether this could be happening in

our Primo medium requires further investigation. Of course, this could be a difference in their gene expression but bulk RNA-seq analysis revealed *MIXL1* expression in the *MIXL1*-GFP negative population. Beyond just the targeted genes and their genetic backgrounds these two reporter lines differ in another aspect which is the fluorescent proteins that are expressed. While eGFP is 35 times brighter than wild type GFP, Venus is a further 55% brighter than eGFP, albeit with a slower maturation time (Iizuka et al., 2011; Zhang et al., 1996). It could be that the use of Venus as a reporter allows for greater sensitivity when assessing low level gene expression by flow cytometry. The use of both *MIXL1*-eGFP and *T*-Venus reporter lines coupled with single cell transcriptomic analysis may help resolve the disparity we see between these reporter lines.

In summary, we have transcriptionally characterised the *MIXL1*(**+**)/SSEA-3(**+**) human PSC substate by both bulk and single cell transcriptomic techniques. The fraction has been assessed from both standard culture systems and our defined induction system. I have been able to model the dynamic process of early differentiation in human PSC, either sporadically in self renewal conditions or by directed differentiation. This model has allowed us to identify where our *MIXL1*(**+**)/SSEA-3(**+**) fractions are situated in this continuum from pluripotency to differentiation, and the important gene expression changes during this process. Comparison between the *MIXL1*(**+**)/SSEA-3(**+**) fractions seen in the two culture systems, demonstrate similar gene expression profiles and overlapping placement in this continuum. This data has provided insight into the genes which impact on the functional characteristics exhibited by these fractions. This data, represents the first transcriptionally profile of a “trapped” mesoderm biased subset of human PSC.

6. Final Discussion

6.1. Endogenous and induced heterogeneity within the stem cell compartment.

The idea that human PSC exists as a heterogenous population with differing propensities to differentiation is not a new one. However, what I have shown here is not only the characterisation of "biased" substate but also mechanisms for its creation, manipulation and maintenance. The development of Primo medium provides useful tool in investigating heterogeneity with a defined system which promotes the presence of mesoderm biased substates.

Traditionally human PSC were derived on a layer of fibroblast feeder cells, be it mouse or human origin. Over time culture systems were developed to be feeder free, but the growth of human PSC on a feeder layer still offers benefits and insight into important properties of human PSC. Here I have utilised this feeder culture system to disentangle the heterogenous population that exists and analyse the progression from pluripotency, through priming and finally committed differentiation.

Controlling heterogeneity is a difficult task and not the actual premise for creating our biased medium. Heterogeneity is an ever-expansive terminology, and technological advancements in assessing cellular states, has further increased the aspects and differences that define heterogeneity within cell populations. I exploited the heterogeneity that exists with in standard culture conditions to identify a particular mesoderm biased substate. When I moved to the creation of cell populations which reside in that substate, I attempted to manipulate the environment to cause cells to transition into the state I identified. However, attempting to put cell in a uniformed state could be difficult. Although I grow cells in the same medium, their microenvironment can be drastically different. Cell density is a key example of this in Primo, with densely packed areas leading to differentiation, possibly through cell to cell contact interactions. Observing *MIXL1*-GFP expression in colonies, reveals expression in the denser centre of the colonies. This dynamic could be crucial for the application of Primo medium. Therefore, the Primo medium should not be seen as controlling heterogeneity but rather exploiting it to guide cells towards the mesodermal lineage.

While bulk RNA sequencing gave us valuable insights into the substates defined by *MIXL1* and or SSEA-3 expression it did not speak to the heterogeneity that might still be present in the substates. For this we used single cell qPCR to assess the heterogeneity on a single cell basis. I saw further heterogeneity within our *MIXL1*(**+**)/SSEA-3(**+**) substate with cells varying in their correlation to differentiated and undifferentiated states, creating a continuum between these

two states. Single cell RNA-sequencing could further expand on the qPCR analysis we performed. Being able to assess the global transcriptome on a single cell basis will help to resolve the heterogeneity even further and even identify unknown substates within the cultures, as seen in respect to *GATA6*(**+**) substates (Alison et al, 2018).

In our Primo medium while I have created a *MIXL1*(**+**)/SSEA-3(**+**) population that resembles the one identified in standard culture on feeders, there is a *MIXL1*(-)/SSEA-3(**+**) that differs significantly from the population found in standard culture. This fraction in Primo shows upregulation of genes associated with gastrulation and early differentiation not seen in its MEF/KOSR counterpart. This fraction could contain both pristine stem cells and lineage primed cells, which globally are raising the expression values of these genes. The *MIXL1*(-)/SSEA-3(**+**) fraction from Primo was assessed by bulk RNA-sequencing and, in future, utilising a single cell approach will help to resolve the heterogeneity further. Data obtained using the H9 T-Venus reporter and differentiation bias assessment of Primo *MIXL1*(**+**)/SSEA-3(**+**) state implies the cells are within a biased state even though they are *MIXL1*-GFP negative.

6.2. LPA action and implications for cell to cell contact in human PSC maintenance and differentiation.

Cell to cell contact has been heavily implicated in the survival of human PSC, especially when plated as single cells (Barbaric et al., 2014). Less investigated, yet often a critical parameter, is the effect cell density has on human PSC differentiation efficiency (Gage et al., 2013; Kempf et al., 2016; Tonge and Andrews, 2010). Cell to cell contact activates the Hippo pathway and LPA is a known antagonist of this pathway (reviewed in Yu et al, 2013). An active Hippo pathway may facilitate differentiation of human PSC. Directed differentiation protocols for human PSC have become much more efficient in recent years, many no longer relying on EB aggregation as previously. An interesting observation, is the use of ROCKi (Y-27632) either at the beginning or throughout differentiation. Although designed to boost cell survival (Watanabe et al., 2007), classically ROCKi has been used as an activator of the Hippo pathway (Kono et al., 2014; Wada et al., 2011). When ROCK inhibition was applied to developing mouse embryos, they demonstrated decreased YAP nuclear localisation but enhanced differentiation toward ICM over trophectoderm lineages. If the activation of the Hippo pathway does facilitate differentiation this could go some way to explaining the increase efficiency in newer 2D differentiation protocols. It has been demonstrated that ROCKi primes cells particularly towards mesoderm (Maldonado et al., 2016). Our data from Primo medium indicated a differentiation consequence to single cell plating, whether this effect is a result of being single cells or the presence of ROCKi requires further investigation.

The exact mechanism of how LPA blocks differentiation remains theorised but as yet unproven. The proposed mechanisms include the increase in cytoplasmic YAP/TAZ protein (Azzolin et al., 2014), the increase in nuclear YAP (Lian et al., 2010), the increase in TAZ or an increase in Akt signalling. YAP/TAZ has shown to be incorporated beta-catenin destruction complex, thus an increase in YAP/TAZ would lead to further degradation of Beta-catenin. Over expression of YAP has been shown to prevent CHIRON-induced differentiation of human PSC (Qin et al., 2016). Although this model works in other systems the work of Blauwkamp et al, 2012 in human ESC would imply something different. They used LPA and WNT3a in combination with a hES reporter line carrying the WNT response element TCF driving GFP expression. They noted the LPA was able to prevent differentiation of the cells but did not prevent GFP expression from TCF binding of beta catenin. This implies WNT signalling is active and beta-catenin is not degraded in the cytoplasm before entering the nucleus.

Our data indicates YAP nuclear localisation may play a pivotal role in the prevention of differentiation and maintenance of pluripotency. Analysis of YAP promoter binding in mouse PSC identifies pluripotency-associated genes as direct targets (Lian et al., 2010). The localisation in human PSC culture is variable however, crucially depending on cell density (Hsiao et al., 2016) . The localisation of YAP also, demonstrates a species difference between human and mouse, with respect to the inner cell mass, with predominately nuclear in human (Qin et al., 2016) and cytoplasmic in mouse (Nishioka et al., 2009) embryos . However, in mouse embryos cell to cell contact and subsequently YAP localisation, plays a crucial role in the development of trophectoderm (Nishioka et al., 2009) and also the formation of the epiblast and primitive endoderm (Lorthongpanich et al., 2013).

A proposed model could incorporate cell to cell contact and the localisation of YAP into the development of the early embryo (Figure 6-1). In the inner cell mass (ICM) of the early embryo, cell numbers are lower and spatially less restricted. The localisation of YAP is nuclear, as the Hippo pathway is inactive. The cells of the inner cell mass begin a process of proliferation and secretion of endogenous differentiation factors such as WNT and BMPs (Graham et al., 2014; Price et al., 2013). At this point nuclear YAP is promoting an active pluripotency network.

As the embryo develops and cells of the ICM increase in number, they also become spatially restricted and have increased cell to cell interactions. When embryos move past the cell 64-cell stage, cells begin expressing primitive endoderm and epiblast associated genes (Artus et al., 2011; Chazaud et al., 2006; Kurimoto et al., 2006). At this stage the expression of these markers appears as a ‘salt and pepper’ distribution amongst the cells. Further development sees the segregation of the primitive endoderm and epiblast, as Nanog is downregulation in the primitive endoderm (Graham et al., 2014). As the cell to cell interactions increase, the Hippo pathway

becomes active, restricting YAP to the cytoplasm. Inhibition of the Hippo pathway during this period of time results in compromised formation of the primitive endoderm (Lorthongpanich et al., 2013).

In this model YAP safeguards pluripotency through the proliferative stage of development until the ICM is an appropriate size for differentiation to progress (Figure 6-1). Interestingly, simple aggregation of mouse PSC is sufficient to repress *Nanog* and induce primitive endoderm formation (Hamazaki et al., 2004). Aggregation is also important step in the formation of embryoid bodies. The aggregation may be increasing the cell to cell interactions, and Hippo pathway activation, to allow differentiation to take place. Conversely, our system aims to exploit the YAP safeguard by inhibiting the Hippo pathway and preventing differentiation.

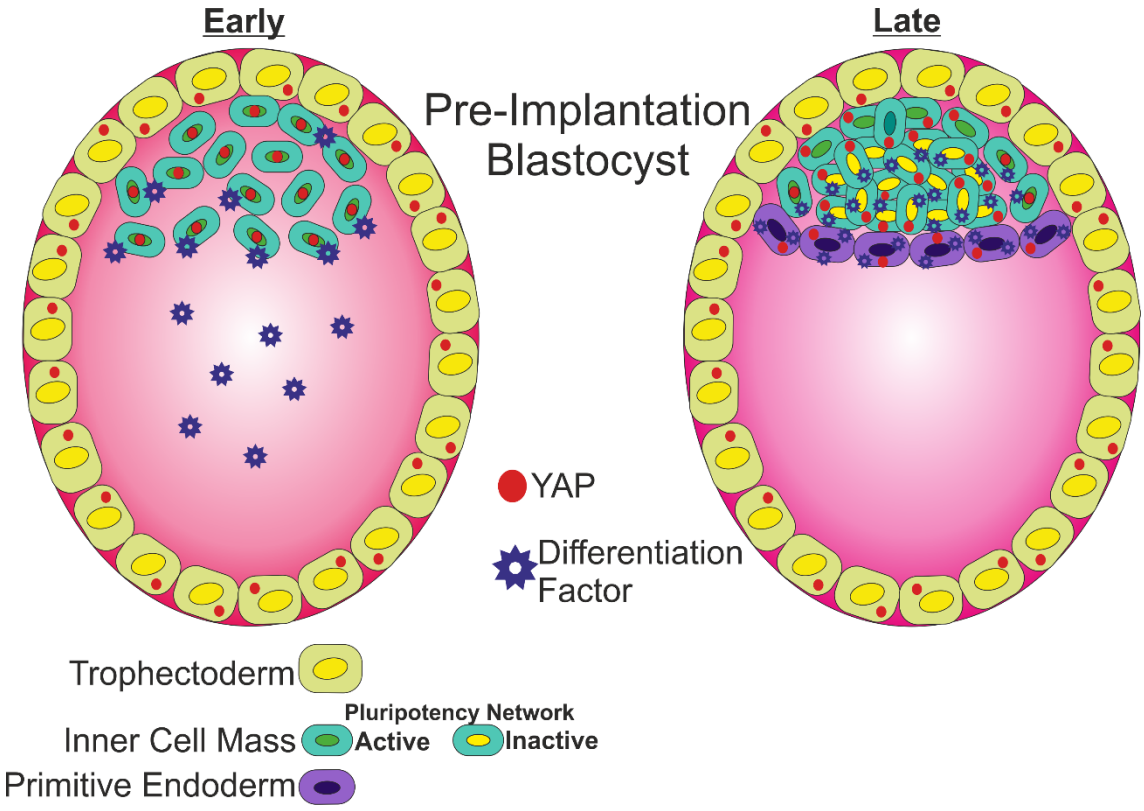


Figure 6-1 YAP in Early Embryo

A proposed model of YAP's role in safeguarding pluripotency in the developing embryo. In the early inner cell mass, cell numbers are low and spatially less restricted, resulting in nuclear YAP. In the late blastocysts cell numbers have increase and are spatially restricted resulting in cytoplasmic YAP. At this point the pluripotency network is down regulated and the primitive endoderm forms

6.3. Further Optimisation of Primo medium

While the addition of LPA does offer us a method of attenuating the differentiation signal from GSK3β inhibition, it does have its drawbacks. An initial issue is that it requires the presence of bovine serum albumin to bind to, for stability, trafficking and solubility purposes. This could compromise the mediums use in xeno-free applications. Interestingly when I tried a LPA variant bound to human serum albumin (Huzzah LPA, Avanti Lipids) I found widespread death and non-*MIXL1***+** differentiation (data not shown).

The second drawback comes from its method of action. Whilst CHIRON acts internally on cells, LPA has to bind to surface receptors to activate its signalling cascade. This means the presentation of these surface receptors is crucial to LPAs action. Etoc et al, 2016 illustrated how human PSC colony morphology can affect the presentation of BMP surface receptors. The same could be true for LPA receptors. What I saw at low levels of LPA was very concentrated *MIXL1* expression in densely packed areas, highlighted again in our microwell assessment. This could be as cells begin to compact in colonies the LPA signalling is weakened, and eventually blocked entirely or overcome by cell to cell contact, leaving the CHIRON to have unimpeded effects on differentiating the cells. Therefore, cell density and confluency is crucial during culturing in this medium: I found as cells became highly confluent they tended to be more differentiated. It is possible that using a recombinant WNT ligand approach would circumvent this issue, but the benefit of CHIRON is a more uniformed differentiation approach as ligand/receptor interactions are bypassed.

With any manipulation system control is key. I found that I need tighter control over the activity of WNT signalling. I implemented a similar approach to Baseline Activation first proposed by Hackland et al, 2017. The addition of the WNT secretion inhibitor, IWP-2, helped to stabilise cultures in priming medium and maintain the expression of SSEA-3, BF4 and CD9.

The idea of using a Baseline Activation approach was considered because early differentiation leads to the secretion of WNT ligands, as seen by the expression on WNT3 in our single cell transcriptomics. To this end I also thought about the secretion of BMP ligands from early differentiating cells, again as was indicated by increases in BMP2/4 in our single cell transcriptomics. I also cultured in the presence of DMH1, a BMP receptor inhibitor: while colonies appeared pristine, proliferation was severely impeded and eventually death began occurring in the colonies (data not shown). It is possible that low level BMP signalling is required for proliferation in our primed state or equally that DMH1 has off target effects, perhaps affecting TGFβ signalling. The use of recombinant NOGGIN instead may aid the medium to

alleviate the effects of BMP secretion, as this has been used in conjunction with hES culture previously (Xu et al., 2005).

LPA is a fairly stable molecule, but in the presence of cells it has a rather short half-life of ~2 hours (Zhao et al., 2005). This is because cells secrete Lipid phosphate phosphatases (LPP) which rapidly degrade LPA (Zhao et al., 2005) . There are chemical inhibitors such as sodium orthovanadate (Simon et al., 2002) and propranolol (Holinstat et al., 2007; Roberts et al., 1998) which competitively inhibit LPPs. A further optimisation may utilise an inhibitor of LPPs. Using an LPP inhibitor may mean that the amount of LPA or equally the amount of CHIRON might need to be reoptimized as you are effectively changing the balance of LPA versus CHIRON by preventing LPAs decay.

Much like how CHIRON replaces the need for recombinant WNT ligands in our system, I sought to find a chemical compound which could mimic the effects of LPA addition. I tried a chemical inhibitor of MST1/2 kinase, an upstream activator of the Hippo pathway (Fan et al., 2016), a TAZ activator, ethacridine (Kawano et al., 2015)and an Akt activator, SC79 (Jo et al., 2012). LPA has been shown as an effective inhibitor of the Hippo pathway (Yu et al., 2012) by inhibiting the actions of firstly MST1/2 and subsequently LATS2 kinase, resulting in increased YAP/TAZ levels. I used a specific chemical inhibitor for MST1/2, XMU-MP-1, to mimic the effects of LPA on the Hippo pathway (Fan et al., 2016). Our early test with this inhibitor resulted in cell death (data not shown), so the level may have to be optimised for the system or might not be applicable in these conditions.

Akt signalling has been shown to be an active switch between the binding of Smad2/3 to pluripotency factors or differentiation factors including *MIXL1* (Singh et al., 2012). LPA has been shown to have strong and substantial effects on P13K/Akt signalling reviewed in Riaz et al., 2016. LPA addition may increase Akt signalling and promote the binding of Smad2/3 to pluripotency factors over forming a complex with beta-catenin over differentiation factors (Singh et al., 2012). TAZ also has a relationship with Smad2/3, it serves to shuttle Smad2/3 into the nucleus when activated by TGFβ signalling (Varelas et al., 2008). Early assessment with Akt activator, SC79 on cells resulted in good cell growth but little effect in attenuating CHIRON-induced differentiation. The TAZ activator, ethacridine, also had little effect in attenuating CHIRON-induced differentiation. It could be that LPA cannot be replaced by chemical addition or equally the effect is caused by a cumulative impact on all of these aspects and requires combination of these inhibitors to be added.

6.4. Future uses for “Primo” medium.

Primo medium might prove a useful tool in studying the interconversion of human PSC, both forward into a lineage primed state but also backwards into a more pristine stem cell state. Here we have investigated transcriptional changes that are occurring in the media, in respect to priming and reversion. However, changes to epigenome, proteome, metabolic state and other cellular aspects could be investigated. Epigenetic alterations, mediated by polycomb, have been previously investigated with respect to lineage priming (Collinson et al., 2016; Dillon, 2012; Illingworth et al., 2016). Assessing the epigenetic landscapes during lineage priming and reversion, might provide key insights to the developmental process and how epigenetic changes facilitate differentiation. Combining the analysis of cellular aspects during lineage priming may help to resolve the critical events which ultimately lead to commitment of human PSC.

Another aspect of research that Primo medium could prove useful, is that of translational medicine. If this medium can improve the yield, efficiency, or time requirements of directed differentiation towards mesodermal derivatives that could have a huge impact on translation medicine. Although directed differentiation protocols have been developed to be fairly efficient, there still remains significant cell line to cell line variation, with some cell lines performing very poorly in differentiations towards a particular lineage (Osafune et al., 2008). This can be seen as an inherent bias within cell lines, possibly arising from the variation in genetic background. Nevertheless, if culturing cells in Primo medium could shift this bias towards mesoderm, even in lines particularly biased in another direction, this could help to uniform directed differentiation protocols. Future treatment derived from patient specific iPS cells, could be hampered by the same differentiation bias between lines and the use of Primo medium might go some way to alleviating that difference. This would establish Primo medium as a critical component in translational therapies related to mesodermal derivatives.

While I have investigated the dynamics of LPA prevention of WNT induced differentiation, possibly through YAP localisation, I have not investigated whether this system could be adapted for other differentiation signals and their subsequent differentiated progeny. Whether LPA addition can prevent directed differentiation towards ectoderm and endodermal lineages remains to be seen. If LPA is able to block these other lineages as well it provides a basis for creating lineage biased trapping systems for other lineages. Conversely, if LPA is unable to do so, this provides an interesting investigative prospect as to why and how the prevention of differentiation is limited to mesoderm.

Our system is based on the concept of baseline activation, proposed in Hackland et al, 2017. Hackland et al, 2017 however, focused on controlling BMP signalling during differentiation, using a Top Down Inhibition system. This system saturates media with BMP ligands and attenuates the signal by inhibiting BMP receptor response. Although the system was used in the context of differentiation toward neural crest, it might be possible to be used in parallel with LPA to create a lineage biased condition for lineages dependent on BMP signalling mediated differentiation.

Endoderm and mesoderm arise from a common precursor, the mesendoderm (Technau and Scholz, 2003). During early differentiation towards both lineages the expression of many genes is shared including that of *MIXL1*. In our neutral EB assay, cells from Primo medium displayed some upregulation of endoderm associated genes, but strong upregulation of mesoderm associated genes. While I was assessing bias under neutral conditions, slight modifications could efficiently generate endoderm cells. High levels of Activin A have been shown to be important in the formation of endoderm (D'Amour et al., 2005). By simply adding high activin A into the EB medium we may be able to shift the bias towards mesodermal lineages.

With that in mind, one could also think of an alteration to Primo in which the TGFβ is replaced for Activin A in E8 medium. Activin A has been to maintain self renewal of human PSC under the correct conditions (James et al., 2005; Vallier et al., 2005; Xiao et al., 2006). Nodal addition can replace TGFβ within E8 medium (Chen et al., 2011), whether Activin A could be used in the same manner is yet to be seen. With Activin A present in the culture medium, we might be able to shift the differentiation bias to endoderm as opposed to mesoderm.

6.5. Modelling stem cell commitment

The point of commitment to differentiation is difficult to define, and the bias of a cell is often difficult to ascertain. In our system, I saw a marked reduction of cloning efficiency when cells accrued high *MIXL1*-GFP expression. This gives us a commitment model based on *MIXL1*-GFP and SSEA-3 expression (Figure 6-2A). Cells move from pristine pluripotency, through biasing whilst retaining the ability to revert back to a pristine state. However, eventually cells reach a point of commitment where they cannot return to pristine pluripotency, through standard means, and progress to being fully differentiated.

In our system we are attempting to push cells towards the point of commitment without passing it. I used LPA and modulation of WNT signalling to move cells to different substates within the stem cell compartment without passing the line of commitment (Figure 6-2B). Our use of IWP-2 to inhibit the secretion of WNT ligands, highlights how close to the commitment point we are pushing the cells. As cells begin the differentiation process, endogenous signalling perpetuates this process towards a given lineage. Thus, blocking the endogenous signalling has shifted the commitment point of these cells. Further to this, signal modulation could shift the commitment point further than I have seen in our system and trap cells in a *MIXL1* high state. Therefore, our model of commitment fits the data I have assessed here, but in future could look very different.

Genetic changes can alter classical points of commitment as well, such as the loss of SSEA-3 (Enver et al., 2005). Monitoring genetic stability is a problematic yet necessary task for accurate research, in particular with regard to human PSC, as genetic changes may have functional consequences, in particular causing a resistance to committed differentiation. While wide varieties of chromosomal alterations are found in human PSC, there are some alterations more common than others. From the few samples which we analysed we saw no karyotypically changes in cells grown in Primo conditions. Unfortunately, our sample size for this investigation is relatively small especially when compared to the investigations in conventional human PSC (Amps et al., 2011; Baker et al., 2016; Taapken et al., 2011) In these large-scale studies ~12.90% of hES cultures assessed were karyotypically abnormal (Taapken et al., 2011). This highlights the importance of monitoring the genetic stability of cells grown in Primo in future work and the optimisation of the culture conditions to limit any recurrent genetic changes that may arise.

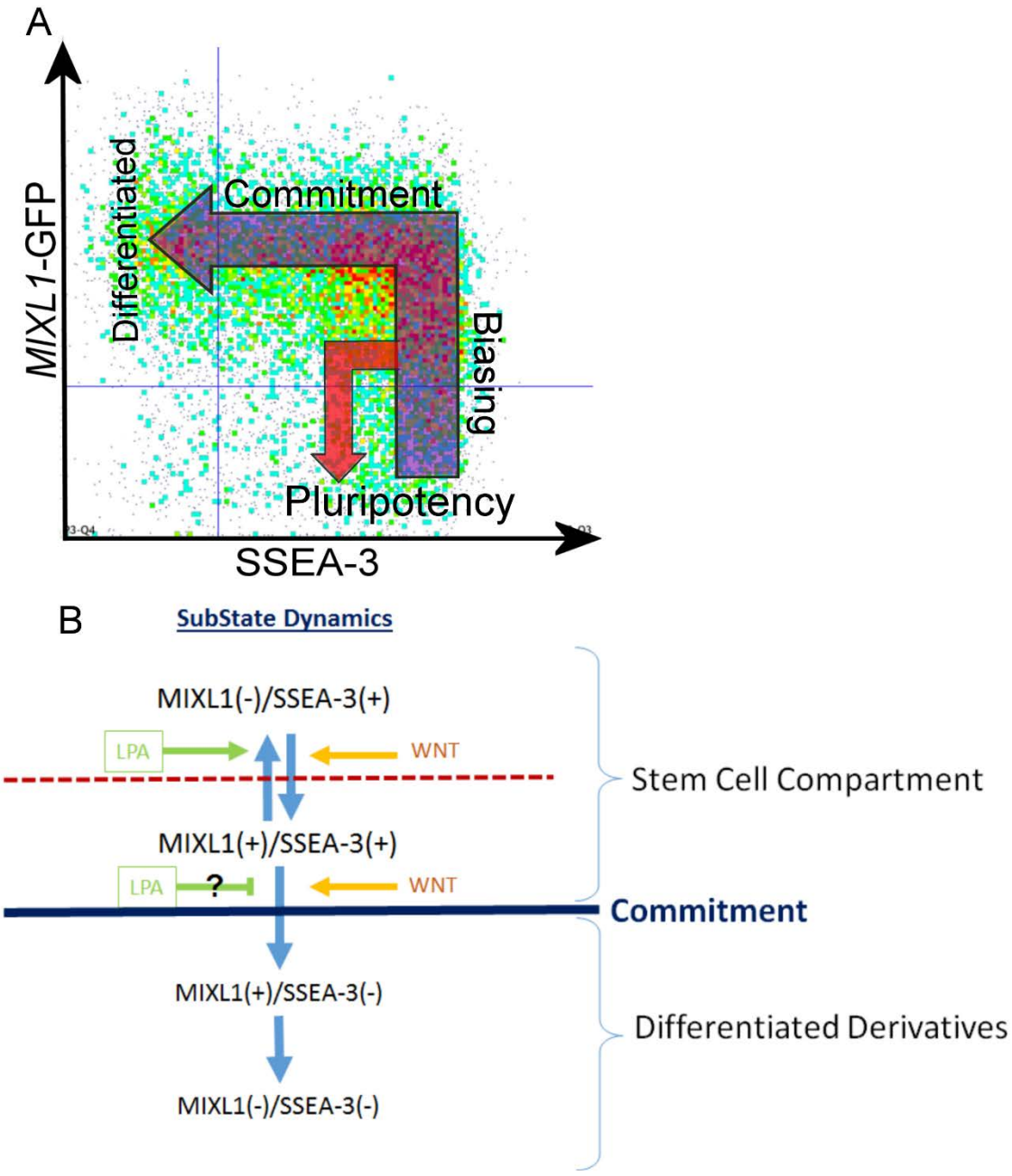


Figure 6-2 Proposed models of transition from pluripotency through lineage biasing towards eventual committed differentiated derivatives

A) A FACS density plot of *MIXL1*-GFP versus SSEA-3 from cells grown in E8V with addition of 3µM CHIR99021. Overlaid on top is the predicted change in expression of *MIXL1*-GFP and SSEA-3as cells transition from pluripotency towards differentiation (purple arrow) and the transition from lineage biased substates back to pluripotency (red arrow). **B)** A schematic of the Substate dynamics and how factors such as WNT and LPA may contribute to changes in *MIXL1* and SSEA-3 expression within the stem cell compartment and differentiated derivatives.

7. Concluding Remarks

In this study, I sought to characterise substates of human PSC found in conventional *in vitro* systems, that might exist as an intermediate state towards human mesoderm development *in vivo*. Utilising a combination of a fluorescent reporter for an early lineage marker and pluripotency-associated surface marker, I identified a particular substate. This substate expressed *MIXL1* whilst retaining the expression of the pluripotency-associated marker, SSEA-3. The state exhibited characteristics corresponding to undifferentiated cells, examined through the generation of single cell clonal lines. However, when assessed further differences between fractions defined by *MIXL1* expression were apparent in relations to differentiation. Examining the transcriptome, by bulk and single cell approaches, highlighted co-expression of pluripotency and differentiation associated gene regulatory networks within the *MIXL1*(**+**)/SSEA-3(**+**) substate.

There was an apparent loss of the substate when cultured defined conditions, which served as a platform to begin investigating how to generate and "trap" cells in a similar state to that identified in feeder systems. Our approach balances pro-self renewal and pro-differentiation environmental cues. Inducing a state which can be propagated long term by harnessing the self-renewal properties of human PSC, but when tested to differentiate more readily forms mesoderm. Furthermore, I demonstrated that this state, can be modulated and reverted back to a more pristine pluripotent state, by simple changes in growth medium.

Transcriptional comparisons between the newly generated substate and endogenously occurring substate from standard culture conditions, shows abundant overlap in gene expression. This implies, at least transcriptionally, these states are very similar. While present sporadically in standard culture system, our work demonstrates the first time human PSC have been propagated in a lineage bias state. Finally, the development of our optimised "Primo" medium has opened new avenues of research in the stem cell compartment not easily investigated under standard conditions. The medium allows for investigative research to be performed on the dynamics of substate inter-conversion. In terms of translational medicine, the medium might aid the differentiation towards mesodermal derivatives for translational applications.

8. References

Allison, T.F., Andrew, J.H.S., Anastassiasdis, K., Sloane-Stanley, J., Biga, V., Jones, M., Coca, D., Stavish, D., Frith, T.R., Hackland, J.O*., et al.* (2018a). Identification and single cell functional characterisation of a novel endodermally biased pluripotent sub-state in human embryonic stem cells. Submitted Manuscript.

Allison, T.F., Smith, A.J.H., Anastassiadis, K., Sloane-Stanley, J., Biga, V., Stavish, D., Hackland, J., Sabri, S., Langerman, J., Jones, M*., et al.* (2018b). Identification and Single-Cell Functional Characterization of an Endodermally Biased Pluripotent Substate in Human Embryonic Stem Cells. Stem Cell Reports *10*, 1895-1907.

Amit, M., Carpenter, M.K., Inokuma, M.S., Chiu, C.-P., Harris, C.P., Waknitz, M.A., Itskovitz-Eldor, J., and Thomson, J.A. (2000). Clonally Derived Human Embryonic Stem Cell Lines Maintain Pluripotency and Proliferative Potential for Prolonged Periods of Culture. Developmental biology *227*, 271-278.

Amps, K., Andrews, P.W., Anyfantis, G., Armstrong, L., Avery, S., Baharvand, H., Baker, J., Baker, D., Munoz, M.B., Beil, S*., et al.* (2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nature biotechnology *29*, 1132-1144.

Andrews, P.W., Banting, G., Damjanov, I., Arnaud, D., and Avner, P. (1984a). 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.

Andrews, P.W., Goodfellow, P.N., Shevinsky, L.H., Bronson, D.L., and Knowles, B.B. (1982). Cell-surface antigens of a clonal human embryonal carcinoma cell line: morphological and antigenic differentiation in culture. International journal of cancer *29*, 523-531.

Andrews PW, G.P., Bronson DL. (1983). Cell surface characteristics and other markers of differentiation of human teratocarcinomas in culture. In Teratocarcinoma Stem Cells, M.G. Silver LM, Strickland S, ed. (Cold Spring Harbor Conferences on Cell Proliferation), pp. 579-590.

Andrews, P.W., Meyer, L.J., Bednarz, K.L., and Harris, H. (1984b). Two monoclonal antibodies recognizing determinants on human embryonal carcinoma cells react specifically with the liver isozyme of human alkaline phosphatase. Hybridoma *3*, 33-39.

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

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development *17*, 126-140.

Azzolin, L., Panciera, T., Soligo, S., Enzo, E., Bicciato, S., Dupont, S., Bresolin, S., Frasson, C., Basso, G., Guzzardo, V*., et al.* (2014). YAP/TAZ incorporation in the beta-catenin destruction complex orchestrates the Wnt response. Cell *158*, 157-170.

Babraham-Bioinformatics (2018). SeqMonk: A tool to visualise and analyse high throughput mapped sequence data.

Baker, D., Hirst, Adam J., Gokhale, Paul J., Juarez, Miguel A., Williams, S., Wheeler, M., Bean, K., Allison, Thomas F., Moore, Harry D., Andrews, Peter W*., et al.* (2016). Detecting Genetic Mosaicism in Cultures of Human Pluripotent Stem Cells. Stem Cell Reports *7*, 998-1012.

Barbaric, I., Biga, V., Gokhale, P.J., Jones, M., Stavish, D., Glen, A., Coca, D., and Andrews, P.W. (2014). Time-Lapse Analysis of Human Embryonic Stem Cells Reveals Multiple Bottlenecks Restricting Colony Formation and Their Relief upon Culture Adaptation. Stem Cell Reports *3*, 142-155.

Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M.C., Tassani, S., Piva, F*., et al.* (2013). An estimation of the number of cells in the human body. Annals of Human Biology *40*, 463-471.

Blakeley, P., Fogarty, N.M.E., del Valle, I., Wamaitha, S.E., Hu, T.X., Elder, K., Snell, P., Christie, L., Robson, P., and Niakan, K.K. (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. Development *142*, 3151-3165.
Blauwkamp, T.A., Nigam, S., Ardehali, R., Weissman, I.L., and Nusse, R. (2012). Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. Nature communications *3*, 1070.
Blum, M., Gaunt, S.J., Cho, K.W.Y., Steinbeisser, H., Blumberg, B., Bittner, D., and De Robertis, E.M. (1992). Gastrulation in the mouse: The role of the homeobox gene goosecoid. Cell *69*, 1097-1106.
Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H*., et al.* (2011). Reference Maps of Human ES and iPS Cell Variation Enable High-Throughput Characterization of Pluripotent Cell Lines. Cell *144*, 439-452.
Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. Dev Cell *35*, 366-382.
Bouwmeester, T., Kim, S.-H., Sasai, Y., Lu, B., and Robertis, E.M.D. (1996). Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. Nature *382*, 595-601.
Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A*., et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.
Cai, H., Chen, H., Yi, T., Daimon, C.M., Boyle, J.P., Peers, C., Maudsley, S., and Martin, B. (2013). VennPlex--a novel Venn diagram program for comparing and visualizing datasets with differentially regulated datapoints. PloS one *8*, e53388.
Canham, M.A., Sharov, A.A., Ko, M.S.H., and Brickman, J.M. (2010). Functional Heterogeneity of Embryonic Stem Cells Revealed through Translational Amplification of an Early Endodermal Transcript. PLOS Biology *8*, e1000379.
Canton, I., Warren, N.J., Chahal, A., Amps, K., Wood, A., Weightman, R., Wang, E., Moore, H., and Armes, S.P. (2016). Mucin-Inspired Thermoresponsive Synthetic Hydrogels Induce Stasis in Human Pluripotent Stem Cells and Human Embryos. ACS Central Science *2*, 65-74.
Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J*., et al.* (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biology *7*, R100.
Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nature biotechnology *27*, 275-280.
Chan, Y.-S., Göke, J., Ng, J.-H., Lu, X., Gonzales, Kevin Andrew U., Tan, C.-P., Tng, W.-Q., Hong, Z.-Z., Lim, Y.-S., and Ng, H.-H. (2013a). Induction of a Human Pluripotent State with Distinct Regulatory Circuitry that Resembles Preimplantation Epiblast. Cell stem cell *13*, 663-675.
Chan, Y.S., Goke, J., Ng, J.H., Lu, X., Gonzales, K.A., Tan, C.P., Tng, W.Q., Hong, Z.Z., Lim, Y.S., and Ng, H.H. (2013b). Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. Cell stem cell *13*, 663-675.
Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. Developmental Cell *10*, 615-624.
Chen, B., Dodge, M.E., Tang, W., Lu, J., Ma, Z., Fan, C.W., Wei, S., Hao, W., Kilgore, J., Williams, N.S*., et al.* (2009a). Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. Nature chemical biology *5*, 100-107.
Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden, S.E., Diol, N.R., Propson, N.E*., et al.* (2011). Chemically defined conditions for human iPSC derivation and culture. Nature methods *8*, 424-429.
Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009b). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic acids research *37*, W305-W311.

Chen, S.S., Fitzgerald, W., Zimmerberg, J., Kleinman, H.K., and Margolis, L. (2007). Cell-cell and cell-extracellular matrix interactions regulate embryonic stem cell differentiation. Stem cells (Dayton, Ohio) *25*, 553-561.
Chu, L.-F., Leng, N., Zhang, J., Hou, Z., Mamott, D., Vereide, D.T., Choi, J., Kendziorski, C., Stewart, R., and Thomson, J.A. (2016). Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biology *17*, 173.
Cohen, M.A., Itsykson, P., and Reubinoff, B.E. (2010). The role of FGF-signaling in early neural specification of human embryonic stem cells. Developmental biology *340*, 450-458.
Collinson, A., Collier, A.J., Morgan, N.P., Sienerth, A.R., Chandra, T., Andrews, S., and Rugg-Gunn, P.J. (2016). Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells. Cell reports *17*, 2700-2714.
Condic, M.L. (2014). Totipotency: What It Is and What It Is Not. Stem Cells and Development *23*, 796-812.
Cooper, S., Bennett, W., Andrade, J., Reubinoff, B.E., Thomson, J., and Pera, M.F. (2002). Biochemical properties of a keratan sulphate/chondroitin sulphate proteoglycan expressed in primate pluripotent stem cells. Journal of anatomy *200*, 259-265.
Covey, T.M., Kaur, S., Tan Ong, T., Proffitt, K.D., Wu, Y., Tan, P., and Virshup, D.M. (2012). PORCN Moonlights in a Wnt-Independent Pathway That Regulates Cancer Cell Proliferation. PloS one *7*, e34532.
D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nature biotechnology *23*, 1534-1541.
da Cruz, L., Fynes, K., Georgiadis, O., Kerby, J., Luo, Y.H., Ahmado, A., Vernon, A., Daniels, J.T., Nommiste, B., Hasan, S.M.*, et al.* (2018). Phase 1 clinical study of an embryonic stem cell-derived retinal pigment epithelium patch in age-related macular degeneration. Nature biotechnology *36*, 328-337.
Daheron, L., Opitz, S.L., Zaehres, H., Lensch, M.W., Andrews, P.W., Itskovitz-Eldor, J., and Daley, G.Q. (2004). LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. Stem cells (Dayton, Ohio) *22*, 770-778.
Davidson, K.C., Adams, A.M., Goodson, J.M., McDonald, C.E., Potter, J.C., Berndt, J.D., Biechele, T.L., Taylor, R.J., and Moon, R.T. (2012). Wnt/β-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. Proceedings of the National Academy of Sciences *109*, 4485-4490.
Davis, R.P., Ng, E.S., Costa, M., Mossman, A.K., Sourris, K., Elefanty, A.G., and Stanley, E.G. (2008). 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.
Dillon, N. (2012). Factor mediated gene priming in pluripotent stem cells sets the stage for lineage specification. BioEssays : news and reviews in molecular, cellular and developmental biology *34*, 194-204.
Dorey, K., and Amaya, E. (2010). FGF signalling: diverse roles during early vertebrate embryogenesis. Development (Cambridge, England) *137*, 3731-3742.
Draper, J.S., Pigott, C., Thomson, J.A., and Andrews, P.W. (2002). Surface antigens of human embryonic stem cells: changes upon differentiation in culture. Journal of anatomy *200*, 249-258.
Draper, J.S., Smith, K., Gokhale, P., Moore, H.D., Maltby, E., Johnson, J., Meisner, L., Zwaka, T.P., Thomson, J.A., and Andrews, P.W. (2004). Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. Nat Biotech *22*, 53-54.
Ealy, M., Ellwanger, D.C., Kosaric, N., Stapper, A.P., and Heller, S. (2016). Single-cell analysis delineates a trajectory toward the human early otic lineage. Proceedings of the National Academy of Sciences *113*, 8508-8513.
Eiselleova, L., Matulka, K., Kriz, V., Kunova, M., Schmidtova, Z., Neradil, J., Tichy, B., Dvorakova, D., Pospisilova, S., Hampl, A.*, et al.* (2009). A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells. Stem cells (Dayton, Ohio) *27*, 1847-1857.

Enver, T., Pera, M., Peterson, C., and Andrews, P.W. (2009). Stem cell states, fates, and the rules of attraction. Cell stem cell *4*, 387-397.
Enver, T., Soneji, S., Joshi, C., Brown, J., Iborra, F., Orntoft, T., Thykjaer, T., Maltby, E., Smith, K., Abu Dawud, R.*, et al.* (2005). Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. Human molecular genetics *14*, 3129-3140.
Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154.
Fan, F., He, Z., Kong, L.L., Chen, Q., Yuan, Q., Zhang, S., Ye, J., Liu, H., Sun, X., Geng, J.*, et al.* (2016). Pharmacological targeting of kinases MST1 and MST2 augments tissue repair and regeneration. Science translational medicine *8*, 352ra108.
Fazeli, A., Liew, C.G., Matin, M.M., Elliott, S., Jeanmeure, L.F., Wright, P.C., Moore, H., and Andrews, P.W. (2011). Altered patterns of differentiation in karyotypically abnormal human embryonic stem cells. The International journal of developmental biology *55*, 175-180.
Fenderson, B.A., Andrews, P.W., Nudelman, E., Clausen, H., and Hakomori, S. (1987). Glycolipid core structure switching from globo- to lacto- and ganglio-series during retinoic acid-induced differentiation of TERA-2-derived human embryonal carcinoma cells. Developmental biology *122*, 21-34.
Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proceedings of the Japan Academy Series B, Physical and biological sciences *85*, 348-362.
Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A.*, et al.* (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282-286.
Gage, B.K., Webber, T.D., and Kieffer, T.J. (2013). Initial Cell Seeding Density Influences Pancreatic Endocrine Development During in vitro Differentiation of Human Embryonic Stem Cells. PloS one *8*, e82076.
Garcia-Gonzalo, F.R., and Izpisúa Belmonte, J.C. (2008). Albumin-Associated Lipids Regulate Human Embryonic Stem Cell Self-Renewal. PloS one *3*, e1384.
Ghosheh, N., Olsson, B., #xf6, rn, Edsbagge, J., #xfc, ppers-Munther, B., Van Giezen, M., Asplund, A., Andersson, T.B.*, et al.* (2016). Highly Synchronized Expression of Lineage-Specific Genes during In Vitro Hepatic Differentiation of Human Pluripotent Stem Cell Lines. Stem Cells International *2016*, 22.
Giacomelli, E., Bellin, M., Orlova, V.V., and Mummery, C.L. (2017). Co-Differentiation of Human Pluripotent Stem Cells-Derived Cardiomyocytes and Endothelial Cells from Cardiac Mesoderm Provides a Three-Dimensional Model of Cardiac Microtissue. Current protocols in human genetics *95*, 21.29.21-21.29.22.
Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M.K., Itskovitz-Eldor, J.*, et al.* (2004). Differences between human and mouse embryonic stem cells. Developmental biology *269*, 360-380.
Gobaa, S., Hoehnel, S., Roccio, M., Negro, A., Kobel, S., and Lutolf, M.P. (2011). Artificial niche microarrays for probing single stem cell fate in high throughput. Nature methods *8*, 949.
Gokhale, P.J., Au-Young, J.K., Dadi, S., Keys, D.N., Harrison, N.J., Jones, M., Soneji, S., Enver, T., Sherlock, J.K., and Andrews, P.W. (2015). Culture Adaptation Alters Transcriptional Hierarchies among Single Human Embryonic Stem Cells Reflecting Altered Patterns of Differentiation. PloS one *10*, e0123467.
Golos, T.G., Giakoumopoulos, M., and Garthwaite, M.A. (2010). Embryonic stem cells as models of trophoblast differentiation: progress, opportunities, and limitations. Reproduction (Cambridge, England) *140*, 3-9.
Graham, S.J.L., Wicher, K.B., Jedrusik, A., Guo, G., Herath, W., Robson, P., and Zernicka-Goetz, M. (2014). BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. Nature communications *5*.
Graham, V., Khudyakov, J., Ellis, P., and Pevny, L. (2003). SOX2 Functions to Maintain Neural Progenitor Identity. Neuron *39*, 749-765.

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., and Nichols, J. (2016). Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. Stem Cell Reports *6*, 437-446.
Hackland, J.O.S., Frith, T.J.R., Thompson, O., Marin Navarro, A., Garcia-Castro, M.I., Unger, C., and Andrews, P.W. (2017). Top-Down Inhibition of BMP Signaling Enables Robust Induction of hPSCs Into Neural Crest in Fully Defined, Xeno-free Conditions. Stem Cell Reports *9*, 1043-1052.
Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R.*, et al.* (2009). Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. Cell stem cell *5*, 597-609.
Hamazaki, T., Oka, M., Yamanaka, S., and Terada, N. (2004). Aggregation of embryonic stem cells induces Nanog repression and primitive endoderm differentiation. Journal of Cell Science *117*, 5681-5686.
Han, D.W., Tapia, N., Joo, J.Y., Greber, B., Araúzo-Bravo, M.J., Bernemann, C., Ko, K., Wu, G., Stehling, M., Do, J.T.*, et al.* (2010). Epiblast Stem Cell Subpopulations Represent Mouse Embryos of Distinct Pregastrulation Stages. Cell *143*, 617-627.
Hart, A.H., Hartley, L., Sourris, K., Stadler, E.S., Li, R., Stanley, E.G., Tam, P.P., Elefanty, A.G., and Robb, L. (2002). Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo. Development *129*, 3597-3608.
Hayashi, K., de Sousa Lopes, S.M.C., Tang, F., Lao, K., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell stem cell *3*, 10.1016/j.stem.2008.1007.1027.
Henderson, J.K., Draper, J.S., Baillie, H.S., Fishel, S., Thomson, J.A., Moore, H., and Andrews, P.W. (2002). Preimplantation Human Embryos and Embryonic Stem Cells Show Comparable Expression of Stage-Specific Embryonic Antigens. Stem cells (Dayton, Ohio) *20*, 329-337.
Holinstat, M., Voss, B., Bilodeau, M.L., and Hamm, H.E. (2007). Protease-activated receptors differentially regulate human platelet activation through a phosphatidic acid-dependent pathway. Molecular pharmacology *71*, 686-694.
Hough, S.R., Laslett, A.L., Grimmond, S.B., Kolle, G., and Pera, M.F. (2009). A continuum of cell states spans pluripotency and lineage commitment in human embryonic stem cells. PloS one *4*, e7708.
Hough, Shelley R., Thornton, M., Mason, E., Mar, Jessica C., Wells, Christine A., and Pera, Martin F. (2014). Single-Cell Gene Expression Profiles Define Self-Renewing, Pluripotent, and Lineage Primed States of Human Pluripotent Stem Cells. Stem Cell Reports *2*, 881-895.
Hsiao, C., Lampe, M., Nillasithanukroh, S., Han, W., Lian, X., and Palecek, S.P. (2016). Human pluripotent stem cell culture density modulates YAP signaling. Biotechnology journal *11*, 662-675.
Huelsken, J., and Behrens, J. (2002). The Wnt signalling pathway. Journal of Cell Science *115*, 3977-3978.
Iizuka, R., Yamagishi-Shirasaki, M., and Funatsu, T. (2011). Kinetic study of de novo chromophore maturation of fluorescent proteins. Analytical Biochemistry *414*, 173-178.
Illingworth, R.S., Holzenspies, J.J., Roske, F.V., Bickmore, W.A., and Brickman, J.M. (2016). Polycomb enables primitive endoderm lineage priming in embryonic stem cells. eLife *5*.
Jackson, S.A., Schiesser, J., Stanley, E.G., and Elefanty, A.G. (2010). Differentiating Embryonic Stem Cells Pass through 'Temporal Windows' That Mark Responsiveness to Exogenous and Paracrine Mesendoderm Inducing Signals. PloS one *5*, e10706.
James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development *132*, 1273-1282.
Jo, H., Mondal, S., Tan, D., Nagata, E., Takizawa, S., Sharma, A.K., Hou, Q., Shanmugasundaram, K., Prasad, A., Tung, J.K.*, et al.* (2012). Small molecule-induced cytosolic activation of protein kinase Akt rescues ischemia-elicited neuronal death. Proceedings of the National Academy of Sciences of the United States of America *109*, 10581-10586.

Kameda, T., and Thomson, J.A. (2005). Human ERas gene has an upstream premature polyadenylation signal that results in a truncated, noncoding transcript. Stem cells (Dayton, Ohio) *23*, 1535-1540.
Kannagi, R., Cochran, N.A., Ishigami, F., Hakomori, S., Andrews, P.W., Knowles, B.B., and Solter, D. (1983). Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells. The EMBO journal *2*, 2355-2361.
Katoh, M., and Katoh, M. (2006). CER1 is a common target of WNT and NODAL signaling pathways in human embryonic stem cells. Int J Mol Med *17*, 795-799.
Kawano, S., Maruyama, J., Nagashima, S., Inami, K., Qiu, W., Iwasa, H., Nakagawa, K., Ishigami-Yuasa, M., Kagechika, H., Nishina, H.*, et al.* (2015). A cell-based screening for TAZ activators identifies ethacridine, a widely used antiseptic and abortifacient, as a compound that promotes dephosphorylation of TAZ and inhibits adipogenesis in C3H10T1/2 cells. Journal of biochemistry *158*, 413-423.
Kempf, H., Olmer, R., Haase, A., Franke, A., Bolesani, E., Schwanke, K., Robles-Diaz, D., Coffee, M., Göhring, G., Dräger, G.*, et al.* (2016). Bulk cell density and Wnt/TGFbeta signalling regulate mesendodermal patterning of human pluripotent stem cells. Nature communications *7*, 13602.
Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. (2013a). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology *14*, R36.
Kim, H., Wu, J., Ye, S., Tai, C.-I., Zhou, X., Yan, H., Li, P., Pera, M., and Ying, Q.-L. (2013b). Modulation of β-catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. Nature communications *4*, 2403.
Kim, J.B., Sebastiano, V., Wu, G., Arauzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D.*, et al.* (2009). Oct4-induced pluripotency in adult neural stem cells. Cell *136*, 411-419.
Kinder, S.J., Tsang, T.E., Quinlan, G.A., Hadjantonakis, A.K., Nagy, A., and Tam, P.P. (1999). The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. Development *126*, 4691-4701.
Kleinsmith, L.J., and Pierce, G.B. (1964). Multipotentiality of Single Embryonal Carcinoma Cells. Cancer research *24*, 1544-1551.
Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K., and Herrmann, B.G. (2017). Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. Dev Cell *42*, 514-526.e517.
Kohler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. 1975. Journal of immunology (Baltimore, Md : 1950) *174*, 2453-2455.
Kojima, Y., Kaufman-Francis, K., Studdert, Joshua B., Steiner, Kirsten A., Power, Melinda D., Loebel, David A.F., Jones, V., Hor, A., de Alencastro, G., Logan, Grant J.*, et al.* (2014). The Transcriptional and Functional Properties of Mouse Epiblast Stem Cells Resemble the Anterior Primitive Streak. Cell stem cell *14*, 107-120.
Kono, K., Tamashiro, D.A.A., and Alarcon, V.B. (2014). Inhibition of RHO-ROCK signaling enhances ICM and suppresses TE characteristics through activation of Hippo signaling in the mouse blastocyst. Developmental biology *394*, 142-155.
Kuijk, E.W., van Tol, L.T.A., Van de Velde, H., Wubbolts, R., Welling, M., Geijsen, N., and Roelen, B.A.J. (2012). The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. Development *139*, 871-882.
Kurek, D., Neagu, A., Tastemel, M., Tüysüz, N., Lehmann, J., van de Werken, Harmen J.G., Philipsen, S., van der Linden, R., Maas, A., van Ijcken, Wilfred F.J.*, et al.* (2015). Endogenous WNT Signals Mediate BMP-Induced and Spontaneous Differentiation of Epiblast Stem Cells and Human Embryonic Stem Cells. Stem Cell Reports *4*, 114-128.
Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K.D., Yamada, R.G., Ueda, H.R., and Saitou, M. (2006). An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic acids research *34*, e42-e42.
Lang, S., Ugale, A., Erlandsson, E., Karlsson, G., Bryder, D., and Soneji, S. (2015). SCExV: a webtool for the analysis and visualisation of single cell qRT-PCR data. BMC Bioinformatics *16*, 320.

Latinkic, B.V., and Smith, J.C. (1999). Goosecoid and mix.1 repress Brachyury expression and are required for head formation in Xenopus. Development *126*, 1769-1779.
Li, J., Wang, G., Wang, C., Zhao, Y., Zhang, H., Tan, Z., Song, Z., Ding, M., and Deng, H. (2007). MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. Differentiation *75*, 299-307.
Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M.A., Goldstein, L.S.B., Abujarour, R.*, et al.* (2010). The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. Genes & development *24*, 1106-1118.
Lian, X., Bao, X., Al-Ahmad, A., Liu, J., Wu, Y., Dong, W., Dunn, Kaitlin K., Shusta, Eric V., and Palecek, Sean P. (2014). Efficient Differentiation of Human Pluripotent Stem Cells to Endothelial Progenitors via Small-Molecule Activation of WNT Signaling. Stem Cell Reports *3*, 804-816.
Lian, X.J., Hsiao, C., Wilson, G., Zhu, K.X., Hazeltine, L.B., Azarin, S.M., Raval, K.K., Zhang, J.H., Kamp, T.J., and Palecek, S.P. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. Proceedings of the National Academy of Sciences of the United States of America *109*, E1848-E1857.
Lim, S.M., Pereira, L., Wong, M.S., Hirst, C.E., Van Vranken, B.E., Pick, M., Trounson, A., Elefanty, A.G., and Stanley, E.G. (2009). Enforced Expression of Mixl1 During Mouse ES Cell Differentiation Suppresses Hematopoietic Mesoderm and Promotes Endoderm Formation. Stem cells (Dayton, Ohio) *27*, 363-374.
Lindsley, R.C., Gill, J.G., Kyba, M., Murphy, T.L., and Murphy, K.M. (2006). Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. Development *133*, 3787-3796.
Lippmann, E.S., Estevez-Silva, M.C., and Ashton, R.S. (2014). Defined Human Pluripotent Stem Cell Culture Enables Highly Efficient Neuroepithelium Derivation Without Small Molecule Inhibitors. Stem cells (Dayton, Ohio) *32*, 1032-1042.
Loh, K.M., Ang, L.T., Zhang, J.Y., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C.Y.Y., Nichane, M.*, et al.* (2014). Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. Cell stem cell *14*, 237-252.
Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M.*, et al.* (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell *166*, 451-467.
Loh, Y.H., Yang, L., Yang, J.C., Li, H., Collins, J.J., and Daley, G.Q. (2011). Genomic approaches to deconstruct pluripotency. Annual review of genomics and human genetics *12*, 165-185.
Lorthongpanich, C., Messerschmidt, D.M., Chan, S.W., Hong, W., Knowles, B.B., and Solter, D. (2013). Temporal reduction of LATS kinases in the early preimplantation embryo prevents ICM lineage differentiation. Genes & development *27*, 1441-1446.
Lowe, L.A., Yamada, S., and Kuehn, M.R. (2001). Genetic dissection of nodal function in patterning the mouse embryo. Development *128*, 1831-1843.
Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. PLoS computational biology *13*, e1005457.
Ludwig, T.E., Levenstein, M.E., Jones, J.M., Berggren, W.T., Mitchen, E.R., Frane, J.L., Crandall, L.J., Daigh, C.A., Conard, K.R., Piekarczyk, M.S.*, et al.* (2006). Derivation of human embryonic stem cells in defined conditions. Nature biotechnology *24*, 185-187.
Maaten, L.v.d., and Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research *9*, 2579-2605.
Maldonado, M., Luu, R.J., Ramos, M.E., and Nam, J. (2016). ROCK inhibitor primes human induced pluripotent stem cells to selectively differentiate towards mesendodermal lineage via epithelial-mesenchymal transition-like modulation. Stem cell research *17*, 222-227.
Martin Gonzalez, J., Morgani, S.M., Bone, R.A., Bonderup, K., Abelchian, S., Brakebusch, C., and Brickman, J.M. (2016). Embryonic Stem Cell Culture Conditions Support Distinct States Associated with Different Developmental Stages and Potency. Stem Cell Reports *7*, 177-191.
Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 7634-7638.

McFadden, D.G., Barbosa, A.C., Richardson, J.A., Schneider, M.D., Srivastava, D., and Olson, E.N. (2005). The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. Development *132*, 189-201.
Mendjan, S., Mascetti, Victoria L., Ortmann, D., Ortiz, M., Karjosukarso, Dyah W., Ng, Y., Moreau, T., and Pedersen, Roger A. (2014). NANOG and CDX2 Pattern Distinct Subtypes of Human Mesoderm during Exit from Pluripotency. Cell stem cell *15*, 310-325.
Mikawa, T., Poh, A.M., Kelly, K.A., Ishii, Y., and Reese, D.E. (2004). Induction and patterning of the primitive streak, an organizing center of gastrulation in the amniote. Developmental dynamics : an official publication of the American Association of Anatomists *229*, 422-432.
Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C.H., Miyazono, K*., et al.* (1997). TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. The EMBO journal *16*, 5353-5362.
Narayanan, K., Lim, V.Y., Shen, J., Tan, Z.W., Rajendran, D., Luo, S.C., Gao, S., Wan, A.C., and Ying, J.Y. (2014). Extracellular matrix-mediated differentiation of human embryonic stem cells: differentiation to insulin-secreting beta cells. Tissue engineering Part A *20*, 424-433.
Neely, M.D., Litt, M.J., Tidball, A.M., Li, G.G., Aboud, A.A., Hopkins, C.R., Chamberlin, R., Hong, C.C., Ess, K.C., and Bowman, A.B. (2012). DMH1, a Highly Selective Small Molecule BMP Inhibitor Promotes Neurogenesis of hiPSCs: Comparison of PAX6 and SOX1 Expression during Neural Induction. ACS Chemical Neuroscience *3*, 482-491.
Nichols, J., and Smith, A. (2009a). Naive and Primed Pluripotent States. Cell stem cell *4*, 487-492.
Nichols, J., and Smith, A. (2009b). Naive and primed pluripotent states. Cell stem cell *4*, 487-492.
Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.
Nimmo, R.A., May, G.E., and Enver, T. (2015). Primed and ready: understanding lineage commitment through single cell analysis. Trends in cell biology *25*, 459-467.
Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N*., et al.* (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Dev Cell *16*, 398-410.
Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nature biotechnology *26*, 313-315.
Paige, S.L., Osugi, T., Afanasiev, O.K., Pabon, L., Reinecke, H., and Murry, C.E. (2010). Endogenous Wnt/β-Catenin Signaling Is Required for Cardiac Differentiation in Human Embryonic Stem Cells. PloS one *5*, e11134.
Pearce, J.J., and Evans, M.J. (1999). Mml, a mouse Mix-like gene expressed in the primitive streak. Mechanisms of development *87*, 189-192.
Pereira, L.A., Wong, M.S., Lim, S.M., Sides, A., Stanley, E.G., and Elefanty, A.G. (2011). Brachyury and Related Tbx Proteins Interact with the Mixl1 Homeodomain Protein and Negatively Regulate Mixl1 Transcriptional Activity. PloS one *6*, e28394.
Pereira, L.A., Wong, M.S., Mossman, A.K., Sourris, K., Janes, M.E., Knezevic, K., Hirst, C.E., Lim, S.M., Pimanda, J.E., Stanley, E.G*., et al.* (2012). Pdgfrα and Flk1 are direct target genes of Mixl1 in differentiating embryonic stem cells. Stem cell research *8*, 165-179.
Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, Sarita P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell *165*, 1012-1026.
Piccolo, S., Agius, E., Leyns, L., Bhattacharyya, S., Grunz, H., Bouwmeester, T., and Robertis, E.M.D. (1999). The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. Nature *397*, 707-710.
Pina, C., Fugazza, C., Tipping, A.J., Brown, J., Soneji, S., Teles, J., Peterson, C., and Enver, T. (2012). Inferring rules of lineage commitment in haematopoiesis. Nature cell biology *14*, 287-294.

Pina, C., Teles, J., Fugazza, C., May, G., Wang, D., Guo, Y., Soneji, S., Brown, J., Eden, P., Ohlsson, M.*, et al.* (2015). Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis. Cell reports *11*, 1503-1510.
Price, F.D., Yin, H., Jones, A., van Ijcken, W., Grosveld, F., and Rudnicki, M.A. (2013). Canonical Wnt signaling induces a primitive endoderm metastable state in mouse embryonic stem cells. Stem cells (Dayton, Ohio) *31*, 752-764.
Proserpio, V., and Lönnberg, T. (2015). Single-cell technologies are revolutionizing the approach to rare cells. Immunology And Cell Biology *94*, 225.
Qi, X., Li, T.-G., Hao, J., Hu, J., Wang, J., Simmons, H., Miura, S., Mishina, Y., and Zhao, G.-Q. (2004). BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. Proceedings of the National Academy of Sciences of the United States of America *101*, 6027-6032.
Qin, H., Hejna, M., Liu, Y., Percharde, M., Wossidlo, M., Blouin, L., Durruthy-Durruthy, J., Wong, P., Qi, Z., Yu, J.*, et al.* (2016). YAP Induces Human Naive Pluripotency. Cell reports *14*, 2301-2312.
Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017a). Single-cell mRNA quantification and differential analysis with Census. Nature methods *14*, 309-315.
Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017b). Reversed graph embedding resolves complex single-cell trajectories. Nature methods *14*, 979-982.
Ramirez, J.-M., Bai, Q., Péquignot, M., Becker, F., Kassambara, A., Bouin, A., Kalatzis, V., Dijon-Grinand, M., and De Vos, J. (2013). Side scatter intensity is highly heterogeneous in undifferentiated pluripotent stem cells and predicts clonogenic self-renewal. Stem Cells and Development *22*, 1851-1860.
Ramirez, J.-M., Gerbal-Chaloin, S., Milhavet, O., Qiang, B., Becker, F., Assou, S., Lemaître, J.-M., Hamamah, S., and De Vos, J. (2011). Brief Report: Benchmarking Human Pluripotent Stem Cell Markers During Differentiation Into the Three Germ Layers Unveils a Striking Heterogeneity: All Markers Are Not Equal. Stem cells (Dayton, Ohio) *29*, 1469-1474.
Riaz, A., Huang, Y., and Johansson, S. (2016). G-Protein-Coupled Lysophosphatidic Acid Receptors and Their Regulation of AKT Signaling. International Journal of Molecular Sciences *17*, 215.
Rivera-Perez, J.A., and Magnuson, T. (2005). Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3. Developmental biology *288*, 363-371.
Robb, L., Hartley, L., Begley, C.G., Brodnicki, T.C., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., and Elefanty, A.G. (2000). Cloning, expression analysis, and chromosomal localization of murine and human homologues of a Xenopus mix gene. Developmental dynamics : an official publication of the American Association of Anatomists *219*, 497-504.
Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011a). Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics *27*, 2325-2329.
Roberts, A., Trapnell, C., Donaghey, J., Rinn, J., and Pachter, L. (2011b). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biology *12*, R22.
Roberts, R., Sciorra, V.A., and Morris, A.J. (1998). Human type 2 phosphatidic acid phosphohydrolases. Substrate specificity of the type 2a, 2b, and 2c enzymes and cell surface activity of the 2a isoform. The Journal of biological chemistry *273*, 22059-22067.
Roode, M., Blair, K., Snell, P., Elder, K., Marchant, S., Smith, A., and Nichols, J. (2012). Human hypoblast formation is not dependent on FGF signalling. Developmental biology *361*, 358-363.
Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W., and Staudt, L.M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature *345*, 686-692.
Rosowski, K.A., Mertz, A.F., Norcross, S., Dufresne, E.R., and Horsley, V. (2015). Edges of human embryonic stem cell colonies display distinct mechanical properties and differentiation potential. Scientific Reports *5*, 14218.
Row, R.H., and Kimelman, D. (2009). Bmp inhibition is necessary for post-gastrulation patterning and morphogenesis of the zebrafish tailbud. Developmental biology *329*, 55-63.
Rugg-Gunn, P.J. (2017). Naive pluripotent stem cells as a model for studying human developmental epigenomics: opportunities and limitations. Epigenomics *9*, 1485-1488.

Sahr, K., Dias, D.C., Sanchez, R., Chen, D., Chen, S.W., Gudas, L.J., and Baron, M.H. (2002). Structure, upstream promoter region, and functional domains of a mouse and human Mix paired-like homeobox gene. Gene *291*, 135-147.
Schugar, R.C., Robbins, P.D., and Deasy, B.M. (2007). Small molecules in stem cell self-renewal and differentiation. Gene Therapy *15*, 126.
Sharov, A.A., Masui, S., Sharova, L.V., Piao, Y., Aiba, K., Matoba, R., Xin, L., Niwa, H., and Ko, M.S.H. (2008). Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. BMC Genomics *9*, 269-269.
Shevinsky, L.H., Knowles, B.B., Damjanov, I., and Solter, D. (1982). Monoclonal antibody to murine embryos defines a stage-specific embryonic antigen expressed on mouse embryos and human teratocarcinoma cells. Cell *30*, 697-705.
Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. PLOS Biology *6*, e253.
Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722-737.
Simon, M.F., Rey, A., Castan-Laurel, I., Gres, S., Sibrac, D., Valet, P., and Saulnier-Blache, J.S. (2002). Expression of ectolipid phosphate phosphohydrolases in 3T3F442A preadipocytes and adipocytes. Involvement in the control of lysophosphatidic acid production. The Journal of biological chemistry *277*, 23131-23136.
Singh, A.M., Reynolds, D., Cliff, T., Ohtsuka, S., Mattheyses, A.L., Sun, Y., Menendez, L., Kulik, M., and Dalton, S. (2012). Signaling network crosstalk in human pluripotent cells: a Smad2/3-regulated switch that controls the balance between self-renewal and differentiation. Cell stem cell *10*, 312-326.
Singh, P., and Schwarzbauer, J.E. (2012). Fibronectin and stem cell differentiation – lessons from chondrogenesis. Journal of Cell Science *125*, 3703-3712.
Slack, J.M.W. (1984). From Egg to Embryo: Determinative Events in Early Development (Cambridge University Press).
Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688.
Smith, R.C.G., Stumpf, P.S., Ridden, S.J., Sim, A., Filippi, S., Harrington, H.A., and MacArthur, B.D. (2017). Nanog Fluctuations in Embryonic Stem Cells Highlight the Problem of Measurement in Cell Biology. Biophysical journal *112*, 2641-2652.
Solter, D., and Knowles, B.B. (1978). Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). Proceedings of the National Academy of Sciences of the United States of America *75*, 5565-5569.
Stamos, J.L., and Weis, W.I. (2013). The beta-catenin destruction complex. Cold Spring Harbor perspectives in biology *5*, a007898.
Stein, S., Roeser, T., and Kessel, M. (1998). CMIX, a paired-type homeobox gene expressed before and during formation of the avian primitive streak. Mechanisms of development *75*, 163-165.
Sturn, A., Quackenbush, J., and Trajanoski, Z. (2002). Genesis: cluster analysis of microarray data. Bioinformatics *18*, 207-208.
Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PloS one *6*, e21800.
Suwinska, A. (2012). Preimplantation mouse embryo: developmental fate and potency of blastomeres. Results and problems in cell differentiation *55*, 141-163.
Taapken, S.M., Nisler, B.S., Newton, M.A., Sampsell-Barron, T.L., Leonhard, K.A., McIntire, E.M., and Montgomery, K.D. (2011). Karyotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. Nat Biotech *29*, 313-314.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.
Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.
Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W*., et al.* (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254-1269.
Technau, U., and Scholz, C.B. (2003). Origin and evolution of endoderm and mesoderm. The International journal of developmental biology *47*, 531-539.
Teles, J., Pina, C., Eden, P., Ohlsson, M., Enver, T., and Peterson, C. (2013). Transcriptional regulation of lineage commitment--a stochastic model of cell fate decisions. PLoS computational biology *9*, e1003197.
Teo, A.K., Ali, Y., Wong, K.Y., Chipperfield, H., Sadasivam, A., Poobalan, Y., Tan, E.K., Wang, S.T., Abraham, S., Tsuneyoshi, N*., et al.* (2012). Activin and BMP4 synergistically promote formation of definitive endoderm in human embryonic stem cells. Stem cells (Dayton, Ohio) *30*, 631-642.
Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.
Theunissen, Thorold W., Friedli, M., He, Y., Planet, E., O'Neil, Ryan C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M*., et al.* (2016). Molecular Criteria for Defining the Naive Human Pluripotent State. Cell stem cell *19*, 502-515.
Theunissen, Thorold W., Powell, Benjamin E., Wang, H., Mitalipova, M., Faddah, Dina A., Reddy, J., Fan, Zi P., Maetzel, D., Ganz, K., Shi, L*., et al.* (2014). Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. Cell stem cell *15*, 471-487.
Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science (New York, NY) *282*, 1145-1147.
Tonge, P.D., and Andrews, P.W. (2010). Retinoic acid directs neuronal differentiation of human pluripotent stem cell lines in a non-cell-autonomous manner. Differentiation *80*, 20-30.
Tonge, P.D., Shigeta, M., Schroeder, T., and Andrews, P.W. (2011). Functionally defined substates within the human embryonic stem cell compartment. Stem cell research *7*, 145-153.
Torres-Padilla, M.-E., Richardson, L., Kolasinska, P., Meilhac, S.M., Luetke-Eversloh, M.V., and Zernicka-Goetz, M. (2007). The anterior visceral endoderm of the mouse embryo is established from both preimplantation precursor cells and by de novo gene expression after implantation. Developmental biology *309*, 97-112.
Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature biotechnology *32*, 381-386.
Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotech *31*, 46-53.
Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotech *28*, 511-515.
Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R., Economou, C., Karagianni, E., Zhao, S., Lowell, S*., et al.* (2014). Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. Development (Cambridge, England) *141*, 1209-1221.
Tsukazaki, T., Chiang, T.A., Davison, A.F., Attisano, L., and Wrana, J.L. (1998). SARA, a FYVE Domain Protein that Recruits Smad2 to the TGFβ Receptor. Cell *95*, 779-791.

Valdimarsdottir, G., and Mummery, C. (2005). Functions of the TGFbeta superfamily in human embryonic stem cells. APMIS : acta pathologica, microbiologica, et immunologica Scandinavica *113*, 773-789.
Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. Journal of Cell Science *118*, 4495-4509.
Varelas, X., Sakuma, R., Samavarchi-Tehrani, P., Peerani, R., Rao, B.M., Dembowy, J., Yaffe, M.B., Zandstra, P.W., and Wrana, J.L. (2008). TAZ controls Smad nucleocytoplasmic shuttling and regulates human embryonic stem-cell self-renewal. Nature cell biology *10*, 837-848.
Vega Crespo, A., Awe, J.P., Reijo Pera, R., and Byrne, J.A. (2012). Human Skin Cells That Express Stage-Specific Embryonic Antigen 3 Associate with Dermal Tissue Regeneration. BioResearch Open Access *1*, 25-33.
Vincent, S.D., Dunn, N.R., Hayashi, S., Norris, D.P., and Robertson, E.J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. Genes & development *17*, 1646-1662.
Wada, K.-I., Itoga, K., Okano, T., Yonemura, S., and Sasaki, H. (2011). Hippo pathway regulation by cell morphology and stress fibers. Development *138*, 3907-3914.
Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57-63.
Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. Cell stem cell *10*, 440-454.
Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A.*, et al.* (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell stem cell *7*, 618-630.
Warrier, S., Van der Jeught, M., Duggal, G., Tilleman, L., Sutherland, E., Taelman, J., Popovic, M., Lierman, S., Chuva De Sousa Lopes, S., Van Soom, A.*, et al.* (2017). Direct comparison of distinct naive pluripotent states in human embryonic stem cells. Nature communications *8*, 15055.
Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S., Muguruma, K.*, et al.* (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nature biotechnology *25*, 681-686.
Williams, B.P., Daniels, G.L., Pym, B., Sheer, D., Povey, S., Okubo, Y., Andrews, P.W., and Goodfellow, P.N. (1988). Biochemical and genetic analysis of the OKa blood group antigen. Immunogenetics *27*, 322-329.
Wright, A., Andrews, N., Bardsley, K., Nielsen, J.E., Avery, K., Pewsey, E., Jones, M., Harley, D., Nielsen, A.R., Moore, H.*, et al.* (2011). Mapping the stem cell state: eight novel human embryonic stem and embryonal carcinoma cell antibodies. International journal of andrology *34*, e175-187; discussion e187-178.
Xiao, L., Yuan, X., and Sharkis, S.J. (2006). Activin A maintains self-renewal and regulates fibroblast growth factor, Wnt, and bone morphogenic protein pathways in human embryonic stem cells. Stem cells (Dayton, Ohio) *24*, 1476-1486.
Xu, R.-H., Barron, T.L., Gu, F., Root, S., Peck, R.M., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R.*, et al.* (2008). NANOG is a Direct Target of TGFβ/Activin Mediated SMAD Signaling in Human ES Cells. Cell stem cell *3*, 196-206.
Xu, R.H., Chen, X., Li, D.S., Li, R., Addicks, G.C., Glennon, C., Zwaka, T.P., and Thomson, J.A. (2002a). BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nature biotechnology *20*, 1261-1264.
Xu, R.H., Chen, X., Li, D.S., Li, R., Addicks, G.C., Glennon, C., Zwaka, T.P., and Thomson, J.A. (2002b). BMP4 initiates human embryonic stem cell differentiation to trophoblast. In Nature biotechnology (United States), pp. 1261-1264.
Xu, R.H., Peck, R.M., Li, D.S., Feng, X., Ludwig, T., and Thomson, J.A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nature methods *2*, 185-190.
Yamanaka, Y., Lanner, F., and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development *137*, 715-724.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J*., et al.* (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol *20*, 1131-1139.
Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.
Yu, F.-X., and Guan, K.-L. (2013). The Hippo pathway: regulators and regulations. Genes & development *27*, 355-371.
Yu, F.-X., Zhao, B., Panupinthu, N., Jewell, J.L., Lian, I., Wang, L.H., Zhao, J., Yuan, H., Tumaneng, K., Li, H*., et al.* (2012). Regulation of the Hippo-YAP pathway by G-protein coupled receptor signaling. Cell *150*, 780-791.
Yu, P., Pan, G., Yu, J., and Thomson, J.A. (2011). FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation. Cell stem cell *8*, 326-334.
Zernicka-Goetz, M. (2002). Patterning of the embryo: the first spatial decisions in the life of a mouse. Development *129*, 815-829.
Zhang, G., Gurtu, V., and Kain, S.R. (1996). An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. Biochem Biophys Res Commun *227*, 707-711.
Zhang, H., Fraser, S.T., Papazoglu, C., Hoatlin, M.E., and Baron, M.H. (2009). Transcriptional Activation by the Mixl1 Homeodomain Protein in Differentiating Mouse Embryonic Stem Cells. Stem cells (Dayton, Ohio) *27*, 2884-2895.
Zhang, S., and Cui, W. (2014). Sox2, a key factor in the regulation of pluripotency and neural differentiation. World Journal of Stem Cells *6*, 305-311.
Zhao, Y., Usatyuk, P.V., Cummings, R., Saatian, B., He, D., Watkins, T., Morris, A., Spannhake, E.W., Brindley, D.N., and Natarajan, V. (2005). Lipid phosphate phosphatase-1 regulates lysophosphatidic acid-induced calcium release, NF-kappaB activation and interleukin-8 secretion in human bronchial epithelial cells. The Biochemical journal *385*, 493-502.

9. Appendix

Table 9.1 Single Cell qPCR assay list.

Assay ID	Gene Symbol	Gene Name	RefSeq	Amplicon Length	Detects gDNA	Best Coverage
Hs01060665_g1	ACTB	actin; beta	NM_001101.3	63	Yes	Yes
Hs00154192_m1	BMP2	Bone morphogenetic protein 2	NM_001200.2	60	No	Yes
Hs03676628_s1	BMP4	bone morphogenetic protein 4	NM_130850.2;NM_001202.3; NM_130851.2	116	Yes	Yes
Hs01034913_g1	BMPR1A	bone morphogenetic protein receptor; type IA	NM_004329.2	94	Yes	Yes
Hs00193796_m1	CER1	cerberus 1; DAN family BMP antagonist	NM_005454.2	92	No	Yes
Hs01897804_s1	CITED2	Cbp/p300-interacting transactivator; with Glu/Asp-rich carboxy-terminal domain; 2	NM_001168388.2; NM_001168389.2; NM_006079.4	106	Yes	Yes
Hs00607528_s1	CLDN6	claudin 6	NM_021195.4	154	Yes	Yes
Hs00164004_m1	COL1A1	collagen; type I; alpha 1	NM_000088.3	66	No	Yes

Assay ID	Gene Symbol	Gene Name	RefSeq	Amplicon Length	Detects gDNA	Best Coverage
Hs00607978_s1	CXCR4	chemokine (C-X-C motif) receptor 4	NM_003467.2;NM_001008540.1	153	Yes	Yes
Hs00171876_m1	DNMT3B	DNA (cytosine-5-)-methyltransferase 3 beta	NM_001207055.1; NM_001207056.1; NM_175848.1;NM_175849.1; NM_175850.2;NM_006892.3	55	No	Yes
Hs00172872_m1	EOMES	eomesodermin	NM_001278183.1; NM_001278182.1; NM_005442.3	81	No	Yes
Hs01549976_m1	FN1	fibronectin 1	NM_212482.1;NM_054034.2; NM_002026.2;NM_212478.1; NM_212474.1;NM_212476.1	81	No	Yes
Hs00232764_m1	FOXA2	forkhead box A2	NM_021784.4;NM_153675.2	66	No	Yes

Assay ID	Gene Symbol	Gene Name	RefSeq	Amplicon Length	Detects gDNA	Best Coverage
Hs00255287_s1	FOXD3	forkhead box D3	NM_012183.2	78	Yes	Yes
Hs00173503_m1	FRZB	frizzled-related protein	NM_001463.3	108	No	Yes
Hs00246256_m1	FST	follistatin	NM_006350.3;NM_013409.2	108	No	No
Hs00544355_m1	GAL	galanin/GMAP prepropeptide	NM_015973.3	125	No	Yes
Hs00171403_m1	GATA4	GATA binding protein 4	NM_002052.3	68	No	Yes
Hs00232018_m1	*GATA6*	GATA binding protein 6	NM_005257.4	91	No	Yes
Hs00906630_g1	GSC	goosecoid homeobox	NM_173849.2	100	No	No
Hs00193435_m1	HAS2	hyaluronan synthase 2	NM_005328.2	63	No	Yes
Hs00242160_m1	HHEX	hematopoietically expressed homeobox	NM_002729.4	110	Yes	Yes
Hs00705137_s1	IFITM1	Interferon-induced transmembrane protein 1	NM_003641.3	93	Yes	Yes
Hs01547673_m1	ITGA5	integrin; alpha 5 (fibronectin receptor; alpha polypeptide)	NM_002205.2	54	No	Yes
Hs00761767_s1	KRT19	keratin 19	NM_002276.4	116	Yes	Yes

Assay ID	Gene Symbol	Gene Name	RefSeq	Amplicon Length	Detects gDNA	Best Coverage
Hs00764128_s1	LEFTY1	left-right determination factor 1	NM_020997.3	136	Yes	Yes
Hs00745761_s1	LEFTY2	left-right determination factor 2	NM_001172425.1; NM_003240.3	102	Yes	Yes
Hs00355202_m1	LGALS1	lectin; galactoside-binding; soluble; 1	NM_002305.3	63	No	Yes
Hs00232144_m1	LHX1	LIM homeobox 1	NM_005568.3	60	No	Yes
Hs00702808_s1	LIN28A	lin-28 homolog A (C. elegans)	NM_024674.4	143	Yes	Yes
Hs00430824_g1	*MIXL1*	Mix paired-like homeobox	NM_031944.1	152	No	No
Hs00899658_m1	MMP1	matrix metallopeptidase 1 (interstitial collagenase)	NM_001145938.1; NM_002421.3	64	No	Yes
Hs01548727_m1	MMP2	matrix metallopeptidase 2 (gelatinase A; 72kDa gelatinase; 72kDa type IV collagenase)	NM_004530.4;NM_001127891.1	65	No	Yes
Hs01085598_g1	MYL7	myosin; light chain 7; regulatory	NM_021223.2	74	No	Yes

Assay ID	Gene Symbol	Gene Name	RefSeq	Amplicon Length	Detects gDNA	Best Coverage
Hs04399610_g1	*NANOG*	*NANOG* homeobox	NM_024865.2	101	Yes	No
Hs00378379_m1	NCLN	nicalin	NM_020170.3	65	No	Yes
Hs00415443_m1	NODAL	nodal growth differentiation factor	NM_018055.4	68	No	Yes
Hs00219496_m1	PAF1	Paf1; RNA polymerase II associated factor; homolog (S. cerevisiae)	NM_019088.3;NM_001256826.1	100	No	Yes
Hs04260367_gH	POU5F1	POU class 5 homeobox 1	NM_001173531.1; NM_002701.4;NM_203289.4	77	Yes	Yes
Hs01375212_g1	RPS18	ribosomal protein S18	NM_022551.2	93	Yes	Yes
Hs00183425_m1	SMAD2	SMAD family member 2	NM_001135937.2; NM_001003652.3; NM_005901.5	129	No	No
Hs00195591_m1	SNAI1	snail family zinc finger 1	NM_005985.3	66	Yes	Yes
Hs00751752_s1	SOX17	SRY (sex determining region Y)-box 17	NM_022454.3	149	Yes	Yes

Assay ID	Gene Symbol	Gene Name	RefSeq	Amplicon Length	Detects gDNA	Best Coverage
Hs01053049_s1	*SOX2*	SRY-box2	NM_003106.3	91	Yes	Yes
Hs00610080_m1	T	T; brachyury homolog (mouse)	NM_001270484.1; NM_003181.3	132	No	Yes
Hs00761239_s1	TAGLN2	transgelin 2	NM_001277224.1; NM_001277223.1; NM_003564.2	163	No	Yes
Hs02339499_g1	TDGF1	teratocarcinoma-derived growth factor 1	NM_003212.3;NM_001174136.1	170	Yes	No
Hs00902257_m1	WNT3	wingless-type MMTV integration site family; member 3	NM_030753.4	76	No	Yes

9.1.1. Monocle single cell qPCR Script:

Denotes annotation of steps.

##Set Working Directory

setwd("D:/Users/Dylan/Documents/Google Drive/Meso/Single Cell/Monocle2")

##Load Bioconductor

source("http://bioconductor.org/biocLite.R")

biocLite()

##Install Monocle

biocLite("monocle")

##Load Monocle

```
library(monocle)
##Install dependencies
biocLite(c("DDRTree", "pheatmap"))
##Load Data
mydata <- read.delim("All Samples CT Values trimmed.txt")
mydata <- data.frame(mydata[,-1], row.names=mydata[,1])
mysamples <- read.delim ("samplesall.txt")
mysamples <- data.frame(mysamples[,-1], row.names=mysamples[,1])
row.names(mysamples) <- colnames(mydata)
mygenes <- read.table("genes.txt")
##Create annotated Data frames
pd <- new('AnnotatedDataFrame', data = mysamples)
fd <- new('AnnotatedDataFrame', data = mygenes)
cds <- newCellDataSet(as.matrix(mydata), phenoData = pd, featureData = fd, expressionFamily
= gaussianff())
##Create Cell Hierarchy
cth <- newCellTypeHierarchy()
##Create Cell Types based on Gene expression
MIXL1_id <- row.names(substate(fData(cds), gene_short_name == "MIXL1"))
cth <- addCellType(cth, "MIXLPLUS", classify_func = function(x) { x[MIXL1_id,] < 40 )
PLURI_id <- row.names(substate(fData(cds), gene_short_name == "NANOG"))
cth <- addCellType(cth, "PLURI", classify_func = function(x) { x[PLURI_id,] < 20 & x[MIXL1_id,] >
39 })
##Classify cells based on type
cds <- classifyCells(cds, cth, 0.1)
##Create table and pie chart displaying cell types
table(pData(cds)$CellType)
pie <- ggplot(pData(cds), aes(x = factor(1), fill = factor(CellType))) +
  geom_bar(width = 1)
pie + coord_polar(theta = "y") +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
##Set function for start of pseudotime as MIXL1neg SSEA-3plus
GM_state <- function(cds){if (length(unique(pData(cds)$State)) > 1){T0_counts <-
table(pData(cds)$State, pData(cds)$Group)["NegPlus"]
return(as.numeric(names(T0_counts)[which(T0_counts == max(T0_counts))]))}else
##Order Cells
```

```
cds <- orderCells(cds)
##Order cells by pseudotime and reverse order
cds <- orderCells(cds, root_state = GM_state(cds), reverse = 1)
##Reduce dimensions for DDR tree trajectory
cds <- reduceDimension(cds, max_components = 2, method = 'DDRTree', norm_method = c("none"), pseudo_expr = 0)
##Plot cell trajectory trees with various options
plot_cell_trajectory(cds, color_by = "CellType")
plot_cell_trajectory(cds, color_by = "Group") +scale_color_manual(values = c("red", "blue", "forestgreen", "purple4"))
plot_cell_trajectory(cds, color_by = "Pseudotime")
##Plot cell trajectory trees with Marker gene plotting
plot_cell_trajectory(cds, color_by = "Group", markers = "GATA6") +scale_color_manual(values = c("red", "blue", "forestgreen", "purple4")) +scale_size(trans = "reverse")
##Plot Heatmaps of genes in pseudotime with various options
plot_pseudotime_heatmap(cds, show_rownames = TRUE, num_clusters = 3)
plot_pseudotime_heatmap(cds, show_rownames = TRUE, hmcols = colorRampPalette(c("red","yellow","cyan","blue"))(65))
##Plot branch points as heatmaps with various options
plot_genes_branched_heatmap(cds, branch_point = 1, num_clusters = 2, cores = 1, use_gene_short_name = T, show_rownames = T, norm_method = c("log"))
plot_genes_branched_heatmap(cds, branch_point = 1, cores = 1, use_gene_short_name = T, show_rownames = T, norm_method = c("log"), hmcols = colorRampPalette(c("red","yellow","cyan","blue"))(65))
##Set Pluripotent gene substate
Pluri_genes <- row.names(substate(fData(cds), gene_short_name %in% c("SOX2", "NANOG", "POU5F1", "DNMT3B", "LIN28A", "CLDN6")))
Pluri_substate <- cds[Pluri_genes,]
## Plot Pluripotent genes as branched pseudotime plots
plot_genes_branched_pseudotime(Pluri_substate, color_by = "Group", ncol = 3, nrow = 2) +scale_y_continuous(trans = "reverse") +scale_color_manual(values = c("red", "blue", "forestgreen", "purple4"))
##Set Differentiation gene substate
Diff_genes <- row.names(substate(fData(cds), gene_short_name %in% c("CER1", "EOMES", "FOXA2", "GATA6", "MIXL1", "T")))
Diff_substate <- cds[Diff_genes,]
```

Plot Differentiation genes as branched pseudotime plots

plot_genes_branched_pseudotime(Diff_substate, color_by = "Group", ncol = 3, nrow = 2) **+**scale_y_continuous(trans = "reverse", limits = c(40,0)) **+**scale_color_manual(values = c("red", "blue", "forestgreen", "purple4"))

Plot all genes as branched pseudotime plots

plot_genes_branched_pseudotime(cds, color_by = "Group", ncol = 9, nrow = 5) **+**scale_y_continuous(trans = "reverse", limits = c(40,0)) **+**scale_color_manual(values = c("red", "blue", "forestgreen", "purple4"))

##Save and export Eigen Values for assessment in other software.

S_matrix <- reducedDimS(cds)

write.csv(S_matrix, file = "DDRposition.csv", row.names = TRUE)

##Reduce dimensions for tSNE analysis

cds <- reduceDimension(cds, max_components = 2, num_dim = 6, reduction_method = 'tSNE', verbose = T, norm_method = c("none"), pseudo_expr = 0)

##Cluster Cells

cds <- clusterCells(cds, num_clusters = 3, gaussian = T)

##Plot Cell clusters on tSNE with various options

plot_cell_clusters(cds, 1, 2, 3, color = "Cluster") **+** facet_wrap(~CellType)

plot_cell_clusters(cds, 1, 2, 3, color = "Cluster") **+** facet_wrap(~Group)

##Plot Cell clusters on tSNE with marker gene sizes

plot_cell_clusters(cds, 1, 2, 3, color = "Cluster", markers = "*SOX2*") **+**scale_color_manual(values = c("red", "blue", "forestgreen", "purple4")) **+**scale_size(trans = "reverse")