The interplay between human germline specification and pluripotency factors

Keir Heath Murison

Imperial college London, Institute of Clinical Sciences

Thesis submitted for the degree of Doctor of Philosophy

IMPERIAL

UKRI MRC London Institute of Medical Sciences

Dedication

I wish to dedicate this thesis to Oliver ‘Oli’ Burnham, an incredible man, a wonderful friend, and a simply brilliant centre. I was proud to call you a friend and miss you each day.

Abstract

Pluripotency is regulated by transcription factors (TF) that maintain this cell state. In mouse, the same core pluripotency factors; Oct4, Nanog and Sox2 are expressed in the germline. The expression of these allows for the reversion of primordial germ cells (PGCs), founders of the gametes, into a pluripotent state *in vitro.*

Human primordial germ cells (PGCs) express OCT4 and NANOG but repress SOX2 and upregulate SOX17. Reports of similar conversion of human primordial germ cell-like cells (PGCLCs) or human germline tumours into a pluripotent state, *in vivo* or *in vitro*, require the downregulation of SOX17 and upregulation of SOX2. I hypothesised that overexpression of SOX2 within hPGCLCs might destabilise the germline network and trigger entry into a pluripotent state. I wanted to observe the effect of overexpression of other pluripotency related transcription factors (TFs), NANOG and KLF2, on germ cell-like cells.

To track entry into pluripotency, I utilised a cell line containing a reporter for endogenous SOX2 expression. A set of plasmids that overexpress the different pluripotency TFs of interest were transfected into the SOX2 reporter cells. These cell lines were then used to generate an *in vitro* cell type that resembles early human PGCs, called hPGC-like cells (hPGCLCs). The TFs could be activated and overexpressed at different points of the hPGCLC induction protocol and evidence of SOX2 re-activation could be observed.

I observed that SOX2 overexpression could not activate the endogenous SOX2 gene in any conditions studied, suggesting a dominant mechanism of repression probably mediated by BMP4 signalling. SOX2 could block entry into germline fate when overexpression was initiated at the point of hPGCLC induction but had no effect when triggered after induction. The other TF cell lines showed a failure to activate overexpression in the hPGCLC state, suggesting the inserted cassettes were silenced in hPGCLCs.

Student Declaration

I certify that:

- The thesis being submitted for examination is my own account of my own research.
- My research has been conducted ethically.
- The data and results presented are the genuine data and results actually obtained by me during the conduct of the research.
- Where I have drawn on the work, ideas and results of others this has been appropriately acknowledged in the thesis.
- Where any collaboration has taken place with other researchers, I have clearly stated in the thesis my own personal share in the investigation.

Copyright declaration

Acknowledgements

It is hard to think of my PhD beyond just the last four months of writing, but that would discount the nearly 5 years I have spent as a member of LMS. Indulging myself slightly by looking further back, I want to thank the person that set me on this path. Charlotte Rhead, my biology teacher who, when I said I might study physics at A level, refused to end the conversation until she had convinced me I would go on to study Biochemistry at University. 12 years later I am proud to submit my thesis.

A massive thank you to my best and oldest friends, Suzi, Megan, Sadie, and Piers, who have been on this ride with me throughout those 12 years and have become like family to me. Speaking of family, I want to thank my siblings, Duncan, Emily and Leo who have at times given encouragement, at others caused distress. I'm not sure which I found more disturbing. My parents, Sanchi and Michael, have put up with be becoming a perennial boomerang child, always returning home. You have supported me emotionally and with constant love, and read every word of this thesis even if it sounded like gobbledygook most of the time. You will say it is your job, but it is this love that has carried me through even the darkest of times.

There are many friends and communities to thank for making me laugh, giving me a shoulder to cry on and for keeping me sane over these past years. I must mention my netball teams who have provided the weekly distractions of games, and frequent pub trips. My 'North London' netball community and teams, Quiche, Guns & Throwsies, Knight and Unicorns who have provided countless memories and opportunities. I want to thank Pete, Dan, Kevin, Joe among many, many others from these teams for their friendship. Thank you to my old flatmates, Gabby and Chrissie who put up with me and my mess.

A special mention needs to go to my original London team Netsgronis, née Nets Go. To Emmi, Jake, Chris, Sana, Aimee, MC, Charlotte, Ash, Katie E and Katie P, you are all simply amazing. To Netsgronis final member, I dedicate this thesis to the memory of Oliver Burnham, you are forever missed.

I want to thank my therapist Helen and kinesiologist Anne who have both helped me get through mental and physical health issues.

A huge thank you to all the members of the LMS facilities who helped me with my experiments. Dr Dirk Dormann and Chad Whidling from microscopy, Dr George Young from bioinformatics and Dr Ivan Andrews from genomics. A special thanks to Dr James Elliot and Dr Bhavik Patel for the FACS facility for putting up with all my nonsense and somehow managing to deliver useful samples most weeks. Thanks to my office mates; Ferran, Diachi, Lucas, Xuan, Irina, Illinca, Paul, Steve, and Bryony as well as the wider LMS PhD and post-doc community. Thanks to Dr Alexis Barr and Dr Louise Fets for providing guidance at different times during my PhD. Thanks to Dr Enrique 'Fadri' Martinez-Perez for his support as post-graduate tutor, especially in helping me to manage my migraines and sick leave.

A thanks to all members of the GAP and R&CH groups, past and present. To Dr Lucy Watson, for bringing her enthusiastic, ever curious and kind attitude to the Lab. To Dr Lessly Sepulveda Rincon, my first friend in the LMS, partner in many crimes and sage advisor - when I was prepared to listen.

Finally, to my two supervisors, Professor Petra Hajkova and Dr Camille Dion, who both stepped up into this role under stressful and difficult circumstances. Petra, thank you for guiding me in this last year, giving up your time to help steer this project and reading this thesis in a condensed time.

Camille, I do not have the words to ever express the debt of gratitude I have to you. Even before you become my official supervisor, you taught me everything I know about human cell culture. You have always looked out for me, fought for me, and showed that you cared about me and my project, even at times when I was struggling to deal with it. Our relationship has not always been easy, but I feel we have grown together and formed a formidable team. It has been a genuine pleasure to discuss my project and this thesis over the last four and a half years. I am forever grateful for all you have done for me; I do not think I would have made it this far without you.

Contents

List of Figures

List of Tables

List of abbreviations

2i	2 inhibitor medium
36B4	Acidic ribosomal phosphoprotein P0
4i	4 inhibitor medium
5i/L	5 inhibitor medium + LIF
Abs	Antibodies
AKT	Protein Kinase 3
ALK2	Activin receptor-like kinase-2
Ap2γ	Protein produced by TFAP2C gene
aRB27	RPMI + 1% B27 supplement
BCA	Bicinchoninic acid
bFGF	Basic Fibroblast growth factor
BLIMP1	B lymphocyte-induced maturation protein-1
BMP	Bone morphogenic protein
BMP2	Bone morphogenic protein 2
BMP4	Bone morphogenic protein 4
BMP8b	Bone morphogenic protein 8b
BRAF	proto-oncogene B-raf
BSA	Bovine serum albumin
BTAG	BLIMP1-tdT TFAP2C-eGFP reporter cell line
c-Myc	c-MYC proto-oncogene, bHLH transcription factor
c#	Clone number e.g. clone 1 (c1)
CAG	Cytomegalovirus early enhancer element and chicken beta-actin promoter
cDNA	Complimentary DNA
CDX2	Caudal type homeobox transcription factor 2
CHD1	Chromodomain Helicase DNA Binding Protein 1
CHIR	Chir99021
CmR	Chloramphenicol resistance gene
CMV	Cytomegalovirus

CRIPSR	Clustered Regularly Interspaced Short Palindromic Repeats
d#	Day dox was added to 3D aggregate/2D culture e.g day 1 (d1)
DAPI	4',6-diamidino-2-phenylindole,
DAZL	Deleted In Azoospermia Like
DNMT3a	DNA methyltransferase 3 alpha
DMNT3B	DNA methyltransferase 3 beta
dox	Doxycycline
DTT	Dithiothreitol
dx	No dox added to aggregate/2D culture
E	Embryonic day
E+ I-	EPCAM positive INTERGRIN alpha 6 negative population
E+ I+	EPCAM positive INTERGRIN alpha 6 positive population
E8	Essential 8 medium
EC	Embryonic Carcinoma
EG	Embryonic germ cell
EGF	Epidermal growth factor
eGFP	Enhanced green fluorescent protein
EOMES	Eomesodermin
EpiLCs	Epiblast-like cells
ERK	Extracellular signal-regulated kinases
ESC	Embryonic stem cell
ESSRB	Estrogen related receptor beta
EV	Empty vector
ExE	Extraembryonic ectoderm
FACS	Fluorescence-activated cell sorting
FCS	Foetal calf serum
FGF	Fibroblast growth factor
FGF2	Fibroblast growth factor 2
FGF4	Fibroblast growth factor 4
g	Force of gravity
G418	Geneticin

GATA2	GATA-binding factor 2
GATA3	GATA-binding factor 3
GATA4	GATA-binding factor 4
GATA6	GATA-binding factor 6
GBX2	Gastrulation brain homeobox 2
GCT	Germ cell tumour
GK15	GMEM with 15% KSR medium
GMEM	Glasgow's minimal essential medium
GSK-3	Glycogen synthase kinase 3
H3K4Me3	Histone 3 lysine 4 tri-methylation
H3K27Me3	Histone 3 lysine 27 tri-methylation
H3K36Me2	Histone 3 lysine 36 di-methylation
HDAC	Histone deacetylases
hEG	Human embryonic germ cells
hEGCLC	Human embryonic germ cell-like cells
hEpiLC	Human EpiLCs
hESC	Human ESCs
Hox	Homeobox genes
hPGC	Human PGC
hPGCLC	Human PGCLCs
hPSC	Human PSC
HRP	Horseradish peroxidase
iMeLCs	Incipient mesoderm-like cells
iPSC	Induced pluripotent stem cells
IRES	Internal ribosome entry site
JAK	Janus kinase
JNK	Mitogen-activated protein kinase 8
kD	Kilo Dalton
KDM2B	Lysine demethylase 2B
KLF	Krüpple-like factor
KLF2	Krüpple-like factor 2

KLF4	Krüpple-like factor 4
KLF5	Krüpple-like factor 5
KO	Knock-out
KSR	Knock-out replacement serum
LIF	Leukaemia inhibitory factor
MAP	Mitogen activated protein kinase
mEG	Mouse EG
MEK	Mitogen activated protein kinase kinase
mEpiLCs	Mouse EpiLCs
mEpiSCs	Mouse EpiSCs
mESC	Mouse ESCs
MIXL1	Mix paired-like homeobox 1
mPGC	Mouse PGC
mPGCLC	Mouse PGCLCs
mV	mVenus fluorescent reporter
NANOS1	Nanos C2HC-type zinc finger 1
NANOS3	Nanos C2HC-type zinc finger 3
NLS	Nuclear localisation signal
OCT	Optimal cutting temperature compound
OCT4	Octamer-binding transcription factor 4
OE	Overexpression
OTX2	Orthodenticle homeobox 2
p38	Mitogen-activated protein kinase
PB 3'TR	Piggybac 3 prime terminal repeat
PB 5'TR	Piggybac 5 prime terminal repeat
PBS-T	Phosphate buffered saline + 0.1% Tween-20
PE	Primitive endoderm
PFA	Para-formaldehyde
PI3K	Phosphoinositide 3-kinase
PIWIL1	Piwi like RNA-mediated gene silencing 1
PKC	Protein kinase C

POE	Prolonged overexpression
PRDM1	PR domain zinc finger protein 1 (also known as BLIMP1)
PRDM14	PR domain zinc finger protein 14
PRMT5	Protein arginine methyltransferase 5
PRMT8	Protein arginine methyltransferase 8
PSC	Pluripotent stem cell
qPCR	Quantative polymerase chain reaction
RA	Retinoic acid
RE	Relative expression
REX-1	Zinc finger protein 42
RIPA	Radioimmunoprecipitation assay buffer
ROCK	Rho-associated kinase
ROCKi	Rho-associated kinase inhibitor
RPMI	Roswell Park memorial institute medium
RT-qPCR	Revers-transcriptase quantitative polymerase chain reaction
rtTAM2	Reverse tetracycline-transactivator
SALL4	Spalt like transcription factor 4
SCF	Stem cell factor
SEM	Seminoma tumour
SOX1	SRY-related homeobox gene 1
SOX2	SRY-related homeobox gene 2
SOX3	SRY-related homeobox gene 3
SOX15	SRY-related homeobox gene 15
SOX17	SRY-related homeobox gene 17
SRC	Proto-oncogene tyrosine-protein kinase SRC
Si-RNA	Small interfering RNA
ssRNA-seq	Single cell RNA sequencing
STAT3	Signal transducer and activator of transcription 3
STO	Sandos inbred mice 6-thioguanine-resistant, oubain-resistant
SYCP3	Synaptonemal complex protein 3
T	Brachyury

t2iLGö	Titrated 2 inhibitor +LIF +Gö 6983
TBX3	T-box transcription factor 3
TCF3	Transcription factor 3
tdT	tdTomato fluorescent protein
TE	Trophectoderm
TET1	Ten-eleven translocation methylcytosine dioxygenase
TF	Transcription factor
TFAP2C	Transcription factor AP-2 gamma
TFCP2L1	Transcription factor CP2 like 1
TGF	Transforming growth factor
TNAP	Tissue non-specific alkaline phosphatase
TRE	Tetracycline response element
UTF1	Undifferentiated embryonic cell transcription factor 1
UTX	Ubiquitously transcribed tetratricopeptide repeat gene on the X chromosome
VASA	DEAD-box helicase 4
Wnt	Wingless-related integration site
WT	Wildtype

Chapter 1. Introduction

The germline of mammalian species requires the expression of pluripotency factors (Leitch and Smith, 2013; Reik and Surani, 2015). The role these pluripotency factors play in mouse germline development are well understood, but their role in human germline development and the differences between species are less understood. In this chapter I will describe pluripotency in the context of the early embryo and *in vitro* pluripotent stem cells, and how culture conditions support the expression of the core pluripotency factors. The expression of these core pluripotency factors in early human germline cells, called primordial germ cells (hPGCs), has helped us to study the specification of the human germline *in vivo*. However, most of our understanding of nascently specified hPGCs has come from the *in vitro* cell model, hPGC-Like cells (hPGCLCs). I will detail how this system was established and the gene networks which govern this early germline state. I wanted to further the understanding of the roles different pluripotency factors play in the human germline by challenging an *in vitro* hPGCLCs, with overexpression of three pluripotency factors; *SOX2, NANOG* and *KLF2*. Finally, I will set out my hypothesis for what effects the overexpression of *SOX2*, *NANOG* and *KLF2* will have on early human germline state.

1.1 Mammalian pluripotent stem cells represent an *in vitro* form of the early embryo state

Pluripotency refers to the ability of a cell to form any cell of the embryo proper and ultimately the adult body (Bradley et al., 1984). Another feature of *in vitro* derived pluripotent stem cells (PSCs) is that they show limitless self-renewal (Nichols et al., 2001). The most common methods for deriving PSC are based on two sources, culturing of early embryonic tissue culture to drive embryonic stem cells (ESCs) or through somatic cells reprogramming to derive induced pluripotent stem cells (iPSCs).

The culture conditions which allowed the pluripotency to be captured from embryonic tissue *in vitro* were pioneered in mouse (Evans and Kaufman, 1981). Outgrowths from the inner cell mass of pre-implantation embryo when cultured on STO feeders with 10% foetal calf serum (FCS) gave rise to ESCs. More specifically, the *in vivo* epiblast gives rise to ESCs when correctly

cultured in vitro (Gardner and Beddington, 1988). The segregation of the embryo into the inner cell mass and the trophectoderm (TE) is followed by the separation of the inner cell mass into epiblast and primitive endoderm (PE) (Ralston and Rossant, 2005). TE and PE cells linages do not contribute to the embryo proper (Rossant, 1987), but do form unique cell lines *in vitro* (Ralston and Rossant, 2005), which are not considered pluripotent (Kunath et al., 2005; Tanaka et al., 1998). These results confirmed that the epiblast is the sole source of ESC from the embryo. ESCs demonstrated pluripotency as they formed teratomas, tumours that contain cells from all three germ layers (indicating they are pluripotent), when injected into the flank of mouse and differentiated in embryo bodies *in vitro* (Evans and Kaufman, 1981)

1.1.1 Pluripotent states in mouse; naïve, formative, and primed

It is important to note that pluripotency is not a single state but rather a continuum with three distinct metastable states; naïve, formative, and primed (Morgani et al., 2017; Nichols and Smith, 2009; Smith, 2017). These states have mainly been defined by cell culture systems but correspond to *in vivo* development (Nichols and Smith, 2009). Below, the mechanism that underlies all pluripotent states will be described, before the specific differences between these three states and what they correspond to *in vivo* are discussed. All three states rely on the expression of the core pluripotency factors; *SOX2, OCT4* and *NANOG* to maintain their limitless self-renewal and pluripotent capacity.

1.1.1.1 Naïve; the ground state of pluripotency.

Naïve pluripotency is the most epigenetically primitive state, with low DNA methylation levels and few repressive histone marks (Takahashi et al., 2018), indicating cells in this state are not developmentally specified. The naïve state characterises mouse ESCs, iPSCs, and the preimplantation epiblast (Nichols and Smith, 2009).

Mouse embryonic stem cells (mESCs) were the first to be derived, and they remained the main model of studying pluripotency. Therefore, a number of the discoveries made in this system actually characterise the naïve state, rather than pluripotency in general. Firstly, the

development of a feeder-free culture helped to define the signalling networks required for maintenance of the naïve state. Analysis of the secretome produced by feeders led to the discovery of leukaemia inhibitory factor (LIF), allowing for mESCs to be cultured in serum/LIF, without feeders (Smith et al., 1988; Smith and Hooper, 1987).

The 129 mouse strain, also the original source of teratomas (Stevens, 1958), is a mouse strain that shows higher conversion into mESCs compared to other strains, and is hence termed permissive to mESCs derivation (Kawase et al., 1994). The breakthrough that allowed high conversion of ESCs from multiple mouse lines and other rodents species was 2i culture. GSK-3 inhibitor CHIR99021 (CHIR) and ERK inhibitor PD0325901 are used in combination to sustain ESC identity without growth factors (Ying et al., 2008). The addition of LIF to 2i culture allowed for robust derivation of ESCs from non-competent lines (Kiyonari et al., 2010; Nichols et al., 2009) and other rodent species such as rat (Buehr et al., 2008). Pluripotency factors *Nanog* and *Rex-1* are expressed heterogeneously in serum/LIF cultures but are uniformly expressed higher in 2i and 2i/LIF, suggesting 2i is better at maintaining robust pluripotency (Wray et al., 2010).

1.1.1.2 Primed pluripotency, a cell state ready to enter lineage development.

In mice, derivation of epiblast derived stem cells (mEpiSCs), from the post-implantation embryos (E5.5), (Brons et al., 2007; Tesar et al., 2007), demonstrated a second metastable form of pluripotency existed. These primed cells could undergo differentiation into the three germ layers, indicating they were pluripotent but showed flatter morphology and did use LIF signalling to maintain this cell state (Brons et al., 2007; Tesar et al., 2007). A significant difference in between naïve and primed states is that mEpiSCs cannot integrate into pre-implantation embryos (Tesar et al., 2007). These cells can be incorporated into post-implantation embryos between E6.5 and E8.5, but show no germline transmission unlike naïve mESCs (Huang et al., 2012).

Primed cells are considered to be developmentally downstream of naïve cells, representing a phase just before cells commit to a lineage from the three germ layers (Martello and Smith,

2014; Nichols and Smith, 2009). Epigenetic repressors, such as DNA methylation, Polycombe and mRNA methylation are integral for maintaining the primed state, but their removal enhances the naïve state (Geula et al., 2015; Weinberger et al., 2016). In naïve cells, the action of transcription factors (TF) maintain the state by repressing differentiation genes (Silva and Smith, 2008) while epigenetic regulators trigger the differentiation of this state by silencing pluripotency genes. In primed cells, these regulators are required to block differentiation queues (Weinberger et al., 2016). Primed PSCs express fewer naïve genes, such as *Klf4* and *Essrb* which support the pluripotency network (Daman, 2016; Weinberger et al., 2016), instead expressing lineage genes, such as *OTX2*, priming them for differentiation into germ layers with quicker dynamics than naïve cells (Tsakiridis et al., 2014)

1.1.1.3 Formative pluripotency; an in-between state and the origin of the germline.

Formative pluripotency was only proposed as a hypothesis in 2017 (Smith, 2017). In this essay, Smith proposes that two phases, naïve and primed, are not sufficient to describe the behaviour of intermediate cells, which show unique features. Formative pluripotency was proposed to encompass the events which occur as both naïve mPSCs and the epiblast *in vivo* begin to lose markers of the naïve state and upregulate post-implantation genes but not lineage specification genes. An important feature of this state was the competence to enter the germline. While naïve and primed cells can undergo differentiation into the three germ layers of the embryo, cells that exist in the formative state can only undergo germline specification. This is well defined in mouse as Epi-like stem cells (EpiLCs), a transient cell type generated by stimulating mPSCs with primed conditions, (Hayashi et al., 2011) and specific stages of the mouse epiblast (Ohinata et al., 2009), can enter the germline.

A defined culture method for deriving formative cells both from E5.5 epiblast or naïve mPSCs utilises low nodal signalling using Activin A, Wnt inhibition through XAV939 and inhibition of retinoic acid (RA) signalling (Kinoshita et al., 2021b). The same combination of signalling molecules and inhibitors could also derive formative stem cells from naïve human cells and epiblasts (Kinoshita et al., 2021b).

1.1.2 Derivation of human pluripotent stem cells

Human ESCs (hESCs) were first derived from surplus embryos from IVF (Marshall, 1998). When derived on feeders with FCS, hESCs colonies are flat compared to domed mESCs (Evans and Kaufman, 1981; Thomson et al., 1998). A significant difference between mESCs and hESCs recognised in the first studies on hESCs, was that LIF could not mediate feeder-free culture (Reubinoff et al., 2000b; Thomson et al., 1998). mEpiSCs show similarities to hPSCs (Rossant, 2007) in gene expression and signalling requirements, discussed in 1.1.3.2. This lead to the idea that hPSC and mPSC showed differences in pluripotency rather than species differences (Nichols and Smith, 2009).

1.1.2.1 Resetting primed hPSCs to naïve.

While primed hPSCs are widely used in the field, it is possible to reset these primed PSCs to a naïve state through various methods such as transient overexpression of TFs *NANOG* and *KLF2* (Takashima et al., 2014) or by inhibiting histone deacetylase (HDAC) (Guo et al., 2017). Resetting of hPSC into naïve state was first achieved by overexpressing *KLF2* and *NANOG* (Takashima et al., 2014). Once resetted into a naïve state, hPSCs have slightly altered culture requirements to mouse. 2i/LIF forms the basis but, to liberate these cells from continuous *KLF2* and NANOG overexpression, a lower concentration of CHIR, a GSK-3 inhibitor (1μm compared to 3μM) is used and PKC inhibitor Gö6983 is added, with feeders or pre-coating laminin or Matrigel or Geltrex. This media was termed t2iLGö (Takashima et al., 2014) and it allowed derivation of naïve human ESCs from day 6 postfertilization epiblast (Guo et al., 2016) or using epigenetic resetting through inhibition of HDAC (Guo et al., 2017). Naïve genes *KLF4, KLF5* and *TBX3* among others are upregulated in resetted and naïve epiblast derived cells, as well as a high expression of *NANOG* (Guo et al., 2017; Guo et al., 2016; Takashima et al., 2014).

A second media, termed as 5i/L, was developed by screening signalling molecules which could maintain naïve cells after *NANOG* and *KLF2* overexpression was withdrawn. Inhibiting

GSK3, MEK, BRAF, SRC and ROCK and adding LIF, resetted cells could survive without the requirement of feeders (Theunissen et al., 2014).

The addition of Activin A, FGF to this 5i/L media in feeder culture allows for the generation of naïve hPSCs from primed hPSCs without TF overexpression. This resetting was defined by the cells using the distal enhancer for *OCT4*, a key feature of naïve pluripotency (Yeom et al., 1996). The lack of a TF overexpression or epigenetic inhibition during this process in generating naïve cells, could suggest that the generation of these 'naïve cells' is through biasing survival of the more naïve-like cells in the primed culture. Various signalling cascades will also be triggered or blocked by the 5i/L, which could lead to the emergence and expansion of naïve-like cells which are not fully reset. These 5i/L cells show inactive X chromosome and higher expression of *DNMT3a* and lower of *TET1* which are not features of the naïve cells generated with the TF-based system of resetting (Takashima et al., 2014), suggesting they might not be bona fide naïve cells.

Beyond culture conditions, there are some differences between human and mouse naïve cells. Mouse naïve cells in 2i require *Klf2* expression (Yeo et al., 2014), whereas human cells require *KLF4* and *TFCP2L1* (Takashima et al., 2014). Another naïve factor *ESSRB* can rescue *Nanog* null-mESCs and mediates *Nanog* over-expressions signalling in LIF independence (Festuccia et al., 2012), but is not upregulated in human naïve cells.

All three states are pluripotent as they can form the three-germ cell layers and self-renew in culture indefinitely. Despite differences between these three PSC states, they require to maintain the expression of the core pluripotency regulators, *OCT4, SOX2* and *NANOG*, although they rely on different signalling pathways to maintain these.

1.1.3 *OCT4, SOX2* and *NANOG*, the core transcription factor regulators of pluripotency

Pluripotency is regulated by the expression of TFs, which maintain this cellular state and block differentiation through coordinating a pattern of gene expression. Their combined expression alone can be powerful enough to reset differentiated cells into a pluripotent state (Takahashi and Yamanaka, 2006). Several TFs are associated with the pluripotency state but two are considered core: *SOX2* and *OCT4.* These TFs dimerise when co-expressed and direct the gene expression pattern of pluripotent cells (Boyer et al., 2005; Masui et al., 2007). *SOX2* and *OCT4* are supported by other TFs such as *NANOG* (Boyer et al., 2005; Loh et al., 2006; Pan and Thomson, 2007)and the *KLF* family (Bourillot and Savatier, 2010), which enhance pluripotency but are not required for the maintenance of pluripotency *in vitro* (Chambers et al., 2007; Jiang et al., 2008).

Intensive studies of mESCs paved the way for the development of reprogramming of somatic cells to a pluripotent state, called induced pluripotent stem cells (iPSC) (Takahashi and Yamanaka, 2006). Reprogramming involves the forced overexpression of four pluripotency factors; *OCT4*, *SOX2*, *KLF4* and c-Myc, which act together to drive expression of the endogenous pluripotency factor network (Takahashi and Yamanaka, 2006).

Embryonic and induced PSC have been generated from multiple species such as rat (Buehr et al., 2008; Kawamata and Ochiya, 2010; Li et al., 2008; Liao et al., 2009), cynomolgus monkey (Onozato et al., 2018; Suemori et al., 2001), pig (Choi et al., 2019; Fukuda et al., 2017) and human (Takahashi et al., 2007; Thomson et al., 1998) among many others, and all show limitless self-renewal and the ability to enter all three germ layers. Our ability to derive PSC from both embryonic and somatic sources in various species demonstrates that pluripotency is a not unique to human or mouse. Instead, pluripotency is a wider property of mammals, perhaps universal to mammalian development. While the ability to derive a PSC is proof that a species shows pluripotency, it is possible that other species whose PSCs have not yet been derived do show pluripotency.

1.1.3.1 *OCT4* is foundational for pluripotency.

The importance of Octamer-binding transcription factor 4 (*Oct4*) in early development and pluripotency is underlined in knock-out (KO) mouse embryos, which have no discernible inner cell mass, and hence no pluripotent epiblast (Nichols et al., 1998). Generation of *Oct4* KO mESCs is not possible from these embryos (Nichols et al., 1998). *Oct4* KO in established mESCs leads to a loss of pluripotency and develop into trophectoderm (Niwa et al., 2000). *In vivo,* the role of *Oct4* is to repress TE fate by antagonising the TE transcription factor *Cdx2* (Strumpf et al., 2005) and this antagonism can be mimicked in mESCs (Niwa et al., 2005).

The advent of CRIPSR technology (Jinek et al., 2012) and the granting of permission to use it on human embryos (The Francis Crick Institute, 2016) has provided evidence that *OCT4* is crucial for early stages in the human embryo, as *OCT4* KO embryos failed to develop (Fogarty et al., 2017). Knock-down of *OCT4* in human ESCs leads to dysregulation of the pluripotency network and the upregulation of either primitive endoderm markers after brief (<48 hours) knock-down (Hay et al., 2004) or trophoblast and mesoderm markers during prolonged (>48 hours) knock-down (Zafarana et al., 2009). Overall, these studies showed that the role of *OCT4* is conserved between human and mouse, it maintains pluripotency and blocks entrance into extraembryonic fate.

1.1.3.2 SOX2 is similarly important in establishing and maintaining pluripotency.

Sry-related HMG-box 2 (*SOX2*) is crucial for pluripotency and early embryonic development, as well as being involved in later lineage development. *In vitro*, it has a critical role in somatic reprogramming (Takahashi and Yamanaka, 2006) and in forming epiblasts in mouse embryos (Avilion et al., 2003). In mouse, *Sox2* KO causes embryonic lethality after implantation, while outgrowths of these embryos develop into trophoblast linage rather than epiblast (Avilion et al., 2003). An identical result is seen when *Sox2* is deleted in mESCs, where trophectoderm lineage markers are upregulated. (Masui et al., 2007). In human, knock-down of *SOX2* in hESCs mainly triggers the expression of trophectoderm genes (Adachi et al., 2010; Fong et al., 2008). Unlike for *OCT4*, human embryos have not yet been edited to assess the role of *SOX2*.

The role of *SOX2* therefore seems to be maintaining the pluripotent identity and suppressing the trophectoderm fate.

1.1.3.3 *OCT4* and a *SOX* factor must form a heterodimer to maintain PSCs.

SOX2 and OCT4 are part of the core four reprogramming factors to generate induced pluripotent stem cells from somatic cells, both in humans and mice (Takahashi et al., 2007; Takahashi and Yamanaka, 2006). Co-binding between these two factors into a heterodimer at UTF1-like and FGF4-like motifs is essential for maintenance of pluripotency (Tapia et al., 2015). *Oct4* overexpression can rescue *Sox2* knock-out mESCs, as a major role for S*ox2* in pluripotency is to maintain *Oct4* expression and repress repression of *Oct4*, with other *Sox* factors able to act redundantly at *Oct-Sox* motifs (Masui et al., 2007).

SOX2 overexpression might be expected to drive more a robust pluripotent state resistant to differentiation, by increasing the expression of its targets such as *OCT4* and *NANOG*, but instead, it causes differentiation. In mouse, *Sox2* overexpression triggers the upregulation of genes associated with multiple lineages: ectoderm, mesoderm and extraembryonic tissue (Kopp et al., 2008). There is also evidence that forcing overexpression of *Sox2* may downregulate its endogenous locus, suggesting its expression is under a negative feedback loop (Kopp et al., 2008). Genes targeted by Oct4:Sox2 dimer are also downregulated despite the total Sox2 expression being higher (Kopp et al., 2008).

Human pluripotent stem cells also show this reduction in *OCT4* and *NANOG* expression, when *SOX2* is overexpressed (Adachi et al., 2010). In addition, trophectodermal markers such as CDX2 and Cytokeratin 8 can be identified by immunohistochemistry when *SOX2* is overexpressed, but unlike in mouse, other lineage markers are not detectable (Adachi et al., 2010). In summary, the expression of *SOX2* is tightly regulated to maintain pluripotency without triggering differentiation.

Molecularly, the role of these two TFs, *SOX2* and *OCT4,* is to regulate transcription of a group of other TFs, termed the pluripotency network, along with repressing differentiation genes (Boyer et al., 2005; Chen et al., 2008).

1.1.3.4 *NANOG* supports entry into pluripotency but is dispensable for its maintenance.

NANOG is an interesting pluripotency TF; while its overexpression can directly regulate pluripotency, its expression is not required to maintain this state. It has been identified in mouse ESCs as a factor which could sustain ESCs in undifferentiated state without LIF addition (Chambers et al., 2003; Mitsui et al., 2003). LIF and its signalling are discussed below. In reprogramming experiments, exogenous *Nanog* is dispensable for generating miPSCs (Takahashi and Yamanaka, 2006), but its overexpression can reset the primed state of mouse pluripotent cells, mEpiSCs, back to naïve state (Silva et al., 2009). Furthermore the presence of a *Nanog* allele is required for the final transition of pre-iPSCs in serum/LIF into 2i/LIF and for the transition into the naïve epiblast *in vivo* (Silva et al*., 2009)*.

Overexpression of *NANOG* in hESCs allowed for the removal of feeders and feeder conditioned media (Darr et al., 2006). Genes associated with primitive ectoderm are upregulated in *NANOG* overexpressing cells and endogenous *NANOG* is upregulated during differentiation of embryo bodies from human pluripotent cells (Darr et al., 2006). As noted in Darr et al., 2006, primitive ectoderm genes that are upregulated in human culture may be absent from mouse culture due to the presence of LIF which blocks primitive ectoderm (Shen and Leder, 1992). Knock-down of *NANOG* in human pluripotent cells causes differentiation. The marker genes upregulated in response to *NANOG* knock-down are associated with extraembryonic endodermal and trophectoderm lineages (Hyslop et al., 2005; Zaehres et al., 2005), suggesting that *NANOG* is involved in regulating and suppressing these cell fate decisions.

Triggering of differentiation in response to overexpression is a common feature amongst the pluripotency TFs. *OCT4* overexpression causes endodermal differentiation (Rodriguez et al., 2007), while *NANOG* overexpression directs cells towards primitive ectoderm (Darr et al.,

2006). While all these genes are crucial for entering and maintaining pluripotency, they also play roles in lineage development, such as *OCT4* for endoderm (Aksoy et al., 2013), often with alternative binding partners. The balance of expression altering their balance in pluripotent cells can lead to differentiation.

1.1.3.5 Auxiliary factors; the KLF family in the context of pluripotency.

The 17 members of the Krüpple-like factor (KLF) family play various roles in development and lineage specification. *Klf2, Klf4* and *Klf5* have redundant roles in supporting mESC pluripotency, as only the loss of all three results in differentiation of mESCs (Jiang et al., 2008), with ectoderm markers being upregulated in response. *Klf4* and *Klf5* are upregulated directly by LIF signalling (Niwa et al., 2009), while *Oct4* activates *Klf2* (Hall et al., 2009). These three *Klf*s then activate the core pluripotency TFs *Oct4, Sox2* and *Nanog* to support pluripotency in mPSCs (Hall et al., 2009).

Despite this redundancy, *Klf2* KO causes embryonic lethality by E14.5 in mouse (Kuo et al., 1997). This is due to haemorrhages in both the intra-amniotic and intraembryonic tissues showing *Klf2* has also a role during later developmental stages (Kuo et al., 1997). *Klf2* overexpression can act in similar ways to *Nanog* (Hall et al., 2009) and reset mEpiSCs as well as sustaining mESCs without LIF. Despite it being dispensable for pluripotency *in vivo*, *Klf2* is required for mESCs survival when LIF is removed in 2i conditions, as *Klf2* may repress PGC genes in the absences of LIF, preventing destabilisation of the pluripotent state (Yeo et al., 2014).

Human blastocysts do not express *KLF2*, but instead express *KLF4* and *KLF5* (Blakeley et al., 2015). Therefore, KLF2 appears not to have a role in early human development. However, its combination with *NANOG* overexpression allows to reset the primed hiPSC state into a naïve one (Takashima et al., 2014).

The connection between these four factors when active in a stem cell is displayed in Figure *1*

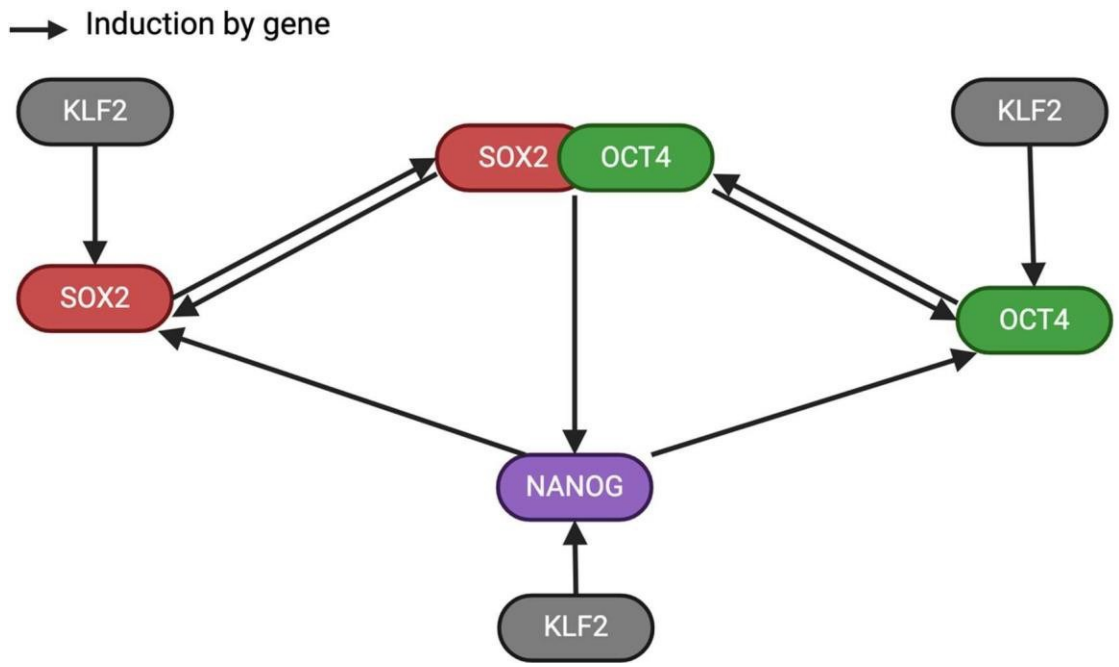

Figure 1. Connection between SOX2, OCT4, NANOG and KLF2 in pluripotency

KLF2 is present in mouse, but it can be replaced by other KLF2 factors in Mouse and Humans

1.1.3 Culture conditions maintain the core pluripotency network.

1.1.3.1 Signalling process in 2i/LIF that support pluripotency network.

The inhibitors in 2i, which target GSK-3 and MEK, work to sustain the pluripotency network and block differentiation which is inherent to mPSCs. GSK-3 blocks the canonical Wnt pathway and its inhibition leads to β-catenin moving into the nucleus in mESCs (Doble et al., 2007). One of the downstream β-catenin effectors is *Tcf3* (Wray et al., 2011), which normally leads to induction of mesoderm or endoderm. However, β-catenin can also bind to Oct4 and increase its activity in a mechanism that is not dependent on its function as a TF (Kelly et al., 2011). These dual outcomes that result from GSK-3 inhibition stabilise the pluripotent potential of mPSC (Kelly et al., 2011). MEK repression is key to maintaining pluripotency as one of the actions of Oct4-Sox2 is to express FGF4 (Yuan et al., 1995) which acts through MEK to trigger differentiation (Kunath et al., 2007). LIF acts through the Jak/Stat pathway (Niwa et

al., 1998) with STAT3 being the key TF to maintain ESC identity (Boeuf et al., 1997). STAT3 drives expression of naïve factors such as; *Klf4, Gbx2* and *Tfcp2I1*, with this last factor being critical for the action of Stat3 for ESC maintenance. (Martello et al., 2013).

With a well-defined culture system and the signalling pathways identified; JAK/STAT3 activation, β-catenin entrance into the nucleus and the blocking of MEK/ERK auto signalling, a computation model for the network of pluripotency factors that maintain this state could be devised (Dunn et al., 2014). This model shows that *Klf4, Nanog, Tfcp2I1, Essrb* and *Gbx2* are all directly upregulated by STAT3 and ß-catenin pathways, these TFs then lead to expression of further TFs such as *Sall4* and *Klf2. Sall4, Tbx2* and *Nanog* positively regulate *Sox2*, while *Klf2, Nanog* and *Tcf3* positively regulate *Oct4*. This provides our mechanistic understanding of the role and importance of these transcription factors and the signalling molecules and pathways which modulate their expression.

1.1.3.2 FGF, nodal and insulin signalling support human primed pluripotency.

2i is unable to support human primed culture, so studies have elucidated the alternative signalling pathways which support human PSC culture.

Human ESCs are supported through FGF and TGF/Nodal/activin A signalling (Vallier et al., 2005). SMAD2/3, downstream effectors of the Nodal/Activin A pathway are phosphorylated and translocated into the nucleus (James et al., 2005). SMAD2/3 directly activates the pluripotency genes such as NANOG to ensure self-renewal and prevents differentiation which is normally triggered through these SMADs (Singh et al., 2012; Vallier et al., 2009; Xu et al., 2008).

FGF works through a number of pathways including PI3K/AKT which leads to GSK-3 phosphorylation (Eiselleova et al., 2009), leading to canonical Wnt signalling through ß-Catenin (Ding et al., 2010). *SOX2* is directly stimulated by AKT signalling, stimulated by FGF and endogenously expressed *PRMT8* in hESCs (Jeong et al., 2017). FGF signalling through

MEK/ERK/MAP kinase cascade has been shown to be required for the maintenance of pluripotency in hPSCs (Haghighi et al., 2018; Li et al., 2007).

Insulin growth factor receptor is essential for hPSC culture; its blocking or knock-down leads to a loss of self-renewal (Wang et al., 2007). FGF and insulin signalling appear to overlap in their triggering of the ERK/MEK/MAP kinase pathway (Eiselleova et al., 2009). Therefore most primed culture media contain insulin or insulin growth factor as well as FGF and some form of nodal signalling agonist (Dakhore et al., 2018). Together, these signalling pathways maintain proliferation and pluripotency in hPSCs (Mossahebi-Mohammadi et al., 2020) and are represented in Figure 2.

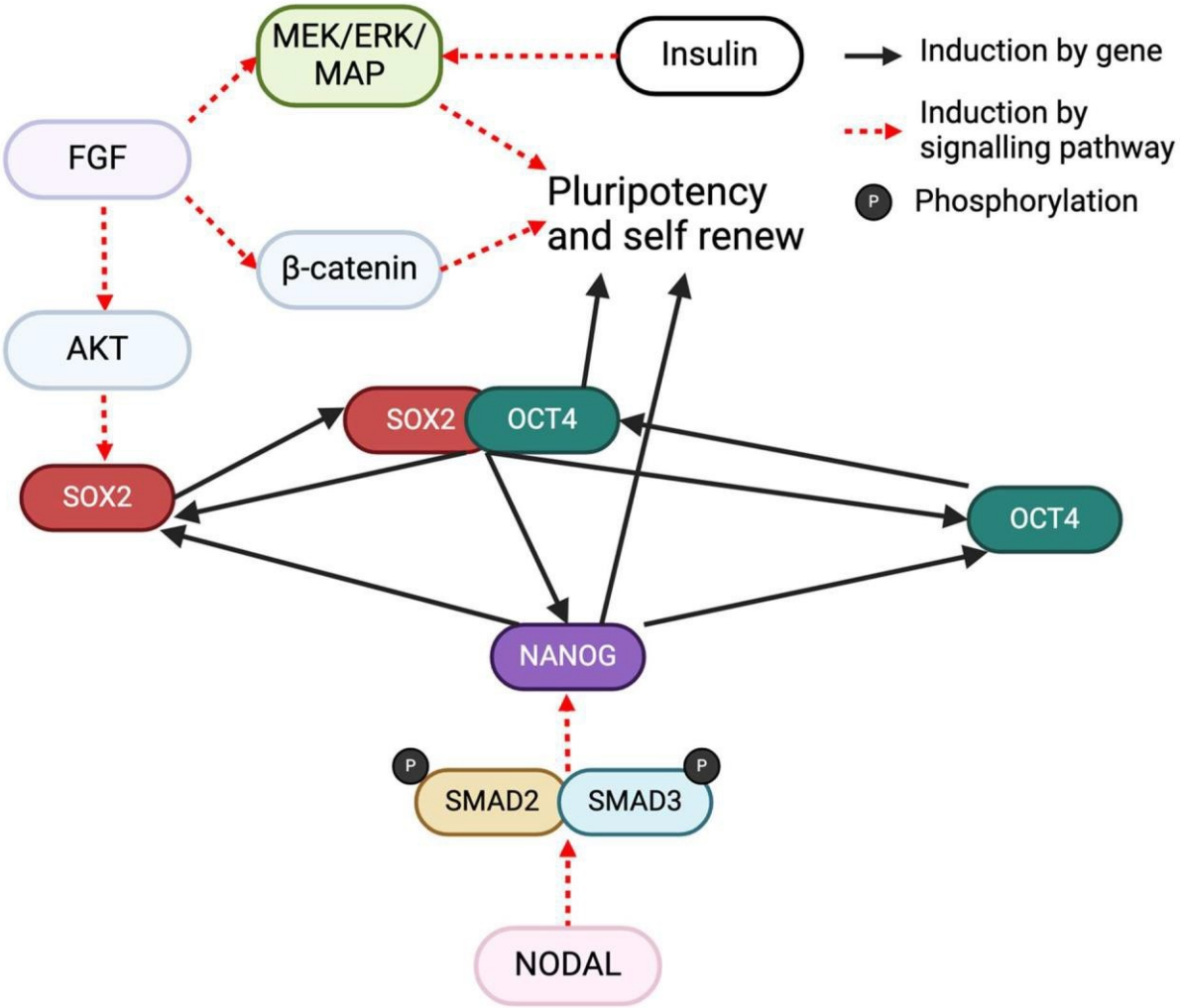

Figure 2 Action of Nodal, FGF and Insulin signalling to maintain primed pluripotency in humans

Of note, in this study I utilised a media termed Essential 8 (E8), which uses FGF2, TGFß and insulin, along with selenium, transferrin, L-ascorbic acid and $NaHCO_3$ (Chen et al., 2011) that maintains hiPSCs in the primed state.

1.2 Pluripotency and the germline; two interconnected cell states

Once the post-implantation embryo has differentiated into the different germ layers and extra-embryonic tissues, the ability to form pluripotent stem cells is lost (Brons et al., 2007). However, a subset of cells within the mouse embryo, the primordial germ cells (PGC), regain the ability to enter pluripotency (Matsui et al., 1992; Resnick et al., 1992). PGCs form the foundation of the germline, the cells that will eventually give rise to gametes (Surani, 2007). Pluripotency factors are important for the germline development, potentially by protecting the PGCs genome from aberrant transcription during the epigenetic reprogramming that occurs during germline development (Leitch and Smith, 2013).

1.2.1 Introduction to the germline and how it is studied.

The germline cycle begins when two haploid gametes from two individuals of the opposite sex fuse to form a single diploid zygote. This zygote develops into the post-implantation embryo, where PGCs are specified. These PGCs then migrate through the developing embryo, reaching the genital ridge where they colonise the gonads. Gonadal PGCs undergo extensive epigenetic and transcriptional changes as they initiate gametogenesis. The resulting gametes can then fuse to form a new embryo and so, propagating the germline cycle (Tang et al., 2016). This germline cycle is shown in Figure 3.

Direct studies of embryonic events in humans are by their nature difficult. Ethical and practical considerations mean access to biological material is limited (The International Society for Stem Cell Research, 2024). Furthermore, genetic studies involving whole organisms are not permitted and primary human embryo culture is limited to 14 days (Appleby and Bredenoord, 2018). In addition to primary embryo culture (Chen et al., 2019; Popovic et al., 2019), human germline development has been studied through analysis of

tissue donated from terminations (Gkountela et al., 2015; Guo et al., 2015; Tang et al., 2015) and through in vitro models of germ cell development derived from pluripotent stem cells culture (Irie et al., 2015; Sasaki et al., 2015). Mouse, non-rodent mammals, and non-human primates have all been studied in more detail with powerful genetic tools and used to further our understanding of human germline development (Tang et al., 2016).

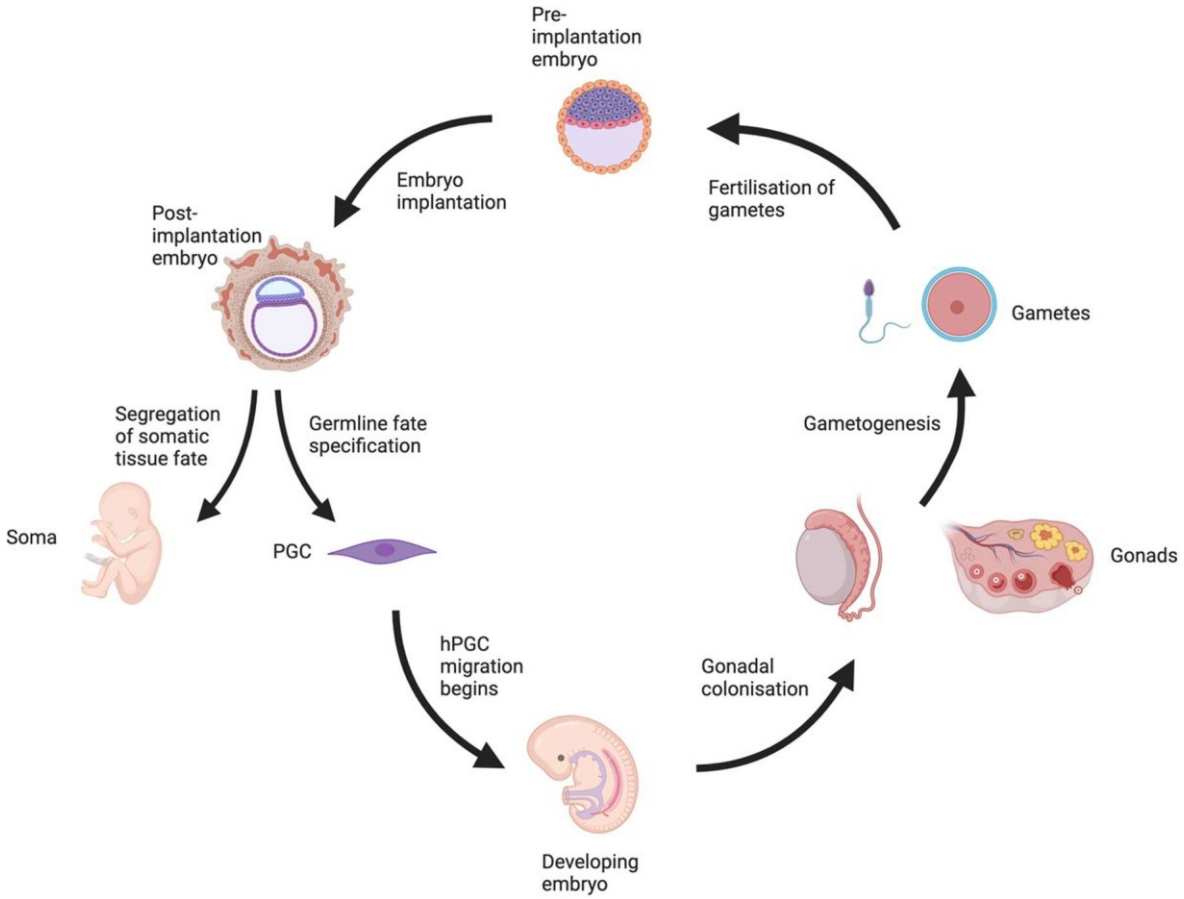

Figure 3 The germline cycle in human.

PGC specified in the post-implantation embryo migrate to the genital ridges, where they undergo gametogenesis to form new gametes. Gametes from two individuals can then fuse to form the pre-implantation embryo and begin the cycle again.

1.2.1.1 Attempts to dissect hPGC specification *in vivo* and *in vitro.*

The foundational cell type of the mammalian germline are the primordial germ cells (PGCs). In mice, PGCs are specified in the early embryo; at embryonic day (E)6.5 from the epiblast (Ohinata et al., 2005), at E11 from the amnion in cynomolgus monkeys (Sasaki et al., 2016)

and in humans, it is estimated to be between week 2-3 post-ferritization (Leitch et al., 2013c; Tang et al., 2016). The exact location and timing of PGC specification in humans is difficult to determine due to ethical considerations in relation to early embryo experiments. A study has identified the presence and location of hPGCs in an embryo between week 2 and 3 post conception (Tyser et al., 2021) within a population which resembles the primitive streak. Without the ability to image a number of embryos over a range of time points, the precise timing of human germline specification may never be understood.

An alternative approach is to use embryo attachment culture, which allows embryos to be cultured until a stage considered day 12 post fertilisation *ex vivo*. These cultures show evidence of hPGCs induction during this time frame. In one study at day 12 post-fertilisation, 2 out of the 26 embryos contained a small number of putative hPGCs that stained positive for human germline markers NANOG, SOX17 and TFAP2C (Chen et al., 2019). Similarly, a second study identified an average of five SOX17/OCT4 positive and GATA6 (an endoderm marker) negative hPGCs at day 12 post fertilisation (Popovic et al., 2019). Both studies identified these putative hPGCs within the NANOG positive epiblast of the embryo outgrowths. Attempts to stimulate PGCLC induction from these outgrowths was unsuccessful as it led the outgrowths to become non-viable (Popovic et al., 2019). This would suggest that hPGCs are specified from the epiblast at around E12 and later, although it is not possible to conclude from these studies if the timings would be different *in vivo*.

To gain a deeper understanding of the mammalian germline development, mouse has been used as the main model organism. The extensive genetic studies performed in mouse paved the way for the development of tools that allowed the derivation of an *in vitro* PGCs model. These primordial germ cell-like cells (PGCLC) (Hayashi et al., 2011) can be derived from PSCs from a number of species such as pig (Wang et al., 2016), cynomolgus monkey (Sakai et al., 2020) including human (Irie et al., 2015; Sasaki et al., 2015)

1.2.1.2 Signalling pathways and transcription factors critical to mouse germline development *in vivo.*

Genetic studies identified Bone morphogenic protein (BMP) signalling to be required for mPGC specification *in vivo*. Deletions of the signalling molecule Bmp4 resulted in mPGC being absent from the embryo (Lawson et al., 1999). Bmp8b deletion reduced mPGC number by half (Ying et al., 2000). Both Bmps are produced in the extraembryonic ectoderm (ExE) (Ying et al., 2000). Expression of the type 1 Bmp receptor Alk2 in the visceral endoderm is also required for mPGC development (de Sousa Lopes et al., 2004) leading to the hypothesis that BMP signalling from the ExE to the epiblast induces germline fate. Deletion of the downstream effectors of the Alk2 receptor, Smad1 (Ohinata et al., 2009; Tremblay et al., 2001) and Smad5 (Chang and Matzuk, 2001), or both (Arnold et al., 2006), results in either a large reduction in mPGC numbers or their total absence in these knock-out embryos. These studies elucidated that all components of the Bmp4 Smad1/5 pathway are involved in specifying mPGC *in vivo*.

BMP signalling specifies mPGCs *in vivo* by activating genes that establish and maintain germ cell identity (Kurimoto et al., 2008; Magnusdottir et al., 2013). Direct genetic analysis revealed Blimp1 is expressed in mPGCs, and its expression restricts the fate of Blimp1 positive cells to germline (Ohinata et al., 2005; Vincent et al., 2005). Blimp1 KO embryos fail to repress somatic Hox gene expression leading to a 6 fold reduction in mPGC numbers (Ohinata et al., 2005) or total lack of mPGCs (Vincent et al., 2005). Tagging Blimp1 helped define the mPGC population in the embryo, making it possible to analyse the transcriptome of mPGCs (Kurimoto et al., 2008). Among many genes to be expressed in mPGCs, the key determinants of germ-cell identity Tfap2c, which encodes Ap2γ, and Prdm14 were identified (Kurimoto et al., 2008) and their role in germ cell maintenance were confirmed through knock-out studies (Magnusdottir et al., 2013; Weber et al., 2010). These extensive *in vivo* studies unravelled the signalling pathways and genes that were essential for mPGC development; allowing for the development of an *in vitro* protocol which could then be translated to human cells.

1.2.2 Development and characterisation of primordial germ cell-like cell

1.2.2.1 Early attempts at derivation of mouse PGCLCs.

An *in vitro* cell type that resembled primordial germ cells was reported in hanging drop embryo body differentiation (Geijsen et al., 2004; Toyooka et al., 2003) or random differentiation of mESCs when growth factors are withdrawn (Hubner et al., 2003). These cells were rare, and the differentiation not directed by defined external signalling molecules. One study used BMP4 signalling to enhance the rate of production of these germ cell like cells (Toyooka et al., 2003). Oocytes (Hubner et al., 2003) and sperm (Toyooka et al., 2003) were able to be derived from these cells, although they were not fertilised into offspring. These early studies demonstrated that in vitro germ cells could be generated from mESCs.

1.2.2.2 Converting mouse epiblast into germ cell like cells.

The culture conditions that allowed the directed differentiation of mPSCs into mPGCLCs were first identified using mouse epiblast (Ohinata et al., 2009). Bmp4 addition to floating cultures of E6.0 epiblasts, separated from the visceral endoderm, induced Blimp1 expression in all cells of the epiblast (Ohinata et al., 2009). This study also confirmed that Bmp4 was required for induction of germ cell fate, as Bmp8b had no effect on Blimp1 expression. Bmp2 could also induce Blimp1 expression, but Blimp1 expression was weaker compared to the addition of Bmp4. A combination of Bmp4, Lif, Scf and Egf gave the highest induction efficiency and survival of mPGCLCs from the epiblast, the non Bmp4 cytokines are thought to enhance mPGCLC survival. mPGCLCs derived from this culture system expressed mouse germline genes and had epigenetic signatures similar to PGCs at the equivalent stage (Ohinata et al., 2009). Finally, mPGCLCs derived from epiblasts could give rise to sperm which had correct imprinting patterns and rescues the fertility of germ-cell deficient mice when injected in the neonatal testis (Ohinata et al., 2009).

1.2.2.3 Derivation of mPGCLCs from mPGCs instead of mouse epiblast.

The discovery that Bmp4 was able to convert cultured post-implantation epiblast to functional mPGCLCs (Ohinata et al., 2009) allowed for the generation of mPGCLCs from mPSCs (Hayashi et al., 2011). The same condition could be used to generate mPGCLC from mouse epiblast could be used to induce mouse Epi-like cells (mEpiLCs) into mPGCLCs. Similar to the experiments with mouse epiblast, floating cultures of mEpiLCs were aggregated in low attachment plates were used to encourage differentiation.

Neither mESCs or mEpiSCs were able to form robust numbers of mPGCLCs in response to the same signalling, and mEpiLCs could only respond on day 2 of their culture (Hayashi et al., 2011). This phenocopies the strict developmental timings of the epiblast (Ohinata et al., 2009). mEpiLCs are derived from mESCs by addition of Activin A and bFGF (Hayashi et al., 2011) and they are hypothesised to represent a formative state of pluripotency, where the naïve network begins to be dismantled (Smith, 2017).

1.2.2.4 The transition from mouse to human.

Spontaneous differentiation of human embryonic stem cells (hESCs) into putative hPGCLC were observed in hESCs culture and embryo body formation (Clark et al., 2004), with BMP addition to media leading to an increase in the percentage of cells which displayed germline gene expression (Kee et al., 2009; Kee et al., 2006). As with mouse, hPGCLCs from these studies were rare, and the differentiation was triggered using an undefined media, containing FCS making it difficult to define the necessary and sufficient signalling pathways for germline induction. Even with BMP4 addition, less than 0.5% of cells expressed specific surface markers for germ cell identity such as c-KIT (Gkountela et al., 2013).

1.2.2.5 The development of the *in vitro* hPGCLC system.

In 2015, both Saitou and Surani groups published a defined system to differentiate hPSCs into hPGCLCs, using the same protocol as mPGCLCs. When aggregated in low attachment plates stimulated with BMP4, LIF, SCF, EGF and ROCKi conventionally cultured hPSC can differentiate into hPGCLCs at higher frequencies than the spontaneous differentiation, around 20% (Sasaki et al., 2015). However, in first instance this protocol resulted in many of the hPSCs dying during differentiation. The two groups developed a 'pre-induction' step before the aggregation of hPSCs in the hPGCLC induction media, which increased both the conversion percentage of the cells into hPGCLCs and the survival rate of these hPGCLCs (Irie et al., 2015; Sasaki et al., 2015).

Irie and colleagues induced hPGCLC straight from an alternative culture condition for hPSCs called '4i' (inhibitors to MEK, GSK3β, p38 and JNK (Gafni et al., 2013)) (Irie et al., 2015). *NANOS3* was used as an internal marker for germ cell identity, with its expression tracked through an mCherry reporter gene. Cells positive for mCherry and the surface marker tissue-nonspecific alkaline phosphatase (TNAP), a pluripotency marker, were identified as hPGCLCs (Irie et al., 2015).

The gene expression pattern of *NANOS3*/TNAP positive cells was similar to the germ cell tumour (GCT) Tcam-2 cells, but also clustered closely with week 7 gonadal hPGCs. The genes which are expressed in both hPGCLCs and gonadal hPGCs are: *SOX17, TFAP2C, BLIMP1, NANOS3, OCT4, NANOG, KLF4, UTF1* and *GATA 4,* while *SOX2* and *KLF2* are repressed in both cell types (Irie et al., 2015). The markers of late germline, *DAZL, VASA, PIWIL1*, *SYCP3* are not expressed in hPGCLCs suggesting hPGCLCs are positioned earlier in the development of the human germline, most likely reflecting pre-migratory hPGCs (Irie et al., 2015).

In the second study from Sasaki and colleagues, conventionally cultured hPSCs were first pre-treated with culture media which supports the expression of mesoderm lineage markers, *T, MIXL1* and *EOMES*. Briefly, hPSCs are dissociated into single cells and then replated onto fibronectin in medium containing ACTIVIN A, CHIR99021 and ROCK inhibitor to induce a

transient state called incipient mesoderm like cells (iMeLCs). Stimulation for 48 hours upregulates mesoderm genes but still maintain core pluripotency factors, suggesting these iMeLCs are representative of the early primitive streak (Sasaki et al., 2015). These iMeLCs are then aggregated with BMP4, LIF, SCF and EGF in low-attachment plates, leading to the formation of hPGCLCs over 4 days of culture, and sustain in these aggregates for up to 12 days (Sasaki et al., 2015). Two fluorescent reporter genes were fused to key early germline genes; *BLIMP1* was fused to tdT (BT) and *TFACP2C* to eGFP (AG) allowing the tracking of expression of both these genes in individual cells.

Cells which expressed both reporter genes were termed BTAG positive and were designated as hPGCLCs. The transcriptome of the BTAG positive cells produced from iMeLCs clusters closer to the TNAP/*NANOS3* positive cells, than the hiPSCs or iMeLCs, suggesting both methods yield bona fide hPGCLCs (Sasaki et al., 2015).

The authors of this study demonstrated that ACTIVIN A and CHIR (or specifically WNT3 signalling) are both required for iMeLCs, but additions of BMP4 during iMeLC culture abolishes iMeLCs competence for hPGCLCs. BMP4 stimulation of iMeLCs must occur after around 42 hours of iMeLC culture, suggesting that if BMP4 is added too early, it leads to an alternative cell fate rather than the germline (Sasaki et al., 2015) The role of BMP4, LIF, SCF and EGF during the hPGCLC aggregate were also elucidated. All four cytokines produce 42% BTAG positive cells at day 2, falling to 36% at day 4 and to 21% at day 8. Only BMP4 is able to induce the expression of the BTAG reporters, but without the other cytokines, the percentage of BTAG positive cells falls from 32% at day 2 of aggregation to 13% at day 4 and only 0.2% at day 8. SCF in combination show similar percentages of all four cytokines: 34% at day 2, 29% at day 4 and 11% at day 8. BMP4 is essential for hPGCLC induction but the other cytokines, LIF, SCF and EGF increase hPGCLC survival and are required for longer-term culture within the aggregate (Sasaki et al., 2015).

BMP4 signalling in aggregates of competent cells, such as 48 hours iMeLCs or 4i cells, leads to the formation of hPGCLCs, presumably through a convergent mechanism. For my experiments, I used the protocol pioneered by Sasaki et al.2015, which involves inducing hPSCs into incipient mesoderm-like cells (iMeLCs).

1.2.3 The transcription factor network which establishes and maintains hPGCLC identity

These two initial studies which developed the hPGCLC system, also identified two factors that were crucial for hPGCLC induction, *SOX17* and *BLIMP1* (Irie et al., 2015; Sasaki et al., 2015). The reproducibility of the hPGCLC induction protocol allowed for the assessment of phenotypes and genetic analysis on how these TFs interact to specify hPGCLCs. Knock-out (KO) studies identified further factors such as *EOMES* and *TFAP2C* that are essential for hPGCLCs (Kojima et al., 2017). This KO study also revealed that the timings and order of the different germline genes are expressed in hPGCLCs (Kojima et al., 2017), which was supported by single cell RNA-sequencing (Chen et al., 2019). The importance of *SOX17* and *TFAP2C* was underlined as, when combined with *GATA3,* forced expression of these three TFs allowed for direct induction of hPGCLCs from aggregates of iMeLCs without BMP4 (Kojima et al., 2021). Combining these studies, the timings, and the role the different factors expressed in hPGCLCs was unravelled.

1.2.3.1 CHIR induces expression *EOMES,* which is required for *SOX17* activation.

This pre-induction step, either iMeLC or 4i, appears to be required for hPGCLC induction because it induces the expression of specific lineage genes. *EOMES*, a key transcription factor in endoderm specification (Arnold et al., 2008), has been identified as a critical gene that must be expressed at the iMeLC stage (Kojima et al., 2017). *EOMES* expression between 24-48 hours after iMeLCs induction is required for hPGCLC induction and can replace signalling from ACTIVIN A or CHIR, although at a lower efficiency (Kojima et al., 2017). The other lineage genes induced by CHIR, such as *T* and *MIXL1,* are therefore important in specification or survival of hPGCLCs (Kojima et al., 2017). When combined with ACTIVIN A, *EOMES* overexpression can fully replace the WNT signalling activated by CHIR stimulation (Kojima et al., 2017). As the '4i' media also contain CHIR (Gafni et al., 2013), and cells cultured in this media express *EOMES* (Sasaki et al., 2015), the inclusion of this inhibitor is at least partly responsible for allowing 4i cells to direclty respond to BMP signalling to induce hPGCLCs.

Variation in hPGCLC induction efficiency between differenet hPSC lines is well characterised (Yokobayashi et al., 2017). Higher expression of lineage markers for endoderm and mesoderm, *EOMES*, *T* and *MIXL1*, and lower expression ectodermal genes *SOX3* and *CHD1* (Acloque et al., 2011) are observed in iMeLCs from hiPSCs which show higher competence for hPGCLC induction (Chen et al., 2019; Yokobayashi et al., 2017).

1.2.3.2 *SOX17* is critical for establishing human germline fate.

SOX17 KO cells within the hPGCLC aggregate express *TFAP2C* during the day 1 and day 2, suggesting *TFAP2C* induction is independent of *SOX17* (Kojima et al., 2017). However, sustaining expression of *TFAP2C* does seem to require *SOX17* expression (Irie et al., 2015). *BLIMP1* is not expressed in *SOX17* KO cells, above the levels observed in the iMeLC (Kojima et al., 2017), suggesting the full activation of *BLIMP1* in hPGCLCs is dependent on *SOX17* expression (Irie et al., 2015). The later marker *NANOS3* shows normal expression at day 1 but falls rapidly by day 2 in these KOs. Pluripotency factor *NANOG* is quickly downregulated upon hPGCLC induction in *SOX17* KO, suggesting a major role for *SOX17* is to maintain *NANOG* expression. A second key pluripotency factor *OCT4* shows robust expression in day 1 and day 2 of induction (Kojima et al., 2017), but its expression is not detected in day 5 aggregates (Irie et al., 2015).

EOMES KO leads to *SOX17* and *BLIMP1* failing to be induced (Kojima et al., 2017). *TFAP2C* however, can be induced on the day 1 and day 2, but is not detected at day 4 (Kojima et al., 2017). Due to the failure of *SOX17* expression in *EOMES* KO these two KOs appear to phenocopy each other (Kojima et al., 2017). EOMES KO does have larger effect than *SOX17* KO on reducing expression of linage markers *T* and *MIXL1*, although neither are required for hPGCLC induction (Kojima et al., 2017). *EOMES* role to directly activate *SOX17* is observed in hiPSCs (Kojima et al., 2017), although *SOX17* is only expressed at low levels in iMeLCs (Chen et al., 2019; Kojima et al., 2017; Sasaki et al., 2015), indicating that further mechanisms are needed to induce *SOX17* expression in hPGCLCs. Direct *SOX17* activation can overcome *EOMES* KO, but the resulting hPGCLCs aggregates are smaller, suggesting *EOMES* also has a role in increasing proliferation in hPGCLCs (Kojima et al., 2017). *EOMES* is expressed higher in

iMeLCs derived from cell lines which generate higher proportions of hPGCLCs (Yokobayashi et al., 2017), demonstrating *EOMES* importance in the iMeLC state for inducing hPGCLCs.

1.2.3.3 TFAP2C does not activate SOX17, but both factors maintain each other.

In *TFAP2C* KO, *SOX17* and *BLIMP1* expression can be detected at day 2 with RT-qPCR but is decreased compared to WT by day 4 (Kojima et al., 2017). However, TF network analysis on ssRNA-seq data comparing *TFAP2C* KO and WT suggests *TFAP2C* lies upstream of *SOX17* as it binds to the *SOX17* promoter region. *TFAP2C* probably does have a role in maintaining *SOX17* expression, but it is likely not critical for its induction, based upon more direct gene expression analysis on cells that are *BLIMP1* positives (Kojima et al., 2017).

Similar to *SOX17* KO, *NANOS3* and *NANOG* are downregulated in *TFAP2C* KO, suggesting the requirement for both factors in maintaining expression of the germline network. Again, there is disagreement as to the effect on *OCT4* expression in *TFAP2C KO*, as immunofluorescence show an absence of OCT4 from hPGCLC aggregates (Chen et al., 2018), but qPCR data from a different study suggests *OCT4* is expressed in cells *BLIMP1* positive, *TFAP2C* KO cells (Kojima et al., 2017). The differences reported in *TFAP2C's* roles may be due to the enrichment of *BLIMP1* positive cells in the qPCR analysis (Kojima et al., 2017), which have presumably activated *SOX17* expression and show more germline related characteristics.

1.2.3.4 *BLIMP1* guards the germline from somatic differentiation.

BLIMP1 KO cells in the hPGCLC aggregate fail to upregulate the later hPGCLC marker *NANOS3* (Irie et al., 2015; Sasaki et al., 2015) and cells which are TNAP positive are only weakly positive for *TFAP2C* (Irie et al., 2015). Both groups reported *SOX17* is expressed in the *BLIMP1* KO (Sasaki et al., 2015), although one study showed that *SOX17* expression is reduced to half the levels observed in the WT hPGCLCs (Irie et al., 2015), suggesting *BLIMP1* is downstream of *SOX17* but *BLIMP1* has a role in maintaining *SOX17* expression. *TFAP2C* is expressed in *BLIMP1* KO cells, indicating it is not induced by *BLIMP1.*

1.2.3.5 TFAP2C, SOX17 and BLIMP1 activate the downstream germline genes and repress somatic fates

Principle component analysis of a series of KO hPGCLCs shows that *EOMES* and *SOX17* KO first fall out of germline identity after day 2 (Irie et al., 2015; Kojima et al., 2017). These KO express vascular genes related to mesoderm (Kojima et al., 2017), suggesting an important role for *SOX17* in the repression of the mesoderm lineage. *TFAP2C* KO progresses slightly further but express neural and ectodermal genes and downregulate *SOX17* and BLIMP1(Kojima et al., 2017). *BLIMP1* KO fall out at between day 2-4 and begin to express *HOX* genes and other endoderm genes induced by *SOX17* (Irie et al., 2015; Kobayashi et al., 2017). All knockouts can form day 2 hPGCLCs but fail to maintain this identity by day 4, showing the signalling from BMP4 is enough to specify hPGCLC fate but cannot sustain it.

The cells of the aggregate within the *TFAP2C* and *SOX17* KO both show successful SOX2 repression, in response to the BMP4 addition. *SOX2* expression is higher in *BLIMP1* KO cell, suggesting it might act as a repressive factor for *SOX2* expression. However, *SOX2* expression is lower in these *BLIMP1 KO; TFAP2C* positive cells compared to the iMeLC state (Sasaki et al., 2015), suggests there are further mechanisms which repress this factor.

1.2.3.6 *GATA* TFs act as pioneering factors to drive expression of early germline genes

In iMeLCs system, driving *SOX17* expression does not induce germline specification (Kojima et al., 2021), nor does driving any combination of *TFAP2C*, SOX17 and *BLIMP1.* Instead, GATA2 or GATA3 is required alongside SOX17 and TFAP2C to drive germline induction and to repress *SOX2* (Kojima et al., 2021). The developmental timings of hPGCLCs induced by their forced expression are slightly altered compared to hPGCLCs generated with BMP4. The day 1 TFs derived hPGCLCs have a transcriptome more similar to day 2 BMP4 generated hPGCLCs, suggesting that *GATA2* or *GATA3* is one of the first key downstream effectors of BMP signalling (Kojima et al., 2021). This is supported by the observation that expression of *GATA3* only needs to occur in day 1 for correct hPGCLC induction (Kojima et al., 2021).

Removal of *GATA2* and/or *GATA3* does not completely abolish hPGCLC competency. However, it reduces the induction efficiency with less than 2% cells that can form hPGCLCs (Kojima et al., 2021). This could be redundancy of other *GATA* TFs, or it might speak to the role of these GATA as pioneering factors, which open the chromatin to allow for gene expression (Sanalkumar et al., 2014). The rare hPGCLCs in these double KO therefore could be that in a few cells, BMP signalling can activate *SOX17* and *TFAP2C* without *GATA* TFs.

Taken together a stepwise scheme of inducing hPGCLC from iMeLCs is as follows: *EOMES* expression in the iMeLCs state and *GATA3* (or GATA*2*) in day 1 of hPGCLC activates *SOX17* in response to BMP4 signalling; *SOX17* probably co-binds with *OCT4*; evidence for this interaction comes from germline tumours (Jostes et al., 2020) and other developmental lineages (Aksoy et al., 2013; Stefanovic et al., 2009), but has not been shown directly in hPGCLCs. Independently, BMP signalling along with GATA3 activates *TFAP2C*. *BLIMP1* is then activated by *SOX17* and together these factors trigger *NANOS3* expression, maintain *OCT4* and *NANOG* expression. A further mechanism, perhaps involving BLIMP1, represses SOX2 expression. This model of hPGCLC induction is recapitulated in Figure 4.

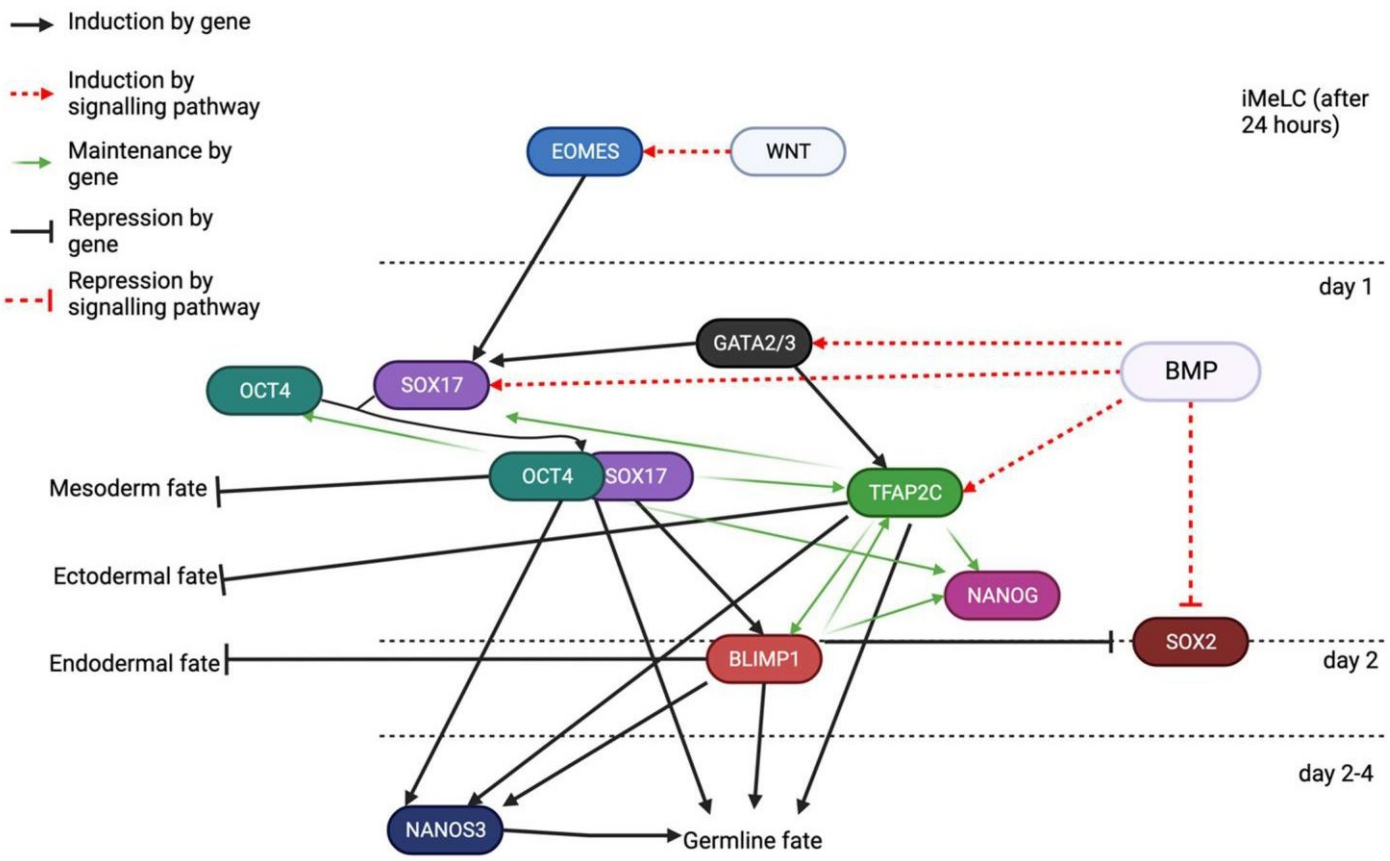

Figure 4 specification and maintenance of hPGCLC state from iMeLCs based upon KO, ssRNA-seq and TF overexpression studies. EOMES is expressed in the iMeLC state, which along with BMP4 induces SOX17. SOX17 induces BLIMP1. BMP4 also induces TFAP2C. The three core factors then induce other germline factors and maintain the germline network.

1.2.3.7 Other genes implicated in hPGCLC induction

A number of other genes are implicated in hPGCLCs such as *PRDM14* (Pierson Smela et al., 2019; Sybirna et al., 2020) whose knock-out reduces the induction efficiency of hPGCLCs. Genes such as *NANOG, KLF4* and *NANOS1* are downregulated in *PRDM14* KO hPGCLCs (Sybirna et al., 2020) suggesting it does have a role in hPGCLC state maintenance. However, hPGCLCs can be formed from these *PRDM14* KO cells, suggesting it is not required for hPGCLCs induction, but does aid it.

SOX15 KO causes reduction of hPGCLC cell numbers in the aggregate at day 6 and later, although this effect is only strong when the KO occurs before or at the point of hPGCLC induction. Overexpression of *SOX15* does increase survival of day 8 hPGCLCs (Pierson Smela et al., 2019) suggesting that once established, hPGCLCs do utilise *SOX15* for prolonged survival.

Epigenetic modifiers of both DNA methylation and histone modifications have been shown to play a key role in the induction, by allowing the correct expression of germline genes and repressing destabilising gene expression. TET enzyme expressions are required for mediating the epigenetic changes in the promoter of *NANOG* and *SOX17* to trigger robust expression from these loci (Li et al., 2022). A histone demethylase *KDM2B* is also implicated in hPGCLC induction as its knock-out leads to a reduction in hPGCLC conversion efficiency (Yuan et al., 2021). This appears to be due to general dysregulation of the transcriptome, presumably as histones maintain the active H3K4Me3 and H3K36Me2 histone marks, preventing silencing of multiple genes (Yuan et al., 2021).

Core pluripotency factors *NANOG* and *OCT4* are expressed in all stages of the hPGCLC protocol, in hPSCs, iMeLCs and hPGCLCs, and their levels do rise slightly in hPGCLCs (Irie et al., 2015; Kojima et al., 2017; Sasaki et al., 2015). A few naïve related genes are also upregulated in hPGCLCs compared to hPSCs, including *KLF4* and *TFCP2L1* (Sasaki et al., 2015). Directly testing the requirement for *NANOG* and *OCT4* is difficult as they maintain hPSCs and so generating KO hPSCs that can undergo hPGCLC development is not possible.

1.3 Connecting pluripotency and germline specification

1.3.1 Pluripotent states and germline induction.

As discussed above, *in vitro* germline specification through the PGCLC protocol begins with pluripotent stem cells. However, the naïve state, in mouse or human, does not give rise to PGCLCs in response to BMP signalling; ESCs must first be differentiated into EpiLCs (Hayashi

et al., 2011; von Meyenn et al., 2016). The EpiLCs cell type is considered to be in the 'formative' state of pluripotency (Smith, 2017), which sits between naïve and primed (Figure 5). Formative stem cells derived from naïve mPSC have been shown to be able to be directly specified into mPGCLCs (Kinoshita et al., 2021a), supporting the notion that the germline is specified from this pluripotency state. Human naïve PSCs can also be differentiated into hEpiLC that can form hPGCLCs when aggregated in hPGCLC induction media (von Meyenn et al., 2016).

hPSCs are able to enter hPGCLC fate directly, although at a very low efficiency, suggesting hPSC might be less primed than mEpiSCs. This is despite the presumption that hPSCs are thought to exhibit a primed pluripotent character similar to mouse EpiSCs, which cannot respond to BMP signalling to produce mPGCLCs. However, it is becoming evident that hPSCs do not exhibit an identical primed state to mEpiSCs, and their ability to directly enter the germline fate may suggest they more closely resemble the formative state (Smith 2017, Kinoshita et al., 2021).

In Figure 5, iMeLCs and '4i' hPSCs are shown between the formative and primed states, but their exact location in the pluripotent spectrum remains unclear and they may not represent any state that is present *in vivo*. '4i' cells were initially described as being in the naïve state (Gafni et al., 2013), however they are very similar to conventional primed hPSCs when global gene expression is considered (Irie et al., 2015). The PGCLC-competent state of the iMeLC is transient and appears to vary between different hPSC lines (Yokobayashi et al., 2017), but the expression of lineage marker *EOMES* is critical (Chen et al., 2019; Yokobayashi et al., 2017).

There are still many unanswered questions when considering human pluripotent states, including how representative these *in vitro* culture conditions are to the embryo and how they relate to germ cell induction. It could be that in humans, hPSCs are close enough to the formative state that stimulation Wnt signalling is able to push cells into a receptive state for germline entry. The position of the different *in vitro* PSCs from mouse and human, and the derivates EpiLCs and iMeLCs, within the pluripotency continuum are displayed in Figure 5.

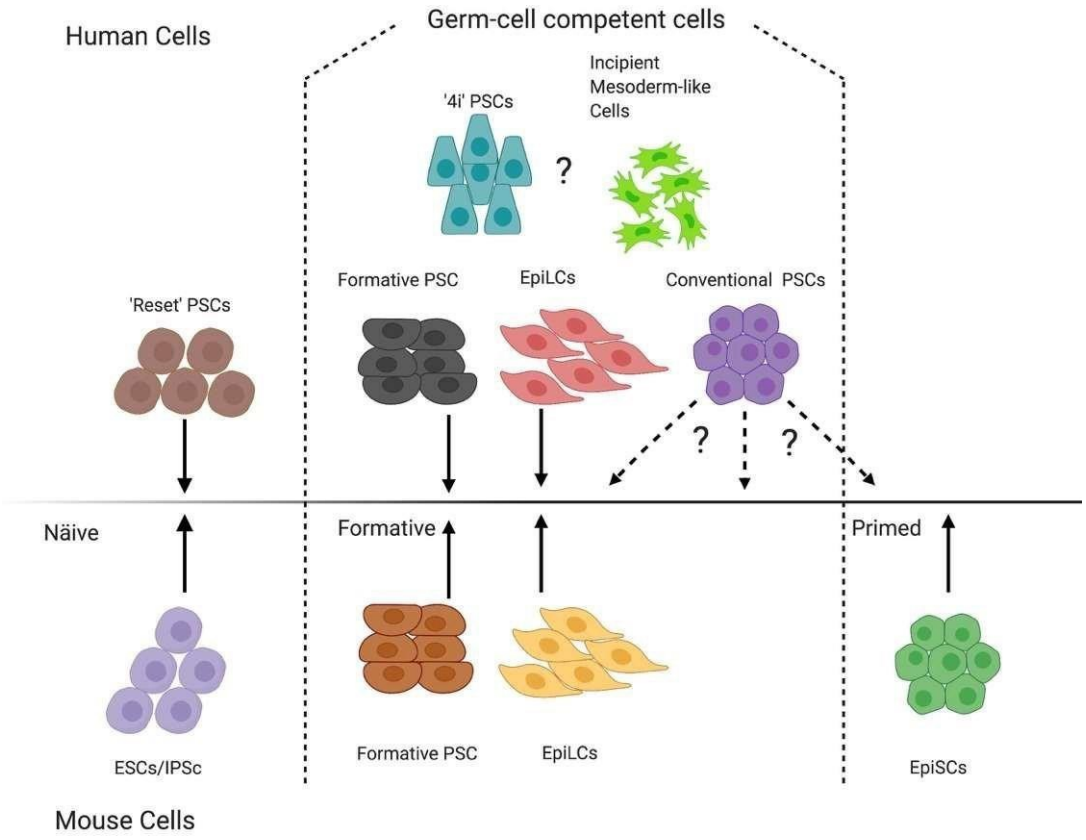

Figure 5 In vitro pluripotent stem cells and their derivatives within the pluripotency continuum. From left to right, naïve, formative, and primed cells displayed. Top panel; human cells. Position of conventional, 4i and iMeLCs are unclear, but based on their entry into the germline, they probably fall within the formative state. Lower panel; mouse cells.

1.3.2 Germ cells which 'regain' pluripotency: Embryonic germ cells (EGs)

Despite the apparent unipotency, mouse PGCs retain a pluripotent 'capacity' or latent pluripotency after being specified while they migrate (Leitch and Smith, 2013; Saitou and Yamaji, 2012). In mice, this latent pluripotency allows mPGCs to undergo conversion to pluripotent stem cell lines termed embryonic germ cells (EGs) *in vitro* (Matsui et al., 1992; Resnick et al., 1992).

1.3.2.1 Development of defined culture system.

When cultured in defined conditions these mEGs cells are equivalent to mESC in their transcriptional state, epigenetic landscape and functional ability to colonise the epiblast of a pre-implantation embryo (Leitch et al., 2013c). Some mEG lines do exhibit erasure of genomic imprints, but this is not a universal finding (Leitch et al., 2013a; Shovlin et al., 2008). Of note, mESCs have also been demonstrated to exhibit imprint instability (Humpherys et al. 2001).

As discussed above, *Sox2*, *Oct4* and *Nanog* are considered the 'core' pluripotency factors in mammals (Figure 1) (Chen et al., 2008; Nichols and Smith, 2012). Expression of these key pluripotent genes in mPGCs is thought to be central to their ability to convert to mEGs (Leitch et al., 2013d; Santagata et al., 2007), albeit this has only been tested directly for *Sox2* (Campolo et al., 2013). *Oct4* is expressed throughout mPGC development and *Sox2* and *Nanog* are upregulated shortly after specification (Kurimoto et al., 2008; Sekita et al., 2016). Each of these play key roles in mPGC development (Campolo et al., 2013; Kehler et al., 2004; Zhang et al., 2018), and their continued expression might explain why the conversion from germline to pluripotency is possible in mouse.

1.3.2.2 Historical attempts to derive human EGCs.

As of today, human PGCs have hitherto not been converted into stable pluripotent stem cell lines. While human EGs (hEGs) have been reported in the literature (Shamblott et al., 1998) these primary cultures are not able to be maintained long term or undergo freeze-thaw cycles (Turnpenny et al., 2006).

Pluripotent 'capacity' can be defined as how readily a cell type can regain pluripotency. In this sense, current evidence indicates that mPGCs have a higher pluripotent capacity compared to hPGCs, as conversion can be achieved with high efficiency (and in defined conditions) (Leitch et al., 2013b) whereas no such system has been established in human. As

no hEGs have been generated, the lack of successful conversion of *in vivo* cells is represented by a cross in the germline cycles represented in Figure 6.

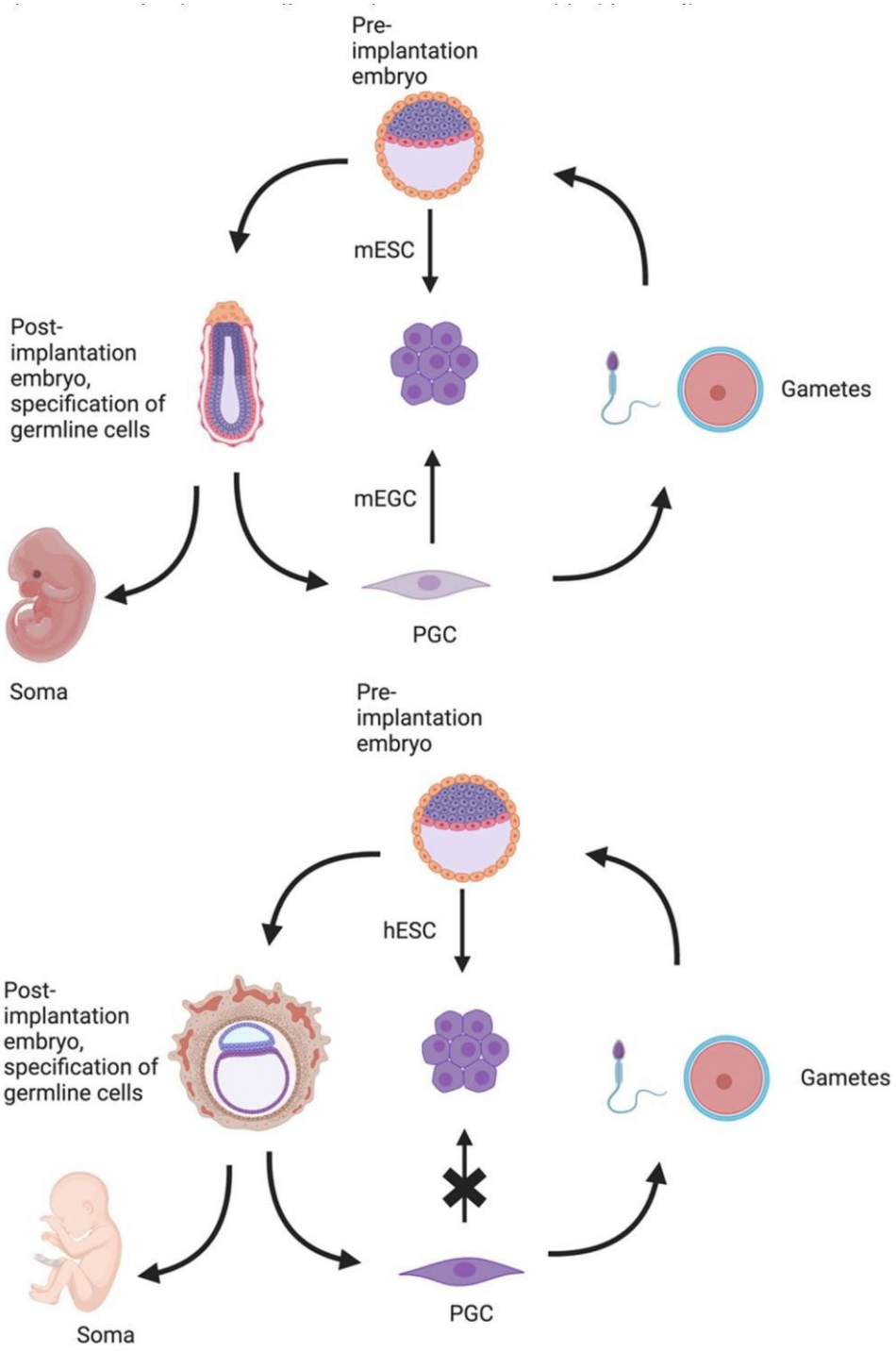

Figure 6 Mouse and Human germline cycle

1.3.2.3 Reprogramming of hPGC-like germ cell tumour to an EG-like germ cell tumour.

Pluripotent conversion of human germ cell has been observed to occur in the context of human germ cell tumours (GCTs), suggesting an *in vitro* conversion of hPGC/LCs could be possible. There are two main types of GCTs, embryonic carcinomas (EC) and seminoma (SEM). *SOX2* is expressed in EC tumours, like in mouse EG cells, but absent from SEM tumours which express SOX17 like PGCs (Santagata et al., 2007). EC is often found in mixed nonseminomatous GCTs, along with teratoma cells differentiated into the three germ-layers; leading to the suggestion that EC take on a pluripotent capacity which gives rise to tumours with mixed histology (Santagata et al., 2007). The same is not true for seminoma, suggesting that pluripotent conversion of PGCs can only occur in human when *SOX2* is expressed. ECs therefore may represent an *in vivo* reprogramming of hPGCs into a pluripotent state.

Intriguingly, SEM cells can convert to an EC fate by changing the micro-environment of the cells (Nettersheim et al., 2015). Xenografting a SEM cell line called TCam-2 into the flank or brain of mice resulted in conversion into an EC tumour, expressing *SOX2* and repressing *SOX17* (Nettersheim et al., 2011). A similar result can be achieved by treating with Tcam-2 cells with Noggin, an inhibitor for BMP signalling, which led to upregulation of *SOX2* and repression of *SOX17* (Nettersheim et al., 2015). BMP signalling is presumably provided by the addition of FCS to the TCam-2 culture.

Expression of *SOX2* is essential for this process, as SOX2 KO TCam-2 cells cannot undergo the transition to EC state, even when xenographed (Nettersheim et al., 2016a). BMP signalling is reduced in the microenvironment, but upregulation of NODAL and WNT are not observed in TCam-2 *SOX2* KO cells, suggesting these signalling pathways are downstream of *SOX2* (Nettersheim et al., 2016a). An alternative system used retinoic acid to re-active *SOX2* in TCam-2 cells (Kushwaha et al., 2016). RA addition caused UTX-mediated removal of the repressive mark H3K27me3 at the *SOX2* locus and causing *SOX2* reactivation. *OCT4* and *NANOG* were repressed in response to RA treatment, and TCam-2 cells differentiate into neuronal lineages (Kushwaha et al., 2016). The difference in the differentiation between these two treatments, noggin vs RA, might be due to retinoic acid signalling triggering stem

cell differentiation (Gudas and Wagner, 2011) meaning RA treatment caused further effects beyond *SOX2* derepression in TCam-2 cells.

1.3.2.4 Reports of hEGC(LC)s from hPGCLC culture.

A clear difference between the mouse and human germline, and potential clue for the lack of hPGC conversion into hEG, is the downregulation of *SOX2* in hPGCs (Perrett et al., 2008). Instead, as mentioned above, *SOX17* plays a key role in human PGC specification and regulation of germ cell fate (Irie et al., 2015; Tang et al., 2015). As observed in TCam-2 reprogramming studies *in vivo* and *in vitro*, repression of *SOX17* and upregulation of *SOX2* is essential for these Tcam-2 cells to obtain an EC-like fate. Therefore, resetting of hPGC(LCs) will no doubt involve the re-activation of *SOX2* and probably the repression of *SOX17*.

Maintaining hPGCLCs in culture has recently been achieved. Murase et al., 2020 were able to maintain a stable *in vitro* culture of hPGCLCs for 120 days. Cells were cultured on inactivated SCF expressing m220 feeders with soluble SCF, Forskolin (a cAMP activator) and bFGF in a basal media containing KSR and FCS. Interestingly, this study also showed that even without the addition of cytokines cells proliferated, although to a lesser extent, suggesting the feeders themselves are able to promote proliferation. While hPGCLCs expand in this system, only a diminishing fraction of the cells maintains their PGCLC identity, falling from 77% after 20 days in culture to 50% after 60 days. Cells that lost their *AP2γ* expression upregulated *SOX2* expression, suggesting they could have reverted to or regained a pluripotency. Alternatively, these *SOX2* expressing cells could be differentiated cells as *SOX2* is expressed in neural lineages derived from mouse and human stem cells (Tchieu et al., 2017; Zhao et al., 2004).

A 'feeder-free' system has been established using conditioned media. Similar to early studies, SCF, EGF and LIF are included in the media, which has been conditioned by STO feeders for 24 hours (Kobayashi et al., 2022). Presumably the feeders release a, or a set, of soluble cytokines which maintain hPGCLC identity. Removal of the conditioned media, and addition of SCF and FGF2 led to the formation of SOX2 positive cells that behaved like stem cells

(Kobayashi et al., 2022). There were termed human embryonic germ cell-like cells (hEGCLCs), to recognise their *in vitro* origins.

1.3.2.5 Differences between *SOX2* and *SOX17.*

While belonging to the same superfamily, *SOX17* cannot functionally replace *SOX2* in the pluripotent network, in contrast with *SOX1* and *SOX*3 (Jauch et al., 2011; Nakagawa et al., 2008; Sarkar and Hochedlinger, 2013). A lack of endogenous *SOX2* expression in human could explain why hPGCs fail to convert to a pluripotent state, as this is an essential factor for human pluripotent stem cell establishment and self-renewal (Fong et al., 2008).

A commonality between both factors is their co-binding to *OCT4* that has been visualised through X-ray crystallography (Palasingam et al., 2009). The binding motifs of these complexes are different; *SOX17*/OCT4 binds the 'compressed' motif where there is no space between the SOX motif and the OCT4 motif, while *SOX2/OCT4* binds to the 'canonical' motif which contains a small space piece of DNA (Jauch et al., 2011; Jostes et al., 2020). Both factors co-operatively bind to OCT4 on their preferred motif, but SOX2 is not able to bind to the compressed motif due to its larger size causing steric exclusion on the motif when bound to OCT4 (Jauch et al., 2011). SOX17 can bind to the canonical motif but only in an additive binding with OCT4, meaning this binding is not favoured compared to binding at the compressed motif. By altering the structure of *SOX17* to swap its preference from the compressed to the canonical motif, *SOX17* was able to reprogramme miPSCs and hiPSCs in place of *SOX2* (Hu et al., 2023; Jauch et al., 2011). Thus, the different actions of *SOX2* and *SOX17* are caused by their preference for motifs in the genome, rather than them having different roles in regulating transcription. This preference for different motifs is also observed in endoderm specification (Aksoy et al., 2013).

As discussed above, *SOX17* is critical for hPGCLC induction and their maintenance. Comparing the different roles of *SOX2* and *SOX17* in GCTs provides context for these factors within the background of the human germline and how the placement of the compressed and canonical motifs within the genome alters the genes which are affected by the different SOX factors.

Within the *SOX2* positive EC cells and the *SOX17* positive SEM cells, both SOX factors regulate genes of the pluripotency network such as *NANOG* and *LIN28A* but only *SOX2* in EC cells appears to directly regulate *SOX2* and *OCT4* (Jostes et al., 2020). Despite not binding to as many pluripotency regulatory elements as *SOX2*, SOX*17* is crucial for maintaining pluripotency in SEM cells as its deletion downregulates *OCT4, NANOG* and others in TCam-2 cells (Jostes et al., 2020). Interestingly, *SOX17/OCT4* complexes appeared to bind more canonical motifs in TCam-2 cells compared to hESCs which had been differentiated into different lineages (Jostes et al., 2020).

SOX factor binding to motifs with co-binding to OCT4 is important for pluripotency in the context of PSCs, PGCs, PGCLCs, SEM and ECs. Which factor is expressed depends on species and cell state. The expression of *SOX2* in mPSCs, hPSCs, mPGCs and hECs is contrasted with *SOX17* in hPGC(LC)s and Tcam-2 cells. *SOX17,* as well as maintaining pluripotency in 'normal' human germline cells, also plays a significant role in defining and specifying these cells as germline cells. In contrast, *SOX2* can regulate a few germline genes in the context of 'reset' germ cells ECs, although the majority of genes it appears to directly regulate are more conventional pluripotency genes (Jostes et al., 2020).

Mouse germline utilising *SOX2* to maintain pluripotency uses a different germline network where *Blimp1* appears to restrict mouse germline fate along with *Prdm14* (Hayashi et al., 2011), while in humans this restriction is brought about by *SOX17* and *TFAP2C* (Kojima et al., 2017). SOX17 also defines the germlines of *Cynomolgus* monkey (Sasaki et al., 2016) and Pig (Kobayashi et al., 2017). Human, pig and monkey all develop as bilaminar discs, while mouse develop as egg cylinders (Alberio et al., 2021) and this deviation in embryology may explain why different mammals use different SOX factors, although the germlines of other bilaminar disc mammals such as rabbit or cow are yet to be discovered.

SOX17 dual role is to induce germline fate and maintain pluripotency through binding to compressed motifs and in the germline the canonical motifs, meaning *SOX2* expression is not required. Mouse germline does not have these dual acting SOX factors, so uses the pluripotency related factor *Sox2* to bind at canonical sites to maintain pluripotency but requires further factors to specify the germline.

1.3.2.6 The importance of pluripotency transcription factors in hPGCLCs.

The importance of pluripotency in the mammalian germline has been recognised, although mainly analysed in mouse (Leitch and Smith, 2013). It is clear from the hPGCLC experiments and from *in vivo* tissue analysis that expression of *OCT4* and *NANOG* is universal in hPGC (Gkountela et al., 2015; Guo et al., 2015; Tang et al., 2015; Tyser et al., 2021) and hPGCLCs (Chen et al., 2019; Irie et al., 2015; Kojima et al., 2017; Sasaki et al., 2015). These pluripotency factors and others play a role in the early embryo and are essential for the survival and self-renewal of PSCs, which hPGCLCs are derived from. Understanding the roles, they play in hPSCs and how they might therefore challenge or disrupt the germline network will be important for understanding the effects of overexpression.

1.4 Aims and hypothesis

In this project, my aim was to see how the human germline network in hPGCLCs would react to the overexpression of different pluripotency transcription factors, *SOX2, NANOG* and *KLF2.* These three TFs will be overexpressed from a randomly integrated plasmid using PiggyBac system (Takashima et al., 2014). As *SOX2* re-activation appears to be critical for the entry to pluripotency, its endogenous expression will be tracked with a reporter gene.

Exogenous *SOX2* could act as a pioneering factor (Vanzan et al., 2021) to activate its endogenous locus, although it might be antagonised by the co-binding between *SOX17* and *OCT4,* preventing this activity. *NANOG* plays important roles in both germline and pluripotency, so it will be interesting to see if in the context of hPGCLCs it enhances the germline fate or reverts these cells to a pluripotent fate. Finally, *KLF2* shows the potential to reset human primed PSCs to naïve in combination (Takashima et al., 2014) with *NANOG*, but as *NANOG* is expressed higher in hPGCLCs (Kojima et al., 2017) *KLF2* on its own might be able to reset hPGCLCs into a naïve state.

Chapter 2. Materials and Methods

2.1 hiPSC culture

The UHi001-A also known as SOX2-tdTomato hiPSCs (Balboa et al., 2017) and the clones derived from this line were cultured in E8 essential media (A1517001; Thermo Fisher Scientific) on 1% Geltrex (A1413302; Life Technologies) coated 35 mm cell-culture plates (351008, SLS) or 6 well cell-culture plates (10119831, Thermo Fisher Scientific). Every 4-5 days, cells were passaged using 0.5 mM EDTA/DPBS (15575020; Life Technologies) (14190094, Life Technologies) and replated as small clumps at 1:10 or 1:20 dilution, medium was changed daily. The 1383D2 also known as BTAG hiPSCs (Sasaki et al., 2015) were cultured in Stemfit media (SFB-04-CT; ambsio) on 1% Geltrex coated cell-culture plates. Every 4-5 days, cells were passaged using ESGRO Complete Accutase (SF006; Millipore) and replated as single cells at 1:20 dilution with 10 μM ROCK inhibitor, Y-27632 (72304; Stemcell Technologies). Medium was changed the day after splitting to remove Y-27632, and then every other day. All cell lines were incubated in 5% CO_2 and 5% O_2 at 37^{o}C.

2.2. Generation of inducible TF cell line

2.2.1 Transformation of bacteria

pPB hCMV1IV hKLF2, pPB hCMV1IV hNANOG, pPBCAG-cHA-IN+ rtTAM2 and pBase helper plasmid were gifts from A. Smith (University of Exeter) (Takashima et al., 2014) and pPB hCMV1IV hSOX2 was a gift from H. Niwa (Kumamoto University). Manufacturer's instructions for DH5α (EC0112, Life Technologies) were followed for transformation. In brief, 20 μl competent DH5α and 0.1 ng plasmid DNA were combined for 30 minutes on ice, followed by a heat-shock of 42°C and a final 2-minute incubation on ice. After transformation, 250 μl S.O.C media (supplied with DH5α) was added to cultures and grown shaking at 225 rpm 37°C for 1 hour. Two volumes of transformed culture, 20 μl and 100 μl, were spread onto 100ug/ml ampicillin (A5354, Sigma) LB agar (12795027, Thermo Fisher Scientific), grown upside down overnight at 37°C. Single colonies were picked and grown in 3 ml LB medium supplemented 100 μg/ml ampicillin with overnight.

2.2.2 Mini-prep of plasmids

Monarch plasmid miniprep kit (T1010; NEB) was used to purify plasmids from bacterial culture, following manufacturer's instructions. In brief, 1.5 ml of bacterial culture was pelleted at 16,000g for 30 seconds and resuspended in 200 μl plasmid resuspension buffer (B1) through pipetting. To lyse the bacterial cells, 200 μl plasmid lysis buffer (B2) was added, mixed through inversion, and incubated at room temperature for 1 minute, followed by 400 μl plasmid neutralisation buffer (B3), mixed through inversion, and incubated at room temperature for 2 minutes. Lysate was centrifuged at 16,000 g for 5 minutes, supernatant was moved to a supplied spin column and centrifuged at 16,000 g for 1 minute to bind plasmid to the membrane. The membrane was washed with 200 μl plasmid wash buffer 1, centrifuged at 16,000 g for 1 minute. Membranes were washed again with 400 μl plasmid wash buffer 2 and centrifuged at 16,000 g for 1 minute. The column was moved to DNA LoBind tube (E0030108051, SLS), and 30 μl elution buffer, warmed to 50°C, was added directly on top of the membrane and incubated for 1 minute. Centrifugation at 16,000 g for 1 minute was used to elute the plasmid. Plasmid concentration was measured using Nanodrop one (Thermo Fisher Scientific); the quality of the DNA was deemed acceptable if the 260/280 ratio was between 1.6-2. Plasmid's identity was confirmed with sequencing by Genewiz.

2.2.3 Maxi-prep of plasmids

Plasmids were purified using PURE II™ Plasmid Maxiprep Kit (D4202, Zymo) following the manufacturer's instructions. 150 ml of bacteria culture were grown overnight shaking at 225 rpm at 37°C, pelleted at 3,400 g for 10 minutes, and resuspended in 14 ml of ZymoPURE P1 and vortexed. Next, 14 ml of ZymoPURE P2 was added, inverted gently, and incubated for 3 minutes, followed by 14 ml ZymoPURE P3, inverted gently to form a precipitate. Lysate was added to a ZymoPURE syringe filter with Luer Lock attached and allowed to settle for 5 minutes. Luer lock was then removed, the syringe placed on top of a 50 ml falcon tube and plunger pushed to filter out precipitate. The collected supernatant was mixed through gentle inversion with 14 ml ZymoPURE binding buffer. Zymo-spin V-P column with a 15 ml conical

reservoir was placed in a 50ml falcon. 15 ml of supernatant at a time was placed in the reservoir and DNA bound to the membrane in the column by centrifugation at 500 g for 2 minutes until all supernatant was passed through. The membrane was washed with 5ml ZymoPURE wash 1 and centrifuged at 500 g for 2 minutes; then washed twice more with 5ml ZymoPURE wash 2 and centrifuged at 500 g for 2 minutes. The column was then placed in a collection tube and spun at 10,000 g for 1 minute to dry the membrane. The column was then placed in DNA LoBind tube and DNA eluted with 400 μl ZymoPURE elution buffer through incubation for 2 minutes followed by centrifugation at 10,000 g for 1 minute. Plasmid concentration was measured using Nanodrop one; the quality of the DNA was deemed acceptable if the 260/280 ratio was between 1.6-2. The plasmids' identity was confirmed with sequencing by Genewiz.

2.2.4 Transfection of hiPSCs

hiPSCs were transfected with 0.2 μg of hTF plasmid and 0.2 μg of the rtTAM2 plasmid along with 0.4 μg of pBase helper plasmid per 800,000 hiPSCs, using Human Stem Cell Nucleofector Kit 2 (VPH-5022, Lonza) on the H9 protocol. Transfected cells were replated in warmed E8 media with 10 μM Y-27632 These plasmids integrate into the genome using the pBac transposase system (Chen et al., 2010). After two days, transfected cells were selected for successful transfection and plasmid integration with the cytotoxic agent G418 (10131035, Thermo Fisher Scientific) (100 μg/ml) for 7 days.

2.2.5 Generation of clonal populations of hiPSCs

Clonal cell lines were generated by depositing 1 cell per well of a 96 well plate using FACS Aria III (BD). Cells were maintained in E8 with 50 U/mL penicillin/streptomycin (15070063, Gibco) with 10 μM Y-27632 added just after sorting. Media was changed every 2 days, removing Y-27632 after 2 days. After 10 days, wells that contained single colonies were transferred to one well of a 12 well plate and further expanded. To test for mVenus expression, 1 μM Doxycycline (10592-13-9, Fisher) was added to one well of each clone and

mVenus expression was assessed using florescent microscopy. Alternatively, for the DSOX2 cell line, mVenus positive cells, stimulated with 1μM Doxycycline (10592-13-9, Fisher) were isolated using FACS Aria III (BD). This population was then plated at a low density; 10,000 cells per 35mm dish. Single colonies were then isolated using a needle and placed in a 48 well plate for cell expansion.

2.3 Cell analysis methods

2.3.1 Analysis of SOX2 expression and protein content in hiPSCs

Cells were permeabilised using BD cytofix/cytoperm (554714, BD Bioscience) following the manufacturer's recommendation. Briefly, hiPSCs were dissociated with ESGRO Complete Accutase, pelleted and resuspended in fixation/permebilisation solution for 20 minutes at 4^{o}C, washed with 1 ml BD Perm/Wash buffer twice and stained using 1:100 V450-conjugated anti-SOX2 (56160, BD Bioscience) in 500 μl BD Perm/Wash. Excess antibody was washed with BD Perm/wash twice. Stained and fixed cells were run on a LSR II analyser (BD Bioscience) to measure the intensities of these fluorophores and the tdT reporter.

2.3.2 Analysis of tdT high and tdT low reporter levels in SOX2-tdT hiPSCs

SOX2 hiPSCs were separated based on tdT expression using FACS Aria III (BD). tdT high and tdT low expressing cells were cultured for 10 passages, at each passage cells were dissociated with a TrypLE express (12604013, Life Technologies) to form a single cell suspension, pelleted, and resuspended in FACS buffer. Cells were run on an LSR II analyser (BD Bioscience) to measure the intensity of tdT reporter, while remaining cells were replated for the next passage.

2.3.3 FACS analysis

FACS analysis on the hiPSCs, iMeLCs or hPGCLCs samples was performed using FlowJo (BD, V.10). Forward and side scatters were used to isolate single cells through gating around these

populations. The same strategy was used within the same experimental replicate but altered between them to better capture the populations. When setting gates for positive populations for the various fluorophores, unstained cells were included as a negative control to set the gate correctly, again this changed depending on the experimental run.

2.3.4 Protein extraction and quantification from cell culture

One million cells were harvested using ESGRO Complete Accutase before being spun down at 200 g for 10 minutes. Protein was extracted from pellets using RIPA buffer (R0278, Sigma) for 30 minutes on ice. After centrifugation at max speed (<17,000g) for 30 minutes at 4^0C, supernatant was taken, and protein content measured using Pierce™ BCA protein assay kit (23225, ThermoFisher). Protein standards were prepared using the supplied BSA stock solution and BCA working reagent, prepared using a ratio of 50:1 reagent A to reagent B. This formed a working reagent, 200 μl per sample of this working reagent was combined with 10 μl of standard or protein sample in one well of a 96 well plate. Standards were prepared in triplicate and samples in duplicate. The plate was incubated at 37°C for 30 minutes. Absorbance at 592 nm was measured using FLUOstar Omega (BMG Labtech). Averages across the readings were taken for each standard and sample, with any outliers above or below 10% of the average removed. The blank (0 mg/ml) was subtracted from each standard and sample. A standard curve was generated by plotting mean absorbance of each standard against protein concentration and a linear regression line (y = mx +C) plotted in Microsoft Excel, with the intercept set at 0 absorbance, as this was the blank standard. Therefore, the gradient of the line or m was used to convert absorbance of protein samples to protein concentration of samples. Samples of an appropriate amount of protein were prepared with by diluting in 4X Laemmli buffer (1610747, Biorad) with 200 μM DTT (10162994, Fisher Scientific) and boiled at 95°C for one minute.

2.3.5 Western blot using sodium dodecyl sulphate polyacrylamide gel electrophoresis

A mini protean tgx gels Any kD 10-well (4569033, BioRad) in 1X Tris/Glycine/SDS buffer (1610732, BioRad) was loaded with 20ug of protein per sample and 7.5 µl Precision Plus Protein Dual Color Standards (1610374, BioRad). Gels were run at 150mV for 75 minutes using mini gel tank (Bio-Rad). Proteins were transferred onto Immun-Blot PVDF Membranes (1620174, BioRad) using Tris/Glycine buffer (1610734, BioRad) with 10% methanol at 300 mA.

Membranes were blocked with 3% skim milk powder (84615.05, WVR) in PBS-0.1% Triton (A16046.AE, ThermoFisher) for 3 hours. Membranes were cut before primary antibody incubation, see Table 1. After primary incubation overnight at 4°C, membranes were washed with PBS-0.1% Tween-20 3 times before secondary antibody incubation for 2 hours at room temperature (Table 2). To visualise proteins, membrane was exposed to HRP substrate on Amersham ImageQuant™ 800.

Table 1 Primary antibodies used for western blot visualisation.

Reagent name	Concentration	Supplier/ Catalogue no.
Rabbit Anti-NANOG	0.4 µg/ml	Abcam; #ab109250
Goat Anti-SOX2	1 µg/ml	Bio-Techne/R&D systems; #AF2018
Rabbit Anti-OCT4	0.75 µg/ml	Abcam; #ab181557
Rabbit Anti-VINCULIN	0.63 ng/ml	Abcam; #ab129002

Table 2 Secondary antibodies used for Western blot visualisation.

Reagent name	Concentration	Supplier/ Catalogue no.
Goat anti-rabbit HRP	0.2 µg/ml	Abcam; ab205718
Donkey anti-goat HRP	0.2 µg/ml	Abcam; ab205723

2.3.6 RNA Extraction

RNA was extracted from hPGCLCs and hiPSCs using RNA total Monarch® Total RNA Miniprep Kit (T2010; NEB). Collected hPGCLCs from FACS were stored in 300 μl DNA/RNA protection reagent at -20°C. Once defrosted, 300 μl RNA lysis buffer was added to each sample. Supernatant was passed through gDNA removal column by centrifugation at 16,000g for 30s and 600 μl EtOH was added to the flow-through. This mixture was transferred to an RNA purification column. RNA was bound to membrane by centrifugation at 16,000 g for 30 seconds, before 500 μl of RNA priming buffer was added to the membrane and the column centrifuged at 16,000 g for 30 seconds. The membrane was washed twice with RNA wash buffer and centrifuged once at 16,000 g for 30 seconds and a second time for 2 minutes. Finally, RNA was eluted with 50 μl Nuclease-free water (T2006-1; NEB). RNA concentration and quality was measured using RNA 6000 pico (5067-1513; Agilent) run on a 2100 Bioanalyzer (Agilent). RNA was used for complementary DNA (cDNA) generation or stored at -80°C.

2.3.7 cDNA generation

PrimeScript RT Reagent Kit with gDNA Eraser (RR047, Takara) was used on extracted RNA to generate cDNA. 500 ng RNA was made up to 8 μl total and was combined with 1 μl gDNA Eraser and 2 μl 5X gDNA Eraser Buffer and incubated at 42°C for 2 minutes. Mixture from this reaction was combined with 4 μl 5X PrimeScript Buffer 2, 1 μl PrimeScript RT Enzyme Mix I, 1.0 μl RT Primer Mix, 4 μl RNase Free dH2O and incubated at 37°C for 15 minutes. Reaction was inactivated by incubation at 85°C for 5 seconds.

2.3.8 Reverse Transcriptase-qPCR

Reactions were set up using master mix as in Table 3. 9 μl of master mix was added to each well per Multiplate™ 96-Well PCR Plates (MLL9651, BioRad) along with 1 μl cDNA. The primers were designed to only recognise the endogenous forms of each gene and are

detailed in Table 4. The qPCR reaction was recorded by C1000 touch Thermo Cycler (BioRad). An initial 95°C step for 10 mins to activate the Taq is followed by 39 cycles of extensions:

95 °C for 15s

60 °C for 30s

72 °C for 30s

Intensity of Sensi-dye was recorded after this step.

After a final 60 °C step for 30s, the final intensity of the dye is recorded. The cycle where the dye intensity reached half of its maximum intensity was calculated and given as quantification cycle (cq) value.

Table 3 Master mix for qPCR

Reagent name	Volume	Supplier/ Catalogue no.
2X SensiMix	5µl	Bioline Reagents Ltd; #QT650-05
10µM Forward primer	0.5µl	IDT; Custom order
10µM Reverse primer	0.5µl	IDT; Custom order
Nuclease-free H20	3µl	Thermo Fisher Scientific; #AM9937

Table 4 Primers used in qPCR

Primer name	Sequence
SOX2-F	TCAGGAGTTGTCAAGGCAGAGAAG
SOX2-R	GCCGCCGCCGATGATTGTTATTAT
OCT4-F	CTTGCTGCAGAAGTGGGTGGAGGAA
OCT4-R	CTGCAGTGTGGGTTTCGGGCA
NANOG-F	TTTGGAAGCTGCTGGGGAAG
NANOG-R	GATGGGAGGAGGGGAGAGGA
NANOS3-F	CCCGAAACTCGGCAGGCAAGA
NANOS3-R	AAGGCTCAGACTTCCCGGCAC
SOX17-F	GAGCCAAGGGCGAGTCCCGTA
SOX17-R	CCTTCCACGACTTGCCCAGCAT
TFAP2C-F	CGCTCATGTGACTCTCCTGACATCC
TFAP2C-R	TGGGCCGCCAATAGCATGTTCT

BLIMP1-F	CGGGGAGAATGTGGACTGGGTAGAG
BLIMP1-R	CTGGAGTTACACTTGGGGGCAGC
KLF2 -F	GTCCTTCTCCACTTTCGCCA
KLF2 -R	GGAGCGCGAGAAGGGAATG
tdT -F	GGCGAGGAGGTCATCAAAGAG
tdT-R	GATGACGGCCATGTTGTTGTC
36B4-F (Housekeeping)	TCTACAACCCTGAAGTGCTTGAT
36B4-R (Housekeeping)	CAATCTCGAGACAGACACTGG

2.3.9 Relative expression calculations

To calculate the relative expression of each gene with respect to the control without doxycycline, the raw ct values of the house keeping gene (heq) is taken from the experimental gene (ecq) to give a CTE value. The CTE value of the control sample (CCTE) is taken from the experimental sample (ECTE) and then raised to the exponential 2. This gives a value for the relative expression (RE) of the experimental gene in the experimental sample compared to control sample.

$$CTE = ecq - heq$$

$$RE = 2^{(\mathrm{ECTE\text{-}CCTE})}$$

2.4 Induction of human primordial germ cell like cells

2.4.1 hPGCLC induction in 3D aggregates

The method used for generation of hPGCLCs in 3D culture has been described previously (Sasaki et al., 2015) with some variations. Briefly, hiPSCs grown in 6 well plates were dissociated with 1 ml TrypLE express to form a single cell suspension. The reaction was quenched with 4 ml TrypLE wash (Table 5).

Table 5 TrypLE wash

Reagent name	Concentration	Supplier/ Catalogue no.
7.5% Bovine serum albium Fraction V	1.3%	Life Technologies; #15260037
Glasgow's Minimal Essential Media (GMEM)	Up to 500ml	Life Technologies; #21710082

hiPSCs cell suspensions were pelleted at 200 g for 3 minutes, and resuspended in 1ml GK15, (Table 6), supplemented with reagents for iMeLC induction, see Table 7. To check cell viability, 10 ml hiPSCs dilution were mixed in a 1:1 dilution with 0.4% trypan blue stain (T10282; Thermo Fisher Scientific), and then 10 ml of this solution was pipetted into a Countess Cell Counting Chamber Slides (C10228; Life Technologies) using the Countess II system (Thermo Fisher Scientific). Once counted, 50-70x10^3 cells per ml dilutions were made up using GK15 with iMeLC reagents and 1ml per well was plated on a fibronectin (FC010; Millipore) coated 12-well cell culture plate.

Table 6 GK15 media

Reagent name	Concentration	Supplier/ Catalogue no.
Sodium Pyruvate	1 mM	Life Technologies; #11360039
L-glutamine	2 mM	Life Technologies; #25030024
Non-essential ammino acids	0.1 mM	Life Technologies; #11140035
2-Mercaptoethanol	0.1 mM	Life Technologies; #31350010
KSR	15 ml	Life Technologies; #A3181502
GMEM	Up to 100 ml final volume	Life Technologies; #21710082

Table 7 Reagents for iMeLC induction

Reagent name	Concentration	Supplier/ Catalogue no.
Activin A	50 ng/ml	QKine; #Qk001
CHIR99021 (CHIR)	3 µM	Cambridge Bioscience; #SM13
Y-27632	10 µM	Stemcell Technologies; #72304

After 60 hours differentiation, iMeLCs were dissociated with 0.4ml TrypLE express to form a single cell suspension. The reaction was quenched with 1.6ml TrypLE wash, (Table 5). Cells were counted again in the method described in the paragraph above. 3 x 10^3 of iMeLCs per well were aggregated into Nunclon Sphera-Treated, U-Shaped-Bottom Microplate (174925, Thermo Fisher Scientific) in 100 µl GK15 supplemented with reagents for hPGCLC induction, (Table 8).

Table 8 Reagents for hPGCLC induction in 3D aggregates

Reagent name	Concentration	Supplier/ Catalogue no.
Bone Morphogenic Protein 4 (BMP4)	200 ng/ml	Bio-Techne/R&D systems; #314-BP
Leukaemia Inhibitory Factor (LIF)	1,000 U/ml	Merck-Millipore; #LIF1005
Stem Cell Factor (SCF)	100 ng/ml	Bio-Techne/R&D systems; #255-SC
Epidermal Growth Factor (EGF)	50 ng/ml	Bio-Techne/R&D systems; #236-EG
Y-27632	10 µM	Stemcell Technologies; #72304

After 4 days, aggregates were dissociated using 0.25% Trypsin/PBS (15090046, Life Technologies) at 37°C, shaking at 1,400rpm. After digestion (~15-30 minutes) trypsin was neutralised using 0.5% BSA/PBS (15260037, Life Technologies) and cell suspension passed through a cell strainer (10585801, Fisher) to remove clumps and debris. Cells were pelleted at 200 g for 10 minutes; supernatant removed. Pelleted cells were stained with BV421-conjugated anti-human CD49f (B313623, BioLegend) and APC-conjugated anti-human CD326 (324207, BioLegend) in FACS buffer [0.1% BSA/PBS] for 30 minutes at room temperature and with gentle shaking. The volume of antibodies varied, depending on number of cells stained; in brief 5 µl of each antibody was added to 100 µl FACS buffer per single 96-well plate. Cells were sorted using FACS Aria III (BD) for cells that expressed both markers or the endogenous fluorophores in the BTAG cells, which were recovered for further analysis.

2.4.2 hPGCLC culture in 2D culture

The method used for generation of hPGCLCs in 3D culture has been described previously (Overeem et al., 2023). Briefly, hiPSCs grown in 6 well plates were dissociated with ESGRO Complete Accutase. hiPSCs were pelleted at 200 g for 3 minutes, supernatant removed and resuspended in 1ml E8 essential media with 2% Geltrex added. Cells were counted again in the method earlier described 2.4.1. 60,000 cells/cm^2 were plated onto Geltrex coated plates in E8 essential media with 2% Geltrex and 10 µM Y-27632. After 24 hours, this media was replaced by 2D hPGCLC induction media, aRB27 (Table 9) with added cytokines, (Table 10),

with 2% Geltrex (A1413302; Life Technologies). Media was changed every day with 2% Geltrex (A1413302; Life Technologies) included in the media until the 3rd day post plating. The final two days, only the hPGCLC induction media was added to the well.

Table 9 aRB27 media

Reagent name	Volume/Concentration	Supplier/ Catalogue no.
RPMI 1640 + glutamax	100 ml	Thermo Fisher Scientific; #61870036
B27 supplement	1 ml	Cambridge Bioscience; #SM13
Non-essential ammino acids	0.1 mM	Life Technologies; #11140035

Table 10 Reagents for hPGCLC induction in 2D culture

Reagent name	Concentration	Supplier/ Catalogue no.
Bone Morphogenic Protein 4 (BMP4)	10 ng/ml	Bio-Techne/R&D systems; #314-BP
Leukaemia Inhibitory Factor (LIF)	1000 U/ml	Merck-Millipore; #LIF1005
Stem Cell Factor (SCF)	50 ng/ml	Bio-Techne/R&D systems; #255-SC
Epidermal Growth Factor (EGF)	50 ng/ml	Bio-Techne/R&D systems; #236-EG
Y-27632	10 µM	Stemcell Technologies; #72304

After 5 days, a single cell suspension of the 2D cultures was prepared by adding ESGRO Complete Accutase to each well and incubating at 37°C, along with mechanical disruption through pipetting. Accutase was quenched with PBS (14190094; Life Technologies). Cells were pelleted at 200 g for 10 minutes; supernatant removed. The same method for staining and FACS used on the 3D aggregates was used on the 2D cells to isolate the hPGCLCs derived through this culture system.

2.4.3 Transcription factor overexpression activation during hPGCLC induction

Doxycycline (10592-13-9, Fisher) concentrations between 1-50 μM were added at the times shown in section 3.3. The same method for isolation described above were used to stain and sort cells for hPGCLC surface markers using FACS. tdTomato, reporter for endogenous SOX2, and mVenus, reporter for transcription factor, intensities were also recorded.

2.4.4 Cyrosectioning of hPGCLCs

hPGCLC aggregates were fixed in 4% PFA for 25 minutes, then dehydrated in 30% sucrose overnight at 4^{o}C until the aggregates had sunk to the bottom of the well. 3 to 4 aggregates were then transferred per 7 X 7 X 5 mm mould (720-0820, VWR), embedded in OCT (361603E, VWR) and snap frozen on dry ice. Embedded aggregates were fully frozen overnight at -200C. 10mm sections of these embedded aggregates were cut using a CM3050 S cryostat (Lecia) and sections placed on SuperFrost® Plus slides (MIC3040, SLS).

2.4.5 Immunohistochemistry on cryosections

Excess OCT was washed off with DPBS. Slides were bordered using ImmEdge Pen (H-4000, Vector laboratories). Tissue was permeabilised using PBS + 0.1% Triton 100X for 30 minutes, followed by blocking with 200 μl blocking buffer for 2 hours, see

Table 11, covered with a microscope cover slip (631-1574, VWR). 100 μl primary Abs were applied overnight at 4^{o}C covered with a coverslip, (Table 12). Slides were then washed 3 times with 250 μl PBS-0.1% tween-20 for 10 minutes each time. 100 μl of secondary antibodies was applied to slides, (Table 13), covered with a coverslip for 2 hours at room temperature. Slides were then washed 3 times with 250 μl PBS-T for 10 minutes each time. To stain DNA, 150 μl 0.1μg/ml DAPI solution (62248, Life Technologies) was applied to slides; before a final wash with PBS for 10 minutes. Excess PBS was wiped away gently with tissue and air dried for 10 minutes. To fix coverslips to slides, 50 μl Vectashield Vibrance (H-1800-

10, Vector Labs) was added in three drops to make up 50 µl to each slide to affix a coverslip. Slides were stored at 4°C.

Table 11 Blocking buffer for IHC

Reagent name	Concentration	Supplier/ Catalogue no.
PBS	-	Life Technologies; #14190094
Tween-20	0.3%	ACROS organics; #233362500
Triton X100	0.1%	Thermo Fisher; #A16046.AE,
Normal serum (Donkey)	3%	Generon; #SUD004
7.5% Bovine serum albumin Fraction V	1%	Life Technologies; #15260037

Table 12 Primary antibodies used for IHC visualisation

Reagent name	Concentration	Supplier/ Catalogue no.
Rabbit Anti-OCT4	0.2 µg/ml	Abcam; #ab181557
Sheep Anti-mVenus	0.04 µg/ml	Bio-rad; 4745-1051
Mouse Anti-AP2	0.4 µg/ml	Santa-Cruz; sc-12762
Rabbit Anti-SOX2	0.2 µg/ml	Proteintech; #20118-1-AP

Table 13 Secondary antibodies used for IHC visualisation

Reagent name	Concentration	Supplier/ Catalogue no.
Alexa fluor 647 Donkey Anti-mouse	0.2 µg/ml	Abcam; #ab150107
Alexa flour 594 Donkey Anti-rabbit	0.2 µg/ml	Thermo Fisher Scientific; R37119
Alexa flour 488 Donkey Anti-sheep	0.2 µg/ml	Thermo Fisher Scientific; A11015

Chapter 3. Generation and characterisation of cells lines overexpressing pluripotency related transcription factors

3.1 Introduction

Human primordial germ cell like cells (hPGCLCs) can be derived from hiPSCs, allowing study on early human germline development. In this study, I wanted to observe what effect overexpression of different pluripotency related transcription factors would have on germline induction. Therefore, I transfected hiPSCs with a set of plasmids containing these TFs, which could be overexpressed by the addition of an antibiotic called doxycycline (dox). Once generated, these hiPSCs could then be used in the hPGCLCs induction protocol and dox can be added at different points during this induction to trigger the overexpression of the TFs.

The three pluripotency related TFs I chose to overexpress were *SOX2*, *NANOG* and *KLF2*. The effect of these factors being overexpressed in hiPSCs are detailed in chapter 1, section 1.1.3.

3.1.1 Summary of findings in regard to pluripotency TF overexpression

SOX2 and *NANOG* have both been overexpressed in hESCs either on feeders or in feeder conditioned media, while *KLF2* overexpression in hPSCs has not been published. *NANOG* overexpression in hESCs caused differentiation towards primitive ectoderm and allowed for the growth and passaging of hESCs without feeders (Darr et al., 2006). *SOX2* overexpression triggered trophectoderm differentiation in hESCs (Adachi et al., 2010).

In this chapter I will compare the phenotype of *SOX2* and *NANOG* overexpression in feeder free conditions, as well as pioneer the effect of *KLF2* overexpression in hiPSCs.

3.2 Specific aims

1. Generate clonal hiPSC cell lines which over express a pluripotency TF and reporter gene in response to dox addition.
2. Analyse the effects of over expression of each pluripotency TF in these hiPSCs in feeder free conditions.
3. Confirm the genomic integrity of the main clones used in this study.

3.3 Results

3.3.1 Characterisation of SOX2-tdTomato endogenous reporter cell line

For this study, I have used a human induced pluripotent stem cell (hiPSC) line, a kind gift from the Otonkoski lab containing a tdTomato reporter linked to the one copy of endogenous transcript of *SOX2* but cleaved from the protein by a T2A site after translation (Balboa et al., 2017). This fluorescent protein reporter allows real-time quantitative analysis of *SOX2* expression dynamics. These cells are herein termed endogenous SOX2-tdT. In standard Essential E8 (E8) culture conditions, the SOX2-tdT hiPSCs grow as typical, well demarcated human PSC colonies (A) (Viswanathan et al., 2014). The expression of the SOX2 reporter (tdTomato protein) can be visualised readily using fluorescence microscopy B) which is localised to the nucleus due to the presence of a nuclear localisation signal (NLS).

The tdT reporter shows a speckled pattern in this colony, which was analysed using flow cytometry to measure the intensity of the reporter in each cell (C). The SOX2-tdT reporter displayed two florescence intensities, a 'high' and a 'low' within the cell population. These were designated tdT high and tdT low populations. Both populations had a higher tdT florescence intensity than a cell line which contains no tdT fluorophore (C). Both populations disappear upon hPGCLCs induction (D), a cell type that does not express SOX2 (Irie et al., 2015; Sasaki et al., 2015).

The flow cytometry analysis showed a bimodal pattern that was not present in the original study (Balboa et al., 2017), leading me to try and determine the origin of these two populations. To determine if both these populations expressed SOX2 protein, SOX2-tdT cells were stained with a SOX2 antibody to measure the protein content and the intensities of tdT using flow cytometry (E). Both tdT populations stained highly for SOX2 protein suggesting that regardless of whether a cell was tdT high or tdT low, it expresses SOX2 and if a cell was tdT negative, the cell was not expressing SOX2.

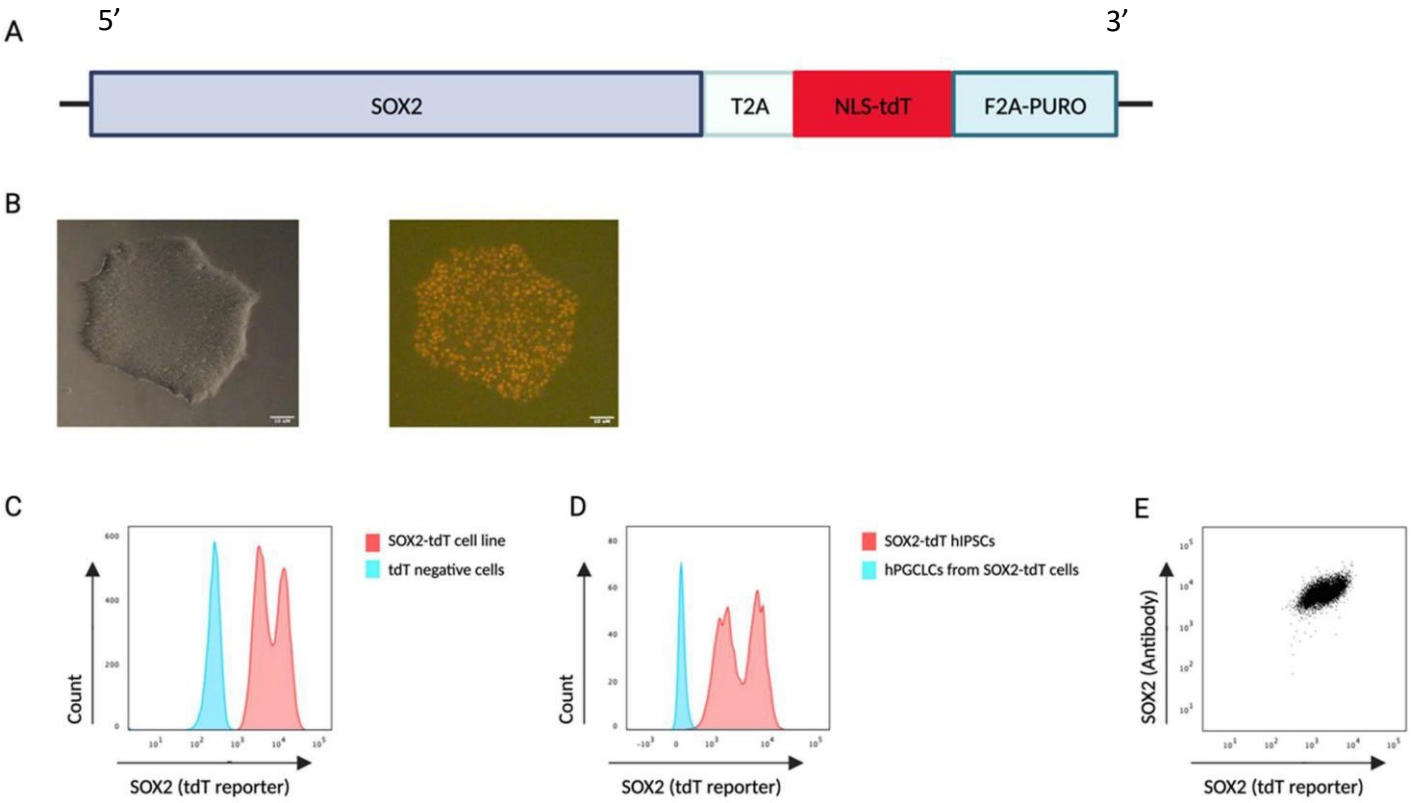

Figure 7 Culturing of SOX2 tdTomato (A) Map of SOX2 locus with the cassette used to target the 3’ end of the SOX2 gene to form the fusion (B) Representative phase contrast image of hiPSC colonies in E8 and the same field imaged under fluorescence showing SOX2 tdT reporter, Scale bar = 10mm (C) Flow cytometry analysis of SOX2-tdT hiPSC and a cell line which contains no such reporter (D) Histograms of hiPSCs and hPGCLCs both from SOX2-tdT cell line (E) Plot of tdT reporter vs. an internal stain of SOX2 protein on SOX2-tdT hiPSCs.

3.3.2 Expression dynamics of *SOX2*-tdT reporter.

To determine if the bimodal population was due to long-term culture of SOX2-tdT cells, high tdT cells and low tdT cells were separated by FACS and replated (Figure 8 A & B). The tdT intensity was measured by flow cytometry in each population, at each passage and for 10

passages (Figure 8 C). It was clear from the histograms that the percentage of cells that expressed high tdT never increased, in either separated populations, but the percentage of Low tdT cells did increase. The percentage of cells within the 'low' gate in each population was measured at each passage and plotted (Figure 8 D) to determine how this population increased during culture. Unsorted or whole population cells were included from passage 4 to assess variation that occurred in the culture without separation of the two populations during 6 passages.

A percentage of tdT low cells emerged from the sorted tdT high population even at passage 1. The percentage of these tdT low cells rose with each passage suggesting they emerged from the high population and increased in frequency within the culture. It was not possible to determine if the tdT low cells which emerged from the high population were not from contamination during FACS. Their number increased during the 10 passages, suggesting that they may have had a competitive advantage, regardless of how they emerged. The low population in contrast never regained any tdT high cells and even appeared to eliminate the small percentage of tdT high cells that were present in the first passage. Overall, this suggests that the tdT low population in the SOX2-tdT probably emerged from an initially high population, then expanded in the culture, either through competitive advantage, or more tdT highs becoming low.

To determine if there was a correlation between tdT intensity and SOX2 protein levels, cells were fixed, stained for SOX2 protein, and analysed by flow cytometry. However, when fixed, it appeared that cells lost the bimodal pattern of the reporter expression and in general had slightly lower reporter intensity overall (Figure 9 A). Nevertheless, there still appeared to be a range of tdT reporter intensity in the fixed cell population; these were designated as high and low (Figure 9 B). The intensity of the *SOX2* protein within these two populations was similar, as shown in the histograms and scatter plots in Figure 9 B (middle and right panel).

Live cells were sorted and pelleted to extract total RNA and protein. Western blots on these sorted tdT high cells, tdT low cells, and whole populations (unsorted cells) showed they all contained SOX2 protein (Figure 9 C) regardless of the level of reporter gene.

To determine if the difference in tdT intensities was due to differences in expression from the two SOX2 alleles, gene expression was directly measured using RT-qPCR analysis. SOX2 showed a slight increase in expression in tdT high cells compared to tdT low cells (relative expression mean SOX2 = 1.21) but similar relative expression of OCT4 (relative expression mean OCT4 = 0.945). The expression of tdT was higher in 'highs' and significantly higher than the relative expression of SOX2. This was unexpected as the reporter is directly fused to the end of one of the SOX2 alleles; and so, the expression of tdT RNA can only come from the expression from that SOX2 locus. This suggests that while the tdT high cells and tdT low cells express a similar amount of SOX2 RNA and protein, the tdT high cells have higher expression from the tagged locus and the tdT low cells have higher expression from the untagged. The emergence of the tdT low cells from the tdT high population (Figure 8 C) suggests that expression from the untagged locus may be preferential or even provide a competitive advantage allowing this population to increase in proportion. This switching of alleles in PSCs has been observed at the *Nanog* allele in mouse ESCs (Miyanari and Torres-Padilla, 2012).

Overall, despite the bimodal expression of the reporter, the molecular analysis suggests this reporter cell line is a good model for tracking SOX2 expression accurately. The difference between 'high' and 'low' tdT levels can be discounted when considering the amount of SOX2 RNA within single cells and it is clear that SOX2 protein is present in both the 'highs' and 'lows' expression, and hence the presence of the reporter indicates the presence of SOX2 within the cell.

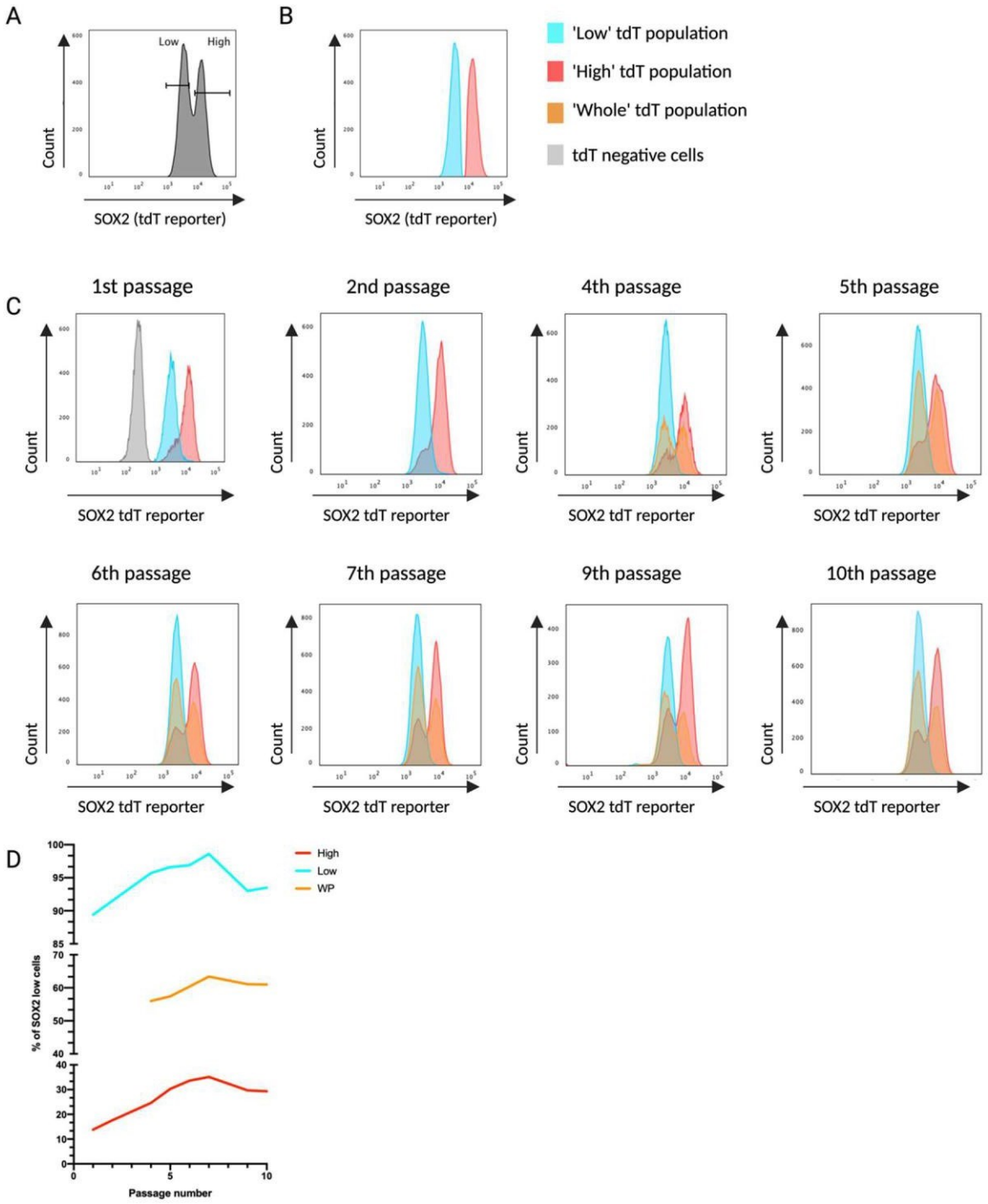

Figure 8 Emergence of the 'low' tdT population from the 'high' tdT populations over 10 passages (A) Flow cytometry gates used to sort 'high' and 'low' populations (B) Resulting populations from this separation (C) Flow cytometry analysis showing the tdT intensities in 'high' (red), 'low' (blue), 'whole' populations (orange) and Miff1 (grey) in the 1st passage as a negative control. (D) The percentage of 'low' tdT cells in the 'high' 'whole population (WP)' and 'low' as defined by the low population in the 'WP' histograms

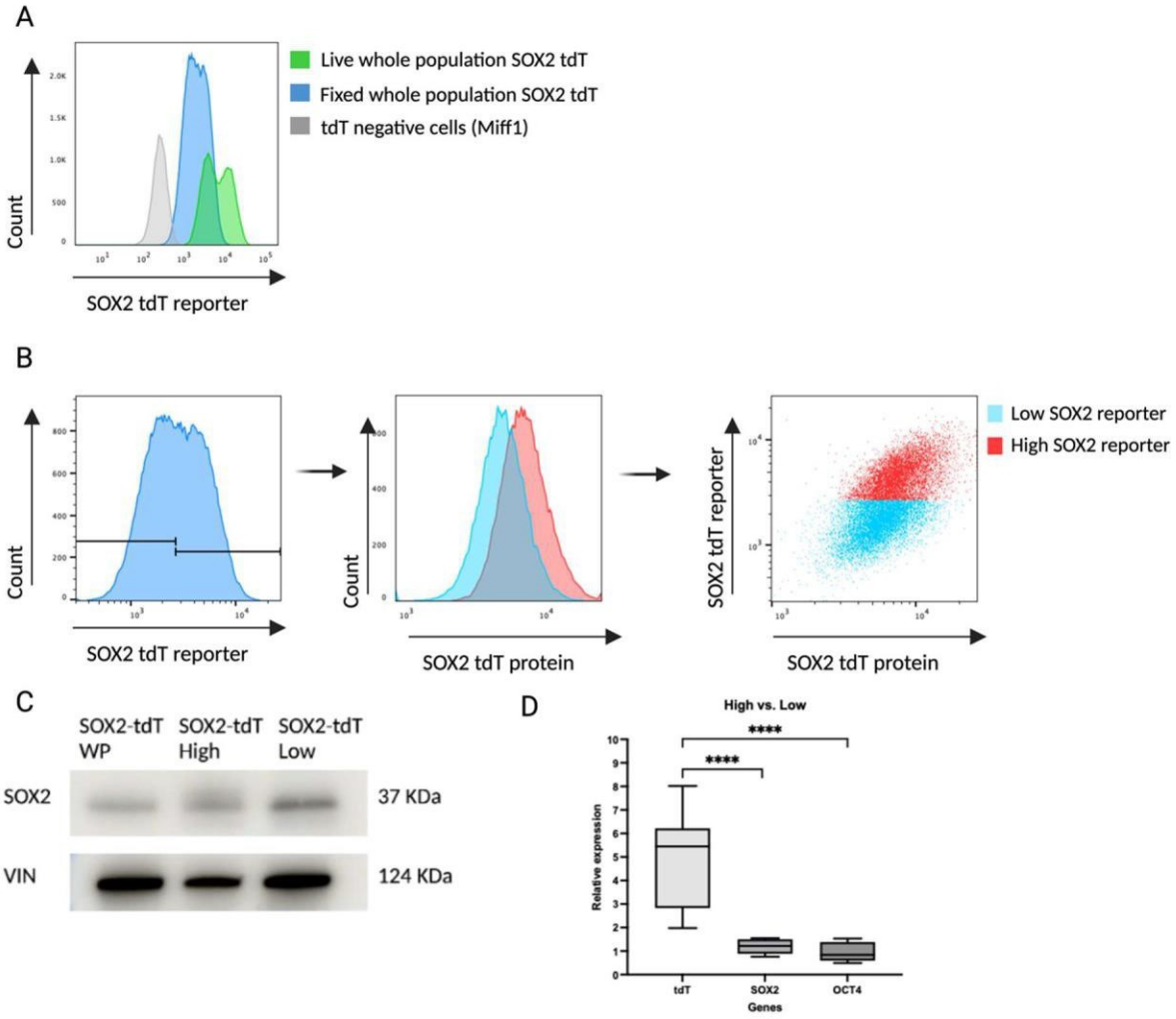

Figure 9 Molecular analysis of 'high' and 'low' tdT populations within SOX2-tdT hIPSC cultures (A) Flow cytometry analysis of tdT reporter in live SOX2-tdT, fixed SOX2-tdT cells and Miff1 (B) Left: Zoomed in section of Flow cytometry analysis of tdT reporter in fixed SOX2-tdT cells with gating for 'high' and 'low' populations. Middle: The SOX2 protein levels in these populations as measured by flow cytometry analysis of a SOX2 antibody. Right: Flow cytometry analysis of SOX2 reporter vs. SOX2 antibody in the 'high' and 'low' populations (C) representative Western blot of SOX2 protein in 'Whole populations', 'High' and 'Low' populations in SOX2-tdT hiPSCs, VIN stands for Vinculin which was used as a loading control (D) qPCR analysis showing relative expression of tdT, SOX2 and OCT4 genes in 'high' cells compared to 'low' cells, n=5, $p < 0.0001$, ordinary one-way ANOVA.

3.3.3 Strategy for inducing overexpression of pluripotency related transcription factors in SOX2-tdT cells

In order to control the overexpression of different pluripotency-related transcription factors, a Tetracycline-on system was used (Takashima et al., 2014); plasmids were a kind gift from A.Smith, Y.Takashima and H.Niwa. One of the TF plasmids containing either SOX2, NANOG, KLF2 or the 'empty vector' (EV) which contained no TF, along with the rtTA-M2 and the pBase helper plasmid (Figure 10 A), were transfected into SOX2-tdT hiPSCs using nucleofection. Cells which had taken up all three and had integrated the rTA-M2 and TF plasmid were then able to respond to doxycycline (dox) through the mechanism shown in Figure 10 B and be resistant to Neomycin selection. Once stimulated by dox, the TF plasmid produces a transcript containing the TF and the mVenus reporter. While transcribed together, the mVenus reporter is translated through an internal ribosome entry site (IRES), its translation and post-translational regulation is independent to the associated TF. The mVenus is not fused to a NLS and its expression will be diffused through the cell whereas the TFs produced will be shuttled into the nucleus to perform its function.

The scheme shown below (Figure 10 C) was then used to obtain clonal cell lines from each transfected pool. For KLF2, NANOG and EV cell lines, the single cell sort method was used to obtain clonal lines. For the SOX2 transfected cells, which will henceforth be known as double SOX2 or DSOX2 due to the presence of two SOX2 reporters, the dox pre-treatment and low-density plating were followed by manual colony picking. I used this method for the DSOX2 as the single cell sorting approach was not always successful for generating clonal cell lines.

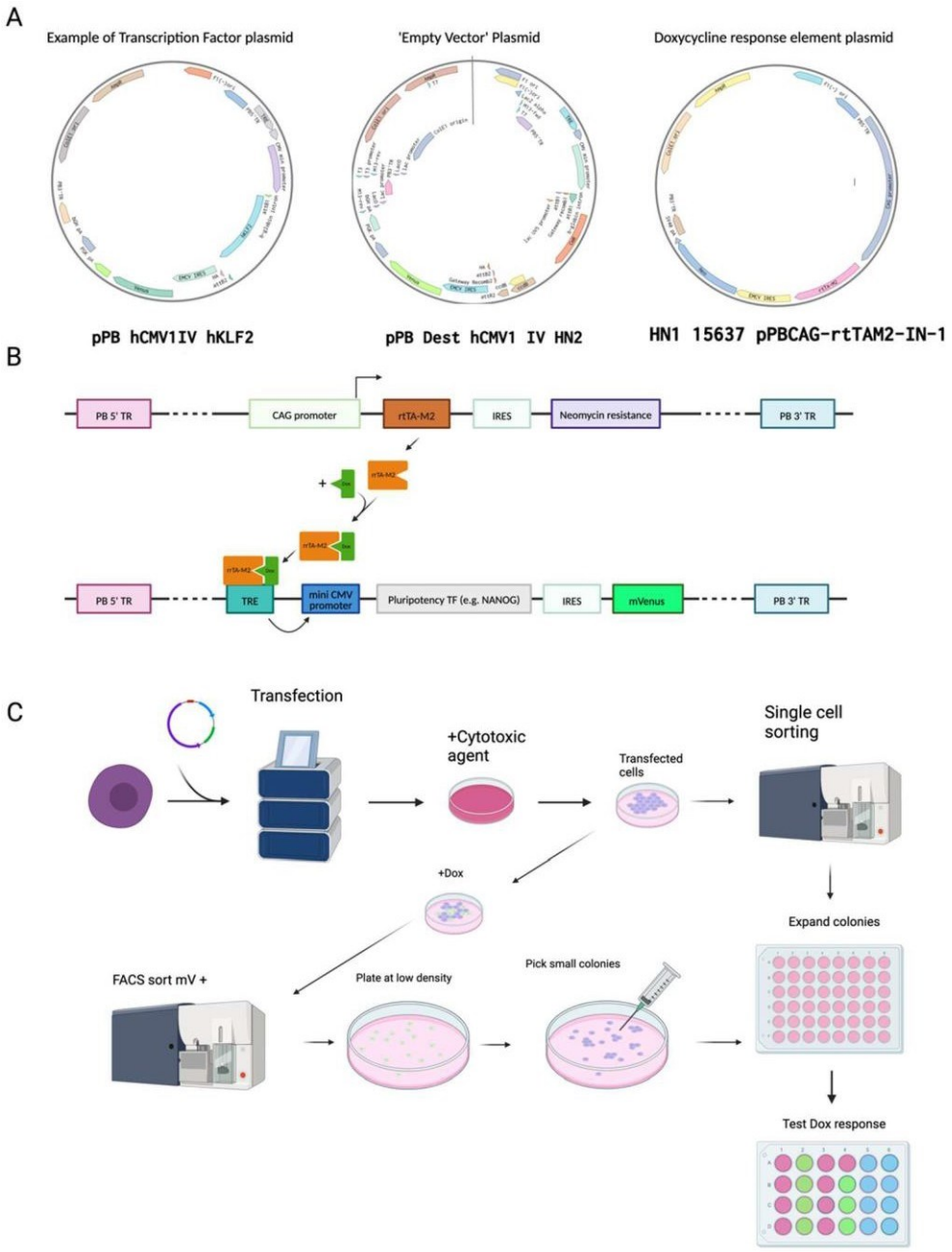

Figure 10 Strategy for generating hIPSCs clones with inducible pluripotency factors (A) Plasmid constructs used in the different transfections of hiPSCs to overexpress the TF of interest. Left: example of plasmid containing the transcription factor of interest (KLF2 shown here) and the mVenus reporter. Middle: 'empty vector' which contains no TF and acts as a negative control. Right: rtTAM2 plasmid which produces the protein responding to dox (B) Schematic showing how the addition of dox to the transfected cells triggers the expression of said TF and mVenus reporter (C) Scheme followed to generate clones containing both rtTAM2 and TF plasmids. Two approaches (single sort or manual picking) were used post transfection and are shown through the diverging paths.

3.3.4 Generation and characterisation of DSOX2 overexpressing cells

The SOX2-tdT cell line was transfected with SOX2 overexpressing plasmid, and clones from these cells were termed Double SOX2 or DSOX2 as they contained two SOX2 reporters. After 1 day of dox treatment, mVenus was highly expressed and observable with both flow cytometry analysis and fluorescence microscopy (Figure 11 A & B). SOX2 protein was overexpressed in dox treated cells to levels above the baseline in untreated cells as shown by western blot (Figure 11 D). This overexpression is so much higher than the baseline level, the expression of the endogenous *SOX2* is barely detectable in the western blot. The saturation of the HRP reaction in the exogenous bands makes it appear that there is little to no expression of *SOX2* in the non-dox treated cells. This data confirms that dox treatment could induce expression of both reporters and SOX2 after 1 day.

DSOX2 dox treated cells appear to have normal morphology with no evidence of differentiation after 3 days of dox treatment. Cells treated with dox for 3 days continued to grow unchanged after dox is removed from the media. Longer dox stimulation, and therefore SOX2 overexpression appeared to cause cell death. After 5-6 days, hiPSCs are normally confluent, colonies compact and their boundaries are clearly defined (Figure 7 B & Figure 11 A). After 3 days of dox treatment, colonies look less compact than their untreated counterparts (Figure 11 A) and this becomes more pronounced after 5 days of dox treatment (Figure 11 A). After 5 days of dox treatment, DSOX2 cells appear sparse, beginning to detach and may be showing signs of cell death (Figure 11 A). This suggests that sustained overexpression of SOX2 in hiPSCs causes cell death.

DSOX2 clone 1 (c1) was selected as the main clone for this study. It had a detectable mVenus expression as measured by flow cytometry analysis and an increase in SOX2 protein in response to dox addition (Figure 11 C & D). To confirm this clone's genomic integrity, low pass whole genome sequencing was performed with the result analysed by Dr George Young, LMS genomic facility (Figure 11 E). Sequencing showed no major duplications or deletions in the genome of DSOX2 c1 and confirmed these cells were male XY genotype.

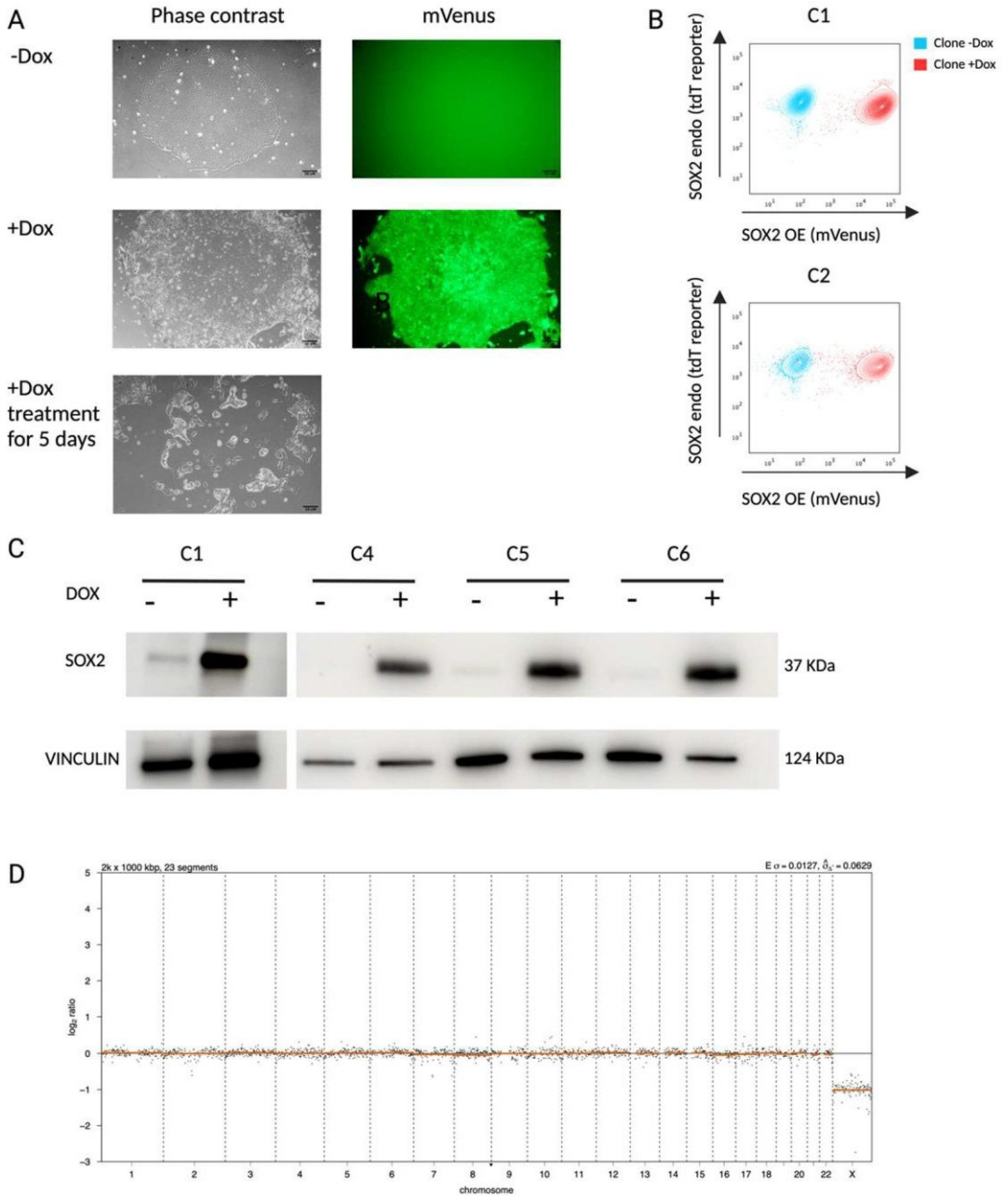

Figure 11 Characterisation of DSOX2 clonal cell lines. (A) Phase contrast and fluoresce images to show the expression of the mVenus reporter. Top panel: no dox added to the culture. Middle panel: 1mM dox added for 3 days. Bottom panel: 1 mM dox added for 5 days Scale bars= 50mm (B) Flow cytometry analysis of mVenus and tdT reporters in two DSOX2 clones after treatment with (red) or without (blue) 1mM dox for 1 day (C) Western blot of four DSOX2 clones after treatment with or without 1mM dox for 1 day (D) Low-pass whole genome sequencing of DSOX2 clone 1 to show genome integrity of main clone used in further experiments. Y axis, Log2 ratio number of reads within a locus vs. mean number of reads within whole genome, meaning a value of 0 indicates normal copy number and -1 (as for the X chromosome) means half copy number (i.e. copy number of 1).

3.3.5 Generation and characterisation of NANOG overexpressing cells.

The SOX2-tdT cell line was transfected with NANOG overexpressing plasmid, (Figure 10 A), and clones from these cells were termed NANOG OE cells. NANOG OE cells which were treated with dox upregulated both the NANOG protein and the mVenus reporter used to track transgene expression from the inserted locus. Flow cytometry analysis and fluorescent microscopy both showed clear mVenus expression in response to dox (Figure 12 A & B). This expression was accompanied by an upregulation and production of NANOG protein (Figure 12 C), which was most pronounced in NANOG OE c26. Therefore, this clone was used for the rest of the experiments in this study. This clone's genomic integrity was confirmed through low pass whole genome sequencing, analysis performed by Dr George Young, LMS genomic facility (Figure 12 D). No deletions or duplications were detected in this sequencing and the XY genotype was returned as expected.

Confluent hiPSC colonies normally present as a tight cluster of cells with a well-defined edge which appears raised from the Geltrex coating (Figure 7 B & Figure 12 A). In comparison, when stimulated with 1 μM dox, NANOG OE cells appear flatter, the boundaries of each cell are more distinct and do not form tight colonies (Figure 12 A). Additionally, cells separated from the colonies appear to survive well, suggesting dox-treated NANOG OE cells may survive as single cells when passaging, a hypothesis which I tested later on in this study. This altered morphology is reversible, with NANOG OE cells resembling WT hiPSCs colonies once dox has been removed and cells passaged using EDTA.

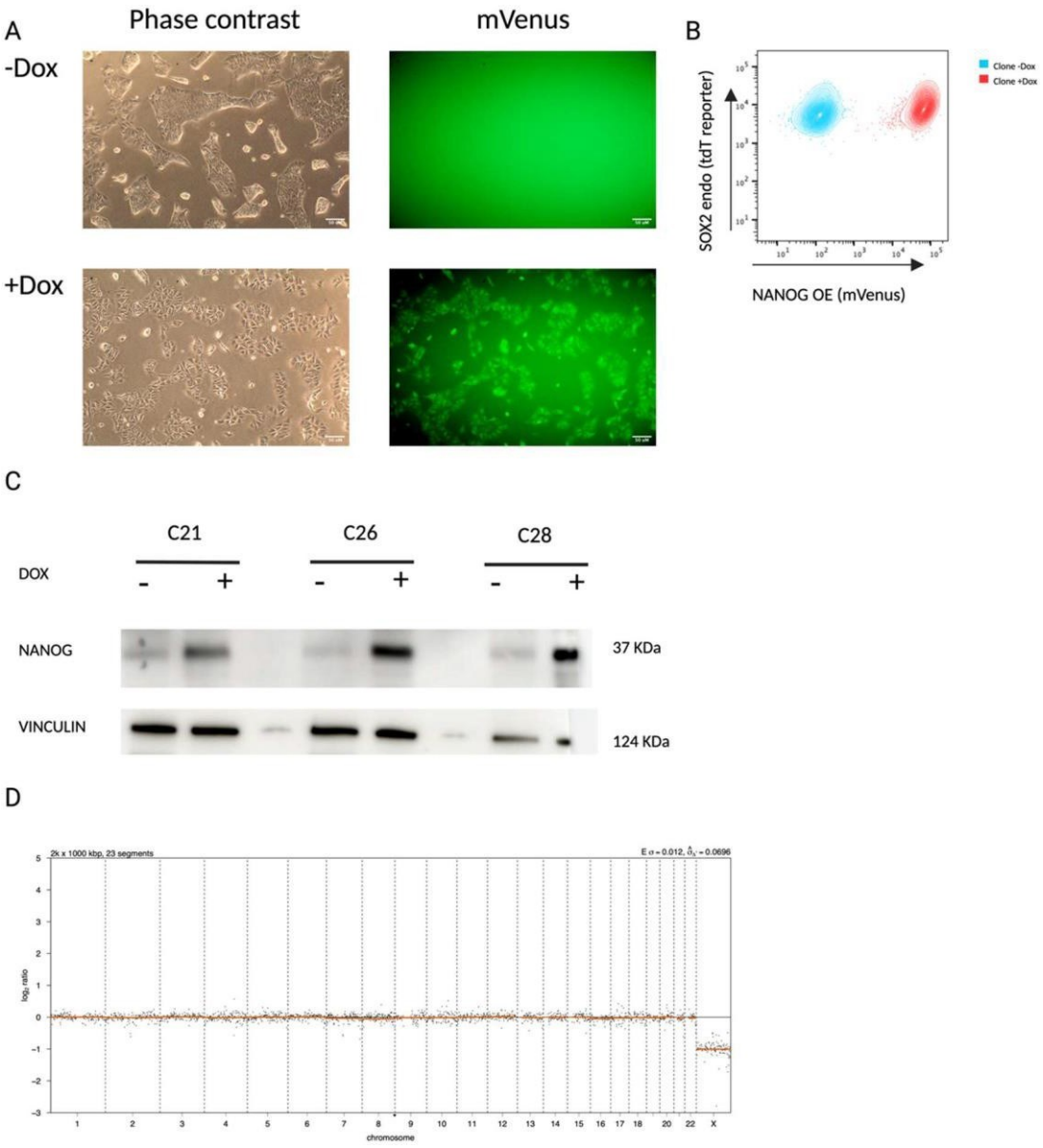

Figure 12 Characterisation of NANOG clonal cell lines (A) Images of Phase contrast and the sample field imaged to show expression of mVenus reporter. Top: no dox added to culture. Bottom: 1mM dox added for 3 days Scale bars = 50mm (B) Flow cytometry analysis of mVenus and tdT reporters in NANOG clone 26 after treatment with (red) or without (blue) 1mM dox for 1 day (D) Western blot of 3 NANOG clones after treatment with or without 1mM dox for 1 day (E) Low-pass whole genome sequencing of NANOG clone 26 to show genome integrity of main clone used in further studies.

hiPSCs and hESCs tend to die if plated as single cells with reports suggesting that any clump of 10 cells or less is susceptible to differentiation (Reubinoff et al., 2000a). I have used the compound Y-27632, a Rho kinase inhibitor (ROCKi), which has been shown to support single

cell plating of hESCs and hiPSCs (Watanabe et al., 2007). Dox treated NANOG OE cells appeared to survive as single cells outside of the colonies (Figure 12 A).

To study the effects of NANOG overexpression and how it may support survival of single cell plated hiPSCs, I set up an experiment to test the ability of dox treated cells to survive passaging as single cells without ROCKi. I set up four initial conditions, NANOG c26 or the negative control 'EV' cell lines, (described in section 3.3.7) were either pre-treated with 1μm dox, termed treated, or not treated with dox, termed untreated. These four conditions were then dissociated into single cells using Accutase and replated into three conditions, +ROCKi, +dox or -dox. Media change was then performed daily. +ROCKi was used as a positive control for cell viability under normal single cell plating conditions. It is worth noting that the inhibitor is removed from the media after the 1st day as it is standard practice in hiPSC culture.

After 5 days, wells were assessed for evidence of cell growth, and representative colonies in each well were photographed by phase contrast microscopy. All cell lines were viable to be passaged as all wells which contained ROCKi grew colonies. Wells in which neither ROCKi nor dox were added gave no colonies after 5 days, as expected. However, pre-treated NANOG cells which were re-plated with dox were able to form colonies. This was the only well not treated with ROCKi that gave rise to colonies. Treated 'EV' cells with dox did not form colonies showing that this was not due to pre-treatment and continued stimulation with dox, but the overexpression of NANOG in the pre-treated cells. Interestingly, NANOG cells that were not pre-treated with dox were not able to be passaged as single cells even when dox was added. This suggests NANOG must be overexpressed prior to single cell dissociation and replating (Figure 13).

Comparisons between the effectiveness of NANOG overexpression and the standard ROCKi addition for single cells hiPSCs passaging could not be measured in this experiment, as the cell numbers added to each well were not controlled. A colony forming assay where a known number of cells are plated per well before staining with alkaline phosphatase to check for pluripotent colonies would need to be performed to answer this question.

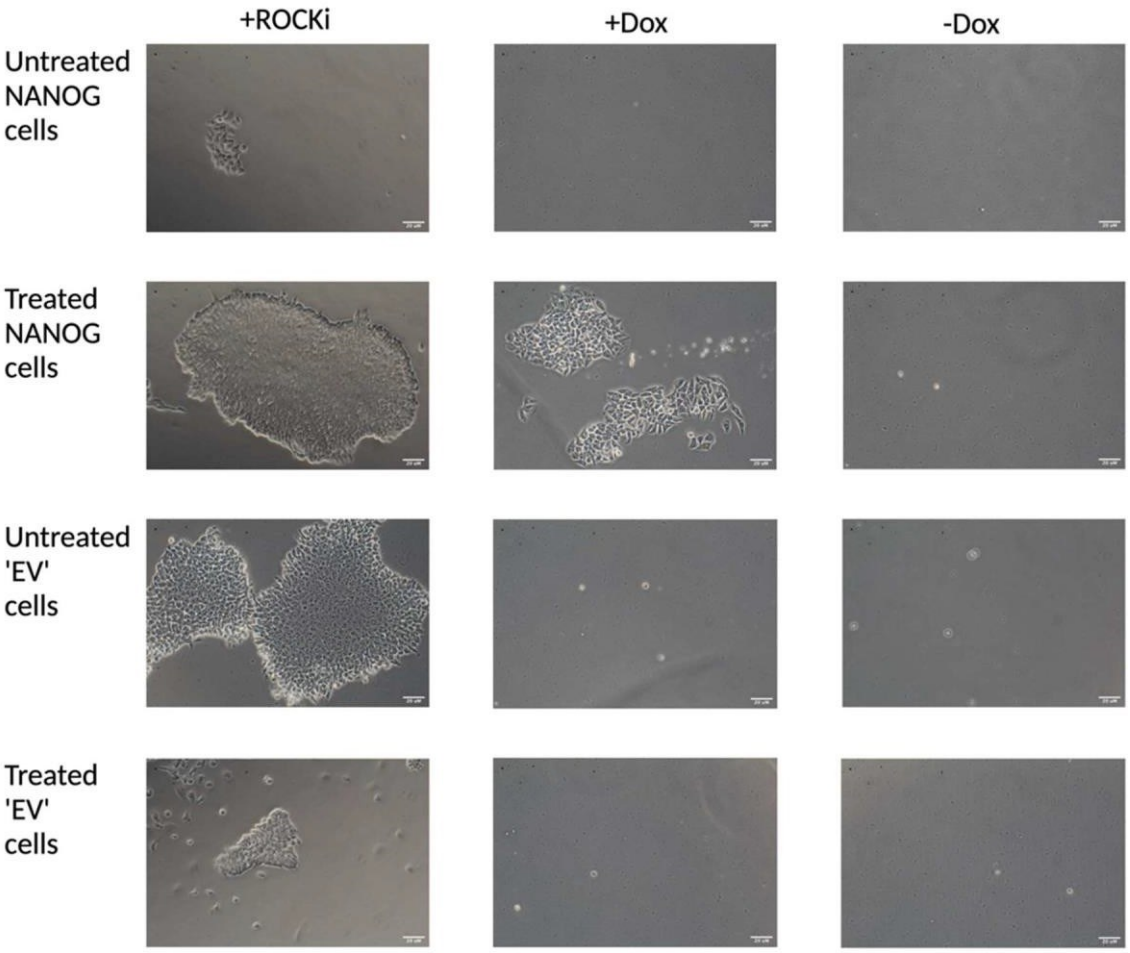

Figure 13 NANOG over-expression in hIPSCs allows for single cell passaging without the support of Y-27632. Representative images of colonies formed under different conditions using the NANOG OE and' EV' cell lines. Scale bars =20mm.

3.3.6 Generation and characterisation of KLF2 overexpressing cells.

The SOX2-tdT cell line was transfected with KLF2 overexpressing plasmid, and clones from these cells were termed KLF2 OE cells. As in the NANOG and DSOX2 cells, the mVenus reporter can be visualised using flow cytometry analysis and fluorescent microscopy after 1 day of 1μM dox treatment of KLF2 OE cells (Figure 14 A & B). To determine if the expression of the *KLF2* was upregulated, RT-qPCR was used, as there currently is no acceptable KLF2 antibody available. Dox treated KLF2 OE cells showed a relative KLF2 expression of 53 times

higher than untreated cells, showing dox was driving a substantial increase KLF2 (Figure 14 C). *SOX2* and *OCT4*, the two core pluripotency factors, were also included in this analysis as the morphological changes suggested KLF2 could be driving differentiation of hiPSCs. RT-qPCR showed there was a much as a 2-fold reduction in *OCT4* and 4-fold reduction in *SOX2* expression in dox treated KLF2 OE cells compared to untreated (Figure 14 C). This suggests KLF2 overexpression may cause an exit from pluripotency.

Low pass whole genome sequencing showed that over the majority of the chromosomes, there were no deletions or duplications (Figure 14 D). On chromosome 2 is a small region where the sequencing average was slightly below the expected value of 0. A full deletion would have an average score of -1, but the average score at this region was -0.2482576. Additionally, this spans across the centromere of chromosome 2, which is a region difficult to sequence due to repetitive sequences and satellite DNA sequences. The location and low score indicate this is probably a sequencing error rather than a true deletion in the chromosome; additionally, if the centromere was deleted, the whole of chromosome 2 would be lost due to its inability to be segregated during mitosis.

KLF2 overexpression had a striking effect on hiPSCs morphology, which appears to be irreversible after three days of dox treatment. The effect of *KLF2* overexpression appears to differ depending on the size and compaction of the colony. When already in a large and compacting colony, *KLF2* overexpression does not appear to affect this compaction but does appear to promote the cells to grow densely and on top of each other inside the colony (Figure 14 A, middle panel). While this behaviour has been observed in very confluent stem cell colonies, it is more focused on the centre of these colonies and not to the extent observed in the KLF2 overexpression cells. Comparatively, cells which are sparse when treated with dox, overexpressing *KLF2*, spread out and show clear morphological changes. These cells appear spindlier and more elongated, often existing as single cells rather than in colonies (Figure 14A, lower panel).

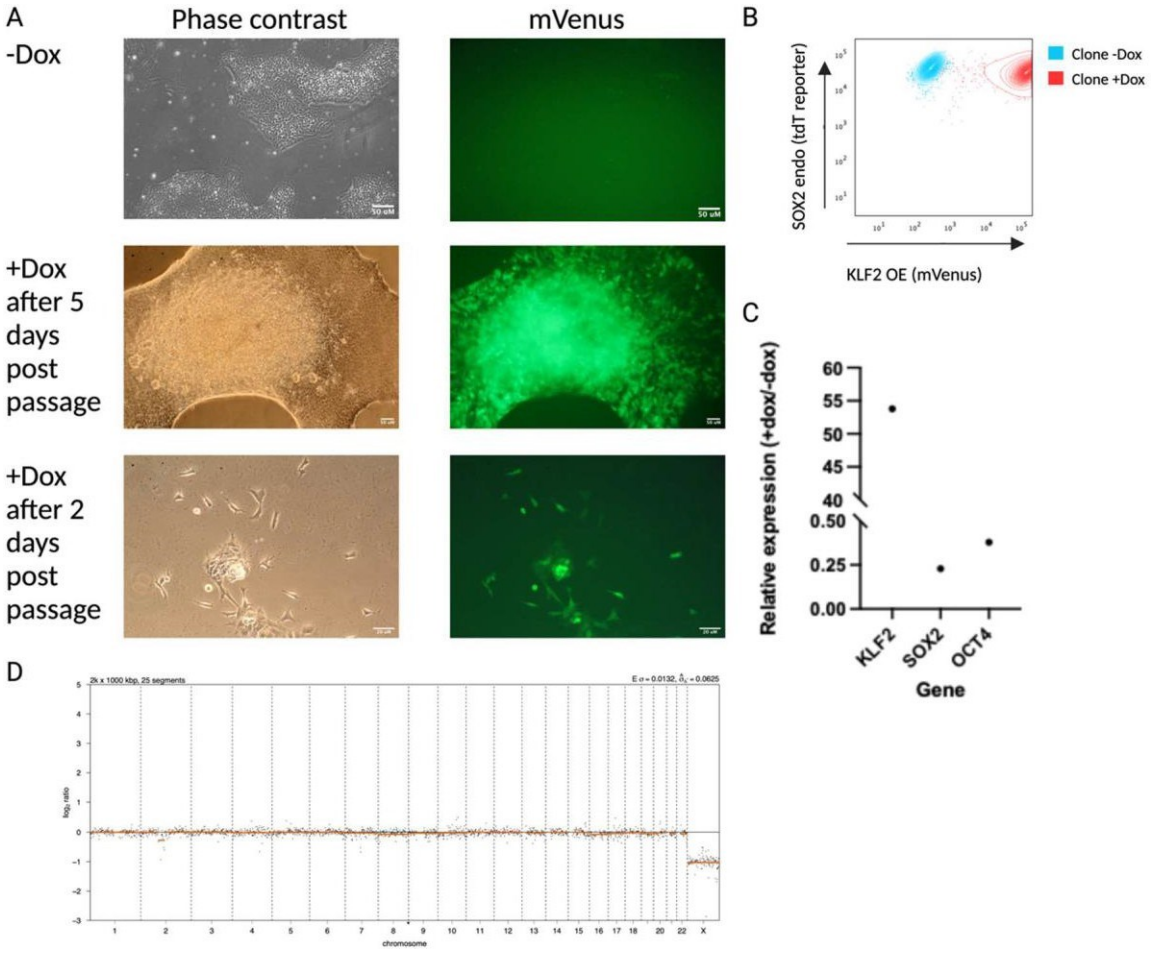

Figure 14 Characterisation of KLF2 clonal cell lines (A) Images of phase contrast and the sample field imaged to show expression of mVenus reporter. Top: no dox added to culture. Scale bars = 50mm Middle: 1mM dox added for 3 days; two days post split. Scale bars = 50mm Bottom: 1mM dox added at split. Scale bars = 20mm (B) Flow cytometry analysis of mVenus and tdT reporters in KLF2 clone 7 after treatment with or without 1mM dox for 1 day (C) RT-qPCR on KLF2 clone 7 with or without 1mM dox for 3 days on KLF2, SOX2 and OCT4, n=1 (D) Low-pass whole genome sequencing of KLF2 clone 7 to show genome integrity of main clone used in further experiments.

The observation that KLF2 appeared to cause differentiation and loss of pluripotency was studied through longer exposure to dox treatment. Briefly, KLF2 c7 cells were treated with prolonged 1μM dox for two passages, termed KLF2 “prolonged overexpression” (POE), before being split into two wells: one with dox (POE +dox) and one without (POE -dox). After reaching confluency, cells were analysed through flow cytometry (Figure 15 A). KLF2 c7 cells

untreated and treated for 1 day with 1μM dox were also included in this analysis for comparison.

Contour plots of these 4 conditions showed that the untreated and 1 day treated cells were both tdT positives, with the treated cells also being strongly mVenus positive (Figure 15 B). Intriguingly, KLF2 POE +dox, which had been treated for 3 passages, appeared to contain three populations: a tdT positive/mVenus negative population, a tdT negative/mVenus negative population and a heterogenous population which was tdT negative but had a range of mVenus intensities. These individual populations were plotted on histograms to visualise the intensity of each reporter (Figure 15 C). Analysis of KL2 POE -dox showed two populations present: the tdT positive/mVenus negative and tdT positive/mVenus negative populations. The lack of dox stimulation explains the absence of mVenus positive population as visualised in the histograms (Figure 15 D), showing the transgene is quickly inactivated upon dox withdrawal.

The SOX2-tdT reporter was only maintained in a subset of cells following the prolonged dox treatment for 2 passages ~ 10 days (Figure 15 B). These cells were also not mVenus positives, suggesting the transgene was not active in these cells even when they were being stimulated by dox. After initial treatment with dox, KLF2 cells show high expression of tdT and mVenus. The downregulation of tdT during the prolonged dox treatment suggests KLF2 over-expression is causing a loss in SOX2 expression, consistent with differentiation. The subset of cells which retain tdT reporter expression, but not mVenus, may have silenced the transgene early during the prolonged dox treatment; supporting the idea KLF2 expression is the driver of the loss of tdT in the tdT negative cells.

This experiment also highlighted a potential limitation of the system, the silencing of the transgene after prolonged dox treatment (Figure 15 C). After 3 passages with dox, the mVenus intensities were universally lower than the 1 day dox treatment. Furthermore, these were not focused in one intensity peak but a range of intensities, including a large proportion the cells being negative for mVenus. This suggests the prolonged dox treatment may lead to a silencing of either the TF locus or the rtTAM-2 locus making the cells unresponsive to dox treatment.

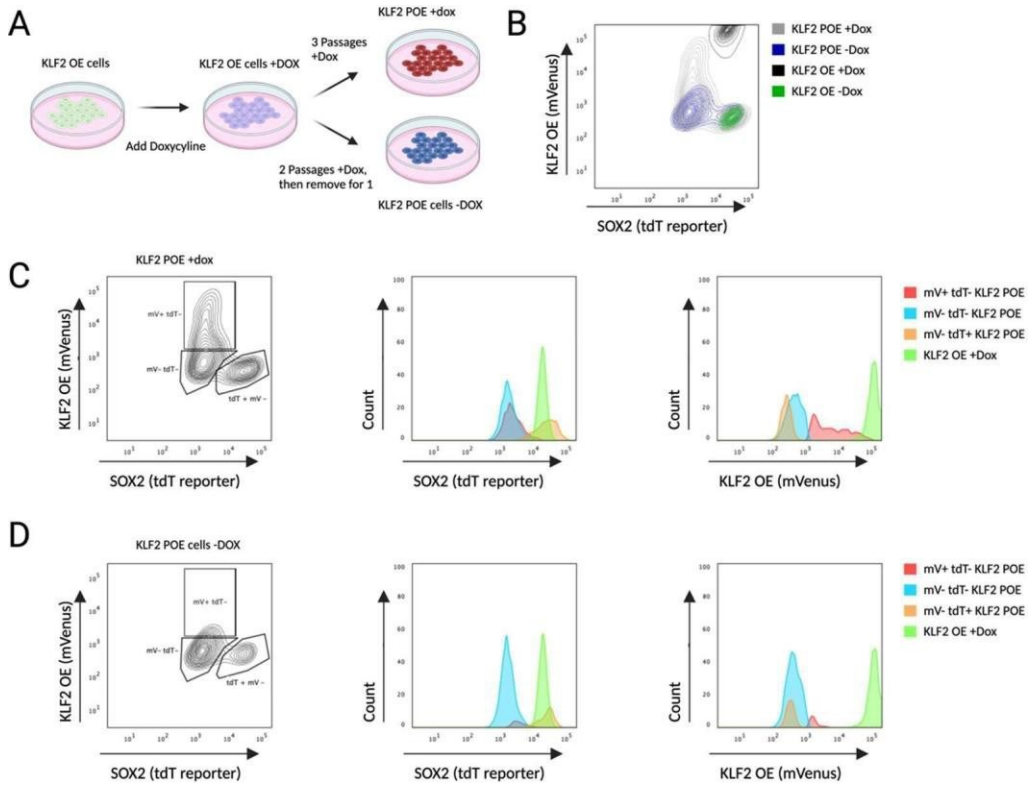

Figure 15 Prolonged overexpression of KLF2 promotes differentiation (A) schematic describing the experimental set up (B) contour plots of four conditions plotting mVenus vs. tdTomato reporters (C) KLF2 prolonged over-expression condition maintained in 1mM dox Left: contour plot of mVenus vs. tdTomato intensities with different populations Middle: Histograms of tdTomato intensity in populations from left panel with KLF2 over expression cells treated with 1mM dox for one day. Right: Histograms of mVenus intensity in the same populations. (D) KLF2 prolonged over-expression with dox removed for one passage Left: contour plot of mVenus vs. tdTomato intensities with different populations gates labelled. Middle: Histograms of tdTomato intensity in populations from left panel with KLF2 over expression cells treated with 1mM dox for one day. Right: Histograms of mVenus.

3.3.7 Generation and characterisation of a control 'empty vector' over expressing transfected cell line.

A negative control cell line was also generated by using the backbone of the TF plasmid which contained no TF but a resistance gene and would still produce the mVenus reporter protein when stimulated with dox. This cell line is termed empty vector or 'EV'. The expression of mVenus could be observed with both flow cytometry analysis and fluorescent microscopy (Figure 16 A & B). This cell line will be used to differentiate between the effects of mVenus expression and of the specific pluripotency TF being studied. The genome integrity of the main clones used in this study, EV c3, was analysed using low pass whole genome sequencing (Figure 16 C). A similar drop in score at a region spanning the chromosome 2 centromere was observed in the KLF2 c7 (Figure 14 D). This deletion spans across the centromere of chromosome 2, which is a region difficult to sequence due to presence of repetitive sequences and satellite DNA. If the centromere was deleted; the whole of chromosome 2 would be lost due to its inability to be segregated during mitosis. Therefore, the small deletion in the sequencing data is probably a sequencing error rather than an actual deletion of DNA within chromosome 2.

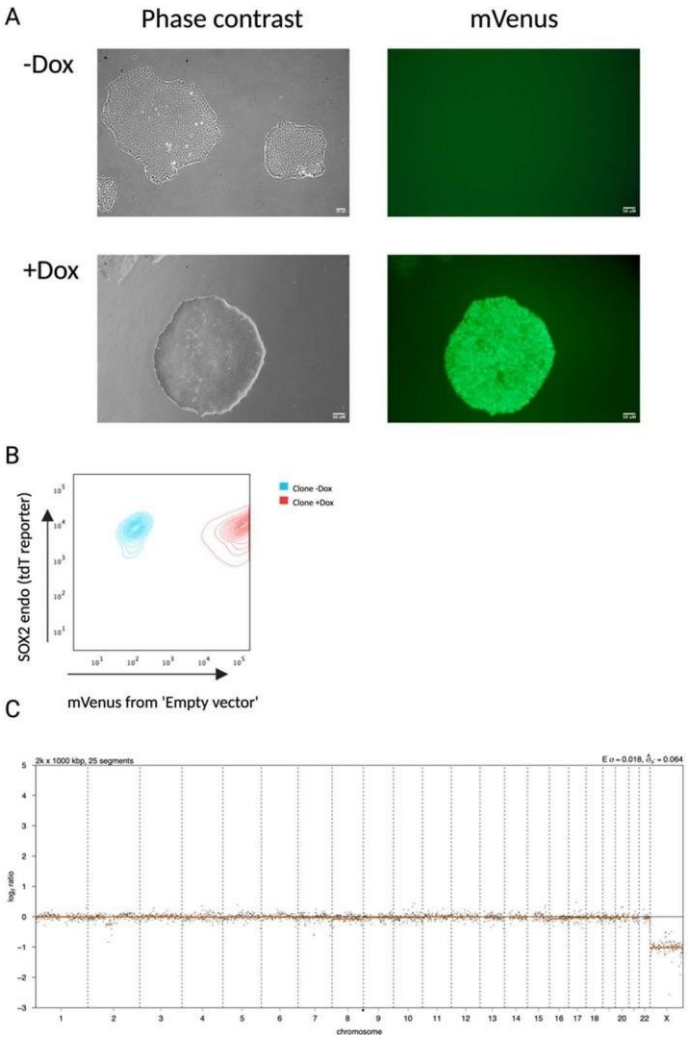

Figure 16 Characterisation of EV clonal cell lines (A) Images of phase contrast and the sample field imaged to show expression of mVenus reporter. Top: no dox added to culture. Middle: 1mM dox added for 3 day. Scale bars=50mm (B) Flow cytometry analysis of mVenus and tdT reporters in EV clone 3 after treatment with or without 1mM dox for 1 day (C) Low-pass whole genome sequencing of EV clone 3 to show genome integrity of main clone used in further studies.

3.4 Discussion

3.4.1 *SOX2* over expression causes cell death in feeder-free human iPSCs

A previous study reported that overexpression *SOX2* caused trophectoderm differentiation when cultured hPSCs were cultured on feeders (Adachi et al., 2010). In my study, when *SOX2* was overexpressed in E8, cultured hiPSCs led to cell death. This discrepancy might be due to my study having higher SOX2 overexpression levels compared to the previous study. Comparing Western blots from my study and the previous study (Adachi et al., 2010) suggests I was able to achieve a much higher level of SOX2 protein expression. It could be the higher level of SOX2 in my study causes cell death, but lower levels of SOX2 produced in the other study causes trophectoderm differentiation. Alternatively, differentiation into trophectoderm might be occurring in my study, and trophectoderm cells could undergo apoptosis in the E8 media. The feeder conditioned media likely contains multiple cytokines which may better support or even trigger trophectoderm differentiation, not present in the E8 media used in my study. Staining's for both apoptotic and trophectoderm markers could help find the answer; however as seen in the KLF2 POE experiment, E8 can support differentiated cell types, the first explanation still seems most likely.

3.4.2 *NANOG* overexpression appears to alter and potentially aid pluripotency.

hiPSCs with *NANOG* overexpression gave comparable results to previous published studies which used feeder media (Darr et al., 2006).This study showed that single cell passaging and colony formation is possible in *NANOG* overexpressing cells, which normally requires chemical intervention to block the rho-kinase pathway. The altering of morphology could be explained as the upregulation of primitive ectoderm genes observed in the previous study (Darr et al., 2006), as is the observation that these cells retain their pluripotent character and able to revert back to morphologically normal hiPSCs once dox is removed from the culture. It is clear from the experiment that NANOG overexpression does confirm the ability for hiPSCs to survive as single cells, to proliferate and to remain pluripotent.

3.4.3 KLF2 over expression causes irreversible differentiation.

Single *KLF2* overexpression in human pluripotent stem cells has never been reported in the literature. In this study, it appears that the overexpression of *KLF2,* rather than promoting pluripotency as in the mouse, triggers differentiation which is irreversible when prolonged. Only cells which appeared to prevent the transgene expressing *KLF2* were able to retain the expression of the *SOX2-tdT* reporter and therefore remain pluripotent. This is especially striking as in mEpiSCs, a cell type more similar to hiPSCs than mESCs or miPSCs, *Klf2* can reprogramme these cells into a naïve state (Qiu et al., 2015). In hiPSCs therefore, it appears *KLF2* has a dominant effect in triggering differentiation.

Chapter 4. Overexpressing pluripotency-related transcription factors during human germline induction

4.1 Introduction

In chapter 3, I described the generation of four hiPSC lines which overexpressed a different pluripotency transcription factor or a control 'empty vector' when doxycycline was added to the media. This chapter describes experiments where I activated these transgenes before and through the process of hPGCLC derivation. My aim was to trigger the expression of SOX2 from the endogenous locus, that could be tracked using the tdT reporter contained within these cells, as evidence of pluripotent conversion in hPGCLCs. The effect of conversion efficiency into the hPGCLCs caused by the overexpression of the transcription factors was also of interest. In addition, I explored whether the generated hPGCLCs showed a different or altered identity in response to the TF overexpression.

4.1.1 Confirming hPGCLCs are a good model to study early human germline development

The transcriptome of human PGCs has been reported in a number of studies (Guo et al., 2015; Tang et al., 2015). However, there is limited data for early (pre-gonadal) human PGCs. Nevertheless, hPGCLCs do appear more similar to the earlier PGC stages (Irie et al., 2015). They do not express a number of genes characterised as late PGC markers such as *DAZL*, *DDX4* and *SYCP3* (Tang et al., 2015), the latter gene being required for meiosis in male mice (Yuan et al., 2000). Direct comparison to mouse suggests hPGCLCs are similar to pre-migratory mPGCs at E6.5-7.5, around human Week 2-3 (Tang et al., 2015). However, the correlations are weak, and it is possible that this may not be the most biologically relevant comparison. Overall, hPGCLCs have been designated as being newly formed or nascent, a hypothesis that is supported by the identification of *SOX17* expressing cells in the primitive streak of a week 2 embryo, which expressed genes found in hPGCLCs such as *NANOG*, *NANOS3* and *DND1* (Tyser et al., 2021).

4.1.1.1 Pushing hPGCLCs to later developmental stages *in vitro*

Human PGCLCs can be aggregated with mouse embryonic gonadal cells to induce early oocytes or prespermatogonia respectively (Hwang et al., 2020; Yamashiro et al., 2020). After 4 months of co-culture of hPGCLCs and female embryonic gonadal cells, termed "xenogenic reconstituted ovaries (xrOvaries)" (Yamashiro et al., 2018) , the hPGCLC daughter cells undergo genome-wide DNA demethylation, including erasure of paternal and the majority of maternal imprints. Whether this process follows the *in vivo* mechanism remains unclear. Gene expression profiles change throughout the process, eventually upregulating genes associated with the RA response pathway while downregulating *TFAP2C*. This study provides strong evidence that hPGCLCs are therefore capable of progressing developmentally, adding confidence that they are bona fide germ cells (Yamashiro et al., 2018). However, this is an inefficient and undefined system, and neither meiotic progression or RA response has been confirmed.

4.1.2 Strategy and justification for overexpressing pluripotency related TFs in hPGCLCs

The hPGCLC system is described in detail in Chapter 1 (Introduction). Briefly in this study, hiPSCs are first differentiated into a transitional stage termed iMeLCs, before these cells are aggregated in low-attachment culture and stimulated with BMP4 (Sasaki et al., 2015). Around 2-30% of cells within these aggregates take on germ cell identity, which can then be identified through the use of specific surface markers. I established that this protocol could be followed to produce hPGCLCs from the different cell lined described in Chapter 3. Subsequently, I optimised the concentration of dox added to these cultures to activate the transgene in as many cells of the aggregate as possible. I then performed several rounds of hPGCLC generation from each hiPSC line, applied the optimised dox concentration and measured the conversion into hPGCLCs, the activity of the transgene and any effect on endogenous *SOX2* expression.

The rationale for choosing *SOX2, NANOG* and *KLF2* was detailed in chapter 1. *SOX2* shows positive feedback in the context of human pluripotency (Boyer et al., 2005). *NANOG* directly positively regulates *SOX2* expression in human pluripotency (Boyer et al., 2005) and in combination with *KLF2,* resets human primed cells into a naïve state (Takashima et al., 2014).

This chapter also describes experiments using *NANOG* overexpression to induce hPGCLC without the use of BMP4, as overexpression of *Nanog* in mouse was sufficient to induce mPGCLCs from mEpiLCs (Murakami et al., 2016).

4.2 Specific aims

- Generate hPGCLCs from the *SOX2*-tdT hiPSC lines as well as the derivatives of this line generated in Chapter 3.
- Use these cells to overexpress *SOX2, NANOG* and *KLF2* during hPGCLC induction.
- Study the effects that this overexpression has on the process of hPGCLC induction and on hPGCLC identity.
- Attempt to induce hPGCLC using *NANOG* overexpression alone.

4.3 Results

4.3.1 Generation of hPGCLCs from *SOX2-tdT* hiPSCs line

My aim was to establish whether the *SOX2-tdT* (Balboa et al., 2017) cell line and its derivatives could form hPGCLCs using the established protocol (Sasaki et al., 2015). This protocol is graphically described in Figure 17 A, and images of the three stages of culture, hiPSC, iMeLC and hPGCLC aggregate in Figure 17 B. Cells aggregated into low attachment wells cultured in hPGCLC specification media will be called "hPGCLC aggregates" in this study.

After 4 days, these aggregates can be dissociated enzymatically using trypsin, the resulting cell suspension is stained for cell surface markers and individual hPGCLCs are separated through FACS (Figure 17 C). These cell surface markers are EPCAM and INTEGRINα6, established to stain 98.9% of cells with hPGCLCs identity (Sasaki et al., 2015). These individually separated cells will be called "hPGCLCs" in this study.

I wanted to confirm that when I performed the hPGCLC derivation protocol it was successful in my hands. I used an established reporter cell line containing two endogenous reporters: BLIMP1-tdTomato, AP2γ-GFP (BTAG) (Figure 17 D). BLIMP1 and AP2γ are lineage markers for the early germline and the expression of these genes has been shown to identify hPGCLCs (Sasaki et al., 2015) and therefore, BTAG cells which are double positive for both tdT and GFP are considered to be hPGCLCs. I was able to generate tdT and GFP double positive hPGCLCs from this cell line (Figure 17 E), confirming I was able to generate hPGCLCs. 97% of the double positive tdT/GFP cells of aggregates are also double positive for the two surface markers selected: EPCAM and INTEGRIN α6 (Figure 17 E), confirming that these surface markers can identify hPGCLCs.

hPGCLC generation shows high clonal and experimental variation, with high efficiency clones giving a mean of 40% conversion, while lower efficiency clones giving >3% conversion or lower (Yokobayashi et al., 2017). In this study, the average conversion rate of hPGCLCs was 10%.

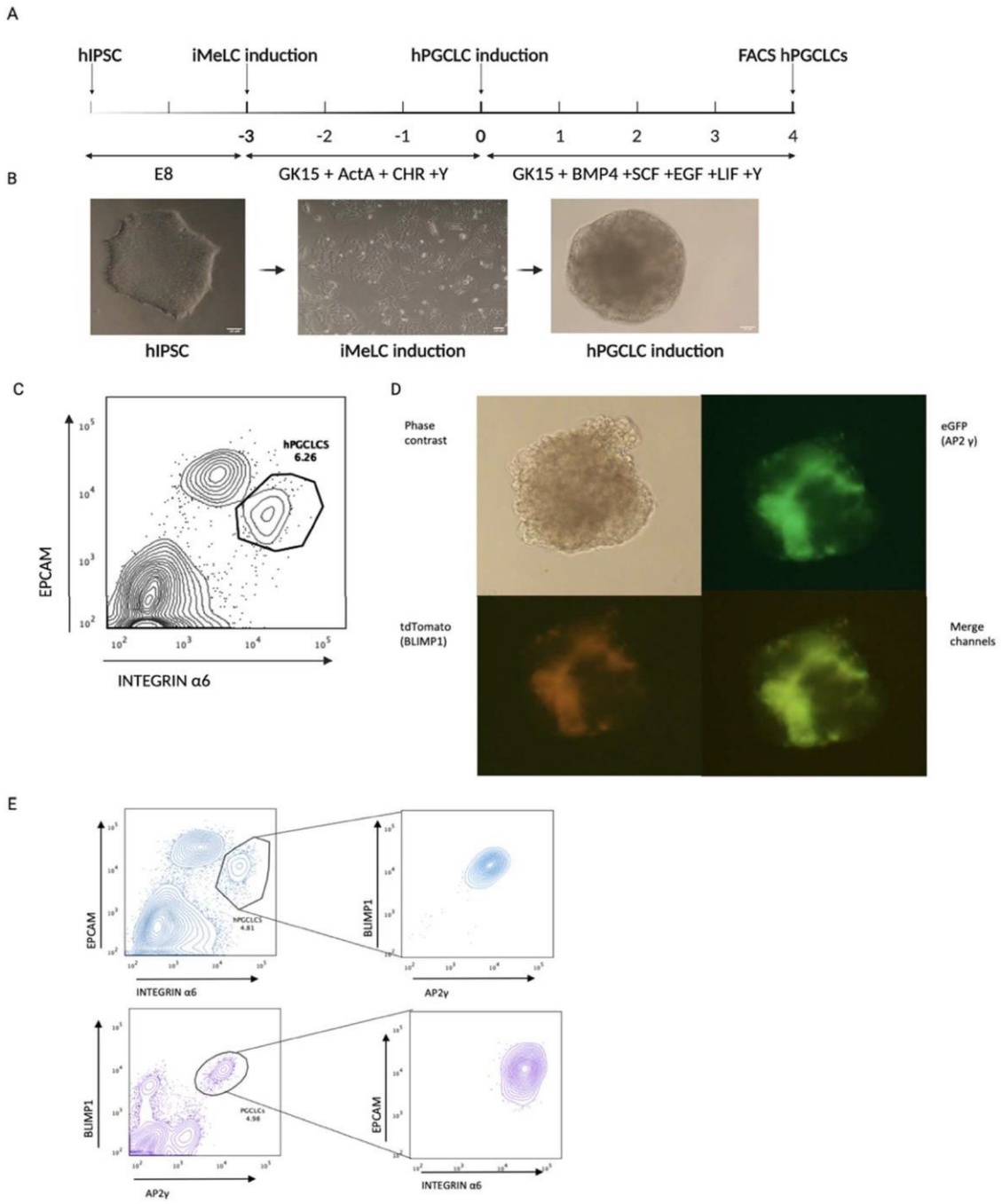

Figure 17 Generating hPGCLCs from the BTAG hIPSC line (A) Schematic of the protocol to generate hPGCLCs from hiPSCs with an intermediate step, iMeLC state (B) Bright field images of the three steps protocol. Left: hiPSCs (scale bar 10mM), middle: iMeLCs (scale bar 10mM), right: hPGCLC aggregate (scale bar 50 mM) (C) Flow cytometry analysis of digested aggregates stained for the surface markers EPCAM and INTEGRINa6 (D) Representative images of a hPGCLC aggregate made from the BTAG cell line which contains reporter genes for BLIMP1 and AP2g (E) Flow cytometry analysis of digested BTAG hPGCLC aggregates stained for the surface markers showing the correlation between reporter genes and surface markers.

4.3.2 Testing dox addition to trigger pluripotency factor overexpression in hPGCLC aggregates.

In order to study the effects that overexpression of the pluripotency-related transcription factors would have during the process of hPGCLC generation, I added dox to the aggregates to trigger expression of the transgene. Dox was added at four different time points; at the point of induction on day 0 (d0) or after induction had taken place at day 1 (d1), day 2 (d2) or day 3 (d3) (Figure 18 A).

The working concentration of dox is 1μM and was sufficient to activate the transgenes in the original study that utilised these plasmids (Takashima et al., 2014) and my hiPSCs. To test the effect of dox addition on the hPGCLCs, 1 μM dox was added to aggregates containing the *KLF2* c7 cells. 1μM dox showed very little activation of the transgene whose expression can be measured within each cell through the expression of the fluorescent reporter mVenus. When 1μM dox was added; <2% of cells within the aggregate were mVenus positive when sorted at day 4 (Figure 18 B). I theorised this may be due to the aggregates laying down ECM which prevented dox from diffusing through them. When preparing aggregates for sorting, even after dissociation with trypsin a visible mass of ECM proteins is left. To increase the concentration gradient between the media and aggregates, higher concentrations of dox - 10μM, 20μM, 50μM or 100μM - was added to aggregates at day 3, one day before sorting (Figure 18 C). Increasing the concentrations of dox, up to 20μM, resulted in more mVenus positive cells (Figure 18 D). The same number of aggregates was collected for each concentration and digested using the same process. Therefore, it was assumed that the number of cells recovered in each sample should be the same. However, the number of cells recovered from the aggregates, which can be measured as the total number of events within the live cell gate, was reduced when concentrations above 20μM were used (Figure 18 E).

This suggests that at concentrations above 20μM dox, there is a toxic effect. Therefore, 20μM was chosen as the working concentration related to hPGCLC induction experiments.

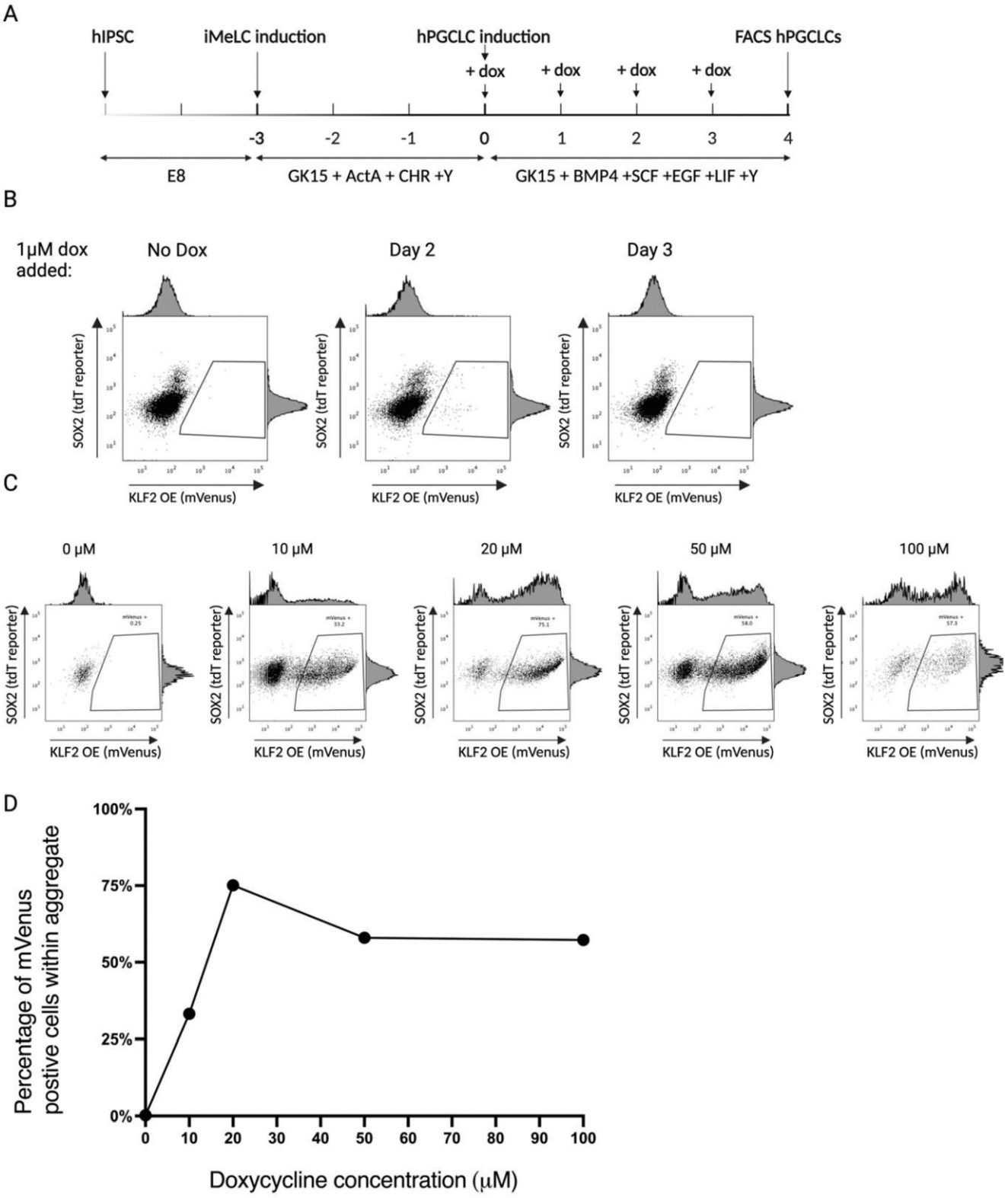

Figure 18 Addition of dox to hPGCLC aggregates triggers transgene expression (A) Schematic showing when dox was added to the aggregates (B) Flow cytometry analysis of mVenus and tdT reporters in digested aggregates when 1 mM dox is added on different days (C) Titration of increasing dox concentrations added on day 3 of hPGCLC culture (D) mVenus positive cells quantified from (C) plotted against dox concentration.

4.3.3 *SOX2* can be overexpressed in hPGCLCs after germline fate induction and appears to cause no change in the derivation efficiency or hPGCLC identity

SOX2 is repressed during hPGCLC specification (Irie et al., 2015; Sasaki et al., 2015). My aim was to re-activate expression of the endogenous *SOX2* gene. I hypothesised that this re-activation would cause reversion to a pluripotent state within the hPGCLCs. My first attempt to re-activate the endogenous *SOX2* locus within hPGCLCs was to use exogenous *SOX2* overexpression. The overexpression of exogenous SOX2 was measured through the mVenus fluorescent reporter and activation from the endogenous SOX2 could be observed through the expression of the tdT fluorescent reporter, (section 3.2). Additionally, I analysed whether the overexpression of SOX2 caused changes to the hPGCLC efficiency and to cell identity.

DSOX2 c1 cell line, which overexpresses SOX2 in response to dox, was used in the hPGCLC protocol above. Dox was added at day 1, day 2 or day 3 (termed d1, d2 or d3 respectively, see Figure 18 A) after hPGCLC induction, analysed by flow cytometry and hPGCLCs collected using FACS. The dox added samples were then directly compared to a no dox (dx) control.

hPGCLC identity can be defined by high expression of surface markers EPCAM and INTEGRIN α6, as shown in Figure 17 E. Expression of these surface markers is measured through staining with fluorescently tagged antibodies, with the higher the expression of a surface marker the higher the intensity recorded by flow cytometry.

When dox was added to the DSOX2 aggregates, there was no observable change in the intensities of the surface markers and therefore there was no change to where the gate was set to define the double positive population of hPGCLCs (Figure 19 A). Similarly, there was little difference in the percentage conversion rate (Figure 19 B). The means of each sample were dx - 6.8%, d1 - 8.6% d2 - 5.6% and d3 - 5.9%, but all had large overlapping ranges. Considering the high level of variability in the PGCLC system between experiments and cell lines (Yokobayashi et al., 2017), this suggests the dox additions did not cause changes in hPGCLC conversion efficiency.

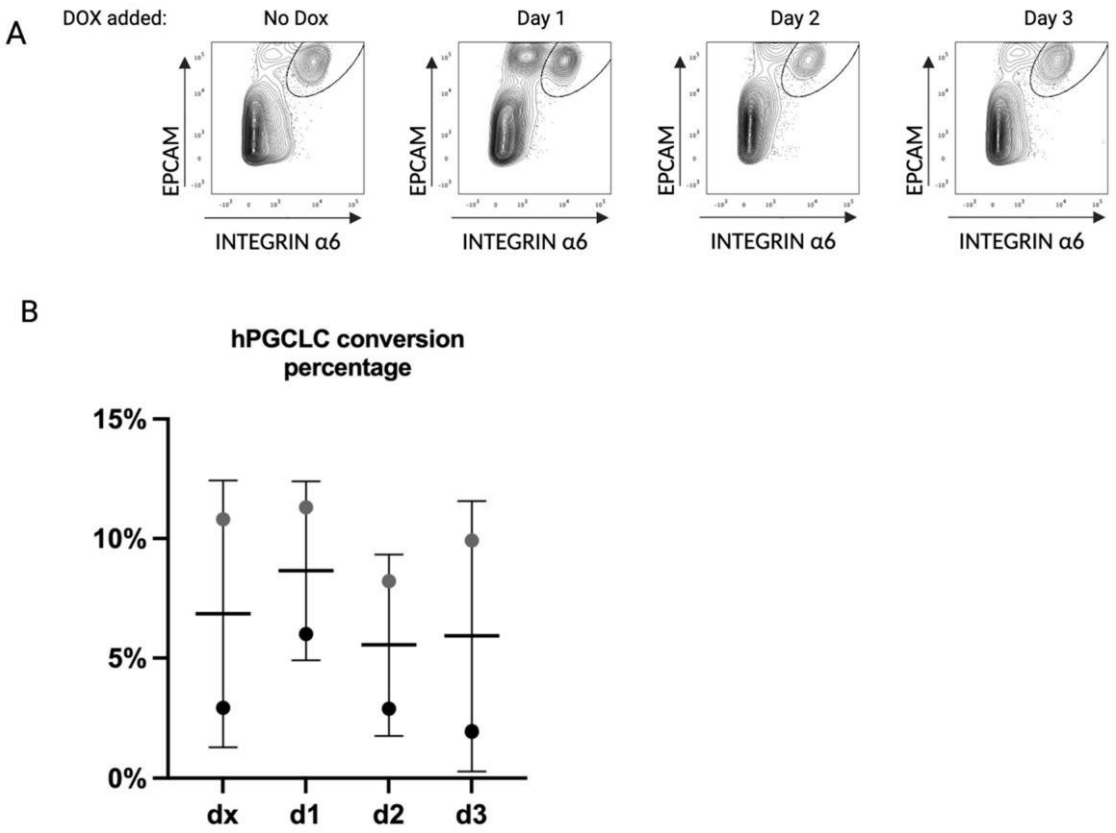

Figure 19 Effect of dox addition to DSOX2 aggregates (A) Representative FACS plots used to sort hPGCLCs from aggregates, based on the two surface markers EPCAM and INTEGRINa6 (B) percentage of cells within hPGCLC gate n=2, mean is represented by the horizontal bar, error bars are standard deviation from the mean.

To analyse the expression of the SOX2 overexpression transgene, the expression of the linked fluorescent reporter mVenus was tracked using flow cytometry. Measuring mVenus showed that the transgene was expressed in a subset of cells within the aggregate in all the samples where dox was added (Figure 20 A upper panel). The percentage of mVenus cells inside the aggregate appeared to decreased the later the dox was added (Figure 20 A upper panel & B upper panel). Without further repeats it is not possible to determine if these differences are simply normal variation between samples. hPGCLCs within the aggregate follow the same pattern as the aggregate as a whole; fewer mVenus positive cells are observed the later dox is added (Figure 20 A & B lower panel). However, the percentage of mVenus positive hPGCLCs was lower compared to the whole suggesting the transgene is more likely to be activated in non-hPGCLCs within the aggregate than hPGCLCs.

When dox is added a day later, the percentage of mVenus positive cells in the aggregate halves, compared to the previous day. However, in hPGCLC population, this rate of reduction is slower with the percentage reducing by one-third compared to the previous day. To normalise the percentage of mVenus positive hPGCLCs to the percentage of mVenus positive all cells in the aggregate, the percentage of mVenus positive hPGCLCs was divided by the percentage of positive cells within the whole aggregate. This gives a more accurate figure for how likely it was for the hPGCLC to activate the transgene based upon how likely it was for any cell of the aggregate to activate it. The calculation shows there might be a slight increase in the likelihood of a hPGCLC being mVenus positive the later dox is added (Figure 20 B right panel) However, this is only a subtle difference between the three days and without further repeats to allow statistical testing, it is not possible to determine if this effect is significant.

As expected, the endogenous SOX2-tdT reporter is fully silenced in hPGCLCs in the control condition, when no dox is added. Despite the overexpression of the exogenous SOX2, the tdT reporter did not show activation in any of the conditions. This shows the endogenous *SOX2* is still repressed in hPGCLCs even in the presence exogenous SOX2 (Figure 20 A).

The presence of mVenus positive cells in hPGCLC population indicates the SOX2 overexpression transgene can be actively expressed in a subset of hPGCLCs (Figure 20 C). SOX2 is repressed in hPGCLCs (Irie et al., 2015; Sasaki et al., 2015) and hPGCs (Perrett et al., 2008). Germ cell tumours which retain their germline identity require repression of *SOX2* repression or differentiate into a neuronal lineage (Kushwaha et al., 2016). Despite this potentially destabilising role of *SOX2*, there appears to be little effect on the hPGCLC conversion efficiency within the aggregate when *SOX2* is overexpressed. This suggests that SOX2 overexpression after induction of hPGCLCs does not impact the process of hPGCLC conversion.

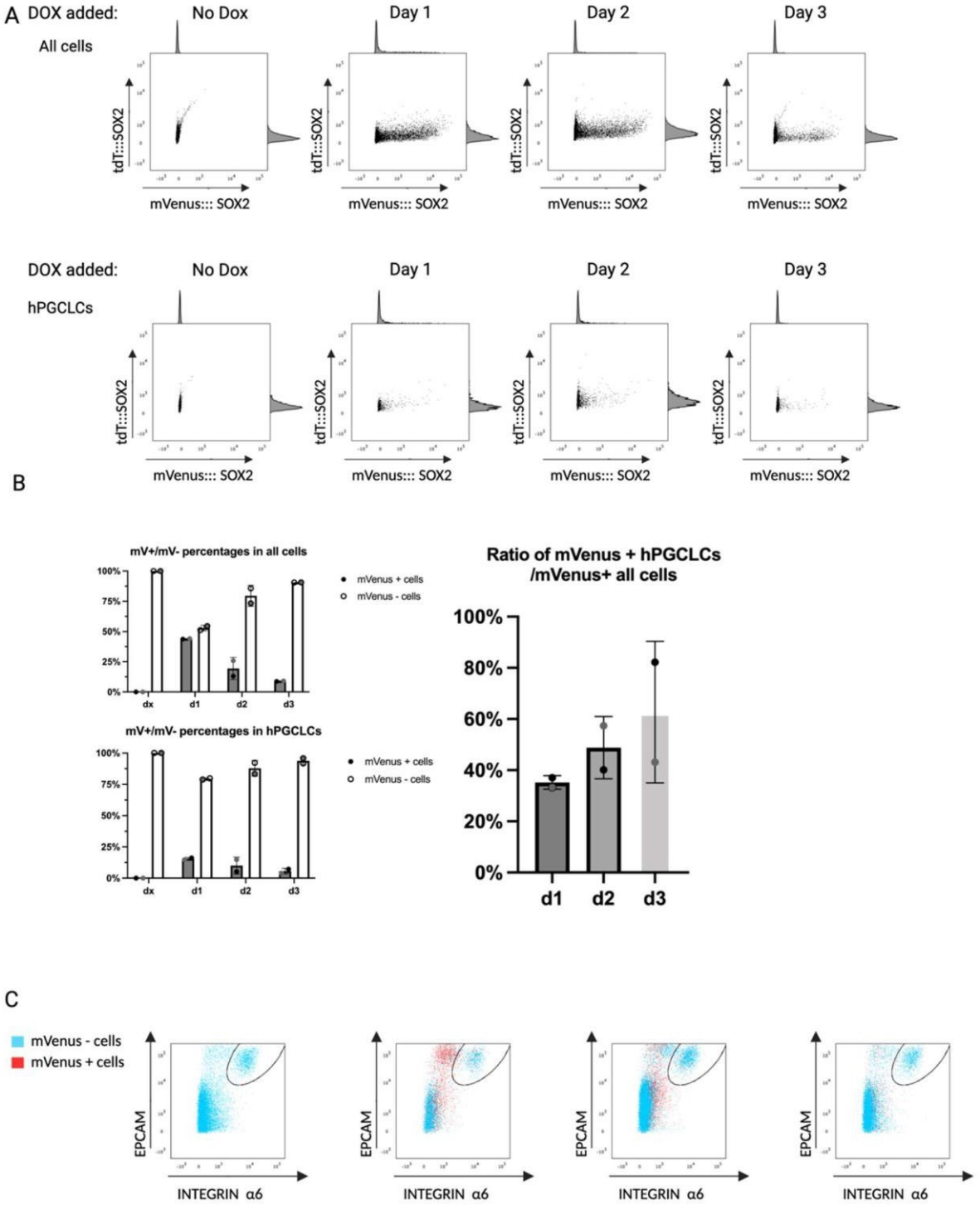

Figure 20 Analysis of mVenus and tdT expression within DSOX2 hPGCLC aggregates by flow cytometry (A) Representative dot plots showing mVenus vs. tdT expression Top: all cells. Bottom: hPGCLCs in the aggregate (B) quantification of mVenus positive and mVenus negative cells in percentages Top: all cells. Bottom: hPGCLCs Right: percentage of mVenus positive hPGCLCs divided by percentage of all cells in the aggregates that are mVenus positive (n=2) (C) FACS plots showing distribution of mVenus positive and mVenus negative cells based on EPCAM and INTEGRINa6 staining's.

To analyse if the overexpression of *SOX2* altered the expression of germline markers, I conducted immunofluorescence on hPGCLC aggregates. Aggregates with dox added at day 2 and day 3 were fixed, cryosectionned and then stained for OCT4 and AP2γ, markers for early germ cell identity (Sasaki et al., 2015). Two different fields from samples where dox was added at day 2 (Figure 21 A) or day 3 (Figure 21 B) are shown. Clusters of cells which are stained for both AP2γ and OCT4 are clearly visible in all fields confirming the presence of hPGCLCs in the aggregates. There are subtle variations in the intensity of stain between the two factors, suggesting there might be variations in expression of the two factors in individual hPGCLCs. Despite this, it appears that the expression of one factor is accompanied by the expression of the other, confirming these cells are hPGCLCs.

The distribution of hPGCLC through the aggregate appears to be non-uniform. Field 1 in day 3 shows a large number of AP2γ+/OCT4+ cells distributed across this section of aggregate, but field 2 shows a lower density of AP2γ+/OCT4+. Previous studies have shown hPGCLCs appear to move into the centre of aggregates by day 4 of culture (Irie et al., 2015). The section of aggregate in the first field in day 3 is larger than the section in field 2; indicating the first field section was towards the centre of the aggregate. As it contained more hPGCLCs and at a higher density; it appears that *SOX2* overexpression does not alter the location of hPGCLCs observed in previous studies.

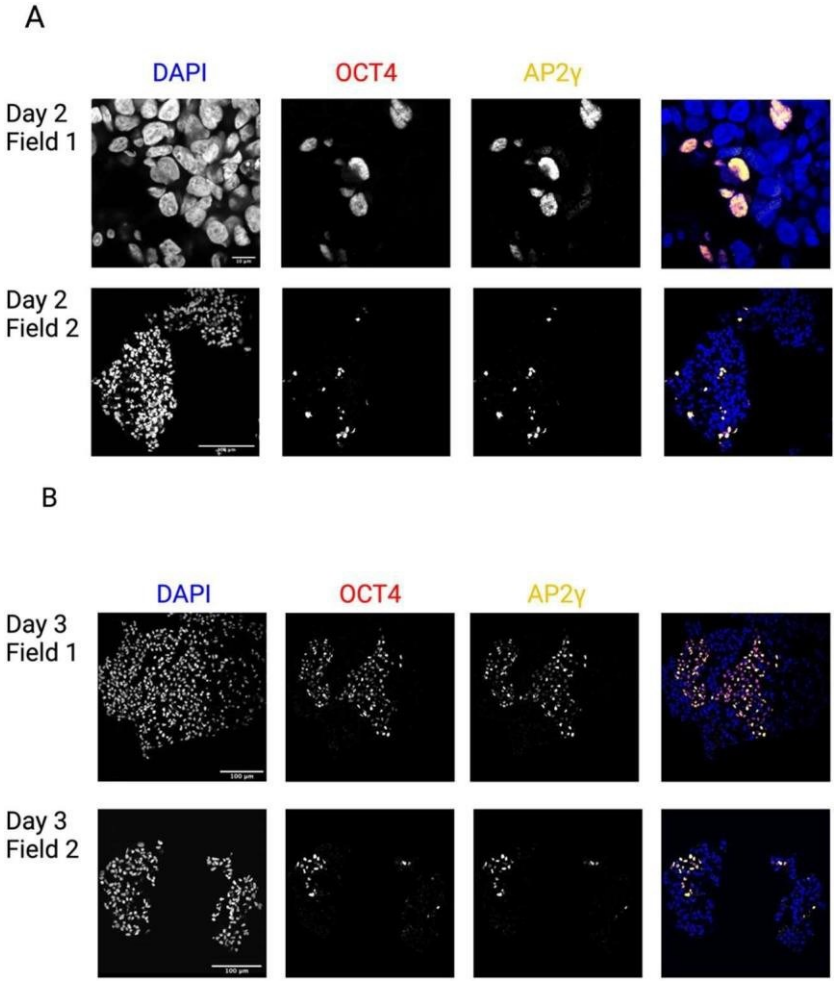

Figure 21 Immunofluorescences of DSOX2 hPGCLC aggregates (A) two fields from aggregates with dox added on day 2, stained with OCT4, AP2g and DAPI with a compositive of the three stains (Scale bar top = 10mM, bottom = 100mM) (B) Two fields from aggregates with dox added on day 3, stained with OCT4, AP2g and DAPI with a compositive of the three stains (Scale bar = 100mM).

4.3.4 hPGCLC generation is blocked when *SOX2* is overexpressed at induction.

In the previous section, I showed that the addition of dox to DSOX2 hPGCLC aggregates had no major effect on conversion efficiency. Subsequently, I wanted to observe the effect of adding dox to the hPGCLC aggregates upon induction or at day 0 (d0), when the cells still had the iMeLC identity.

When 20μM dox was added at induction, the EPCAM positive population appears larger than no dox control (Figure 22 A). The EPCAM/INTEGRINα6 double positive population which represents the hPGCLCs is severely reduced and appears merged into the EPCAM positive population (Figure 22 A). In WT hPGCLC aggregates there are three main populations present in flow cytometry analysis after 4 days of hPGCLC induction: a double negative population, an EPCAM positive population and an EPCAM/INTEGRINα6 double positive population, (which specifies the hPGCLCs, as described previously (Figure 22 E). The flow cytometry analysis suggested that overexpression of *SOX2* at induction leads to the repression of the germline fate and an increase in cells retaining their stem cell identity, as EPCAM is selectively expressed in human pluripotent cells (Lu et al., 2010).

It was easier to activate *SOX2* overexpression when dox was added before germline fate induction compared to after induction. In the experiment where 20 μM dox is added pre-induction, around 58% of all cells in the aggregates are mVenus positives and 57% of hPGCLCs are mVenus positives (Figure 22 A). To normalise the percentage of mVenus positive hPGCLCs to the percentage of mVenus positive all cells in the aggregate, the percentage of mVenus positive hPGCLCs was divided by the percentage of positive cells within the whole aggregate. This resulted in a ratio of 98% for d0 which is higher than the ratios observed when dox is added after induction, which were d1 – 35%, to d2 – 49% and d3 – 62%. This suggests that the *SOX2* overexpression transgene is more responsive in the iMeLC state as there is no difference in likelihood of *SOX2* overexpression being activated in either the hPGCLCs or non-hPGCLCs (Figure 22 A). However, after 24 hours of induction, the likelihood of hPGCLCs activating the transgene is reduced compared to non-hPGCLCs. This would suggest that the hPGCLCs are resistant to expressing the *SOX2* overexpression transgene, either because the overexpression of *SOX2* is detrimental for specifying the germline fate, or because at the transcriptional level the transgene is repressed more within the hPGCLCs compared to the non-hPGCLCs. Alternatively, the addition of dox after 3D aggregation might be preventing it diffusing into the aggregate, which is not an issue when added at the point of aggregation.

To further understand the effect of SOX2 overexpression at induction of hPGCLCs I conducted a titration experiment with a range of dox concentrations. Three concentrations of dox were used: 5 μM, 20 μM and 50 μM. The hPGCLC conversion efficiency when 5μM dox was added was 3%, which reduced to 1.56% when 20μM dox and reduced further to 0.8% when 50μM dox was added (Figure 22 C). There is also no evidence of the endogenous SOX2 locus remaining active (Figure 22), suggesting that BMP4 signalling is sufficient to repress this locus in vitro even with exogenous SOX2 expression. The increase in dox concentration appears to constrain the ability of the aggregates to generate hPGCLCs. The percentage of mVenus positive cells in each concentration was above 90%, although the higher concentration of dox did also appear to increase the intensity of the mVenus fluorescence (Figure 22 B). As the intensity of mVenus should correlate with SOX2 overexpression, as the two genes exist on the same transcript, the high intensities of mVenus in the higher concentration of dox should result in more SOX2 expression. Fewer cells enter the hPGCLC fate when SOX2 overexpression is higher, as in the 20 μM and 50 μM conditions, suggesting the overexpression blocks entrance into the germ cell fate.

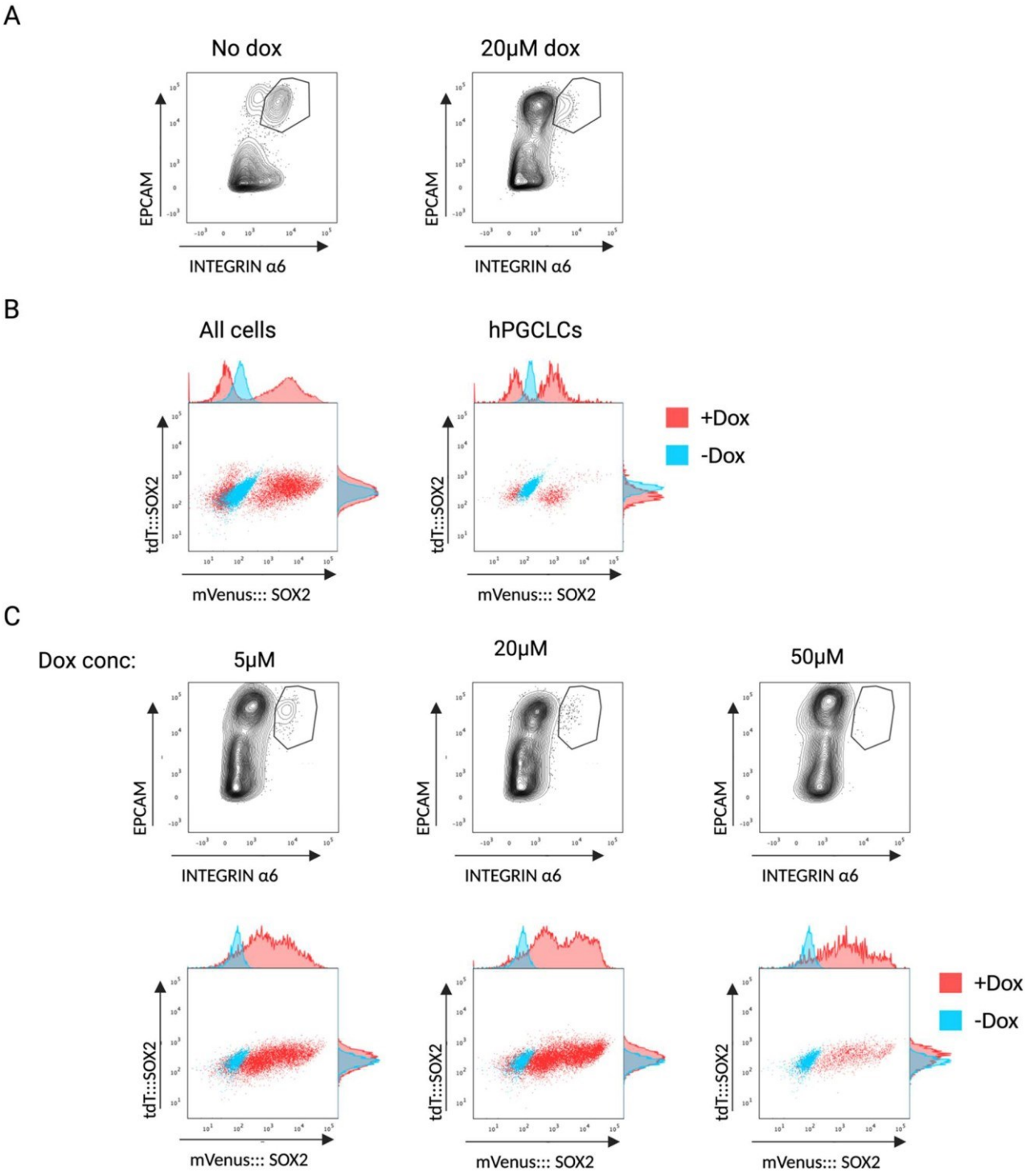

Figure 22 Effect of dox addition at the point of hPGCLC induction (A) first experiment using 20mM dox, flow analysis of surface markers staining and mVenus vs tdT in all cells and hPGCLCs (B) Titration experiments with increasing dox concentrations, flow analysis of surface markers and of mVenus vs tdT in all cells.

4.3.5 NANOG overexpression does not induce differentiation into hPGCLCs

Nanog overexpression in mouse EpiLCs is sufficient to trigger robust mPGCLCs differentiation without BMP signalling (Murakami et al., 2016). To date, no data exists about a similar mechanism in human PGCLC specification, so I attempted to reproduce this experiment using the *NANOG* overexpression cells in the human PGCLC system. The same schematic above (Figure 23 A) was used to induce iMeLCs from the *NANOG* OE hiPSC line. At aggregation and induction of hPGCLCs from the iMeLCs, either dox, BMP4, or both, were added to the media.

NANOG overexpression was observable through the mVenus reporter, but *NANOG* overexpression did not appear to produce hPGCLCs. In the +dox/-BMP4 condition, 100% of the cells were mVenus positive (Figure 23 B), the distribution of EPCAM and INTEGRIN α6 on the FACS plot was different compared to the -dox/+BMP4 condition, the control experiment (Figure 23 A). Instead of the three populations previously described, there are only two distinct cell populations: a small double negative population and a much larger population which has a range of EPCAM positive and INTEGRINα6 positive cells. While 70% cells are in the same gate as the double positive cells defined in the -dox/+BMP4 condition, these cells do not form a separate population characteristic of hPGCLCs (Figure 23 A). Furthermore, the SOX2-tdT reporter is still active in 90% cells of the aggregate and in 100% of the cells within the 'hPGCLC' gate (Figure 23 C & D). Endogenous *SOX2* is repressed in hPGCLCs (Irie et al., 2015; Sasaki et al., 2015) indicating the cells in the -BMP4 condition are not hPGCLCs. These results suggest that in absence of BMP4, hPGCLC differentiation cannot be achieved even with the overexpression of *NANOG*.

To further understand what effects *NANOG* overexpression might be causing in the absence of BMP4, dox and BMP4 were omitted to serve as a negative control. A similar pattern of EPCAM and INTEGRINα6 expression was observed in -dox/-BMP4 as in the +dox/-BMP4 (Figure 23 A). The addition of dox did seem to increase the number of cells in the putative double positive gate from 38% in the -dox and 70% in the +dox. All cells within this hPGCLC gate were SOX2-tdT positive when BMP4 was omitted from the media, again suggesting these cells were not hPGCLCs (Figure 23 C & D). The flow cytometry analysis of -dox and +dox

suggests that without BMP4 the cell aggregate differentiates into an alternative lineage, presumably a neural lineage as this lineage maintains *SOX2* expression (Thomson et al., 2011). It is not possible to say from this data if the overexpression of *NANOG* causes a change in differentiation triggered by the cell aggregation without BMP4.

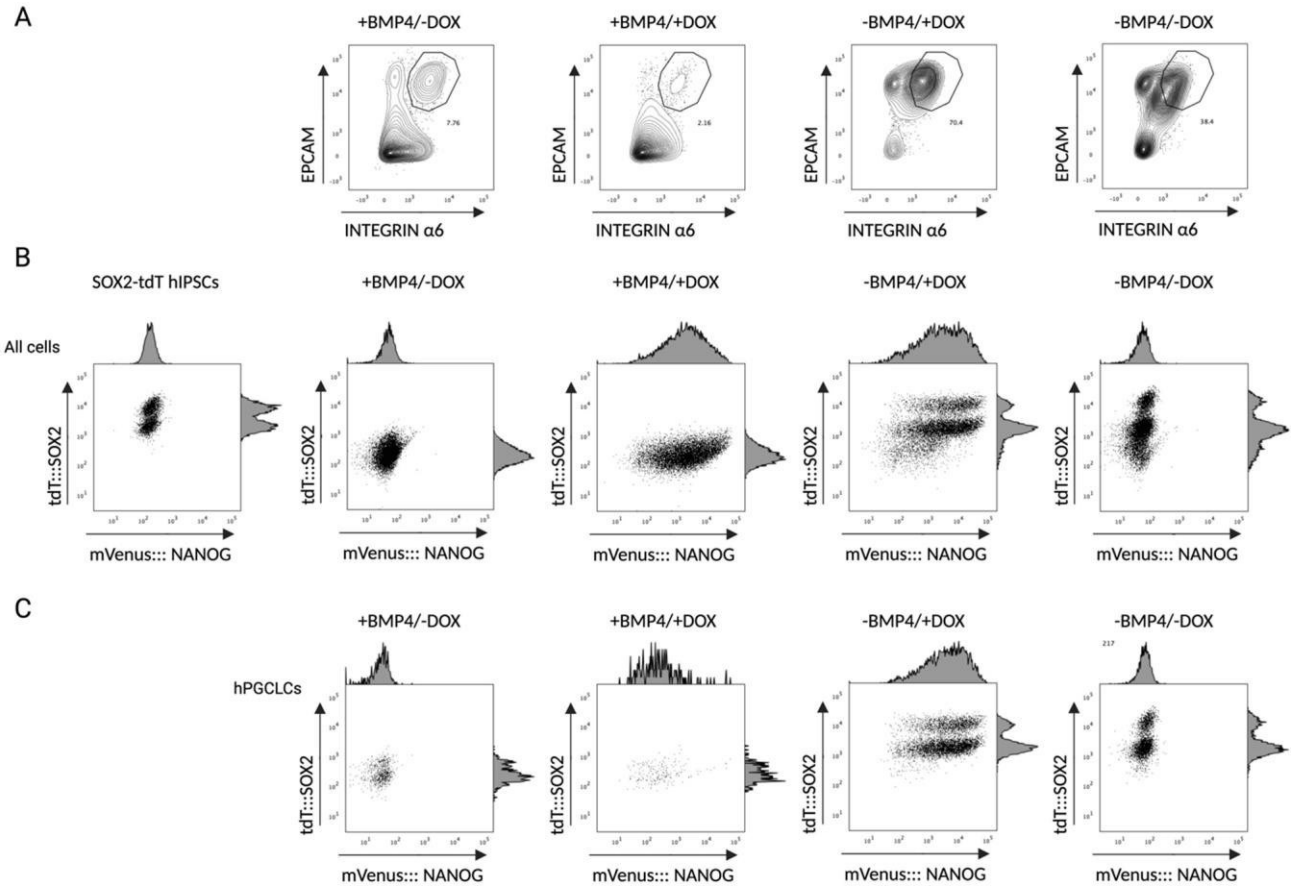

Figure 23 Comparing hPGCLC generation with and without dox, and with and without BMP4 (A) FACS plots of the four conditions trialled in this experiment showing surface marker stains (B) mVenus vs. tdT of all cells in the aggregates along with SOX2-tdT hiPSCs (C) mVenus vs. tdT of cells in hPGCLCs gate from (A).

I wanted first to confirm the cells generated without BMP4 were not hPGCLCs. I performed a preliminary RT-qPCR experiment on RNA from the isolated cell populations (Figure 23). These populations were tdT high and tdT low in the hPGCLC gate from the +dox/-BMP4 condition. Additionally, the EPCAM positive/INTEGRINα6 negative (E+ I-) population and the EPCAM positive/INTEGRINα6 positive (E+ I+) population were sorted from both the +dox -BMP4 and

the -dox -BMP4 samples. Expression of early germline genes *NANOS3, TFAP2C, BLIMP1* and *SOX17* along with pluripotency genes *NANOG* and *OCT4* were measured with RT-qPCR on these populations and expression levels normalised to hPGCLCs derived in -dox/+BMP4.

Expression data from these populations suggests that none of the cells within the hPGCLC gate had hPGCLC identity (Figure 24). The lower relative expression of all germline genes in all conditions when compared with the hPGCLCs from the +BMP4 condition indicates these cells were not hPGCLCs. While this is only a preliminary study with an n of 1, along with the FACS data showing that *SOX2*-tdT was still expressed in the cells within the 'hPGCLC' gate, these two pieces of data suggests that these cells are not hPGCLCs.

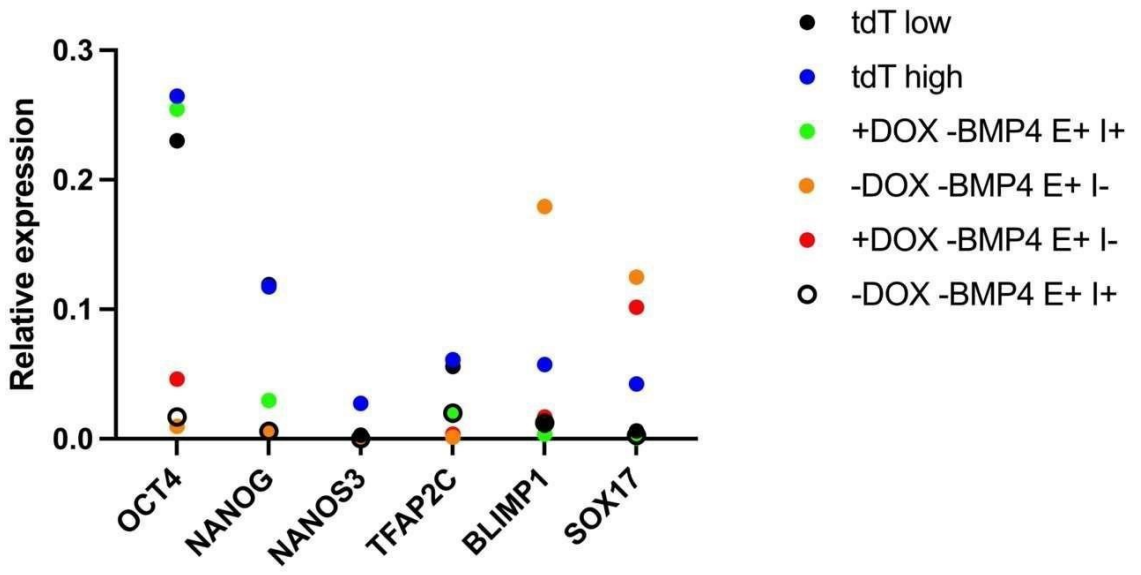

Figure 24 Relative expression of germline and pluripotency genes in the different populations from the experiments in Figure 24, compared to +BMP4 -dox. E+ represents EPCAM positive population, I +/- represents INTEGRINa6 positive or negative population respectively.

4.3.6 NANOG overexpressing cell lines were unable to express mVenus in hPGCLCs after induction.

My second attempt to re-activate the endogenous *SOX2* locus within hPGCLCs was to use exogenous *NANOG* overexpression. The overexpression of exogenous NANOG was measured

through the mVenus fluorescent reporter; activation from the endogenous SOX2 could be observed through the expression of the tdT fluorescent reporter. In addition, I analysed whether the overexpression caused changes to the hPGCLC conversion efficiency and hPGCLC identity.

NANOG c26 hiPSCs line, which overexpresses NANOG in response to dox, was used in the hPGCLC protocol above (Figure 17 A). Dox was added at day 1, day 2 or day 3 (termed d1, d2 or d3 respectively) after hPGCLC induction, analysed by flow cytometry and hPGCLCs collected using FACS. The dox added samples were then directly compared to a no dox (dx) control.

As mentioned above, hPGCLC identity can be defined by high expression of surface markers EPCAM and INTEGRIN α6, as shown in Figure 17 E. Expression of these surface markers is measured through staining with fluorescently tagged antibodies, with intensity recorded by flow cytometry.

There appeared to be no difference to the distribution of cells on the FACS plots (Figure 25 A). The mean conversion percentages into hPGCLCs was dx – 3.7%, d1 - 3.9%, d2 - 4.6% and d3 - 3.9 (Figure 25 B). The ranges of the conversion efficiency percentages in each sample were also overlapping. Using a one-way ANOVA showed there was not significant difference between the different samples, indicating that the addition of dox did not cause any change in conversion efficiency to hPGCLCs.

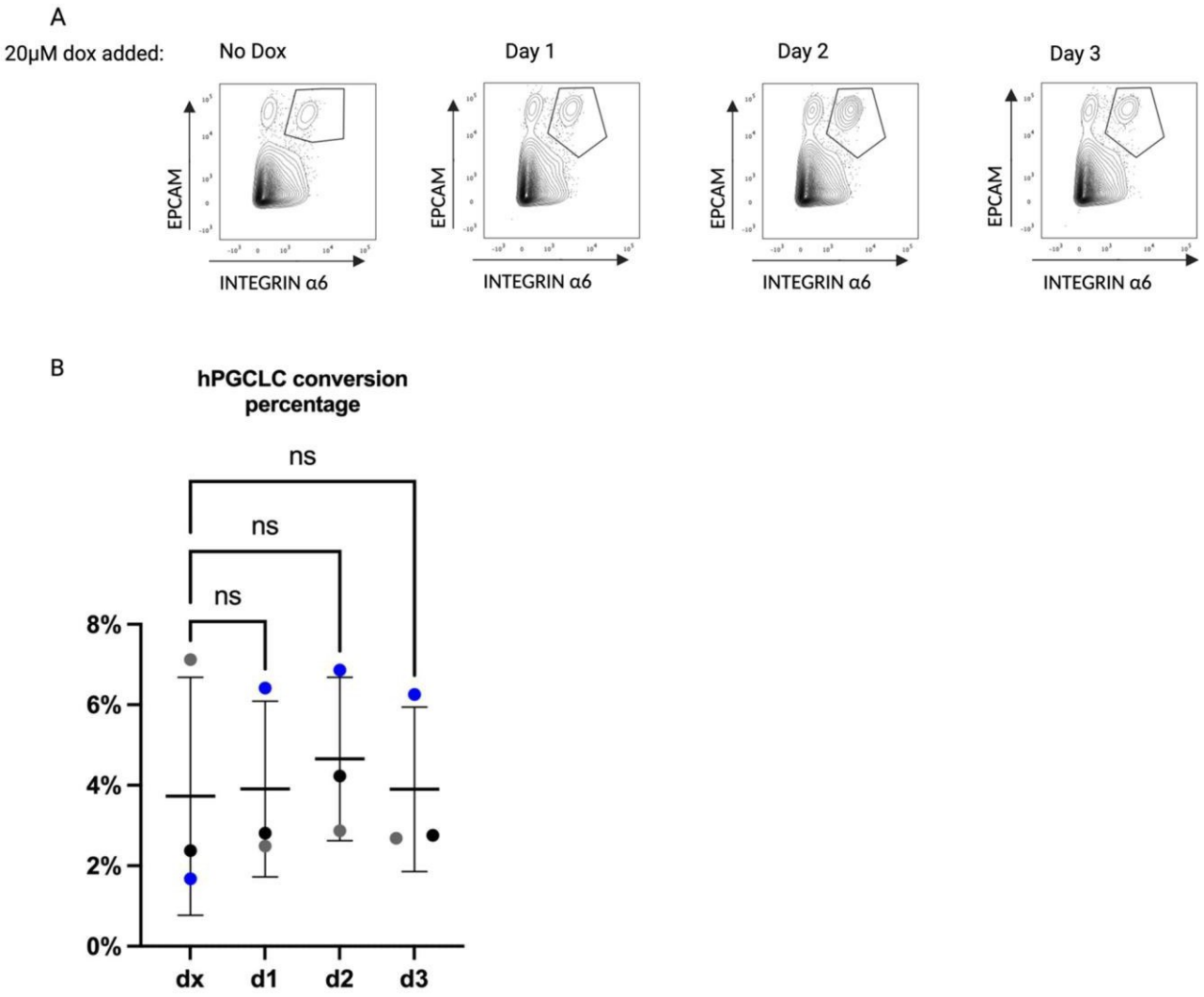

Figure 25 Effect of dox addition to NANOG aggregates (A) Representative FACS plots used to sort hPGCLCs from aggregates (B) Percentage of cells within hPGCLC gate (n=3 biological replicates) p>0.05 ordinary one way ANOVA.

To analyse the expression of the NANOG overexpression transgene, the intensity of the linked fluorescent reporter mVenus was tracked using flow cytometry. Measuring mVenus showed that the transgene was expressed in a subset of cells within the aggregate in all the samples where dox was added (Figure 26 A upper panel). The percentage of mVenus cells inside the aggregate decreased the later the dox was added (Figure 26 A upper panel & B upper panel). The mean percentage of the dox added samples reduced from 48% in d1 to 28% in d2 and 9% in d3. As hPGCLC development continues it appears that all cells of the aggregate, hPGCLCs and non-hPGCLCs, become less likely to activate the transgene.

hPGCLCs within the aggregate did not activate the transgene in larger numbers (Figure 26 A & B lower panel). The mean percentage of mVenus positive hPGCLCs was only d1 – 3.8%, d2 - 1% and d3 - 1% (Figure 26 B lower panel). The lack of mVenus positive hPGCLC can be clearly visualised when the expression of ECPAM and INTEGRIN α6 in the mVenus positive and negative cells within the aggreged are analysed by flow cytometry. Very few mVenus positive cell are inside the hPGCLC gate (Figure 26 C). To normalise the percentage of mVenus positive hPGCLCs to the percentage of mVenus positive all cells of the aggregate, the percentage of mVenus positive hPGCLCs was divided by the percentage of positive cells within the whole aggregate. This normalised ratio between the number of mVenus positive hPGCLCs and all mVenus positive cells in the aggregate is similar, regardless of which day dox was added, ranging between 5-10% (Figure 26 B right panel). This demonstrates that the transgene is resistant to being activated in hPGCLCs derived from the NANOG c26 hiPSC line.

The lack of mVenus positive hPGCLCs from the NANOG c26 hiPSC line was an unexpected result. DSOX2 cells can form mVenus positive hPGCLCs (Figure 26 C). If mVenus or another section of the transgene cassettes were incompatible with hPGCLC identity, the DSOX2 result would not have been possible. This would suggest that either *NANOG* overexpression in hPGCLCs causes differentiation or death, or that the transgene was unable to be activated in this hiPSC line. If NANOG overexpression caused differentiation or death it would lead hPGCLC which had activated the NANOG transgene to be underrepresented in the hPGCLC gate. Alternatively, the cassettes might have inserted into a locus of the genome which, in this clone, is epigenetically repressed during hPGCLC development. If the locus where the cassettes were inserted was tightly repressed in hPGCLCs, when dox was added to the aggregate, transcription from the transgene would be blocked due to chromatin repression of the promoters.

As expected, the endogenous SOX2-tdT reporter is fully downregulated in hPGCLCs in the control condition, when no dox is added. As *NANOG* was not overexpressed in a large number of hPGCLCs, it was not possible to conclude if *NANOG* overexpression could cause *SOX2* re-activation in this cell type (Figure 26 A).

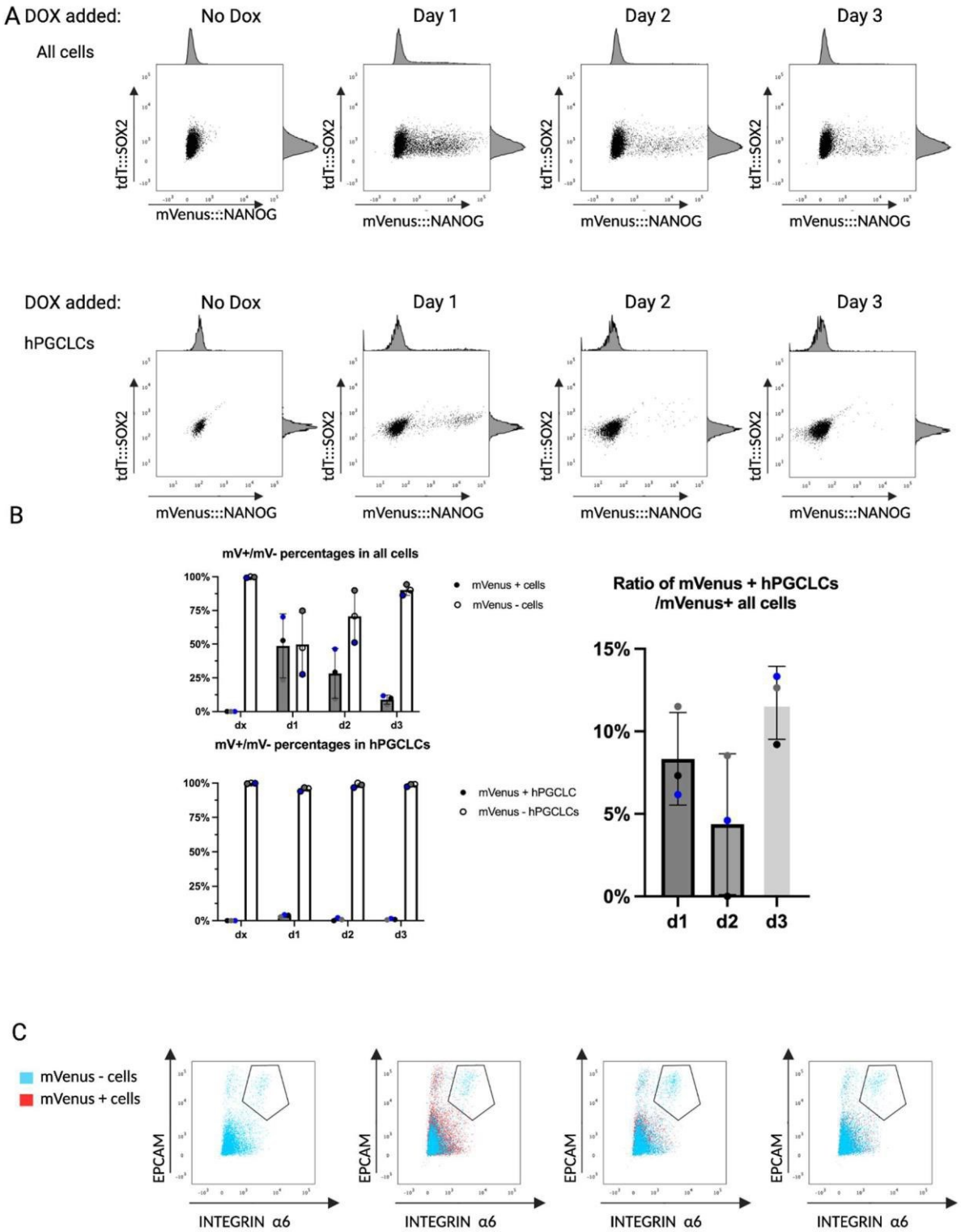

Figure 26 Analysis of mVenus and tdT expression within NANOG hPGCLC aggregates by flow analysis (A) Representative dot plots showing mVenus vs. tdT reporters Top: all cells of the aggregate. Bottom: hPGCLCs in the aggregate (B) quantification of mVenus positive and mVenus negative percentages Top: all cells. Bottom: hPGCLCs Right: percentage of mVenus positive hPGCLCs divided by percentage of all cells in the aggregate that are mVenus positive (n=3) (C) FACS plots showing distribution of mVenus positive and mVenus negative cells based on EPCAM and INTEGRINa6 staining's.

To visualise the location of hPGCLCs and mVenus positive cells within the NANOG c26 hPGCLC aggregates, I performed immunofluorescence on cryosections of these aggregates. mVenus and AP2γ were stained for in aggregates which had had no dox added, or dox added at day 1, day 2 or day 3. hPGCLCs appeared to cluster within the aggregate and could be found on both the external and internal areas of the aggregate (Figure 27). hPGCLCs were present in all 4 conditions and in the days where dox was added, mVenus positive cells were present. mVenus staining was mutually exclusive to AP2γ staining, supporting the FACS data that mVenus expression was not present in hPGCLCs (Figure 27 B C & D). The reduction in mVenus positive cells observed when dox is added on later days could also be observed in the staining's (Figure 27). These staining's therefore supports the FACS data that mVenus positive hPGCLCs could not be obtained. The lack of mVenus positive hPGCLCs did not appear to be due to dox not diffusing through the aggregate as many of the hPGCLC clusters were surrounded by mVenus positive cells.

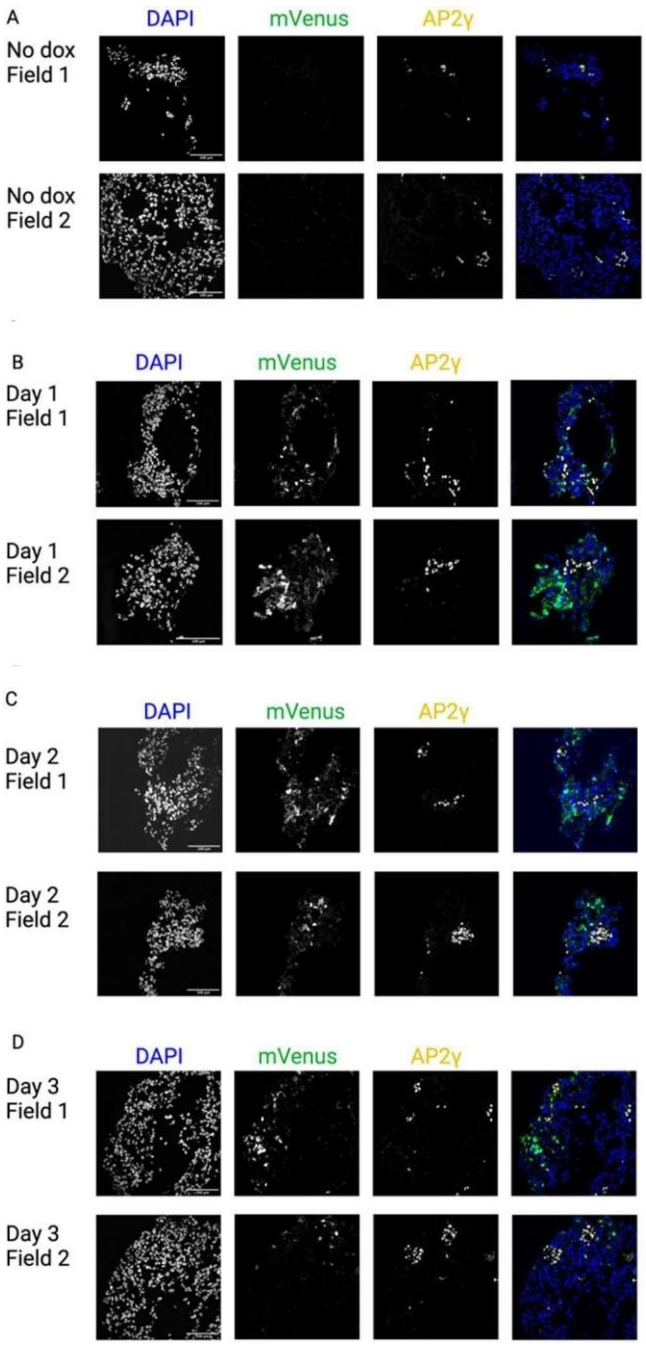

Figure 27 Staining of cryosections of NANOG hPGCLC aggregates (A) two fields from aggregates with no dox added, stained with mVenus, AP2g and DAPI with a composite of the two stains (Scale bar = 100mM) (B) two fields from aggregates with dox added on day 1, stained with mVenus, AP2g and DAPI with a composite of the two stains (Scale bar = 100mM) (C) two fields from aggregates with dox added on day 2, stained with mVenus, AP2g and DAPI with a composite of the two stains (Scale bar = 100mM) (D) two fields from aggregates with dox added on day 3, stained with mVenus, AP2g and DAPI with a composite of the two stains (Scale bar = 100mM).

4.3.7 KLF2 over expressing cell lines were also unable to express mVenus in hPGCLCs after induction

My third attempt to reactivate the endogenous *SOX2* locus within hPGCLCs was to use exogenous *KLF2* overexpression. The overexpression of exogenous KLF2 was measured through the mVenus fluorescent reporter and activation from the endogenous SOX2 could be observed through the expression of the tdT fluorescent reporter. Additionally, I analysed if the overexpression caused changes to the hPGCLC conversion efficiency and hPGCLC identity.

KLF2 c7 hiPSCs line, which overexpresses KLF2 in response to dox, was used in the hPGCLC protocol above (Figure 17 A). Dox was added at day 1, day 2 or day 3 (termed d1, d2 or d3 respectively) after hPGCLC induction, analysed by flow cytometry and hPGCLCs collected using FACS. The dox added samples were then directly compared to a no dox (dx) control.

As mentioned above, hPGCLC identity can be defined by high expression of surface markers EPCAM and INTEGRIN α6, as shown in Figure 17 E. Expression of these surface markers is measured through staining with fluorescently tagged antibodies, with intensity recorded by flow cytometry.

There appeared to be no difference to the distribution of cells on the FACS plots (Figure 28 A). The mean conversion percentages into hPGCLCs were dx – 9.8%, d1 – 10.9%, d2 – 10.1% and d3 – 11.1% (Figure 28 B). The ranges of these percentages were also overlapping, and using a one-way ANOVA showed there was not a significant difference between the different samples, suggesting that the addition of dox didn't cause any change in conversion efficiency to hPGCLCs.

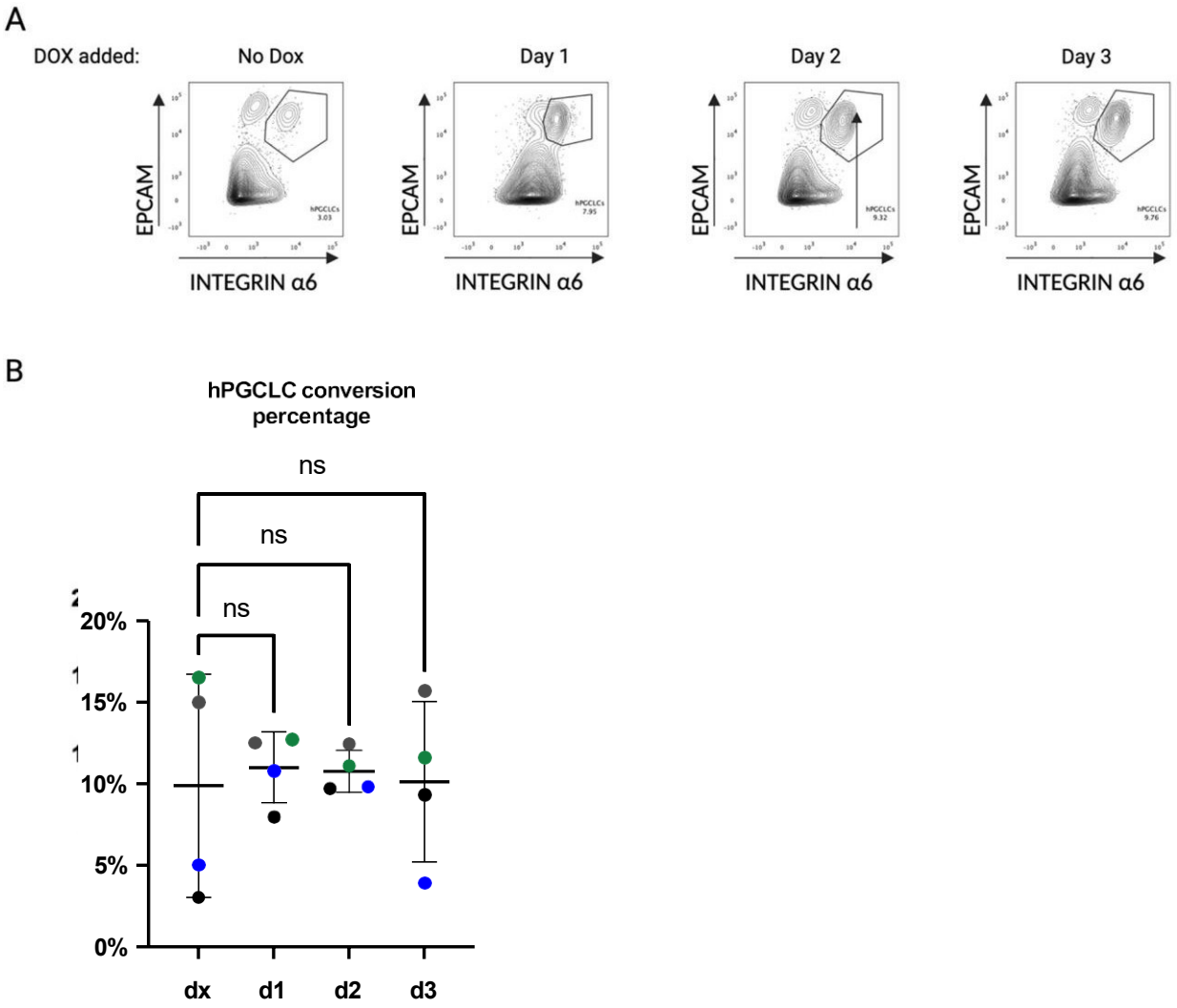

Figure 28 Effect of dox addition to KLF2 aggregates (A) example of FACS plots used to sort hPGCLCs from aggregates (B) percentage of cells within hPGCLC gate n=4 p>0.05 ordinary one-way ANOVA.

To analyse the expression of the KLF2 overexpression transgene, the expression of the linked fluorescent reporter mVenus was tracked using flow cytometry. Measuring mVenus showed that the transgene was expressed in a subset of cells within the aggregate in all the samples where dox was added (Figure 29 A upper panel). The percentage of mVenus cells inside the aggregate decreased the later the dox was added (Figure 29 A upper panel & B upper panel). The mean percentage of the dox added samples reduced from 55% in d1 to 27% in d2 and 13% in d3. As hPGCLC development continues it appears that all cells of the aggregate, hPGCLCs and non-hPGCLCs, become less likely to activate the transgene.

hPGCLCs within the aggregate did not activate the transgene in larger numbers (Figure 29 A & B lower panel). The mean percentage of mVenus positive hPGCLCs was only d1 – 5.4%, d2 – 1.3% and d3 – 0.6% (Figure 29 B lower panel). The lack of mVenus positive hPGCLC can be clearly visualised when the expression of ECPAM and INTEGRIN α6 in the mVenus positive and negative cells within the aggreged are analysed by flow cytometry. Very few mVenus positive cell are inside the hPGCLC gate (Figure 29 C). To normalise the percentage of mVenus positive hPGCLCs to the percentage of mVenus positive all cells in the aggregate, the percentage of mVenus positive hPGCLCs was divided by the percentage of positive cells within the whole aggregate. This normalised ratio between the number of mVenus positive hPGCLCs and all mVenus positive cells in the aggregate is similar regardless of which day dox was added (Figure 29 B right panel), with means between 1.5% - 9%. An ordinary one-way ANOVA showed there was no significant difference between the normalised ratios. Regardless of the day dox is added, the transgene is resistant to being activated in hPGCLCs derived from the KLF2 c7 hiPSC line.

This result is the same as the result observed in the NANOG c26 line. This would suggest that either *KLF2* overexpression in hPGCLCs causes differentiation or death, or that the transgene was unable to be activated in this hiPSC line. If KLF2 overexpression caused differentiation or death it would lead hPGCLC that had activated the KLF2 transgene to be underrepresented in the hPGCLC gate. Alternatively, the cassettes might have inserted themselves into a locus of the genome, which in this clone, is epigenetically repressed during hPGCLC development from hiPSCs. If the locus where the cassettes were inserted were tightly repressed in hPGCLCs, when dox was added to the aggregate, transcription from the transgene would be blocked due to chromatin repression of the promoters.

As expected, the endogenous SOX2-tdT reporter is fully downregulated in hPGCLCs in the control condition, when no dox is added. As *KLF2* was not overexpressed in a large number of hPGCLCs, it was not possible to conclude if *KLF2* overexpression could cause *SOX2* re-activation in this cell type.

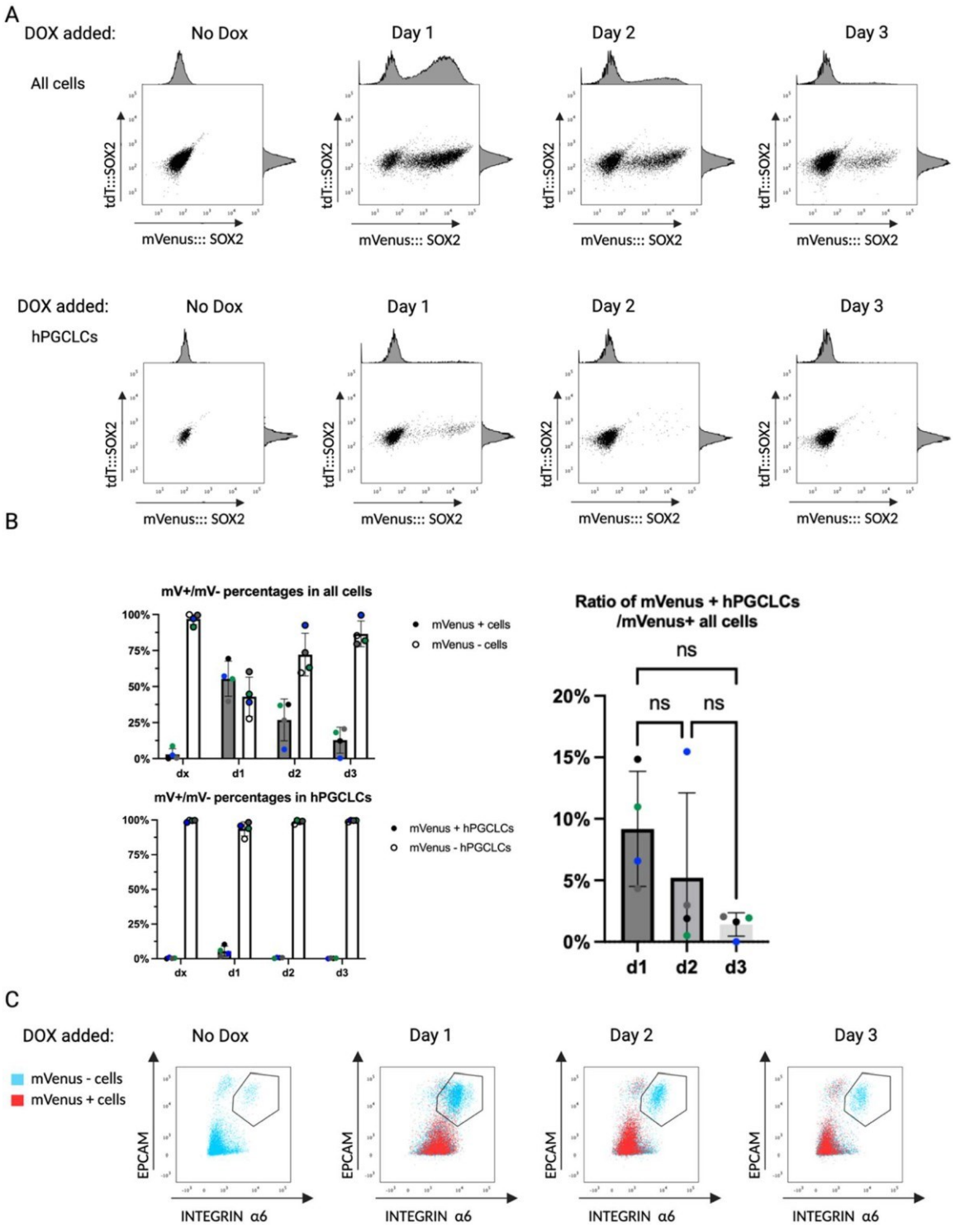

Figure 29 Analysis of mVenus and tdT within KLF2 hPGCLC aggregates by flow analysis (A) example dot plots showing mVenus vs. tdT Top: all cells of the aggregate. Bottom: hPGCLCs in the aggregate (B) quantification of mVenus positive and mVenus negative percentages Top: all cells. Bottom: hPGCLCs Right: percentage of mVenus positive hPGCLCs divided by percentage of all cells in the aggregate that are mVenus positive n=4, $p<0.05$ ordinary one-way ANOVA (C) FACS plots showing

distribution of mVenus positive and mVenus negative cells based on EPCAM and INTEGRINa6 staining).

hPGCLCs were collected for RT-qPCR and aggregates stained for germline markers. Aggregates with dox added at day 2 and day 3 were fixed, cryosectionned and then stained for OCT4 and AP2γ, markers for early germ cell identity (Sasaki et al., 2015). Two different fields from samples where no dox was added (Figure 30 A) and when dox was added at day 1 (Figure 30 B), at day 2 (Figure 30 C) or day 3 (Figure 30 D) are shown. Clusters of cells which are stained for both AP2γ and OCT4 are clearly visible in all fields confirming the presence of hPGCLCs in the aggregates. There are subtle variations in the intensity of stain between the two factors, suggesting there might be variations in expression of the two factors in individual hPGCLCs. Despite this, it appears that the expression of one factor is accompanied by the expression of the other, confirming these cells are hPGCLCs.

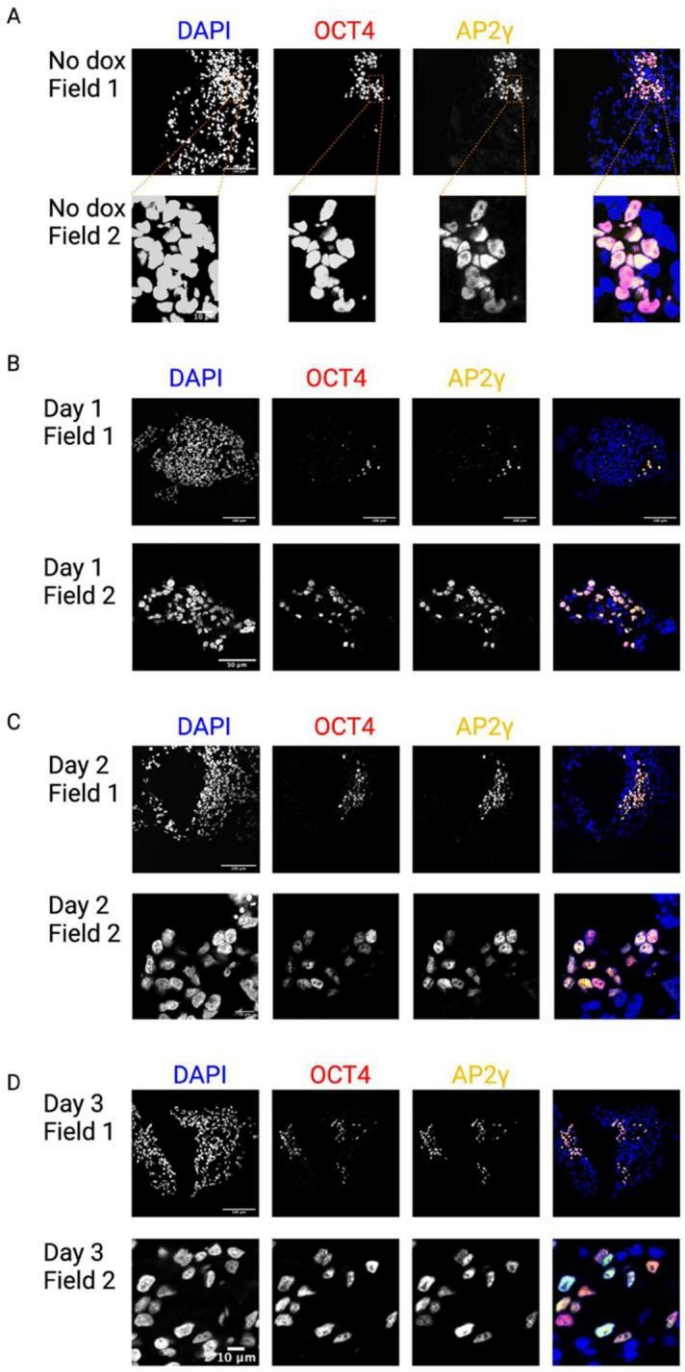

Figure 30 Staining of cryosections of KLF2 hPGCLC aggregates (A) two fields from aggregates with no dox added, stained with OCT4, AP2g and DAPI with a composite of the two stains (Scale bars Top = 100mM, Bottom = 10mM) (B) two fields from aggregates with dox added on day 1, stained with OCT4, AP2g and DAPI with a composite of the two stains (Scale bars top = 100mM, Bottom = 50mM) (C) three fields from aggregates with dox added on day 2, stained with OCT4, AP2g and DAPI with a composite of the two stains (Scale bar Top one = 100mM, Bottom two = 10mM) (D) three fields from aggregates with dox added on day 3, stained with OCT4, AP2g and DAPI with a composite of the two stains (Scale bar Top and Bottom = 100mM, Middle = 10mM).

4.3.8 'EV' cell lines also don't express mVenus in hPGCLCs after hPGCLCs

The lack of mVenus positive hPGCLCs in the NANOG c26 and KLF2 c7 hiPSC lines was unexpected as the DSOX2 c1 hiPSCs were able to form mVenus positive hPGCLCs. To test if the lack of mVenus hPGCLCs in NANOG c26 and KLF2 c7 was due to the pluripotency transcription factor disrupting hPGCLC development and not the transgene failing to be expressed in hPGCLCs, 'EV' c3 hiPSC line was used. The cassettes inserted into these hiPSCs contained the same backbone as the other TF plasmids, but instead of a pluripotency-related TF a chloramphenicol resistance (CmR) gene is placed after the CMV mini-promoter. mVenus is still transcriptionally linked to the CmR and is transcribed and translated in response to dox. This 'EV' line should reveal if either *NANOG* or *KLF2* overexpression is the reason for the lack of mVenus positive hPGCLCs in these lines. If mVenus positive hPGCLCs are produced from the 'EV' hiPSCs, it would suggest that TF overexpression results in the lack of mVenus positive hPGCLCs. However, if there are no or few mVenus positive hPGCLCs produced from the 'EV' hiPSCs, it would suggest the overexpression cassette is being repressed, perhaps epigenetically in hPGCLCs.

EV c3 hPGCLC identity was defined by high expression of surface markers EPCAM and INTEGRIN α6, as shown in Figure 17 E. Expression of these surface markers is measured through staining with fluorescently tagged antibodies, the intensity recorded by flow cytometry. There was no difference in the distribution of these surface markers when dox was added (Figure 31 A). The conversion was lower in the dox added conditions with the percentages dropping from 19.3% in dx to 14.4% in d1 13.2% in d2 and 10.8% in d3 (Figure 31 B). This data would suggest that the addition of dox reduces the conversion efficiency into hPGCLCs in the 'EV' cell line, although no significant conclusions can be drawn as this experiment only had an n of 1.

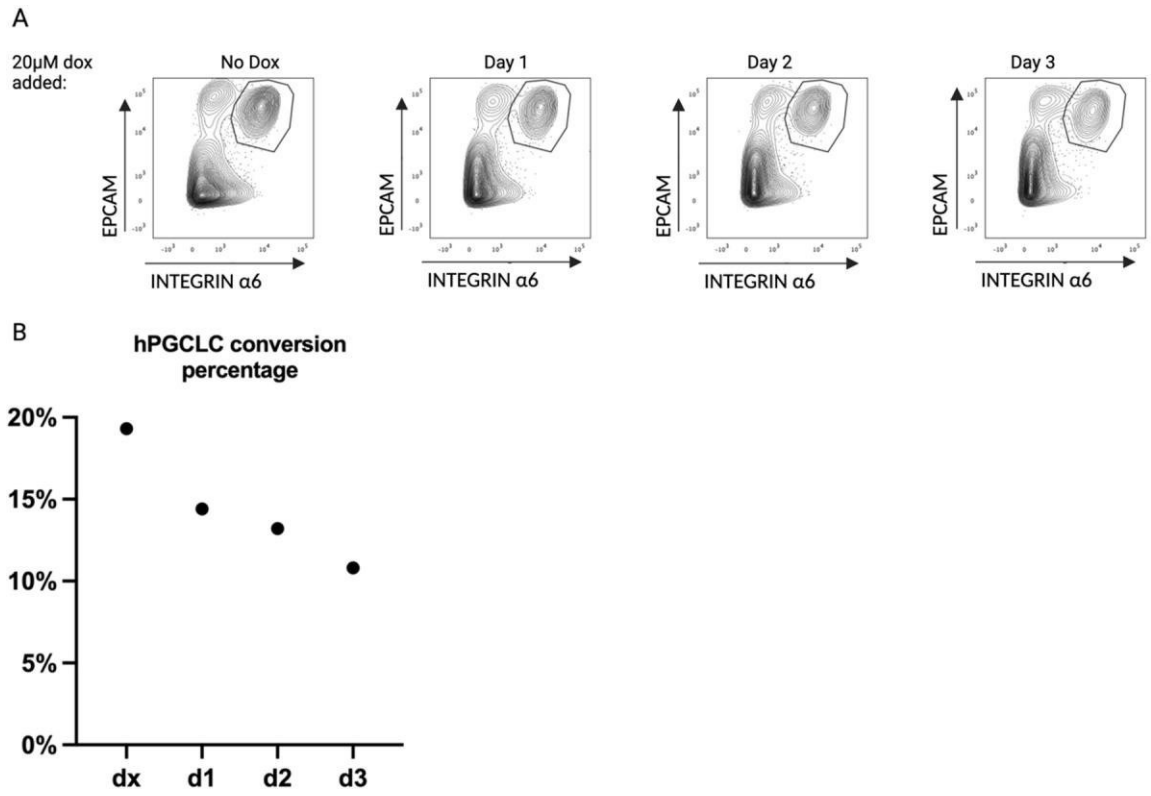

Figure 31 Effect of dox addition to 'EV' aggregates (A) example of FACS plots used to sort hPGCLCs from aggregates (B) percentage of cells within hPGCLC gate n=1.

The expression of the fluorescent reporter mVenus was tracked using flow cytometry, revealing these 'EV' aggregates are more likely to express the transgene. Measuring mVenus showed that the transgene was expressed in more cells in the 'EV' aggregate compared to the other TF aggregates; d1 – 79.5%, d2 -73% and d3- 38.5% (Figure 32 A upper panel & B upper panel), compared to a mean 49% in d1, 26% in d2 and 11% in d3 across the three TF hPGCLCs.

The EV experiment is only conducted with an n of 1, but if the trend of the more cells within 'EV' aggregates activated mVenus compared to the cells of the TF aggregate, it points to two conclusions. Firstly, the higher expression of mVenus; and therefore, high activation of the transgene, could explain why the conversion efficiency reduces in the 'EV' aggregates upon dox addition. High expression of the transgene could be the transcriptional and translational machinery of the cell, perhaps preventing the cells entering hPGCLC fate. Secondly, the higher expression could suggest that the expression of any of the TFs leads to a reduction in

mVenus positive cells within the aggregate. High overexpression of the TF could cause death in the cells of the aggregate, or the lack of overexpression, i.e. the mVenus negative cells could have a competitive advantage as they are not expressing the TF.

hPGCLCs from all four hiPSC show the same pattern of mVenus activation when dox is added to the aggregate later. All cells of the aggregate, hPGCLCs and non-hPGCLCs, become less likely to activate the transgene as hPGCLC development continues. This would suggest that the differentiation of the iMeLCs into any cell type, hPGCLC or non-hPGCLC, results in the repression of the transgene cassette. This could be direct repression of the CAG or CMV promoter or through repression of the locus where the transgene was inserted into the genome.

hPGCLCs within the aggregate did not activate the transgene in larger numbers (Figure 32A & B lower panel). The percentage of mVenus positive hPGCLCs was only d1 – 1.3 %, d2 – 1.3% and d3 – 0.6% (Figure 32B lower panel). The lack of mVenus positive hPGCLC can be clearly visualised when the expression of ECPAM and INTEGRIN α6 in the mVenus positive and negative cells within the aggreged are analysed by flow cytometry. Very few mVenus positive cell are inside the hPGCLC gate (Figure 32C). To normalise the percentage of mVenus positive hPGCLCs to the percentage of mVenus positive all cells in the aggregate, the percentage of mVenus positive hPGCLCs was divided by the percentage of positive cells within the whole aggregate. This normalised ratio between the number of mVenus positive hPGCLCs and all mVenus positive cells in the aggregate is similar, regardless of which day dox was added with the ratio being <2% (Figure 32B right panel).

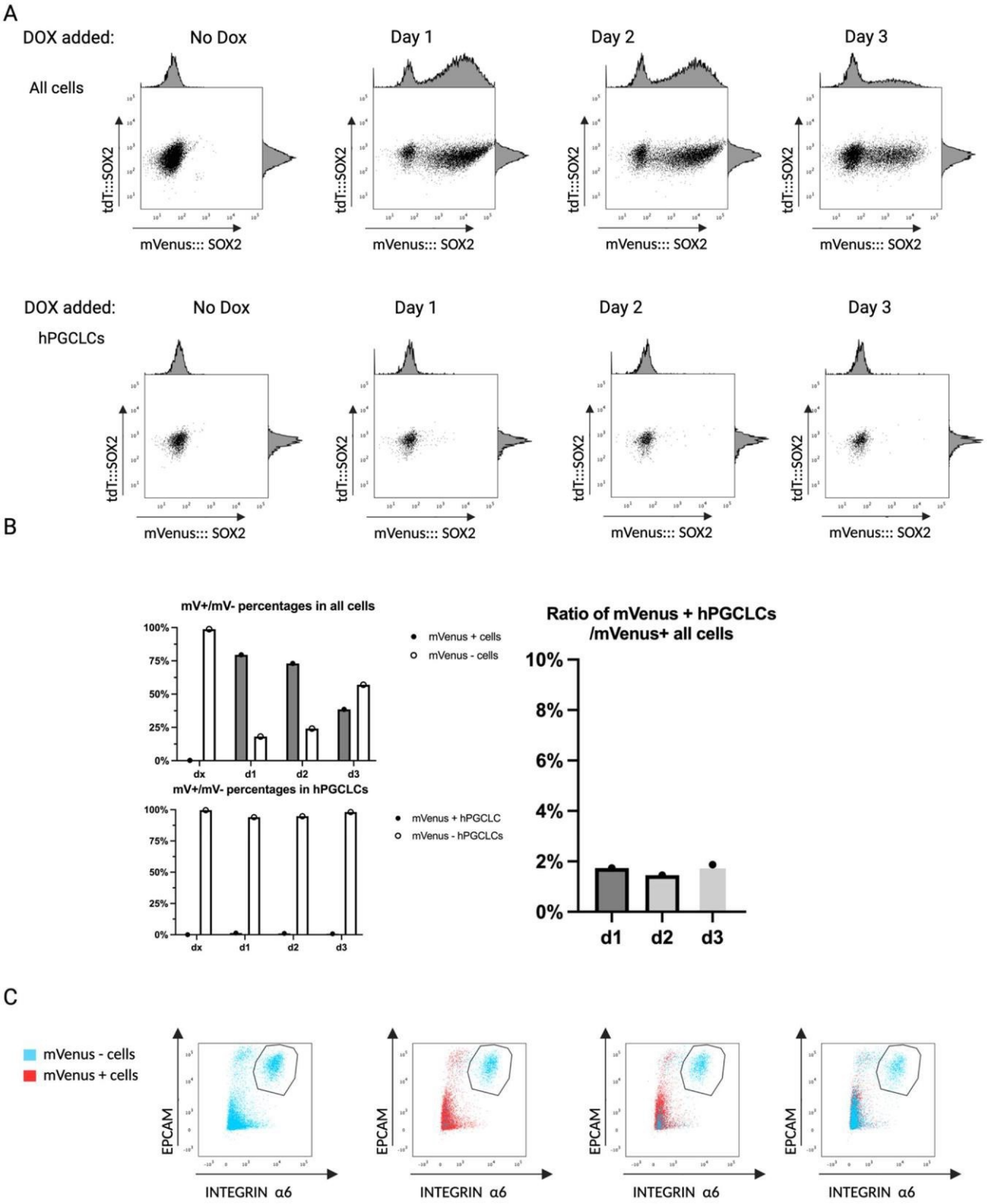

Figure 32 Analysis of mVenus and tdT within 'EV' hPGCLC aggregates by flow analysis (A) example dot plots showing mVenus vs. tdT Top: all cells of the aggregate. Bottom: hPGCLCs in the aggregate (B) quantification of mVenus positive and mVenus negative percentages Top: all cells. Bottom: hPGCLCs Right: percentage of mVenus positive hPGCLCs divided by percentage of all cells in the aggregate that are mVenus positive (C) FACS plots showing distribution of mVenus positive and mVenus negative cells based on EPCAM and INTEGRINa6 staining.

The lack of mVenus positive mVenus hPGCLCs in this 'EV' control suggests that instead of the transcription factor causing differentiation or death in hPGCLCs, the transgene itself is repressed in hPGCLCs (Figure 32 C). Expression of mVenus itself might cause hPGCLCs to die due to cell stress associated with producing the mVenus protein. However, as DSOX2 hPGCLCs were produced at a reasonable ratio to all mVenus positive, around 40% to 60%, contradicting the hypothesis that mVenus was specifically toxic for hPGCLCs. Instead, the most likely explanation for the lack of mVenus positive hPGCLCs in all hiPSC other than the DSOX2 c1 line was because the cassettes randomly inserted into a locus of the genome which is unrepressed in hiPSCs but repressed in hPGCLCs.

4.3.9 mVenus can be activated in NANOG c26 cells before hPGCLC induction, suggesting *NANOG* overexpression is possible in hPGCLCs

NANOG and *KLF2* overexpression could not be achieved from the two hiPSC lines, when dox was added to the hPGCLC aggregate after germline induction. Data from the 'EV' hPGCLCs from the section above suggested that this was because of repression of the inserted cassette in hPGCLCs. As these hiPSCs lines were able to express the TF transgene, I theorised that if dox was added to hPGCLC aggregate at the point of induction, the transgene might be able to be activated.

NANOG c26 hiPSCs line, which overexpresses NANOG in response to dox, were used in the hPGCLC protocol above (Figure 17 A). When dox was added at induction of hPGCLC fate, almost all cells in the aggregate were mVenus positive (Figure 33A, B & C, 3rd panels from left), including any that were in the putative hPGCLC gate. I describe this as a putative gate, as when dox is added to the aggregate, the intensities of surface markers EPCAM and INTEGRIN α6 were variable across the experiments. Flow cytometry analysis can identify the hPGCLC population (Sasaki et al., 2015) but as shown in the -BMP4 experiments from Figure 24 the cells which fall into the double positive EPCAM/INTEGRIN α6 gate are not always hPGCLCs.

High mVenus expression was observed in the 1st and 3rd experiment; suggesting *NANOG* overexpression could be triggered. In these experiments all cells in the aggregate upregulated INTEGRIN α6 and had a wide distribution of EPCAM (Figure 33 A & C, 3rd panels from left). The gate for the first experiment was set on a population with high EPCAM expression, which had slightly lower INTEGRIN α6 expression than the no dox control Figure 33 A, 1st & 2nd panel from left). Cells in this gate highly expressed mVenus suggesting successful activation of the transgene in hPGCLCs when dox was added before induction. In the third experiment, all cells upregulated INTEGRIN α6 compared to the no dox control (Figure 33 C, 1st & 2nd panel from left). Within the EPCAM positive population there appeared to be two populations of different INTEGRIN α6 expression (Figure 33 C, 2nd panel from left & D left panel). mVenus is more highly expressed in the gate where INTEGRIN α6 is expressed at a lower level (Figure 33 C, 4th panel from left and D, right panel); although the difference of INTEGRIN α6 expression was only subtle. Without gene expression analysis it was not possible to say what these populations were or how they different, and if they were hPGCLCs.

The lack of tdT expression, suggesting repression of the endogenous *SOX2* locus would imply that the cells within the double positive gate were in fact hPGCLCs. However, without further analysis, specifically gene expression analysis such as RT-qPCR, it is not possible to discern any difference between these populations, and if one of these defines hPGCLCs exclusively. Because of this, it was not possible to describe the effect that *NANOG* overexpression had on hPGCLC conversion efficiency.

In the second experiment, the addition of dox at induction did not change the distribution of the surface markers and did reduce the number of cells in the hPGCLCs gate compared to the no dox control (Figure 33 B, 1st & 2nd panel from left). These cells showed mVenus expression, although its intensity was much lower in these cells compared to mVenus positive cells from the rest of the aggregate (Figure 33 B, 3rd & 4th panel from left). In this experiment, it appears that *NANOG* expression pattern was more similar to the post-induction experiments (Figure 25 A) to the other pre-induction experiments, where *NANOG* overexpression was absent from hPGCLCs.

Overall, it seems possible that the transgene can be activated in hPGCLCs when dox is added before induction, but its expression is variable between experiments. In two of the experiments *NANOG* overexpression was widely observable, even in the putative hPGCLCs. Furthermore, *NANOG* overexpression caused all cells of the hPGCLCs aggregate to highly express *INTEGRIN 6*, but in the other experiment *INTEGRIN 6* was only expressed in the cells within the hPGCLC gate. In this experiment where *INTEGRIN 6* was not upregulated, the percentage of hPGCLCs was reduced and the expression of mVenus was lower compared with other experiments. These results are intriguing, but without defining the populations on the flow cytometry plots it was not possible to draw further conclusions about the effect *NANOG* has on hPGCLC induction when expressed at induction or what mechanisms might be involved in this phenotype. It was also not possible to conclude if *NANOG* overexpression was not able to activate endogenous *SOX2* expression without first defining if the generated cells were hPGCLCs.

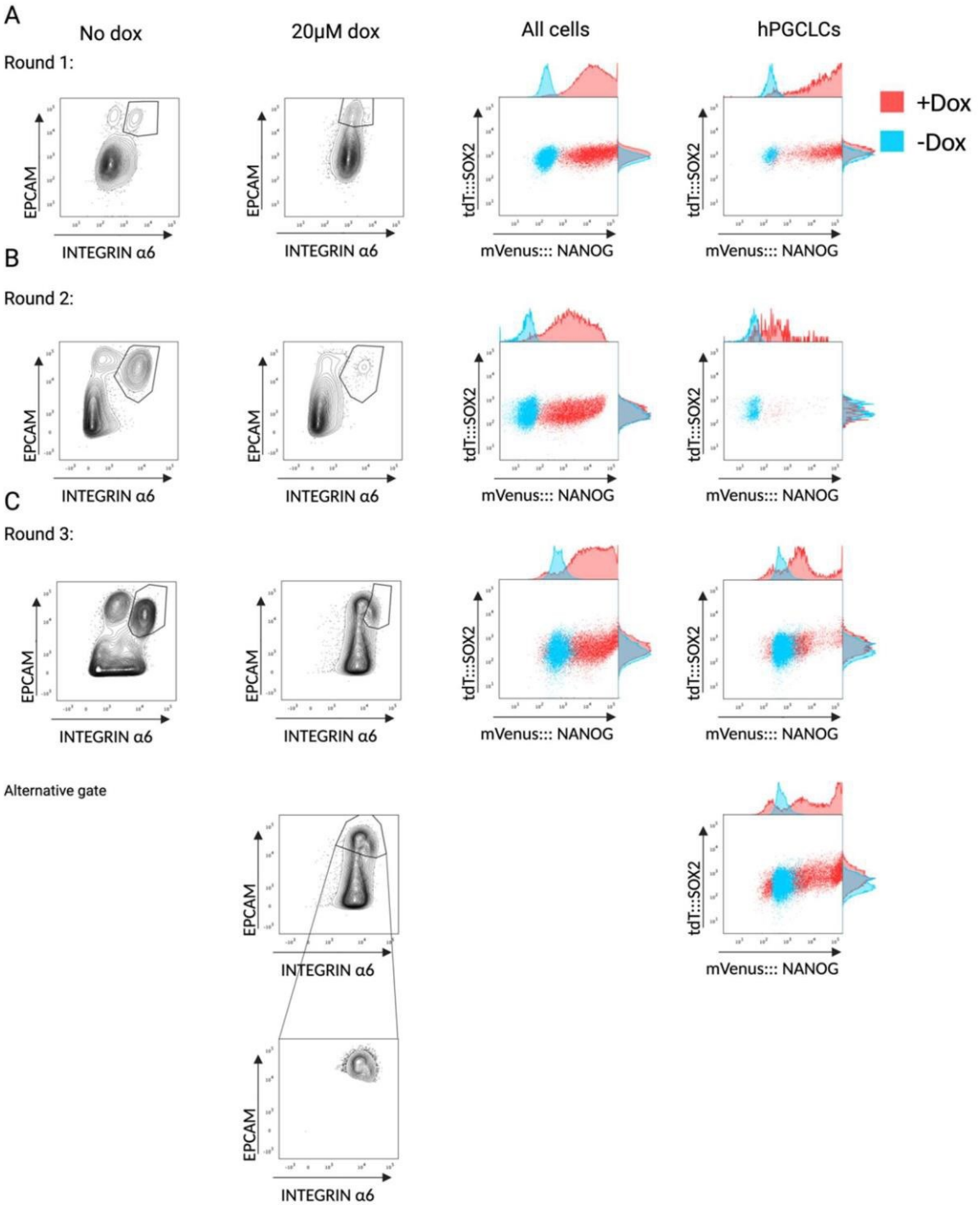

Figure 33 Effect of dox addition at the point of hPGCLC induction in NANOG OE cells (A) First experiment using 20mM dox, Flow analysis of surface markers staining and mVenus vs tdT in all cells and hPGCLCs (B) Second experiment using 20mM dox, Flow analysis of surface markers staining and mVenus vs tdT in all cells and hPGCLCs (C) Third experiment using 20mM dox, Flow analysis of surface markers staining and mVenus vs tdT in all cells and hPGCLCs, Top set of panels use stringent gate,

lower panels show larger alternative gate with this population shown in more detail in the lowest FACS plot.

4.3.10 mVenus can be activated in KLF2 cells when dox is added at hPGCLC induction, suggesting *KLF2* can be overexpressed in hPGCLCs

As *NANOG* could be overexpressed in putative hPGCLCs when dox was added at induction of hPGCLC fate, I wanted to assess if *KLF2* could also be expressed using this method. KLF2 c7, which overexpresses *KLF2* in response to dox, were used in the hPGCLC protocol above (Figure 17 A) with dox added at induction. mVenus was expressed in all cells of the aggregate in the 1st and 2nd experiment, and in the 75% of the cells in the 3rd (Figure 34 A, B & C, 3rd panel from left). INTEGRINα6 was highly upregulated in all the cells of the aggregate compared to the no dox control and had a wide distribution of EPCAM (Figure 34 A, B & C, 1st 2nd panel from left). Despite this change in surface maker distribution, the hPGCLC gate still appears to mark a double positive population in the + dox conditions. The lack of tdT expression shows *SOX2* is repressed in these cells, suggesting these cells are hPGCLCs. I therefore interpreted that these cells are hPGCLCs and the conversion efficiency can be calculated. *KLF2* overexpression at hPGCLCs induction did not significantly influence the conversion percentages as measured by a paired t-test (p=0.5) (Figure 34 D).

KLF2 overexpression, as tracked through the mVenus reporter, was possible in hPGCLCs when dox was added pre-induction. mVenus was highly expressed in hPGCLCs; 100% in the first 2 experiments and 82% in the 3rd. This would suggest *KLF2* could be activated in hPGCLCs.

The presence of mVenus positive hPGCLCs in *KLF2* hPGCLC aggregates through pre-induction dox addition supports the hypothesis that the absence of mVenus positive hPGCLCs is caused by transgene repression associated with differentiation into hPGCLCs. The transgene cassettes are accessible in the iMeLC state, so when dox is added at the induction of hPGCLCs, the transgene begins transcribing in all cells of the aggregate, including those fated to become hPGCLCs.

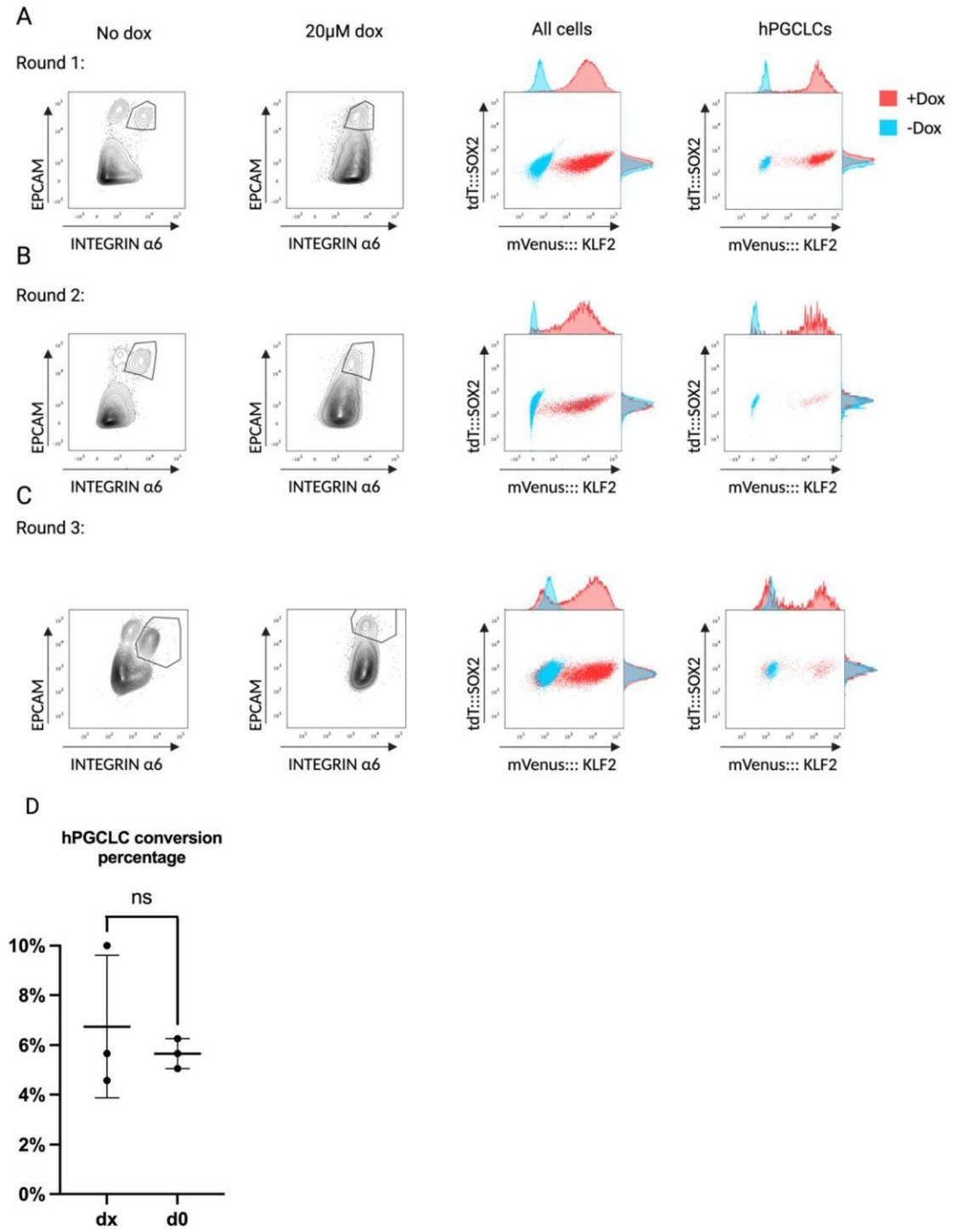

Figure 34 Effect of dox addition at the point of hPGCLC induction in KLF2 OE cells (A) First experiment using 20mM dox, Flow analysis of surface markers staining and mVenus vs tdT in all cells and hPGCLCs (B) Second experiment using 20mM dox, Flow analysis of surface markers staining and mVenus vs tdT in all cells and hPGCLCs (C) Third experiment using 20mM dox, Flow analysis of surface markers staining and mVenus vs tdT in all cells and hPGCLCs, Top set of panels use stringent gate, lower panels show larger alternative gate with this population shown in more detail in the lowest FACS plot.

mVenus was expressed in a higher proportion of all cells within the aggregates, where the mean was 85% across all pre-induction experiments, compared to the post-induction whose mean was 49% in d1, 26% in d2 and 11% in d3. This supports the notion that mVenus becomes harder to express in any cell of the hPGCLC aggregate the longer it has differentiated within the aggregate, regardless of the cell identity.

4.3 Discussion

4.3.1 SOX2 has no effect on hPGCLCs when overexpressed after induction of hPGCLCs, but blocks germline entry when overexpressed before induction of hPGCLCs

It could be expected that exogenous *SOX2* overexpression in hPGCLCs relieve repression of the endogenous *SOX2* locus. However, it appears that SOX2 overexpression is tolerated by hPGCLCs, with no effect on conversion percentages or upregulation of endogenous SOX2. This would suggest that there is another mechanism that is blocking the action of the exogenous SOX2, preventing it from activating the endogenous locus.

In contrast, when SOX2 exogenous overexpression was activated at hPGCLC induction, there is a reduction in hPGCLC conversion efficiency. There is also an increase in EPCAM positive only cells. The higher the dox concentration, the higher the reduction in hPGCLC conversion. This suggests that if SOX2 is overexpressed at the point of hPGCLC induction, it blocks cells from establishing the germline network. Instead, the increase in EPCAM might suggest that cells are being kept in a stem cell state as it is selectively expressed in human pluripotent cells (Lu et al., 2010).

Endogenous *SOX2* is not upregulated in cells of any of the aggregates, regardless of when *SOX2* is overexpressed. This suggests the *SOX2* locus is repressed in all cells that are produced by the aggregate, and this repression cannot be relieved through *SOX2* overexpression.

4.3.2 NANOG alone cannot drive hPGCLC fate in this system

NANOG overexpression was not able to induce hPGCLC without BMP signalling, in contrast to mouse PGCLCs which can be specified through *Nanog* overexpression, even when BMP signalling is blocked (Murakami et al., 2016). There does seem to be an effect on the cells of the aggregate, upregulation of INTEGRIN α6, but this doesn't seem to be germline related. It should be noted that the system used to generate mPGCLCs through Nanog overexpression is different to the one used in this study. Instead of human primed cells being cultured into iMeLCs before aggregation, mouse naïve cells are briefly differentiated into EpiLCs, a formative state (Smith, 2017), before aggregation and NANOG overexpression. In this state, Nanog is able to trigger expression of *Prdm1* and *Prdm14*, through binding at *Prdm1* and *Prdm14* enhancers. If Nanog is overexpressed too early in EpiLCs, 24 hours, it simply re-sets these cells into naïve state as these cells haven't entered into the formative state (Murakami et al., 2016).

It may be possible that human PGCLCs could be induced from NANOG induction if the same protocol is followed. These results do show that NANOG itself is not capable of inducing hPGCLCs in all cell contexts, even in those where BMP signalling is induced/active. As there are a number of ways of inducing hPGCLCs; from iMeLCs (Sasaki et al., 2015), 4i cells (Irie et al., 2015) or EpiLCs (von Meyenn et al., 2016), the context in which NANOG overexpression might trigger germline fate could be instructive in understanding how germline competency is achieved.

4.3.3 Inability of NANOG, KLF2 AND 'EV' clones to generate mVenus positive hPGCLCs could be due to chromatin conformation and positional effects on the inserted cassettes

The inability of the transgene to activate in hPGCLCs in the KLF2 c7, NANOG c26 and EV c3 aggregates limited the effectiveness of this study. DSOX2 c1 cells were able to activate the transgene in hPGCLCs, which suggests that the failure of the transgene to activate in the

other lines is not a property of the constructs themselves. While it is possible that the overexpression of KLF2 and NANOG might be detrimental to hPGCLCs, the similar result from the 'EV' suggests that the inserts are being silenced in hPGCLCs in the cell lines. As these constructs are inserted into the genome randomly using the PiggyBac system (Wilson et al., 2007), the location of each of these cassettes is random. As hPGCLCs are induced, the chromatin conformation might alter and cause repression of either the transcription factor cassette or the transactivator, making the hPGCLCs intransient to the dox stimulation.

Chapter 5. Utilising an alternative hPGCLC protocol for transgene activation

5.1 Introduction

The lack of mVenus positive hPGCLC produced by some of the cell lines created complexity in studying the effects of the different transcription factors. As observed with cyrosectioning and staining of the aggregates, hPGCLCs tend to cluster on the inside of the 3D aggregate. Lower dox concentrations also led to low overall activation of the transgene in the aggregate. I theorised that there might be a penetration issue, where the internal hPGCLCs are less likely to be exposed to the dox in the culture media if it does not diffuse properly into the aggregate.

An alternative system, which grows the cells in a 2D layer, was recently published (Overeem et al., 2023). Instead of a pre-induction step (iMeLCs), hiPSCs are plated in a Geltrex coated well in the primed media with both ROCKi, Y-27632, and 2% Geltrex. After 24 hours, media is changed to the 2D system (Chapter 2 Materials and Method) and further Geltrex is applied over the next two days (Figure 35 A). The concentration of BMP4 is 20 times less than in the 3D system. Despite these differences, large numbers of hPGCLCs and conversion percentages were reported in this system, although there is still variability between cell lines. I hypothesised that as this system was a 2D system where hPGCLCs are produced at the surface of the culture, there would be fewer issues with dox diffusing in the culture to the hPGCLCs.

The 2D system should allow me to probe whether the low percentage of mVenus positive hPGCLCs in the *KLF2, NANOG* and 'EV' cell lines was due to penetration of the dox into the 3D aggregate. It should also provide a different system to study the effects of SOX2 activation in the germline.

5.2 Specific aims

- Compare 2D and 3D hPGCLC systems.
- Attempt to activate transgenes in 2D hPGCLCs from the hiPSC lines which did not express the transgenes in the 3D system.
- Understand if the failure to activate these transgenes is due to dox diffusing into the 3D aggregates

5.3 Results

5.3.1 Dox additions to the 2D hPGCLC system

The 2D protocol (Overeem et al., 2023) has different media and cytokine concentration as well as timings of plating and inducing hPGCLCs. After plating hiPSCs in E8 + Geltrex (+ROCKi) for 24 hours, the hPGCLC induction media +Geltrex is added. hPGCLC media is changed every day, with Geltrex being included on the 1st and 2nd days. Both 3D and 2D systems involve culturing in hPGCLC media for 4 days, but the 3D system requires a pre-induction step, while the 2D system appears to 'prime' hiPSCs with Geltrex in hiPSC media, before chaining into the 2D culture, which includes Geltrex for the first 2 days of hPGCLC culture. To match the dox addition timing to be comparable to the 3D system, dox was added 1, 2 or 3 days before sorting, resulting in dox being added at day 2 (d2), day 3 (d3) or day 4 (d4) following hPGCLC induction (Figure 35 A).

In order to sort hPGCLCs from the non-reporter cell lines, the distribution of *BLIMP1*-tdTomato *AP2*-eGFP cells was used to set the correct gate on the two surface markers, EPCAM and INTEGRINα6. While the pattern of the contour plots was slightly different in the 2D system, there was still a double positive population which could be gated upon. When the double positive population and the negative population are plotted in respect to the internal

germline markers, 95% of *BLIMP1/AP2* double positive population correlated with EPCAM/INTEGRINα6 double positive population (Figure 35 B).

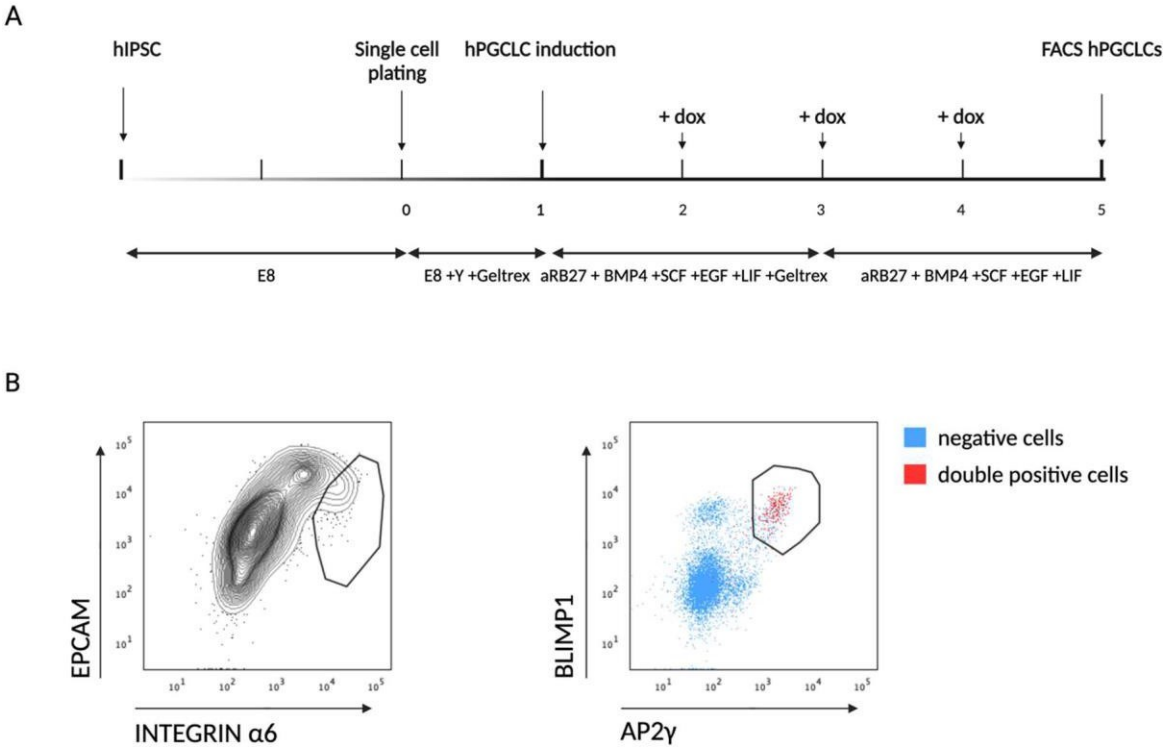

Figure 35 2D hPGCLC system protocol and FACS strategy (A) timeline of 2D protocol with days when dox was added to activate the different transgenes. (B) Gating strategy on surface markers stained BTAG cells on the left and where these cells are placed on a FACS plot based on the internal reporter genes.

5.3.2 SOX2 can be overexpressed in the 2D hPGCLCs

DSOX2 c1 cell line, which overexpresses SOX2 in response to dox, was used in the 2D hPGCLC protocol. As this cell line was able to respond to dox treatment in hPGCLCs, which was not the case for the other cell lines, I wanted to test if these cells retained this ability in the 2D protocol. In a similar manner to the 3D experiments, DSOX2 cells were used to generate hPGCLCs in this 2D system, and 20μM dox added at the days 2, 3 and 4 as indicated (Figure 35 A).

INTEGRINα6 staining was generally higher and EPCAM staining lower in the 2Ds hPGCLCs compared to the 3D hPGCLCs. The 3D NANOG -BMP experiments suggested that a clear separation of the INTEGRINα6 was the most important signifier of hPGCLC identity, (chapter 4, Figure 17). Therefore, the gate to select hPGCLCs in these experiments was set on the population that appeared to be INTEGRINα6 high and EPCAM positive rather than EPCAM high (Figure 36).

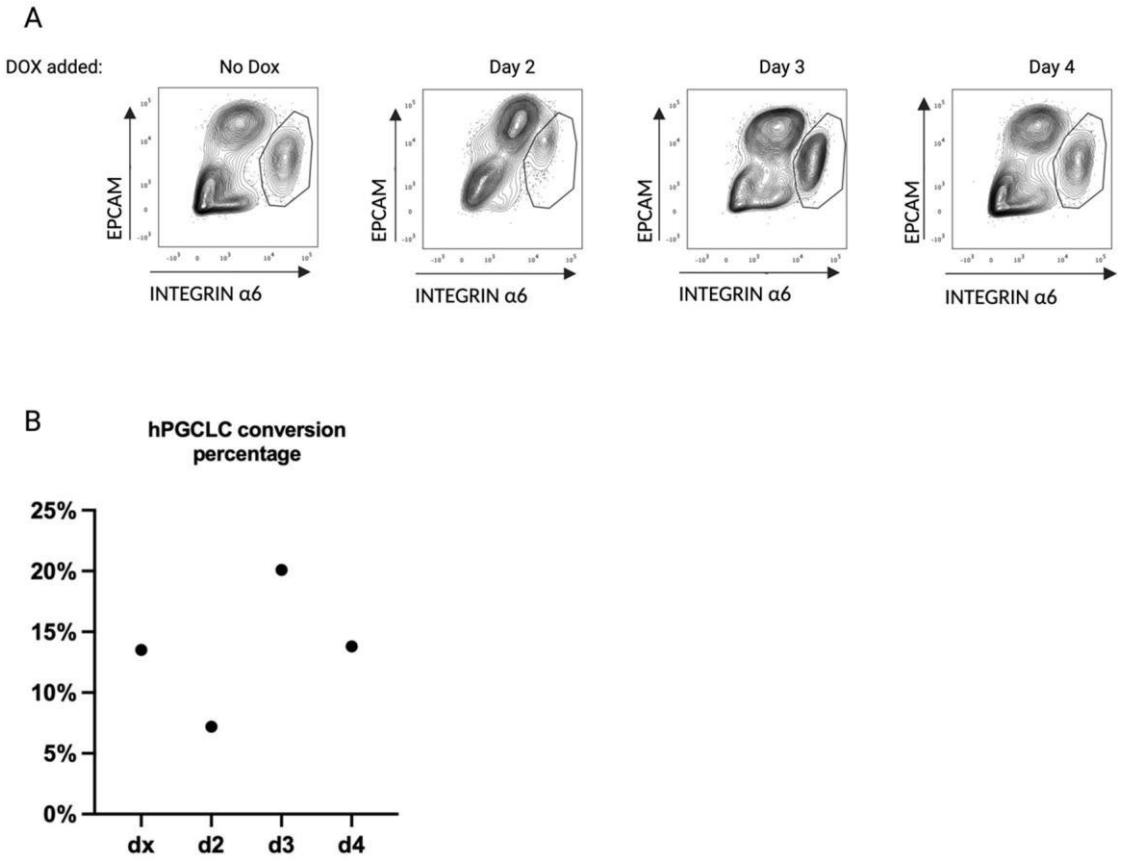

Figure 36 Effect on conversion to hPGCLC in DSOX2 cells using the 2D protocol when dox is added on different days (A) FACS plots of the experiment showing the distribution of surface markers on the cells and the gate used to quantify hPGCLC population (B) quantified percentages of cells within hPGCLC gate in the different samples.

There appeared to be little difference when dox was added at day 3 and day 4 compared to the no dox control. All three had a separate INTEGRINα6 high /EPCAM positive population (the hPGCLCs), an EPCAM positive population and an INTEGRINα6/EPCAM negative population. There was no observable change in the intensities of the surface markers in d3 and d4, compared to the dx conditions (Figure 36 A). In comparison, when dox was added at day 2, there did seem a change in the distribution of the cells. Mainly, the 'hPGCLC' population was less defined and only just separating from the EPCAM positive population (Figure 36 A). This could suggest that the addition of dox at this early stage could have impact on hPGCLC induction, and this is supported by the conversion percentage being markedly reduced, only 7% in the d2 sample compared to 13% in dx, 20% in d3 and 13% in d4. Overinterpretation should be avoided however as this experiment only has a n of 1 and FACS data is not always conclusive.

The number of cells which activated the transgene in the 2D system was higher compared to the 3D system (Figure 37 A). Although direct comparisons are not able to be made as the timings of induction are different, when added the day before sorting the percentage of mVenus positive cells in the 2D was 30% compared to 9% in the 3D system. A similar pattern to the 3D protocol was observed when dox was added on later days, fewer cells turned on mVenus (Figure 37 B). However, the ratio of mVenus positive hPGCLCs compared to all cells was similar at around 30%, regardless of the which day the dox was added. As in the 3D system, SOX2 from the transgene can be expressed in the hPGCLCs generated from the 2D protocol. (Figure 37 C).

Despite the overexpression of the exogenous SOX2, the tdT reporter did not show activation in any of the conditions. This shows the endogenous *SOX2* is still repressed in the 2D hPGCLCs even in the presence exogenous SOX2 (Figure 37 A), a result also observed in the 3D protocol.

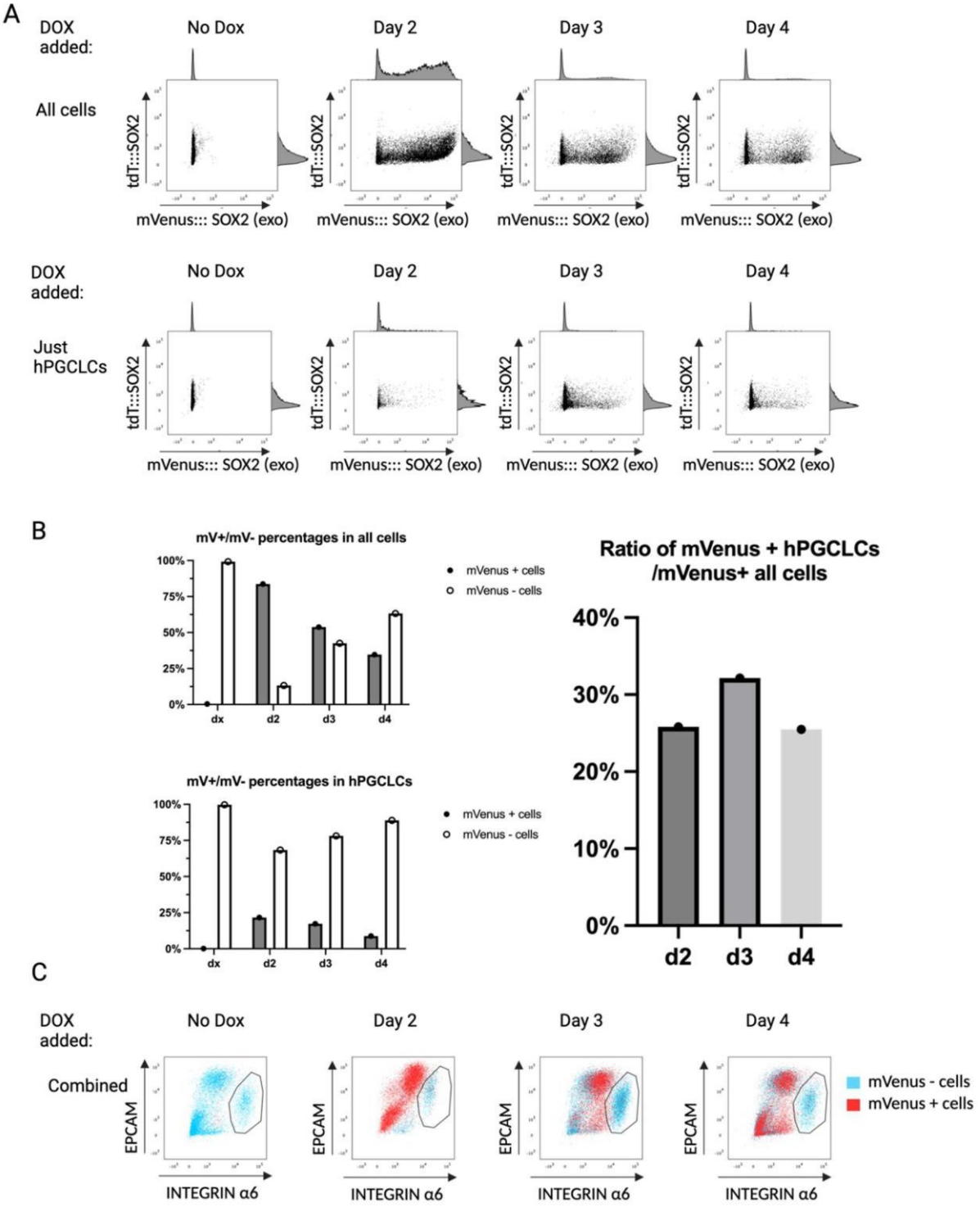

Figure 37 Analysis of mVenus and tdT within DSOX2 hPGCLC 2D cultures by flow analysis (A) dot plots showing mVenus vs. tdT Top: all cells of the aggregate. Bottom: hPGCLCs in the aggregate (B) quantification of mVenus positive and mVenus negative percentages Top: all cells. Bottom: hPGCLCs Right: percentage of mVenus positive hPGCLCs divided by percentage of all cells in the aggregate that are mVenus positive (C) FACS plots showing distribution of mVenus positive and mVenus negative cells based on EPCAM and INTEGRINα6 staining.

The 2D hPGCLC protocol was performed on glass slides to allow immunofluorescence to be conducted (Figure 38). There were issues with staining, perhaps due to the large amount of Geltrex that is added to the media during the protocol, leaving only a few samples and fields possible to be analysed. The no dox and dox added at day 2 were able to be imaged after being stained for mVenus, SOX2 and AP2γ. The transgene was clearly activated when dox was added at day 2, as evidenced by the mVenus and SOX2 are expressed. AP2γ is also highly expressed in a number of cells and co-expressed with SOX2, again supporting the fact the SOX2 transgene can be expressed in hPGCLCs.

Overall, the 2D protocol gives similar results to the 3D protocol. While it does appear the 2D protocol does allow more cells within the culture to activate the transgene; the ratio of mVenus positive hPGCLCs to all mVenus positive cells is slightly reduced in comparison to the 3D.

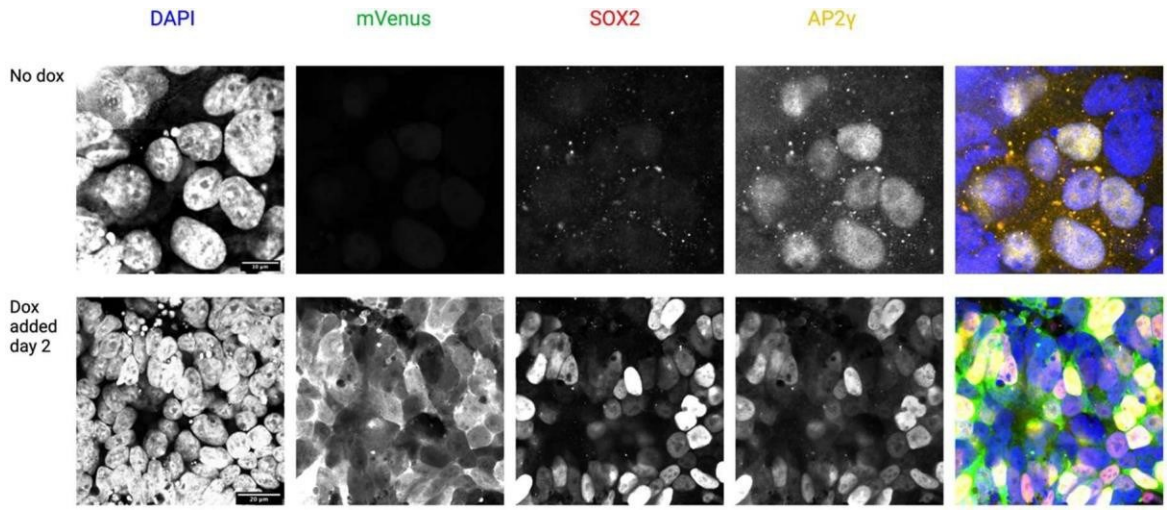

Figure 38 Staining of 2D hPGCLC culture with mVenus, SOX2 and AP2γ top: no dox was added, scale bar = 10mM bottom: dox added at day 2, scale bar = 20mM.

5.3.3 mVenus could be activated in the NANOG OE hPGCLC generated with the 2D protocol, but only when added at day 2 of the protocol

As there were more mVenus positive cells in the 2D protocol, it seemed possible that this protocol would produce more mVenus positive hPGCLCs in the cell lines that did not activate the transgene in the 3D hPGCLCs protocol. The NANOG OE cell line, which overexpresses NANOG in response to dox, was used in the protocol described above (Figure 35 A). The populations on the FACS plots when dox was added were all similar to the no dox control, suggesting there was little effect on the conversion (Figure 39 A). The conversion percentages were also similar, ranging between 10%-14% (Figure 39 B).

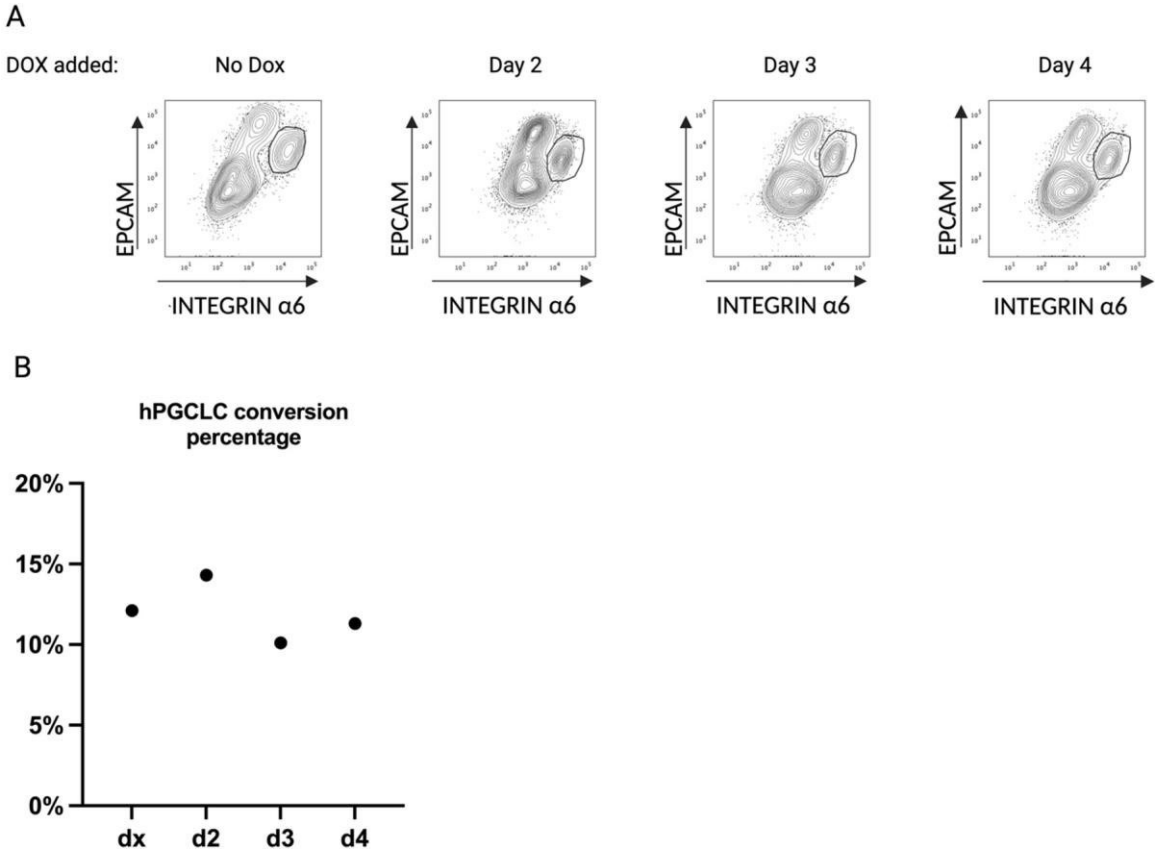

Figure 39 Effect on conversion to hPGCLC in NANOG OE cells using the 2D protocol when dox is added at different days (A) FACS plots of the experiment showing the distribution of surface markers on the cells and the gate used to quantify hPGCLC population (B) quantified percentages of cells within hPGCLC gate in the different samples.

As with the DSOX2 cell line, mVenus activation was higher in the 2D system regardless of which day the dox was added (Figure 40 A). The percentage of mVenus positive NANOG hPGCLCs was one tenth of the percentage of mVenus positive DSOX2 hPGCLCs in the d3 and d4 samples (Figure 40 B). The ratio of mVenus positive hPGCLCs to all mVenus positive cells was also under 5% in d3 and d4 samples (Figure 40 B). At these later times it appears that, as in the 3D system, the NANOG cell line is unable to generate mVenus positive hPGCLCs (Figure 40 C).

There does appear to be a difference when dox is added at day 2, compared to day 3 or day 4 where mVenus is able to be activated in around 10% of hPGCLCs (Figure 40 B), compared to 20% at this stage in DSOX2 cell line. The ratio of mVenus positive hPGCLCs to all mVenus positive cells in this sample was also 3 times higher than the d3 or d4 samples at 15%. This suggests when dox was added at an earlier time point the hPGCLCs were able to activate the transgene better than later time points, but not as well as in the DSOX2 cells (Figure 40 B).

Overall, this data suggests that the similarity in the conversion percentages in the d3 and d4 compared to the dx is due to the transgene not being activated in enough hPGCLCs to cause a biological effect. There are more mVenus positive hPGCLCs in the d2 sample, but this does not appear to affect the conversion percentage. Either the percentage of cells that had activated the transgene was not high enough to cause an effect on conversion, or there is not enough effect on hPGCLC conversion when NANOG is overexpressed in 10% of cells to be detected by flow cytometry analysis.

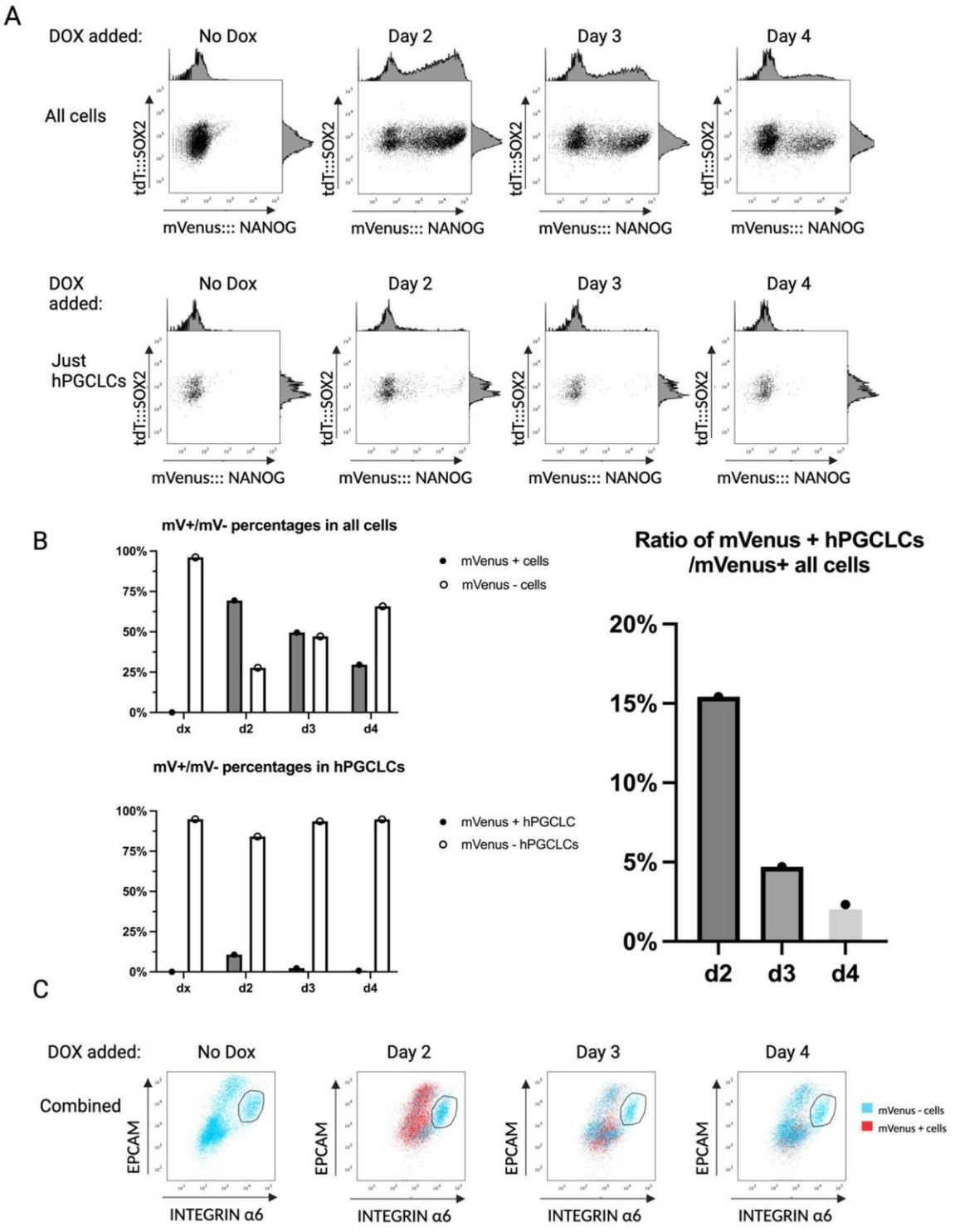

Figure 40 Analysis of mVenus and tdT within NANOG hPGCLC 2D protocol by flow analysis (A) dot plots showing mVenus vs. tdT Top: all cells of the aggregate. Bottom: hPGCLCs in the aggregate (B) quantification of mVenus positive and mVenus negative percentages Top: all cells. Bottom: hPGCLCs Right: percentage of mVenus positive hPGCLCs divided by percentage of all cells in the aggregate that are mVenus positive (C) FACS plots showing distribution of mVenus positive and mVenus negative cells based on EPCAM and INTEGRINa6 staining.

5.3.4 The lack of mVenus positive EV hPGCLC in the d3 and d4 samples suggests the transgenes are not expressed in differentiated hPGCLCs cells.

To confirm that the lack of mVenus positive cells in the NANOG 2D culture was due to a similar reason as in the 3D conditions, i.e. the transgene itself being shut down during hPGCLC differentiation independently of the transcription factor being expressed, the empty vector cell line was used. There was little difference between the FACS plots from the samples where dox was added at different days, but the no dox control appeared to have higher EPCAM expression (Figure 41 A). The conversion percentage for the no dox control was much higher at 27% compared to around 10% in the dox added conditions (Figure 41 B). This would suggest that the addition of dox to the 'EV' hPGCLC culture might inhibit hPGCLC conversion.

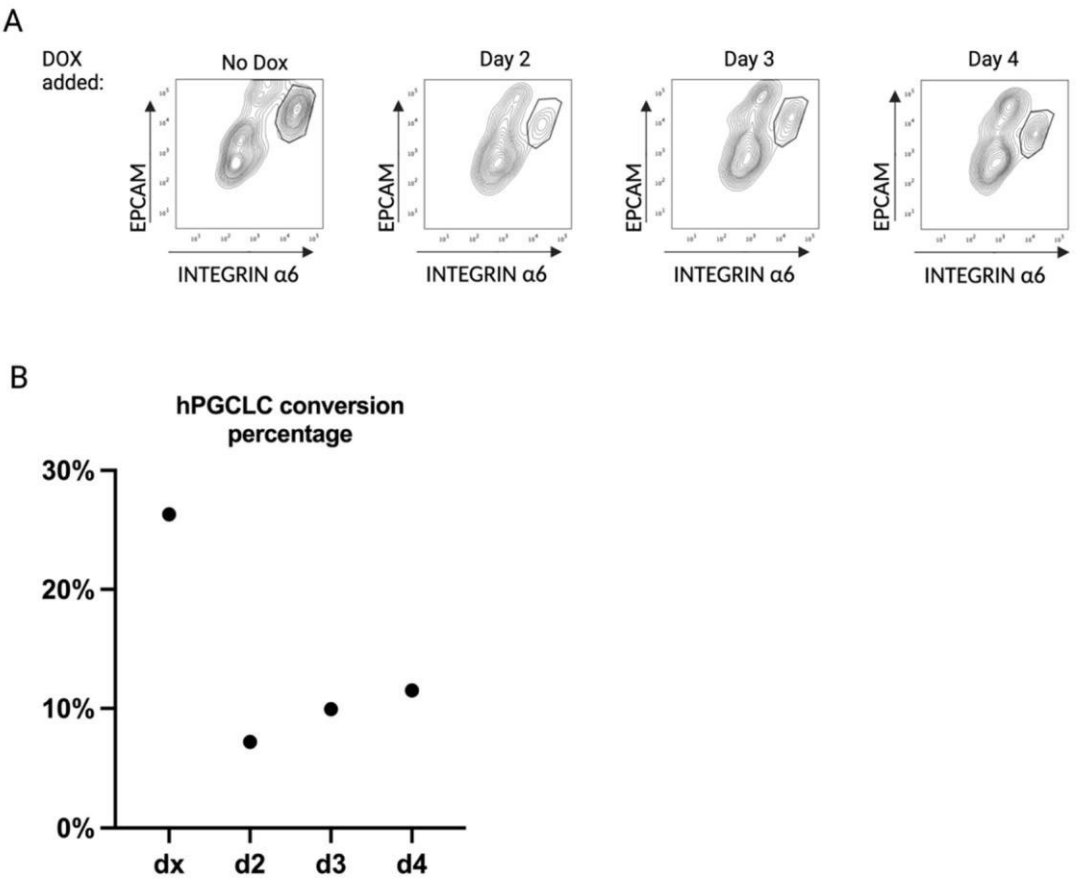

Figure 41 Effect on conversion to hPGCLC in 'EV' cells using the 2D protocol when dox is added at different days (A) FACS plots of the experiment showing the distribution of surface markers on the cells and the gate used to quantify hPGCLC population (B) quantified percentages of cells within hPGCLC gate in the different samples.

The intensity of mVenus was higher in the cells of the EV culture than in TF cell lines, as was the proportion of mVenus positive cells. This proportion was above 75% even when added at day 4 (Figure 42 A & B). This did translate into a 35% of these mVenus positive cells being hPGCLCs in d2 (Figure 42B & C), compared to the DSOX2 or NANOG cell lines. This was reduced to around 10% in d3 and d4 (Figure 42B & C), which is a third of that observed in the DSOX2 cell line (Figure 37 B). The intensity of mVenus also appears to be reduced in d4 compared to d2 or d3 suggesting the transgene was not expressed as highly in this condition (Figure 42 A). The higher levels of mVenus in the EV line compared to the TF lines could explain why adding dox reduced the hPGCLC conversion, as the high levels of mVenus protein

produced could inhibit or cause stress in the hPGCLCs during this protocol, reducing their numbers.

In both the NANOG and EV 2D hPGCLCs systems, mVenus positive hPGCLCs were induced at higher percentages when dox was added at day 2, in comparison to addition at later time points. This is similar to how mVenus positive hPGCLCs could be induced from NANOG 3D aggregates when dox was added at the point of aggregation, (section 4.3.9), but not when dox was added at any later time point. In both the 2D and 3D system, the non-DSOX2 c1 lines form more mVenus positive hPGCLCs when dox is added at earlier time points, when the cells hPGCLCs are less developed and their chromatin conformation is closer to that of the hiPSCs. This would suggest the reason why mVenus positive hPGCLCs cannot be generated when dox is added at these later times is because the more differentiated cells cannot respond to dox.

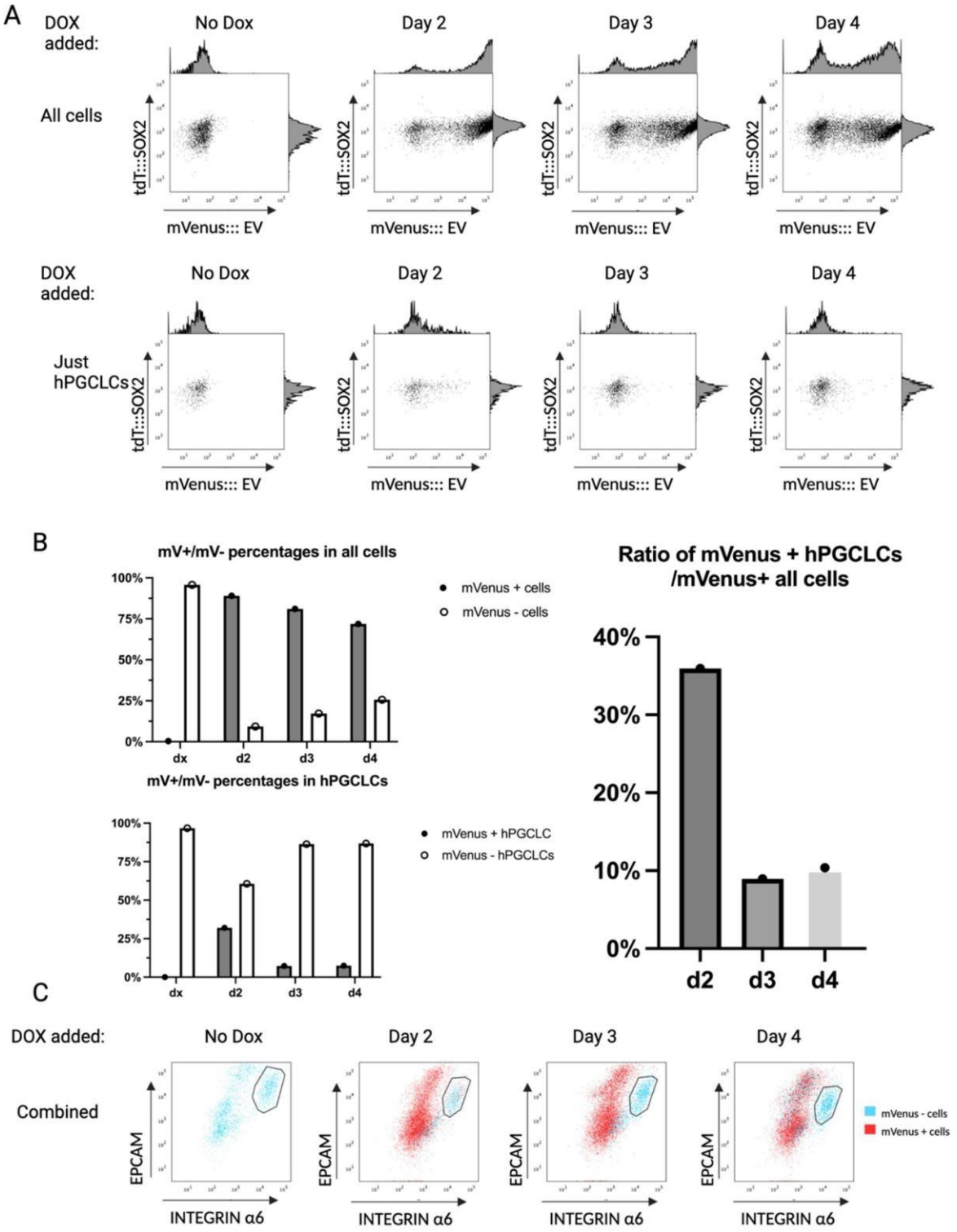

Figure 42 Analysis of mVenus and tdT within EV hPGCLC 2D cultures by flow analysis (A) dot plots showing mVenus vs. tdT Top: all cells of the aggregate. Bottom: hPGCLCs in the aggregate (B) quantification of mVenus positive and mVenus negative percentages Top: all cells. Bottom: hPGCLCs Right: percentage of mVenus positive hPGCLCs divided by percentage of all cells in the aggregate that are mVenus positive (C) FACS plots showing distribution of mVenus positive and mVenus negative cells based on EPCAM and INTEGRINa6 staining.

5.4 Discussion

5.4.1 SOX2 can be overexpressed in the 2D system and appears to have limited effect on hPGCLCs induction

It appeared that while SOX2 could be activated in the hPGCLCs at a reasonable rate, this did not induce the expression of the endogenous SOX2 locus. There is no major difference when dox was added at day 3 or day 4, suggesting the SOX2 overexpression does not affect hPGCLCs at this stage. The reduction in the conversion percentage when dox is added at day 2, along with a slightly altered pattern on the FACS plot, could suggest that at this earlier time point SOX2 overexpression might affect hPGCLC conversion. The reduction in efficiency may be due to a similar result seen in the EV cell line, where the addition of dox reduced the conversion percentage of hPGCLCs accompanied by a high expression of mVenus. The d2 sample has a similarly high mVenus expression to the EV, and the intensity of mVenus was also higher than the d3 or d4 samples. There is a slightly higher expression of EPCAM, a pluripotency related surface marker, in these hPGCLCs suggesting the expression of SOX2 might be maintaining a higher level of pluripotency in this condition. However, the effect is subtle and only based upon a single experiment; further replicates need to be completed to draw more substantial conclusions whether this was due to mVenus toxicity or SOX2 overexpression blocking hPGCLC differentiation.

5.4.2 NANOG cell line do not produce many mVenus positive hPGCLCs

There was a higher proportion of mVenus positive cells in the 2D culture than in the 3D aggregates, suggesting dox was more likely to enter more cells in this culture system. This did appear to increase the proportion of hPGCLCs when dox was added at day 2, but not at day 3 or day 4. This ratio is smaller than in the DSOX2 and the EV cell lines. Either NANOG cells are inherently less likely to turn on the mVenus reporter in the hPGCLCs or the expression of NANOG overexpression may be less compatible with hPGCLCs. The second hypothesis is not supported by the conversion data, as this would suggest a reduction in hPGCLC numbers.

Tracking the fate of mVenus positive cells derived when dox is added at day 2 over the differentiation protocol and comparing this to the EV cell line could help answer these questions.

5.4.3 EV and NANOG lines indicate cells at day 2 of the 2D culture may not have differentiated into germ cells

In the previous chapter, the lack of mVenus positive hPGCLCs from the EV, NANOG and KLF2 cell line was hypothesised to be due to the closing of chromatin during hPGCLCs differentiation. This appears to follow in the 2D protocol with low mVenus positive hPGCLCs in d3 and d4 samples in the NANOG and EV cell lines. However, more mVenus positive hPGCLCs were present in the d2 samples from both these cell lines. At day 2 in the culture, *SOX2* is downregulated while *TFAP2C* and *EOMES* are upregulated, but further genes such as *OCT4* and *SOX17* are not expressed until day 3 (Overeem et al., 2023). This would indicate that before day 2, the cells in the 2D culture are not yet differentiated into hPGCLCs and are able to respond to dox and activate the transgenes. This suggests once a cell commits to becoming a hPGCLC in either protocol, they are not able to respond to dox in these cell lines.

Chapter 6 discussion, conclusions, and future perspectives.

The aim of this study was to use the hPGCLC system to overexpress pluripotency TFs, *SOX2, NANOG* and *KLF2,* in order to see the effect these factors had on human germline fate. This system of generating hPGCLCs while overexpressing *SOX2* was successful in elucidating the role that BMP4 has in repressing the endogenous *SOX2* locus and suggesting it might compete with *SOX17* for *OCT4* during germline induction. However, in the *NANOG* and *KLF2* lines, very few hPGCLCs appeared to turn on the transgene and overexpress these TFs.

6.1 Limitations of this study

A major limitation in this study was the lack of activation of the transgenes in the NANOG c26 or KLF2 c7 cell lines when induced into hPGCLCs. My first hypothesis was that the transcription factor overexpression led to differentiation or death of the hPGCLC that was expressing the transgene. However, when attempting to overexpress a transgene which had no TF and only the mVenus reporter, a similar result was observed, that very few hPGCLC were mVenus positive. To test if this was due to poor penetration of the dox into the 3D aggregate, I repeated the experiments in a 2D system. This 2D system showed a similar result, with very few mVenus positive hPGCLCs being produced in the NANOG c26 and EV c3 cell lines, despite high activation in the non-hPGCLCs from this culture. Overall, this data suggests there was no issue with dox diffusing into the cells and activating the transgene, but the transgenes themselves were unable to respond to the dox activation.

The transgene system involves two cassettes integrated randomly into the genome, as shown in section 3.3.3. The lack of mVenus positive hPGCLCs suggest either the CAG promoter, which constitutively produces the rrTAM-2 protein, or the CMV mini promoter, which is activated by dox binding to the rrTAM-2, are specifically repressed in the hPGCLCs. Epigenetically, hPGCLCs actually show slightly lower DNA methylation (Murase et al., 2020; Sasaki et al., 2015) and lower H3K9Me2 levels (Kobayashi et al., 2022; Sasaki et al., 2015) compared to hiPSCs, but there is an increase of H3K27me3 (Kobayashi et al., 2022). The H3K27Me3 mark is associated with the silencing of PRC complex deposited by EZH2, which has been shown to be crucial for protecting the mouse genome during germ cell

development (Huang et al., 2021). Taken together, it is possible that the cassettes have been inserted into loci which are open in hiPSCs but repressed in hPGCLCs due to H3K27Me3 repressing these loci. The cassettes of the DSOX2 c1 line however appear to be in loci which are accessible in hPGCLCs, hence allowing for the expression of the transgenes in this cell type.

6.1.1 Other methods for generating hiPSCs lines that could overexpress *NANOG* and *KLF2*

During the process of generating the overexpressing hiPSCs lines, I generated a number of different clones for each transcription factor. The most straightforward optimisation step to take would be to screen these clones for their ability to generate mVenus hPGCLCs. Clones which successfully upregulated mVenus in the hPGCLCs could then be used to study the effect of TF activation in hPGCLCs. If these events are rare due to widespread repression of chromatin in hPGCLCs, it may be more sensible to direct the integration of the transgenes. An obvious target would be the human ROSA26 locus, which shows resistance to gene silencing in hiPSC differentiation (Irion et al., 2007).

CMV and CAG has been shown to be repressed in hESCs (Xia et al., 2007), although the constructs used in this study were active in hiPSCs during resetting (Takashima et al., 2014). Therefore, it could be possible that the CMV and CAG promoters are silenced in the hPGCLCs similar to CMV are silenced in hESCs, leading to lower activation of the transgene in the hPGCLCs compared to the non-hPGCLCs. This could be an alternative explanation for the silencing or be additive to the locus being repressed by heterochromatin changes in the hPGCLCs. Using an alternative promoter such as EF1α or PGK, which are repressed less than the CMV and CAG (Xia et al., 2007), could lead to better transgene expression in the hPGCLCs.

An experimental system would be to transfect hPGCLCs with the plasmids described in section 3.3.3 and reprogramme these hPGCLCs into hEGCLCs. hEGCLCs can be derived from hPGCLCs by removing feeder-conditioned media from hPGCLCs cultured on feeders/with feeder conditioned media (Kobayashi et al., 2022). Recently, our laboratory has developed a

defined media for the direct conversion of FACS sorted hPGCLCs into hEGCLCs using the same signalling factors, LIF, SCF, RA, FORSKOLIN, bFGF, CHIR, which reprogramme mPGCs to mEGCs, (Leitch et al., 2013b). These hEGCLCs behave as hiPSCs, including having the ability to generate hPGCLCs themselves. The advantage of transfecting hPGCLCs is that the cassettes should enter open areas of the hPGCLC chromatin. When these transfected hEGCLCs are induced into hPGCLCs, the cassettes should then be an accessible area of chromatin and activate in response to dox treatment. However, no study has yet transfected hPGCLCs so it is unclear if this would be possible.

6.2 Tolerance of *SOX2* overexpression in specified hPGCLCs contrasts its reactivation in Tcam-2 cells

SOX factors play a key role in mammalian germline development; *Sox2* is crucial for the mouse germline but is repressed in the majority of non-rodent species (Mitsunaga and Shioda, 2018). Instead, *SOX17* is upregulated and is essential for human germline development (Irie et al., 2015). By forcing the overexpression of *SOX2* in hPGCLCs, it could be expected that this might destabilise the germ cell network. EC expresses *SOX2* and show stem cell characteristics, expressing *DMNT3B/3L* and shuttling PRDM1 and PRMT5 out of the nucleus (Nettersheim et al., 2016b). Seminomas such as Tcam-2 cells express *SOX17* and retain more germline properties (Jostes et al., 2020). Repression of the *SOX2* locus is relieved by blocking BMP signalling in Tcam-2 cell line (Nettersheim et al., 2015) or grafting SEM tumours into the flank of immunocompromised mice (Nettersheim et al., 2011). Both treatments lead to the generation of an EC-like state, suggesting these germline tumours can undergo a transition to a stem cell fate when *SOX2* is expressed.

However, it appears in my system, that not only is germline identity maintained in hPGCLCs that are overexpressing *SOX2*, but *SOX17* expression is maintained. In my experiments, the overexpression of *SOX2* in the hPGCLCs does not lead to endogenous SOX2 upregulation. This suggests that hPGCLCs overexpressing *SOX2* do not enter a stem cell state observed in Tcam-2 cells. Tcam-2 reprogramming is dependent on the expression of the endogenous

SOX2 (Nettersheim et al., 2016a), whereas in my study I overexpressed *SOX2* in hPGCLC. While I hypothesised it would lead to the repression of the endogenous *SOX2* locus being lifted, my data suggests that this does not occur in specified hPGCLCs. An open question remains about whether hPGCLCs would undergo a conversion to a stem cell state within the aggregate if the endogenous SOX2 locus was activated, as this cannot be achieved by exogenous SOX2 overexpression. The activation of the endogenous locus would be indicative of a wider remodelling of the genetic network of the hPGCLCs and require the co-binding of SOX2 to OCT4 and the SOX2 locus, in contrast to my experiments which just provided excess SOX2 protein. This wider resetting may trigger further events which lead to a pluripotent state.

6.2.1 The formation of the *SOX2/OCT4* may only be possible before germline induction.

After one day of aggregation, qPCR data suggest that the two key factors of germline identity, *TFAP2C*, and *SOX17* are fully upregulated (Kojima et al., 2017; Kojima et al., 2021; Sasaki et al., 2015). The experiments presented in chapter 4 show that *SOX2* represses the germline fate when it is overexpressed at the point of hPGCLC induction. My observation was that mVenus is expressed around 12 to 16 hours after dox addition, suggesting from this point the exogenous *SOX2* protein would be present in the nucleus. It is likely therefore when dox is added at day 1 or later, *SOX2* is not overexpressed until after the germline genes are upregulated. *OCT4* is an important binding partner for both *SOX2* and *SOX17* (Jostes et al., 2020), and the binding of one SOX factor could block the binding of the other. When *SOX2* is overexpressed after induction, the co-binding between *SOX17* and *OCT4* may prevent *SOX2* from co-binding with *OCT4,* which would make it unable to regulate stem cell related genes (Tapia et al., 2015). *SOX2* binds a set of motifs in EC without *OCT4*, however these 'SOX family' motifs are bound by *SOX17* in the Tcam-2 cell line (Jostes et al., 2020). Therefore, even if the exogenously provided *SOX2* is able to bind these motifs, the actions of *SOX2* and *SOX17* at these motifs are the same and so would not disrupt the germline network. Alternatively, *SOX2* overexpression might not be able to override the germline network due

to epigenetic or posttranslational modifications which might occur during the early stages of hPGCLC development which lead to this state being locked in.

However, when SOX2 is overexpressed before induction, it may be able to co-bind with OCT4 and block the germline network from being established. This may occur through blocking the co-binding of SOX17 with OCT4, a crucial factor in setting up the germline network (Irie et al., 2015). Alternatively, *SOX2* overexpression before hPGCLC induction might also be establishing its own network of gene expression, antagonising the germline network. The identity of the cell in the d0 condition was not determined in this study, although the higher proportion of EPCAM positive cells suggests that the overexpression at the point of induction could be driving the cells towards a pluripotency-related state (Lu et al., 2010). While this would suggest the cells within the aggregate are *OCT4* positive, this needs to be confirmed through gene expression analysis such as qPCR on collected cells from the d0 aggregates.

Comparing which *SOX* factor is bound to *OCT4* in the d0 mVenus positive cells versus the d1, d2 and d3 mVenus positive hPGCLCs using immunoprecipitation, would answer whether *SOX2* overexpression at hPGCLC induction directly blocks the binding of *SOX17*. Similarly, it would answer whether *SOX17* expression blocks *SOX2* binding to *OCT4* in the d1, d2 and d3 hPGCLCs, preventing this overexpression from destabilising the germline network. The model would predict in the d0 conditions, *SOX2* is pulled down with *OCT4*, while *SOX17* should be pulled down in the d1, d2 and d3 conditions.

To test if competition for *OCT4* is the cause of the phenotype, providing excess *OCT4* through overexpression could allow for both complexes to form. In the d0 conditions, this could allow *SOX17/OCT4* to form, potentially allowing for hPGCLC induction to occur. In the d1, d2 and d3 conditions, this could allow the *SOX2/OCT4* to form, and help understand if hPGCLCs must repress the formation of this complex in order to gain or maintain germline identity. The model in Figure 47 would currently suggest that providing excess *OCT4* would lead to the formation of the SOX2/OCT4 complex and lead to resetting into a stem cell state. However, if this did not happen, it might suggest that the expression of the germline network and potentially epigenetic or post-translational modification within the germline state block the entry in a stem cell state.

6.2.1.1 Enhancing reprogramming of hPGCLCs with further reprogramming factors

I theorise that *OCT4* does co-bind with *SOX2* in the d0 conditions, and this could be checked through co-IP. The lack of reactivation of the endogenous *SOX2* suggests this complex is not enough to transcriptionally induce expression from this repressed locus. *SOX2/OCT4* has been shown to be a super-pioneering factor, capable of recruiting TET enzymes and blocking the function of DNMT1, (the enzyme involved in DNA methylation maintenance). The binding of *SOX2/OCT4* results in a reduction of DNA methylation where the complex binds (Vanzan et al., 2021). However, chromatin accessibility does not appear to increase with the binding of these TFs alone and perhaps the inclusion of a transcriptional activator such as c-Myc (Araki et al., 2011), as used in the generation of iPSCs (Takahashi et al., 2007; Takahashi and Yamanaka, 2006), could lead to activation of the endogenous *SOX2* locus. As hPGCLC express high levels of *KLF4* and *OCT4* (Sasaki et al., 2015), by expressing the final two factors involved in hiPSC generation, *SOX2* and c-Myc (Takahashi et al., 2007), I theorise these should be sufficient to reprogramme hPGCLC into a stem cell state.

6.2.1.2 Transiently removing SOX17 while overexpressing SOX2 to allow endogenous SOX2 transcription

Challenging this model through SOX17 KO would simply lead to failure to generate hPGCLCs (Irie et al., 2015; Sasaki et al., 2015). *SOX17* expression is required for maintenance of hPGCLCs (Irie et al., 2015; Kojima et al., 2017). Transient knock-down of *SOX17* using inducible siRNA (Mello and Conte, 2004) or degrons (Holland et al., 2012) might be tolerated by hPGCLCs better. This knock-down could be induced after specification on day 2 or day 3. Reducing *SOX17* expression in this way could allow for exogenous *SOX2* protein to co-bind with *OCT4*. Could this lead to a reversion of these hPGCLCs to a stem cell state, similar to the hEGCLCs reported and could this occur even without the *SOX2* overexpression, but *SOX17* knock-down? The reports of robust conversion of hPGCLCs to a stem cell state, termed hEGCLCs, involves the reduction of *SOX17* expression and the upregulation of *SOX2* (Kobayashi et al., 2022) .

6.2.2 BMP4 signalling represses *SOX2* in all cellular contexts

If *SOX2* overexpression at the point of germline induction allows for the exogenously expressed *SOX2* protein to co-bind with *OCT4*, it is insufficient to prevent the repression of the endogenous *SOX2* locus. From my observation this may be potentially due to BMP signalling. The lack of *SOX2* endogenous expression in any cell within the aggregates, hPGCLCs and non-hPGCLCs, from any of the conditions that included BMP4 suggest BMP4 stimulation causes *SOX2* repression. This was directly shown in section 4.3.5, as the cells within the -BMP4 conditions maintained *SOX2*-tdT but cells stimulated with BMP4 repressed the reporter gene, regardless of whether the cell enters germline fate or not. The 2D hPGCLC system, which uses a lower concentration of BMP4 (10ng/ml compared to 200ng/ml in 3D) also displayed a slightly higher expression of tdT in both the hPGCLCs and non-hPGCLCs.

TFAP2C, *SOX17* and *EOMES* KO cells all repress *SOX2* when treated with BMP4, even if they cannot form bona fide hPGCLCs (Irie et al., 2015; Kojima et al., 2017). *BLIMP1* KO cells did express *SOX2* at a higher level than their WT counterparts, but not as high in iMeLCs (Sasaki et al., 2015). hPGCLCs generated from iMeLCs with forced expression of *GATA3, SOX17* and *TFAP2C* also showed higher expression of *SOX2* compared to hPGCLCs generated with BMP4 (Kojima et al., 2021). Overall, this would suggest that BMP4 signalling represses *SOX2* in both hPGCLCs and non-hPGCLCs and that full repression of *SOX2* requires BMP4 signalling in hPGCLCs. In mouse, BMP4 signalling is required for the induction of germline fate, and the activation of *SOX2* in mPGC(LC)s (Hayashi et al., 2011; Ohinata et al., 2009), showing a major difference between the species in how germline-competent cells respond to BMP4 signalling.

SOX2 overexpression cannot relieve this repression, even in a condition where I theorised the overexpressed SOX2 can co-bind with *OCT4*. This suggests the repression of the endogenous *SOX2* locus mediated by BMP4 is greater than the ability of *SOX2/OCT4* to activate this locus. While the exact mechanism by which BMP4 represses SOX2 in the human germline is yet to be uncovered, the downstream effectors of BMP4, phosphorated SMAD proteins have been shown to repress expression of SOX2 in mouse ectodermal tissue (Li et al., 2015). It is

possible that a similar mechanism exits in the human germline, resulting in BMP signalling repressing SOX2 in all cells which are hPGCLC competent. It is intriguing that the action of BMP in the human germline, to repress SOX2, is completely opposite to the action of BMP in the mouse germline which triggers expression of SOX2 and suggests a fundamentally different genetic network exists in these two species which leads to different responses to BMP signalling. Alternatively, *SOX2* overexpression might not be able to override the germline network due to epigenetic or posttranslational modifications which might occur during the early stages of hPGCLC development which lead to this state being locked in

6.2.2.1 Understanding the role of BMP mediated repression of SOX2

The reactivation of *SOX2* locus could still be blocked by BMP4 signalling, even with *SOX17* knock-down. In the reports of converting human germ cells into a stem cell like state, either EC-like cells from Tcam-2 (Nettersheim et al., 2015) or hPGCLCs to hEGCLCs (Kobayashi et al., 2022), blocking or removal of BMP appears to be critical. Blocking continued BMP signalling with noggin, along with *SOX17* knock-down and *SOX2* overexpression, would test if both *SOX17* and BMP signalling represses endogenous *SOX2* to the extent it cannot be reactivated by exogenous *SOX2* overexpression.

This system could provide a useful model for unravelling the repression of *SOX2* in response to BMP signalling. Overexpression of *SOX2* is not able to relieve the repression of the endogenous *SOX2* locus which is caused by BMP4 signalling. Activating *SOX2* overexpression earlier in the system, such as in the last 24 hours of iMeLC development, could also be interesting. If overexpressed *SOX2* protein could sit at the locus before BMP4 signalling, it could antagonise the action of BMP signalling, preventing the repression of the endogenous locus.

6.3 Proposed model for the action of *SOX2* overexpression pre and post germline induction in hPGCLCs.

A model for explaining the observations in this study in the d0, d1, d2 and d3 conditions in the DSOX2 c1 hPGCLC aggregates and results from other studies is shown in Figure 47. When

dox is added at d0, at the point of hPGCLC aggregation, the expression of exogenous *SOX2* binds to *OCT4,* outcompeting *SOX17* for *OCT4*, preventing the formation of the complex. The inability for *SOX17/OCT4* to form prevents the hPGCLC induction in a similar way to the *SOX17* KO (Irie et al., 2015; Kojima et al., 2017). If SOX2/*OCT4* complex binds to the endogenous *SOX2* locus, as it does in hESCs (Boyer et al., 2005), the repression of this locus by BMP4 signaling is dominant, preventing reactivation of this locus.

In the d1, d2 or d3 conditions, *SOX17* is upregulated before the exogenous *SOX2* overexpression but after endogenous *SOX2* repression (Chen et al., 2019), allowing the formation of the *SOX17/OCT4* complex. When the overexpression of exogenous *SOX2* occurs after the induction of *SOX17* expression, it cannot outcompete *SOX17* for *OCT4. SOX2* overexpression cannot block the formation of the *SOX17/OCT4* complex*,* which can then specify the hPGCLCs as in shown in 1.2.3.8 This would suggest the co-binding between *OCT4* and a *SOX* factor is strong and cannot be disrupted by the expression of another. The model would currently suggest that providing excess *OCT4* would lead to the formation of the SOX2/OCT4 complex and lead to resetting into a stem cell state. However, if this did not happen, it might suggest that the expression of the germline network and potentially epigenetic or post-translational modification within the germline state block the entry in a stem cell state.

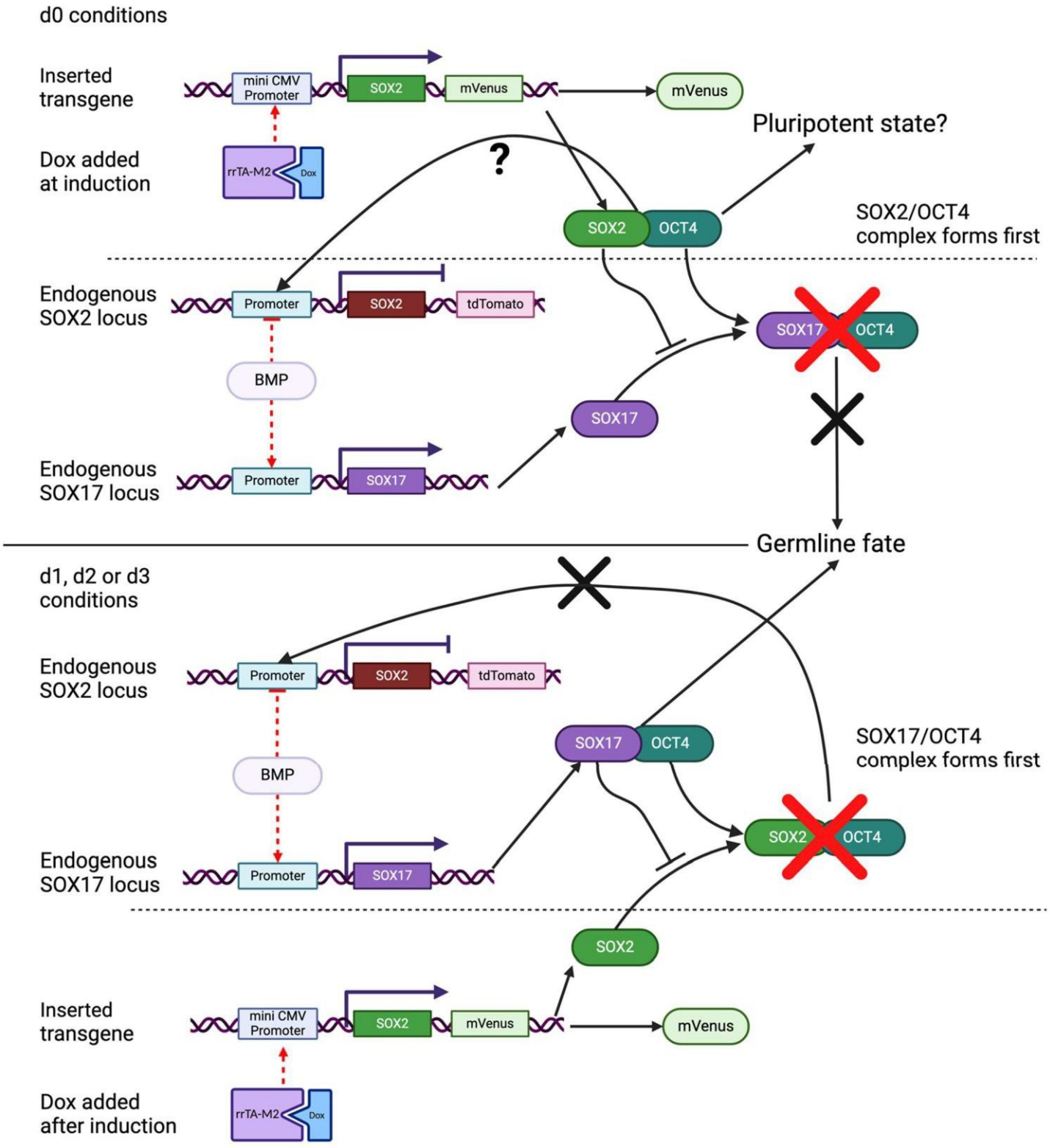

Figure 43 Model for the competition between SOX2 and SOX17 for OCT4. Top panel: d0 conditions where SOX2 is overexpressed before germline induction, resulting in SOX17 not being able to co-bind with OCT4. However, BMP4 still represses endogenous SOX2. Bottom panel: d1, d2 or d3 condition where SOX2 is expressed after germline induction and cannot outcompete SOX17 for OCT4 cobinding. Therefore, the germline network is still induced and not destabilised by SOX2 overexpression

Chapter 7. Bibliography

Acloque, H., Ocana, O.H., Matheu, A., Rizzoti, K., Wise, C., Lovell-Badge, R., and Nieto, M.A. (2011). Reciprocal repression between Sox3 and snail transcription factors defines embryonic territories at gastrulation. Dev Cell 21, 546-558.

Adachi, K., Suemori, H., Yasuda, S.Y., Nakatsuji, N., and Kawase, E. (2010). Role of SOX2 in maintaining pluripotency of human embryonic stem cells. Genes Cells 15, 455-470.

Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G.K., Teo, R., Leng Ng, C.K., Herath, W., Lili, S., et al. (2013). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. EMBO J 32, 938-953.

Alberio, R., Kobayashi, T., and Surani, M.A. (2021). Conserved features of non-primate bilaminar disc embryos and the germline. Stem Cell Reports *16*, 1078-1092

Appleby, J.B., and Bredenoord, A.L. (2018). Should the 14-day rule for embryo research become the 28-day rule? EMBO Mol Med 10.

Araki, R., Hoki, Y., Uda, M., Nakamura, M., Jincho, Y., Tamura, C., Sunayama, M., Ando, S., Sugiura, M., Yoshida, M.A., et al. (2011). Crucial role of c-Myc in the generation of induced pluripotent stem cells. Stem Cells 29, 1362-1370.

Arnold, S.J., Hofmann, U.K., Bikoff, E.K., and Robertson, E.J. (2008). Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. Development 135, 501-511.

Arnold, S.J., Maretto, S., Islam, A., Bikoff, E.K., and Robertson, E.J. (2006). Dose-dependent Smad1, Smad5 and Smad8 signaling in the early mouse embryo. Dev Biol 296, 104-118.

Balboa, D., Weltner, J., Novik, Y., Eurola, S., Wartiovaara, K., and Otonkoski, T. (2017). Generation of a SOX2 reporter human induced pluripotent stem cell line using CRISPR/SaCas9. Stem Cell Res 22, 16-19.

Blakeley, P., Fogarty, N.M., 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.

Boeuf, H., Hauss, C., Graeve, F.D., Baran, N., and Kedinger, C. (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. J Cell Biol 138, 1207-1217.

Bourillot, P.Y., and Savatier, P. (2010). Kruppel-like transcription factors and control of pluripotency. BMC Biol 8, 125.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G.,

Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947-956.
Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature 309, 255-256.
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.
Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J.,Silva, J., McLay, R., Hall, J., Ying, Q.L., and Smith, A. (2008). Capture of authentic embryonic stem cells from rat blastocysts. Cell 135, 1287-1298.
Campolo, F., Gori, M., Favaro, R., Nicolis, S., Pellegrini, M., Botti, F., Rossi, P., Jannini, E.A., and Dolci, S. (2013). Essential role of Sox2 for the establishment and maintenance of the germ cell line. Stem Cells 31, 1408-1421.
Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113, 643-655. 184
Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230-1234.
Chang, H., and Matzuk, M.M. (2001). Smad5 is required for mouse primordial germ cell development. Mech Dev 104, 61-67.
Chen, D., Liu, W., Zimmerman, J., Pastor, W.A., Kim, R., Hosohama, L., Ho, J., Aslanyan, M., Gell, J.J., Jacobsen, S.E., et al. (2018). The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation. Cell Rep 25, 3591-3602 e3595.
Chen, D., Sun, N., Hou, L., Kim, R., Faith, J., Aslanyan, M., Tao, Y., Zheng, Y., Fu, J., Liu, W., et al. (2019). Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors. Cell Rep 29, 4568-4582 e4565.
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. Nat Methods 8, 424-429.
Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional

network in embryonic stem cells. Cell 133, 1106-1117.
Chen, Y.T., Furushima, K., Hou, P.S., Ku, A.T., Deng, J.M., Jang, C.W., Fang, H., Adams, H.P., Kuo, M.L., Ho, H.N., et al. (2010). PiggyBac transposon- mediated, reversible gene transfer in human embryonic stem cells. Stem Cells Dev 19, 763- 771.
Choi, K.H., Lee, D.K., Kim, S.W., Woo, S.H., Kim, D.Y., and Lee, C.K. (2019). Chemically Defined Media Can Maintain Pig Pluripotency Network In Vitro. Stem Cell Reports 13, 221-234.
Clark, A.T., Bodnar, M.S., Fox, M., Rodriquez, R.T., Abeyta, M.J., Firpo, M.T., and Pera, R.A. (2004). Spontaneous differentiation of germ cells from human embryonic stem cells in vitro. Hum Mol Genet 13, 727-739.

Dakhore, S., Nayer, B., and Hasegawa, K. (2018). Human Pluripotent Stem Cell Culture: Current Status, Challenges, and Advancement. Stem Cells Int 2018, 7396905. Daman, K. (2016). States of Pluripotency: Naïve and Primed Pluripotent Stem Cells. In Pluripotent Stem Cells, T. Minoru, ed. (Rijeka: IntechOpen), p. Ch. 3.

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

de Sousa Lopes, S.M., Roelen, B.A., Monteiro, R.M., Emmens, R., Lin, H.Y., Li, E., Lawson, K.A., and Mummery, C.L. (2004). BMP signaling mediated by ALK2 in the visceral endoderm is necessary for the generation of primordial germ cells in the mouse embryo. Genes Dev 18, 1838-1849.

Ding, V.M., Ling, L., Natarajan, S., Yap, M.G., Cool, S.M., and Choo, A.B. (2010). FGF-2 modulates Wnt signaling in undifferentiated hESC and iPS cells through activated PI3- K/GSK3beta signaling. J Cell Physiol 225, 417-428.
Doble, B.W., Patel, S., Wood, G.A., Kockeritz, L.K., and Woodgett, J.R. (2007). Functional redundancy of GSK-3alpha and GSK-3beta in Wnt/beta-catenin signaling shown by using an allelic series of embryonic stem cell lines. Dev Cell 12, 957-971.
Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith,A.G. (2014). Defining an essential transcription factor program for naive pluripotency. Science 344, 1156-1160. 185
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 selfrenewal, survival, and adhesion of human embryonic stem cells. Stem Cells 27, 1847-

1857.
Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154-156.
Festuccia, N., Osorno, R., Halbritter, F., Karwacki- Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell 11, 477-490.
Fogarty, N.M.E., McCarthy, A., Snijders, K.E., Powell, B.E., Kubikova, N., Blakeley, P., Lea, R., Elder, K., Wamaitha, S.E., Kim, D., et al. (2017). Genome editing reveals a role for OCT4 in human embryogenesis. Nature 550, 67-73.
Fong, H., Hohenstein, K.A., and Donovan, P.J. (2008). Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. Stem Cells 26, 1931-1938.
Fukuda, T., Tani, T., Haraguchi, S., Donai, K., Nakajima, N., Uenishi, H., Eitsuka, T., Miyagawa, M., Song, S., Onuma, M., et al. (2017). Expression of Six Proteins Causes Reprogramming of Porcine Fibroblasts Into Induced Pluripotent Stem Cells With Both Active X Chromosomes. Journal of Cellular Biochemistry 118, 537-553.
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.
Gardner, R.L., and Beddington, R.S. (1988). Multi-lineage 'stem' cells in the mammalian embryo. J Cell Sci Suppl 10, 11-27.
Geijsen, N., Horoschak, M., Kim, K., Gribnau, J., Eggan, K., and Daley, G.Q. (2004). Derivation of embryonic germ cells and male gametes from embryonic stem cells. Nature 427, 148-154.
Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S., et al. (2015). Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. Science 347, 1002-1006.
Gkountela, S., Li, Z., Vincent, J.J., Zhang, K.X., Chen, A., Pellegrini, M., and Clark, A.T. (2013). The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. Nat Cell Biol 15, 113-122.
Gkountela, S., Zhang, Kelvin X., Shafiq, Tiasha A., Liao, W.-W., Hargan-Calvopiña, J., Chen, P.-Y., and Clark, Amander T. (2015). DNA Demethylation Dynamics in the Human Prenatal

Germline. Cell 161, 1425-1436. Gudas, L.J., and Wagner, J.A. (2011). Retinoids regulate stem cell differentiation. J Cell Physiol 226, 322-330.
Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. Cell 161, 1437-1452.
Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W., et al. (2017). Epigenetic resetting of human pluripotency. Development 144, 2748-2763.
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.
Haghighi, F., Dahlmann, J., Nakhaei-Rad, S., Lang, A., Kutschka, I., Zenker, M., Kensah, G., Piekorz, R.P., and Ahmadian, M.R. (2018). bFGF-mediated pluripotency maintenance in 186 human induced pluripotent stem cells is associated with NRAS-MAPK signaling. Cell Commun Signal 16, 96.
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.
Hay, D.C., Sutherland, L., Clark, J., and Burdon, T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells 22, 225-235.
Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell 146, 519-532.
Holland, A.J., Fachinetti, D., Han, J.S., and Cleveland, D.W. (2012). Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. Proc Natl Acad Sci U S A 109, E3350-3357.
Hu, H., Ho, D.H.H., Tan, D.S., MacCarthy, C.M., Yu, C.H., Weng, M., Scholer, H.R., and Jauch, R. (2023). Evaluation of the determinants for improved pluripotency induction and maintenance by engineered SOX17. Nucleic Acids Res 51, 8934-8956.
Huang, T.C., Wang, Y.F., Vazquez-Ferrer, E., Theofel, I., Requena, C.E., Hanna, C.W., Kelsey, G., and Hajkova, P. (2021). Sex-specific chromatin remodelling safeguards transcription in germ cells. Nature 600, 737-742.

Huang, Y., Osorno, R., Tsakiridis, A., and Wilson, V. (2012). In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. Cell Rep 2, 1571-1578.

Hubner, K., Fuhrmann, G., Christenson, L.K., Kehler, J., Reinbold, R., De La Fuente, R., Wood, J., Strauss, J.F., 3rd, Boiani, M., and Scholer, H.R. (2003). Derivation of oocytes from mouse embryonic stem cells. Science 300, 1251-1256.

Hwang, Y.S., Suzuki, S., Seita, Y., Ito, J., Sakata, Y., Aso, H., Sato, K., Hermann, B.P., and Sasaki, K. (2020). Reconstitution of prospermatogonial specification in vitro from human induced pluripotent stem cells. Nature Communications 11, 5656.

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

Irie, N., Weinberger, L., Tang, W.W., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. Cell 160, 253-268.

Irion, S., Luche, H., Gadue, P., Fehling, H.J., Kennedy, M., and Keller, G. (2007). Identification and targeting of the ROSA26 locus in human embryonic stem cells. Nat Biotechnol 25, 1477-1482.

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.

Jauch, R., Aksoy, I., Hutchins, A.P., Ng, C.K., Tian, X.F., Chen, J., Palasingam, P., Robson, P., Stanton, L.W., and Kolatkar, P.R. (2011). Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. Stem Cells 29, 940-951.

Jeong, H.C., Park, S.J., Choi, J.J., Go, Y.H., Hong, S.K., Kwon, O.S., Shin, J.G., Kim, R.K., Lee, M.O., Lee, S.J., et al. (2017). PRMT8 Controls the Pluripotency and Mesodermal Fate of 187 Human Embryonic Stem Cells By Enhancing the PI3K/AKT/SOX2 Axis. Stem Cells 35, 2037-2049.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol 10, 353-360.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A

programmable dual- RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.
Jostes, S.V., Fellermeyer, M., Arevalo, L., Merges, G.E., Kristiansen, G., Nettersheim, D., and Schorle, H. (2020). Unique and redundant roles of SOX2 and SOX17 in regulating the germ cell tumor fate. Int J Cancer 146, 1592-1605.
Kawamata, M., and Ochiya, T. (2010). Establishment of embryonic stem cells from rat blastocysts. Methods Mol Biol 597, 169-177.
Kawase, E., Suemori, H., Takahashi, N., Okazaki, K., Hashimoto, K., and Nakatsuji, N. (1994). Strain difference in establishment of mouse embryonic stem (ES) cell lines. Int J Dev Biol 38, 385-390.
Kee, K., Angeles, V.T., Flores, M., Nguyen, H.N., and Reijo Pera, R.A. (2009). Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. Nature 462, 222-225.
Kee, K., Gonsalves, J.M., Clark, A.T., and Pera, R.A. (2006). Bone morphogenetic proteins induce germ cell differentiation from human embryonic stem cells. Stem Cells Dev 15, 831-837.
Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., Lomeli, H., Nagy, A., McLaughlin, K.J., Scholer, H.R., et al. (2004). Oct4 is required for primordial germ cell survival. EMBO Rep 5, 1078-1083.
Kelly, K.F., Ng, D.Y., Jayakumaran, G., Wood, G.A., Koide, H., and Doble, B.W. (2011). betacatenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism. Cell Stem Cell 8, 214-227.
Kinoshita, M., Barber, M., Mansfield, W., Cui, Y., Spindlow, D., Stirparo, G.G., Dietmann, S., Nichols, J., and Smith, A. (2021). Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. Cell Stem Cell 28, 453-471.e458.
Kiyonari, H., Kaneko, M., Abe, S., and Aizawa, S. (2010). Three inhibitors of FGF receptor, ERK, and GSK3 establishes germline-competent embryonic stem cells of C57BL/6N mouse strain with high efficiency and stability. Genesis 48, 317-327.
Kobayashi, M., Kobayashi, M., Odajima, J., Shioda, K., Hwang, Y.S., Sasaki, K., Chatterjee, P., Kramme, C., Kohman, R.E., Church, G.M., et al. (2022). Expanding homogeneous culture of human primordial germ cell-like cells maintaining germline features without serum or feeder layers. Stem Cell Reports 17, 507-521.
Kobayashi, T., Zhang, H., Tang, W.W.C., Irie, N., Withey, S., Klisch, D., Sybirna, A., Dietmann,

S., Contreras, D.A., Webb, R., et al. (2017). Principles of early human development and germ cell program from conserved model systems. Nature 546, 416-420.
Kojima, Y., Sasaki, K., Yokobayashi, S., Sakai, Y., Nakamura, T., Yabuta, Y., Nakaki, F., Nagaoka, S., Woltjen, K., Hotta, A., et al. (2017). Evolutionarily Distinctive Transcriptional 188 and Signaling Programs Drive Human Germ Cell Lineage Specification from Pluripotent Stem Cells. Cell Stem Cell 21, 517-532 e515.
Kojima, Y., Yamashiro, C., Murase, Y., Yabuta, Y., Okamoto, I., Iwatani, C., Tsuchiya, H., Nakaya, M., Tsukiyama, T., Nakamura, T., et al. (2021). GATA transcription factors, SOX17 and TFAP2C, drive the human germ-cell specification program. Life Sci Alliance 4.
Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells 26, 903-911.
Kunath, T., Arnaud, D., Uy, G.D., Okamoto, I., Chureau, C., Yamanaka, Y., Heard, E., Gardner, R.L., Avner, P., and Rossant, J. (2005). Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. Development 132, 1649-1661.
Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development 134, 2895- 2902.
Kuo, C.T., Veselits, M.L., Barton, K.P., Lu, M.M., Clendenin, C., and Leiden, J.M. (1997). The LKLF transcription factor is required for normal tunica media formation and blood vessel stabilization during murine embryogenesis. Genes Dev 11, 2996-3006.
Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev 22, 1617-1635.
Kushwaha, R., Jagadish, N., Kustagi, M., Mendiratta, G., Seandel, M., Soni, R., Korkola, J.E., Thodima, V., Califano, A., Bosl, G.J., et al. (2016). Mechanism and Role of SOX2 Repression in Seminoma: Relevance to Human Germline Specification. Stem Cell Reports 6, 772-783.
Lawson, K.A., Dunn, N.R., Roelen, B.A., Zeinstra, L.M., Davis, A.M., Wright, C.V., Korving, J.P., and Hogan, B.L. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev 13, 424-436.
Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013a). Naive pluripotency is associated with global DNA hypomethylation. Nature structural & molecular biology 20, 311-316.
Leitch, H.G., Nichols, J., Humphreys, P., Mulas, C., Martello, G., Lee, C., Jones, K., Surani,

M.A., and Smith, A. (2013b). Rebuilding pluripotency from primordial germ cells. Stem Cell Reports 1, 66-78.
Leitch, H.G., and Smith, A. (2013). The mammalian germline as a pluripotency cycle. Development 140, 2495-2501.
Leitch, H.G., Tang, W.W., and Surani, M.A. (2013c). Primordial germ-cell development and epigenetic reprogramming in mammals. Curr Top Dev Biol 104, 149-187.
Leitch, H.G., Tang, W.W.C., and Surani, M.A. (2013d). Chapter Five - Primordial Germ-Cell Development and Epigenetic Reprogramming in Mammals. In Current Topics in Developmental Biology, E. Heard, ed. (Academic Press), pp. 149-187.
Li, J., Feng, J., Liu, Y., Ho, T.V., Grimes, W., Ho, H.A., Park, S., Wang, S., and Chai, Y. (2015). BMP-SHH signaling network controls epithelial stem cell fate via regulation of its niche in the developing tooth. Dev Cell 33, 125-135.
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 selfrenewal. Differentiation 75, 299-307.
Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.L., et al. (2008). Germline competent embryonic stem cells derived from rat blastocysts. Cell 135, 1299-1310. 189
Li, Z., Fang, F., Long, Y., Zhao, Q., Wang, X., Ye, Z., Meng, T., Gu, X., Xiang, W., Xiong, C., et al. (2022). The balance between NANOG and SOX17 mediated by TET proteins regulates specification of human primordial germ cell fate. Cell Biosci 12, 181.
Liao, J., Cui, C., Chen, S., Ren, J., Chen, J., Gao, Y., Li, H., Jia, N., Cheng, L., Xiao, H., et al. (2009). Generation of induced pluripotent stem cell lines from adult rat cells. Cell Stem Cell 4, 11-15.
Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38, 431-440.
Lu, T.Y., Lu, R.M., Liao, M.Y., Yu, J., Chung, C.H., Kao, C.F., and Wu, H.C. (2010). Epithelial cell adhesion molecule regulation is associated with the maintenance of the undifferentiated phenotype of human embryonic stem cells. J Biol Chem 285, 8719-8732.
Magnusdottir, E., Dietmann, S., Murakami, K., Gunesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Azim Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat Cell Biol 15, 905-915.
Marshall, E. (1998). A versatile cell line raises scientific hopes, legal questions. Science 282,

1014-1015.
Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J 32, 2561-2574.
Martello, G., and Smith, A. (2014). The nature of embryonic stem cells. Annu Rev Cell Dev Biol 30, 647-675.
Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol 9, 625-635.
Matsui, Y., Zsebo, K., and Hogan, B.L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell 70, 841-847.
Mello, C.C., and Conte, D., Jr. (2004). Revealing the world of RNA interference. Nature 431, 338-342.
Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113, 631-642.
Mitsunaga, S., and Shioda, T. (2018). Evolutionarily diverse mechanisms of germline specification among mammals: what about us? Stem Cell Investig 5, 12.
Miyanari, Y., and Torres-Padilla, M.E. (2012). Control of ground-state pluripotency by allelic regulation of Nanog. Nature 483, 470-473.
Morgani, S., Nichols, J., and Hadjantonakis, A.K. (2017). The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. BMC Dev Biol 17, 7.
Mossahebi-Mohammadi, M., Quan, M., Zhang, J.S., and Li, X. (2020). FGF Signaling Pathway: A Key Regulator of Stem Cell Pluripotency. Front Cell Dev Biol 8, 79.
Murakami, K., Gunesdogan, U., Zylicz, J.J., Tang, W.W.C., Sengupta, R., Kobayashi, T., Kim, S., Butler, R., Dietmann, S., and Surani, M.A. (2016). NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. Nature 529, 403-407.
Murase, Y., Yabuta, Y., Ohta, H., Yamashiro, C., Nakamura, T., Yamamoto, T., and Saitou, M. (2020). Long-term expansion with germline potential of human primordial germ cell-like cells in vitro. The EMBO Journal 39, e104929. 190
Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol 26, 101-106.
Nettersheim, D., Gillis, A.J., Looijenga, L.H., and Schorle, H. (2011). TGF-beta1, EGF and FGF4

synergistically induce differentiation of the seminoma cell line TCam-2 into a cell type resembling mixed non- seminoma. Int J Androl 34, e189-203.
Nettersheim, D., Heimsoeth, A., Jostes, S., Schneider, S., Fellermeyer, M., Hofmann, A., and Schorle, H. (2016a). SOX2 is essential for in vivo reprogramming of seminoma-like TCam-2 cells to an embryonal carcinoma-like fate. Oncotarget 7, 47095-47110.
Nettersheim, D., Jostes, S., Schneider, S., and Schorle, H. (2016b). Elucidating human male germ cell development by studying germ cell cancer. Reproduction 152, R101-113.
Nettersheim, D., Jostes, S., Sharma, R., Schneider, S., Hofmann, A., Ferreira, H.J., Hoffmann, P., Kristiansen, G., Esteller, M.B., and Schorle, H. (2015). BMP Inhibition in Seminomas Initiates Acquisition of Pluripotency via NODAL Signaling Resulting in Reprogramming to an Embryonal Carcinoma. PLoS Genet 11, e1005415.
Nichols, J., Chambers, I., Taga, T., and Smith, A. (2001). Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development 128, 2333-2339.
Nichols, J., Jones, K., Phillips, J.M., Newland, S.A., Roode, M., Mansfield, W., Smith, A., and Cooke, A. (2009). Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. Nat Med 15, 814-818.
Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell 4, 487-492.
Nichols, J., and Smith, A. (2012). Pluripotency in the embryo and in culture. Cold Spring Harb Perspect Biol 4, a008128.
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.
Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev 12, 2048-2060.
Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24, 372-376.
Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature 460, 118-122.
Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell 123, 917-929.

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. Cell 137, 571-584.
Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature 436, 207-213.
Onozato, D., Yamashita, M., Fukuyama, R., Akagawa, T., Kida, Y., Koeda, A., Hashita, T., Iwao, T., and Matsunaga, T. (2018). Efficient Generation of Cynomolgus Monkey Induced Pluripotent Stem Cell-Derived Intestinal Organoids with Pharmacokinetic Functions. Stem Cells Dev 27, 1033-1045. 191
Overeem, A.W., Chang, Y.W., Moustakas, I., Roelse, C.M., Hillenius, S., Helm, T.V., Schrier, V.F.V., Goncalves, M., Mei, H., Freund, C., et al. (2023). Efficient and scalable generation of primordial germ cells in 2D culture using basement membrane extract overlay. Cell Rep Methods 3, 100488.
Palasingam, P., Jauch, R., Ng, C.K., and Kolatkar, P.R. (2009). The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. J Mol Biol 388, 619-630.
Pan, G., and Thomson, J.A. (2007). Nanog and transcriptional networks in embryonic stem cell pluripotency. Cell Res 17, 42-49.
Perrett, R.M., Turnpenny, L., Eckert, J.J., O'Shea, M., Sonne, S.B., Cameron, I.T., Wilson, D.I., Rajpert- De Meyts, E., and Hanley, N.A. (2008). The early human germ cell lineage does not express SOX2 during in vivo development or upon in vitro culture. Biol Reprod 78, 852-858.
Pierson Smela, M., Sybirna, A., Wong, F.C.K., and Surani, M.A. (2019). Testing the role of SOX15 in human primordial germ cell fate. Wellcome Open Res 4, 122.
Popovic, M., Bialecka, M., Gomes Fernandes, M., Taelman, J., Van Der Jeught, M., De Sutter, P., Heindryckx, B., and Chuva De Sousa Lopes, S.M. (2019). Human blastocyst outgrowths recapitulate primordial germ cell specification events. Mol Hum Reprod 25, 519-526.
Qiu, D., Ye, S., Ruiz, B., Zhou, X.,Liu, D., Zhang, Q., and Ying, Q.L. (2015). Klf2 and Tfcp2l1, Two Wnt/beta-Catenin Targets, Act Synergistically to Induce and Maintain Naive Pluripotency. Stem Cell Reports 5, 314-322.
Ralston, A., and Rossant, J. (2005). Genetic regulation of stem cell origins in the mouse embryo. Clin Genet 68, 106-112.
Reik, W., and Surani, M.A. (2015). Germline and Pluripotent Stem Cells. Cold Spring Harbor perspectives in biology 7.

Resnick, J.L., Bixler, L.S., Cheng, L., and Donovan, P.J. (1992). Long-term proliferation of mouse primordial germ cells in culture. Nature 359, 550-551.
Reubinoff, B.E., Pera, M.F., Fong, C.-Y., Trounson, A., and Bongso, A. (2000a). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nature Biotechnology 18, 399-404.
Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., and Bongso, A. (2000b). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nat Biotechnol 18, 399-404.
Rodriguez, R.T., Velkey, J.M., Lutzko, C., Seerke, R., Kohn, D.B., O'Shea, K.S., and Firpo, M.T. (2007). Manipulation of OCT4 levels in human embryonic stem cells results in induction of differential cell types. Exp Biol Med (Maywood) 232, 1368-1380.
Rossant, J. (1987). Development of extraembryonic cell lineages in the mouse embryo. In Experimental Approaches to Mammalian Embryonic Development, R.A. Pedersen, and J. Rossant, eds. (Cambridge: Cambridge University Press), pp. 97-120.
Rossant, J. (2007). Stem cells and lineage development in the mammalian blastocyst. Reprod Fertil Dev 19, 111-118.
Saitou, M., and Yamaji, M. (2012). Primordial germ cells in mice. Cold Spring Harb Perspect Biol 4, a008375.
Sakai, Y., Nakamura, T., Okamoto, I., Gyobu-Motani, S., Ohta, H., Yabuta, Y., Tsukiyama, T., Iwatani, C., Tsuchiya, H., Ema, M., et al. (2020). Induction of the germ cell fate from pluripotent stem cells in cynomolgus monkeys. Biol Reprod 102, 620-638. 192
Sanalkumar, R., Johnson, K.D., Gao, X., Boyer, M.E., Chang, Y.I., Hewitt, K.J., Zhang, J., and Bresnick, E.H. (2014). Mechanism governing a stem cell-generating cis- regulatory element. Proc Natl Acad Sci U S A 111, E1091-1100.
Santagata, S., Ligon, K.L., and Hornick, J.L. (2007). Embryonic stem cell transcription factor signatures in the diagnosis of primary and metastatic germ cell tumors. Am J Surg Pathol 31, 836-845.
Sarkar, A., and Hochedlinger, K. (2013). The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell 12, 15-30.
Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Shiraki, N., Takakuwa, T., et al. (2016). The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. Dev Cell 39, 169-185.
Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H.,

Moritoki, Y., Iwatani, C., Tsuchiya, H., et al. (2015). Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. Cell Stem Cell 17, 178-194.
Sekita, Y., Nakamura, T., and Kimura, T. (2016). Reprogramming of germ cells into pluripotency. World J Stem Cells 8, 251-259.
Shamblott, M.J., Axelman, J., Wang, S., Bugg, E.M., Littlefield, J.W., Donovan, P.J., Blumenthal, P.D., Huggins, G.R., and Gearhart, J.D. (1998). Derivation of pluripotent stem cells from cultured human primordial germ cells. Proc Natl Acad Sci U S A 95, 13726-13731.
Shen, M.M., and Leder, P. (1992). Leukemia inhibitory factor is expressed by the preimplantation uterus and selectively blocks primitive ectoderm formation in vitro. Proc Natl Acad Sci U S A 89, 8240-8244.
Shovlin, T.C., Durcova-Hills, G., Surani, A., and McLaren, A. (2008). Heterogeneity in imprinted methylation patterns of pluripotent embryonic germ cells derived from premigratory mouse germ cells. Developmental biology 313, 674-681.
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.
Silva, J., and Smith, A. (2008). Capturing pluripotency. Cell 132, 532-536.
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.
Smith, A. (2017). Formative pluripotency: the executive phase in a developmental continuum. Development 144, 365-373.
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, 688-690.
Smith, A.G., and Hooper, M.L. (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. Dev Biol 121, 1-9.
Stefanovic, S., Abboud, N., Desilets, S., Nury, D., Cowan, C., and Puceat, M. (2009). Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate. J Cell Biol 186, 665- 673.
Stevens, L.C. (1958). Studies on transplantable testicular teratomas of strain 129 mice. J Natl

Cancer Inst 20, 1257-1275. 193
Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development 132, 2093-2102.
Suemori, H., Tada, T., Torii, R., Hosoi, Y., Kobayashi, K., Imahie, H., Kondo, Y., Iritani, A., and Nakatsuji, N. (2001). Establishment of embryonic stem cell lines from cynomolgus monkey blastocysts produced by IVF or ICSI. Dev Dyn 222, 273-279.
Surani, M.A. (2007). Germ cells: the eternal link between generations. C R Biol 330, 474-478.
Sybirna, A., Tang, W.W.C., Pierson Smela, M., Dietmann, S., Gruhn, W.H., Brosh, R., and Surani, M.A. (2020). A critical role of PRDM14 in human primordial germ cell fate revealed by inducible degrons. Nat Commun 11, 1282.
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.
Takahashi, S., Kobayashi, S., and Hiratani, I. (2018). Epigenetic differences between naive and primed pluripotent stem cells. Cell Mol Life Sci 75, 1191-1203.
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.
Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A., and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. Science 282, 2072-2075.
Tang, W.W., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F., and Surani, M.A. (2015). A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. Cell 161, 1453-1467.
Tang, W.W., Kobayashi, T., Irie, N., Dietmann, S., and Surani, M.A. (2016). Specification and epigenetic programming of the human germ line. Nat Rev Genet 17, 585-600.
Tapia, N., MacCarthy, C., Esch, D., Gabriele Marthaler, A., Tiemann, U., Arauzo-Bravo, M.J., Jauch, R., Cojocaru, V., and Scholer, H.R. (2015). Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. Sci Rep 5, 13533.
Tchieu, J., Zimmer, B., Fattahi, F., Amin, S., Zeltner, N., Chen, S., and Studer, L. (2017). A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages. Cell

Stem Cell 21, 399-410.e397.
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. The Francis Crick Institute. (2016, Feburary). Human Embryo and Stem Cell Laboratory. Retrieved January 2024, from The Francis Crick Institute: https://www.crick.ac.uk/research/labs/kathy-niakan/human-embryo-genomeediting-licence
The International Society for Stem Cell Research. (2024). Laboratory-based Human Embryonic Stem Cell Research, Embryo Research, and Related Research Activities. Retrieved from The International Society for Stem Cell Research: https://www.isscr.org/guidelines/blog-post-title-one-ed2td-6fcdk 194
Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.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 282, 1145-1147.
Thomson, M., Liu, S.J., Zou, L.N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell 145, 875-889.
Toyooka, Y., Tsunekawa, N., Akasu, R., and Noce, T. (2003). Embryonic stem cells can form germ cells in vitro. Proc Natl Acad Sci U S A 100, 11457-11462.
Tremblay, K.D., Dunn, N.R., and Robertson, E.J. (2001). Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. Development 128, 3609-3621.
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 141, 1209-1221.
Turnpenny, L., Spalluto, C.M., Perrett, R.M., O'Shea, M., Hanley, K.P., Cameron, I.T., Wilson, D.I., and Hanley, N.A. (2006). Evaluating human embryonic germ cells: concord and conflict as pluripotent stem cells. Stem Cells 24, 212-220.
Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. Nature 600,

285-289.
Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J Cell Sci 118, 4495-4509.
Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W., Cho, C.H., Martinez, A., Rugg-Gunn, P., et al. (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development 136, 1339-1349.
Vanzan, L., Soldati, H., Ythier, V., Anand, S., Braun, S.M.G., Francis, N., and Murr, R. (2021). High throughput screening identifies SOX2 as a super pioneer factor that inhibits DNA methylation maintenance at its binding sites. Nat Commun 12, 3337.
Vincent, S.D., Dunn, N.R., Sciammas, R., Shapiro-Shalef, M., Davis, M.M., Calame, K., Bikoff, E.K., and Robertson, E.J. (2005). The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. Development 132, 1315-1325.
Viswanathan, P., Gaskell, T., Moens, N., Culley, O.J., Hansen, D., Gervasio, M.K., Yeap, Y.J., and Danovi, D. (2014). Human pluripotent stem cells on artificial microenvironments: a high content perspective. Front Pharmacol 5, 150.
von Meyenn, F., Berrens, R.V., Andrews, S., Santos, F., Collier, A.J., Krueger, F., Osorno, R., Dean, W., Rugg-Gunn, P.J., and Reik, W. (2016). Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification. Dev Cell 39, 104-115.
Wang, H., Xiang, J., Zhang, W., Li, J., Wei, Q., Zhong, L., Ouyang, H., and Han, J. (2016). Induction of Germ Cell-like Cells from Porcine Induced Pluripotent Stem Cells. Sci Rep 6, 27256. 195
Wang, L., Schulz, T.C., Sherrer, E.S., Dauphin, D.S., Shin, S., Nelson, A.M., Ware, C.B., Zhan, M., Song, C.-Z., Chen, X., et al. (2007). Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. Blood 110, 4111-4119.
Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S.-i., Muguruma, K., et al. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nature Biotechnology 25, 681-686.
Weber, S., Eckert, D., Nettersheim, D., Gillis, A.J., Schafer, S., Kuckenberg, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L.H., et al. (2010). Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. Biol Reprod 82, 214-223.

Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Biol 17, 155-169.

Wilson, M.H., Coates, C.J., and George, A.L., Jr. (2007). PiggyBac transposon-mediated gene transfer in human cells. Mol Ther 15, 139-145.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol 13, 838-845.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem Soc Trans 38, 1027-1032.

Xia, X., Zhang, Y., Zieth, C.R., and Zhang, S.C. (2007). Transgenes delivered by lentiviral vector are suppressed in human embryonic stem cells in a promoter-dependent manner. Stem Cells Dev 16, 167-176.

Xu, R.H., Sampsell-Barron, T.L., Gu, F., Root, S., Peck, R.M., Pan, G., Yu, J., AntosiewiczBourget, J., Tian, S., Stewart, R., et al. (2008). NANOG is a direct target of TGFbeta/activinmediated SMAD signaling in human ESCs. Cell Stem Cell 3, 196-206.

Yamashiro, C., Sasaki, K., Yabuta, Y., Kojima, Y., Nakamura, T., Okamoto, I., Yokobayashi, S., Murase, Y., Ishikura, Y., Shirane, K., et al. (2018). Generation of human oogonia from induced pluripotent stem cells in vitro. Science 362, 356-360.

Yamashiro, C., Sasaki, K., Yokobayashi, S., Kojima, Y., and Saitou, M. (2020). Generation of human oogonia from induced pluripotent stem cells in culture. Nat Protoc 15, 1560-1583.

Yeo, J.C., Jiang, J., Tan, Z.Y., Yim, G.R., Ng, J.H., Goke, J., Kraus, P., Liang, H., Gonzales, K.A., Chong, H.C., et al. (2014). Klf2 is an essential factor that sustains ground state pluripotency. Cell Stem Cell 14, 864-872.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development 122, 881-894.

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.

Ying, Y., Liu, X.M., Marble, A., Lawson, K.A., and Zhao, G.Q. (2000). Requirement of Bmp8b for the generation of primordial germ cells in the mouse. Mol Endocrinol 14, 1053-1063.

Yokobayashi, S., Okita, K., Nakagawa, M., Nakamura, T., Yabuta, Y., Yamamoto, T., and Saitou, M. (2017). Clonal variation of human induced pluripotent stem cells for induction into the germ cell fate. Biol Reprod 96, 1154-1166.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev 9, 2635-2645.

Yuan, L., Liu, J.G., Zhao, J., Brundell, E., Daneholt, B., and Hoog, C. (2000). The murine SCP3 gene is required for synaptonemal complex assembly, chromosome synapsis, and male fertility. Mol Cell 5, 73-83. 196

Yuan, W., Yao, Z., Veerapandian, V., Yang, X., Wang, X., Chen, D., Ma, L., Li, C., Zheng, Y., Luo, F., et al. (2021). The histone demethylase KDM2B regulates human primordial germ cell- like cells specification. Int J Biol Sci 17, 527-538.

Zaehres, H., Lensch, M.W., Daheron, L., Stewart, S.A., Itskovitz-Eldor, J., and Daley, G.Q. (2005). High-efficiency RNA interference in human embryonic stem cells. Stem Cells 23, 299-305.

Zafarana, G., Avery, S.R., Avery, K., Moore, H.D., and Andrews, P.W. (2009). Specific knockdown of OCT4 in human embryonic stem cells by inducible short hairpin RNA interference. Stem Cells 27, 776-782.

Zhang, M., Leitch, H.G., Tang, W.W.C., Festuccia, N., Hall-Ponsele, E., Nichols, J., Surani, M.A., Smith, A., and Chambers, I. (2018). Esrrb Complementation Rescues Development of Nanog-Null Germ Cells. Cell Rep 22, 332-339.

Zhao, S., Nichols, J., Smith, A.G., and Li, M. (2004). SoxB transcription factors specify neuroectodermal lineage choice in ES cells. Molecular and Cellular Neuroscience 27, 332-342.