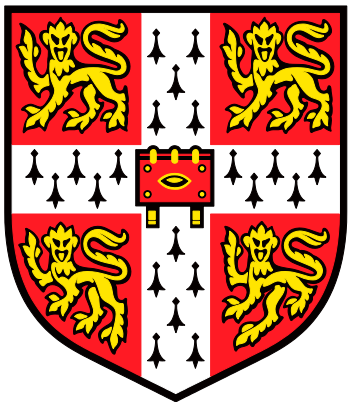
CHARACTERISING THE REPROGRAMMING DYNAMICS BETWEEN HUMAN PLURIPOTENT STATES

Amanda Jayne Collier

Hughes Hall College



University of Cambridge
Babraham Institute

This dissertation is submitted for the degree of doctor of philosophy

July 2018

Declaration

I hereby declare that my thesis entitled ‘Characterising the Reprogramming Dynamics between Human Pluripotent States’ is the result of my own work and includes nothing which is the outcome of work done in collaboration or which has been submitted for a previous degree except when specifically indicated in the text.

This thesis does not exceed the word limit of 60,000 words.

10/01/2019

Signature Date

A. J. Collier

Print

CHARACTERISING THE REPROGRAMMING DYNAMICS BETWEEN HUMAN PLURIPOTENT STATES

Amanda Jayne Collier

Abstract

Human pluripotent stem cells (hPSCs) exist in multiple states of pluripotency, broadly categorised as naïve and primed states. These provide an important model to investigate the earliest stages of human embryonic development. Naïve cells can be obtained through primed-to-naïve reprogramming; however, there are no reliable methods to prospectively isolate unmodified naïve cells during this process. Moreover, the current isolation strategies are incompatible for enrichment of naïve hPSCs early during reprogramming. Consequently, we know very little about the temporal dynamics of transcriptional changes and remodelling of the epigenetic landscape that occurs during the reprogramming process.

To address this knowledge gap, I sought to develop an isolation strategy capable of identifying nascent naïve hPSCs early during reprogramming. Comprehensive profiling of cell-surface markers by flow cytometry in naïve and primed hPSCs revealed pluripotent state-specific antibodies. By compiling the identified state-specific markers into a multiplexed antibody panel, I was able to distinguish naïve and primed hPSCs. Moreover, the antibody panel was able to track the dynamics of primed-to-naïve reprogramming, as the state-specific surface markers collectively reflect the change in pluripotent states.

Through using the newly identified surface markers, I found that naïve cells are formed at a much earlier time point than previously realised, and could be subsequently isolated from a heterogeneous cell population early during reprogramming. This allowed me to perform the first molecular characterisation of nascent naïve hPSCs, which revealed distinct transcriptional changes associated with early and late stage naïve cell formation. Analysis of the DNA methylation landscape showed that nascent naïve cells are globally hypomethylated, whilst imprint methylation is largely preserved. Moreover, the loss of DNA methylation precedes X-chromosome reactivation, which occurs primarily during the late-stage of primed-to-naïve reprogramming, and is therefore a hallmark of mature naïve cells. Using the antibody panel at discrete time points throughout reprogramming has allowed an unprecedented insight into the early molecular events leading to naïve cell formation, and permits the direct comparison between different naïve reprogramming methods. Taken together, the identified state-specific surface markers provide a robust and straightforward method to unambiguously define human PSC states, and reveal for the first time the order of transcriptional and epigenetic changes associated with primed to naïve reprogramming.

Acknowledgements

There are many people that I would like to sincerely thank for their support during my PhD. First and foremost, my supervisor Peter Rugg-Gunn. Throughout the duration of my PhD, he has been a source of endless support, wisdom and supreme mentorship. With his encouragement I have accomplish more than I had ever anticipated possible during my PhD.

Huge thanks goes to all current and past members of the Rugg-Gunn lab for contributing towards a phenomenal PhD experience. Likewise, I would like to thank Sarah Elderkin, Claire Senner, Ferdinand von Meyenn and Stefan Schoenfelder for insightful discussions and valuable mentorship.

I would also like to thank many members of the Babraham Institute, particularly those within the Epigenetics ISP and the superb Babraham core facilities.

Further thanks goes to all of my fantastic collaborators who have helped to make this work possible.

Finally, I am grateful to all of my family, friends and partner for all of their tremendous encouragement. I am endlessly appreciative of the love and inspiration that my mother and partner have provided during this journey.

Table of Acknowledgements

Initial training in techniques and laboratory practice and subsequent mentoring:

Dr Peter Rugg-Gunn
Mr Peter Chovanec
Dr Adam Collinson
Miss Natasha Morgan
Dr Clara Novo
Dr Ferdinand Von Meyenn
Dr Sarah Elderkin
Miss Charlene Fabian
Dr Stefan Schoenfelder
Dr Simon Cook
Dr José Silva

Data obtained from a technical service provider:

Dr Simon Andrews - Bioinformatics training for RNA-seq and ChIP-seq analysis
Dr Laura Biggins – Quantification of X-linked gene expression
Dr Felix Krueger – Quantification of X-linked gene expression
Dr Rachael Walker - Flow cytometry core training and cell sorting
Miss Lynzi Waugh - Flow cytometry core cell sorting
Dr Rebecca Roberts - Flow cytometry core cell sorting
Dr Simon Walker - Imaging facility training
Miss Kristina Tabbada – Sequencing facility
Miss Clare Murnane - Sequencing facility

Data produced jointly:

Dr Ferdinand Von Meyenn - DNA methylation analysis
Dr Sarita Panula - BD Bioscience lyo-plate screen analysis
Mr Peter Chovanec - FlowSOM bioinformatics analysis

Data/materials provided by someone else:

Dr. Fredrik Lanner, Dr. Sarita Panula, Mr. John Paul Schell, Mr. Alvaro Plaza Reyes and Dr. Sophie Petropoulos - Human embryo immunofluorescence

TABLE OF CONTENTS

List of Figures

List of Tables

List of abbreviations

2i Two inhibitors

5i Five inhibitors

BMP bone morphogenic protein

cDNA Complementary DNA

CGI CpG Island

ChIP Chromatin immunoprecipitation

CHIR CHIRON99021

CRISPR Clustered regularly interspaced short palindromic repeats

DNA Deoxyribonucleic acid

dNTP Deoxynucleotide

DOX Doxycycline

EpiSC Epiblast stem cell

ERK Extracellular receptor kinase

ESC Embryonic stem cell

FACS Fluorescence-activated cell sorting

FBS Foetal bovine serum

FGF Fibroblast growth factor

FPKM Fragments per kilobase of transcript per million mapped reads

GFP Green fluorescent protein

H2AK119ub Histone 2A Lysine 119 Mono-ubiquitination

H3K4me1 Histone 3 Lysine 4 Mono-methylation

H3K4me3 Histone 3 Lysine 4 Tri-methylation

H3K9me3 Histone 3 Lysine 9 Tri-methylation

H3K27ac Histone 3 Lysine 27 Acetylation

H3K27me3 Histone 3 Lysine 27 Tri-methylation

hPSC Human pluripotent stem cell

HDAC Histone deacetylase

HMTase Histone methyltransferase

ICM Inner cell mass

iPSC Induced pluripotent stem cell

KLF Krüppel-like family

KO Knockout

LIF Leukemia inhibitory factor

lncRNA Long non-coding RNA

MEF Mouse embryonic fibroblast

NK2 Nanog, KLF2

OSN Oct4, Sox2, Nanog

PBS Phosphate-buffered saline

PCA Principal component analysis

PcG Polycomb Group

PCR Polymerase chain reaction

PD03 PD0325901

PRC1 Polycomb Repressive Complex 1

PRC2 Polycomb Repressive Complex 2

RNA Ribonucleic acid

RPKM Reads per kilobase of transcript per million mapped reads

RT-qPCR Real-time quantitative PCR

STAT Signal transducer and activator of transcription

TF Transcription factor

TGF-β Transforming growth factor beta

ZFP Zinc finger protein

List of Publications

The following publications resulted from the work carried out during my PhD:

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.

Collinson, A., **Collier, A.J***., Morgan, N.P*., Sienerth, A.R., Chandra, T., Andrews, S., and Rugg-Gunn, P.J. (2016). Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells. Cell Reports 17, 2700-2714.

Vallot, C., **Collier, A.J***., Patrat, C*., Huret, C*., Casanova, M., Liyakat Ali, T.M., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P.J., et al. (2017). XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. Cell stem cell 20, 102-111.

Freire-Pritchett, P., Schoenfelder, S., Varnai, C., Wingett, S.W., Cairns, J., **Collier, A.J**., Garcia-Vilchez, R., Furlan-Magaril, M., Osborne, C.S., Fraser, P., et al. (2017). Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. eLife 6.

Collier, A.J*., Panula, S.P*., Schell, J.P., Chovanec, P., Plaza Reyes, A., Petropoulos, S., Corcoran, A.E., Walker, R., Douagi, I., Lanner, F., et al. (2017). Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. Cell stem cell 20, 874-310065408.

Collier, A.J., and Rugg-Gunn, P.J. (2018). Identifying Human Naïve Pluripotent Stem Cells − Evaluating State-Specific Reporter Lines and Cell-Surface Markers. BioEssays 0, 1700239.

*Equal contribution

1. Introduction

Cells of the early mammalian embryo retain the remarkable ability to generate all of the cell types found within the adult body, a potential referred to as pluripotency. As embryonic development progresses, pluripotent cells progressively differentiate into specialised cell types. The transient nature of pluripotency during embryonic development presents a challenge for researchers with the desire to stabilise and maintain this cell type *in-vitro*. Nonetheless, pluripotent stem cell (PSC) lines have been derived from multiple sources such as germ cell tumours, both the pre- and postimplantation-stage embryo, and via somatic cell reprogramming to an induced pluripotent state. These PSCs can be propagated indefinitely *in-vitro* in an undifferentiated state, whilst retaining the capacity to differentiate into all cell lineages of the body. PSCs therefore provide an attractive model to study early embryonic development and subsequent cell fate commitment. The attribute of indefinite self-renewal, whilst retaining the potential to differentiate, is one of the foremost reasons that stem cells receive so much attention. It is hoped that one day they will provide a limitless supply of cells for disease modelling, drug discovery, and the treatment of degenerative diseases. In the subsequent sections, I will provide a brief overview of the different pluripotent cell types and their relationship to embryonic development.

1.1 Mouse Pluripotency and Stem Cells

1.1.1 Mouse Embryonal Carcinoma Cells

In 1954, Stevens and Little hypothesised the existence of pluripotent cells. This statement was based upon the observation that tumours of germ cell origin could be serially transplanted and contained both differentiated and undifferentiated cell types (Stevens and Little, 1954). Further experiments corroborated this finding, and revealed that a single teratocarcinoma cell possessed the capability of deriving a new tumour upon transplantation, which contained cell types from all three germ layers – ectoderm, endoderm and mesoderm (Kleinsmith and Pierce, 1964; Pierce and Verney, 1961). Moreover, the discovery that grafting mouse embryos into adult mice resulted in teratocarcinoma formation, underpinned the concept that pluripotent cells are present in the early embryo (Solter et al., 1970; Stevens, 1970). The putative pluripotent cells within teratocarcinomas were designated Embryonal Carcinoma Cells (ECCs) (Pierce and Verney, 1961). ECCs were subsequently isolated and maintained under culture conditions, serving as the first platform to study embryonic development of mice *in-vitro* (Evans, 1972; Kahan and Ephrussi, 1970; Rosenthal et al., 1970). Further testament to their pluripotent

identity, ECCs were also able to contribute to chimeras upon blastocyst injection (Mintz and Illmensee, 1975; Papaioannou et al., 1975).

1.1.2 Mouse Embryonic Stem Cells

The pioneering work on ECCs laid a foundation for the subsequent isolation of PSCs directly from mouse blastocyst-stage embryos. In 1981, two teams described the ability to derive pluripotent mouse embryonic stem cells (mESCs), by either culturing intact preimplantation mouse blastocysts (Evans and Kaufman, 1981), or from the inner cell mass (ICM) (Martin, 1981); both of which required the support from a layer of mitotically-inactivated mouse embryonic fibroblasts (MEFs). Mouse ESCs are derived from the early epiblast and are able to self-renew indefinitely *in-vitro*. Similar to ECCs, mESCs form teratocarcinomas when ectopically grafted and contribute to chimeric embryos upon blastocyst injection (Bradley et al., 1984; Evans and Kaufman, 1981; Martin, 1981). Importantly, a single mESC was able to generate a chimeric embryo displaying a contribution to multiple germ layers (Beddington and Robertson, 1989), highlighting an essential feature that pluripotent cells retain the ability to differentiate when exposed to different growth conditions. Perhaps the most stringent demonstration of pluripotent potential is via the tetraploid complementation assay. This showed that mESCs can form a whole embryo when aggregated with a tetraploid host, whereby the tetraploid cells contribute to the extraembryonic tissues, whilst the embryo itself and subsequent mouse is derived exclusively from the injected diploid mESCs (Nagy et al., 1993).

1.1.3 The relationship of signaling pathways and alternative pluripotent states in the context of mouse embryonic development

The initial culture conditions used to derive mouse ECCs and ESCs was reliant on the support provided by a layer of MEFs and undefined serum-containing media. Efforts were subsequently taken to understand what factors contained within the serum, or that were produced by MEFs, enabled mESCs to remain in a pluripotent state. These studies revealed that MEFs produce the cytokine leukemia inhibitory factor (LIF), which suppresses spontaneous mESC differentiation (Smith et al., 1988; Williams et al., 1988), and the serum component acts to stimulate the BMP signaling pathway (Ying et al., 2003). LIF is a member of the Interleukin-6 family and acts via the LIF receptor (LIFR), which forms a heterodimer with the LIF coreceptor, GP130. This event triggers the activation of Janus Kinase (JAK), which results in the phosphorylation and activation of the transcription factor Signal Transducers and Activators of Transcription 3

(STAT3) (Boeuf et al., 1997; Niwa et al., 1998). Phosphorylated STAT3 (p-STAT3) can subsequently translocate to the nucleus to activate the transcription of target genes. Serum/LIF conditions are frequently used to maintain mESCs in a pluripotent state; however, a degree of transcriptional heterogeneity has been described within these cultures. Subpopulations of cells co-express pluripotency and lineage-priming genes to varying amounts, which may therefore alter their pluripotent potential (Canham et al., 2010; Hayashi et al., 2008; Toyooka et al., 2008).

To overcome the limitation of using undefined growth conditions, efforts towards maintaining mESCs in defined serum and feeder-free conditions were taken. This identified a combination of low-doses of bone morphogenetic protein 4 (BMP4) and LIF to mimic the signaling cues provided by Serum/LIF (Ying et al., 2003). Subsequently the inhibition of MEK/ERK signaling was found to enhance ESC derivation and culture (Buehr and Smith, 2003; Burdon et al., 1999). The MEK inhibitor PD0325901 blocks the phosphorylation of ERK. In a normal context, activated ERK results in the degradation of key pluripotency transcription factors, such as KLF2 and KLF4 (Dhaliwal et al., 2018; Yeo et al., 2014). This approach was extended with the 3i media formulation that included MEK/ERK inhibition, alongside inhibitors of glycogen synthase kinase-3 (GSK3) and FGFR signaling (Ying et al., 2008). The GSK-3 inhibitor CHIR99021 results in the stabilisation of β-catenin to promote Wnt signaling, which consequently alleviates TCF3 repression of pluripotency factors such as *Oct4*, *Nanog*, and *Rex1* (Martello et al., 2012; Pereira et al., 2006; Sato et al., 2004; Wray et al., 2011). However, it was later shown that FGFR inhibition was dispensable and that culturing mESCs in 2i (dual inhibition of MEK and GSK3) in combination with LIF, collectively referred to as 2i/LIF, resulted in a homogeneous population of cells designated as reaching 'ground-state pluripotency' (Silva et al., 2008; Ying et al., 2008). Moreover, 2i/LIF culture conditions have made it possible to generate germline competent ESCs from rats and recalcitrant mouse strains (Buehr et al., 2008; Czechanski et al., 2014; Li et al., 2008; Nichols et al., 2009). Mouse ESCs derived from preimplantation stage blastocysts (E3.5 – E.45 developmental stage) can be stably maintained in cultured under 2i/LIF or serum/LIF conditions. The resultant cell lines resemble the preimplantation epiblast cells of origin, and have been classified as being in a 'naïve' state of pluripotency (Figure 1.1) (Nichols and Smith, 2009; Ying et al., 2008).

Conversely, an alternative pluripotent cell type can be generated from mouse embryos, termed epiblast stem cells (EpiSCs), which are classified as being in a 'primed' state of pluripotency. These are most commonly derived from the postimplantation-stage (Brons et al., 2007; Tesar et al., 2007), but it was later shown to be possible from the preimplantation-stage as well (Najm et al., 2011). Whilst fulfilling the criteria of pluripotency from the perspective of self-renewal and competent multi-lineage differentiation to all germ layers, EpiSCs are rarely able to contribute to chimeric blastocysts when introduced into preimplantation-stage blastocysts (Brons et al., 2007; Mascetti and Pedersen, 2016a; Tesar et al., 2007). EpiSCs are conventionally derived under culture conditions containing fibroblast growth factor (FGF) and Activin A (collectively F/A) (Kunath, 2011; Nichols and Smith, 2009). Unlike mESCs, EpiSCs cannot be maintained under 2i/LIF conditions and instead undergo widespread cell death and differentiation (Guo et al., 2009). EpiSCs are instead reliant on active TGFβ signalling through Activin/Nodal, as the inhibition of Activin/TGFB receptors results in EpiSCs differentiating into neuronal cell types (Tesar et al., 2007; Vallier et al., 2009). In summary, the acquisition of 'naïve' mESCs and 'primed' EpiSCs are dependent on the extrinsic signals provided by the culture conditions used upon their derivation. This is therefore an important consideration to have in mind when deriving PSCs from other species, as the extrinsic signals may be developmentally stage-specific and may differ from those required for mouse pluripotency.

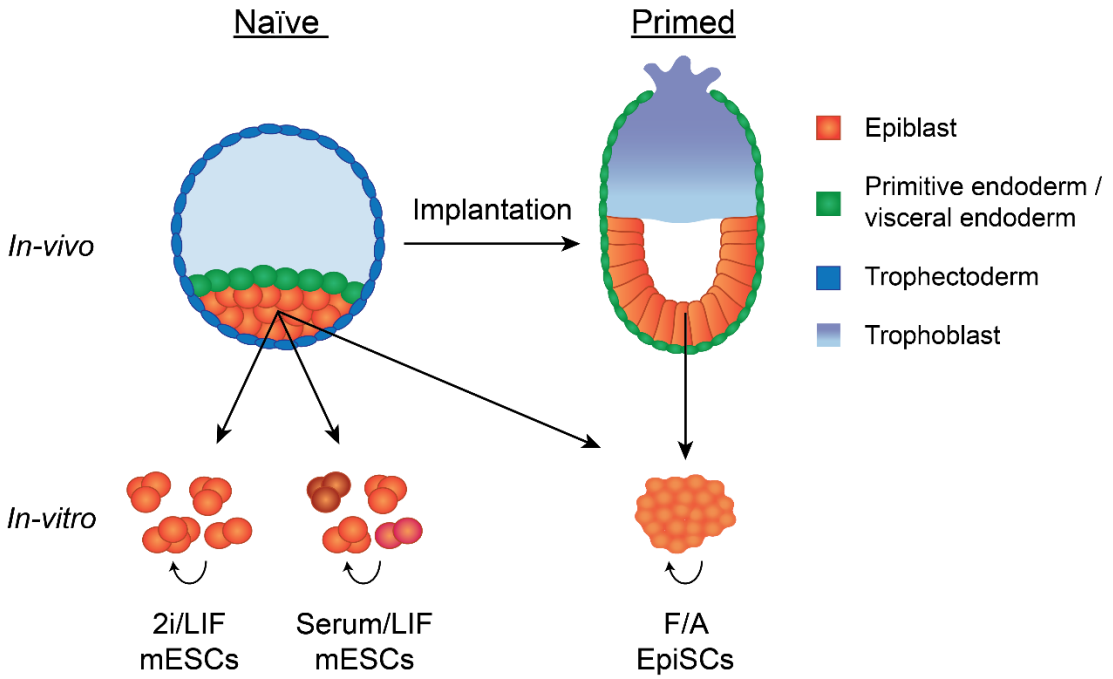


Figure 1.1. Alternative pluripotent cell types in the context of mouse embryonic development
A schematic depiction of the relationship between *in-vivo* pluripotent progression and the respective *in-vitro* cell types isolated from pre- and postimplantation epiblast cells.

1.1.4 Interconvertible pluripotent states

Both naïve mESCs and primed mEpiSCs qualify as pluripotent from the perspective of differentiation potential to all three germ layers, assessed either by the formation of embryoid bodies or teratomas. However, the two cell states are distinct in their signaling requirements and in their developmental potential. Nonetheless, interconversion between the two states is possible by manipulating the cell culture signaling cues provided. Mouse ESCs can be differentiated into mEpiSCs upon exposure to FGF2 and Activin (Guo et al., 2009). However, the conversion in the opposite direction (mEpiSCs to mESCs) can be inefficient, inferring the presence of a molecular barrier. This barrier can be overcome to some extent via the overexpression of naïve-specific transcription factors *Esrrb*, *Klf2*, *Klf4*, *Nanog*, *Nr5a2*, or *Tfcp2l1*, which can enhance the conversion process (Festuccia et al., 2012; Guo and Smith, 2010; Guo et al., 2009; Hall et al., 2009; Hanna et al., 2009; Martello et al., 2013; Zhou et al., 2010). Two further reports have indicated that it may be possible to derive mESCs from mEpiSCs without transgene induction; instead, the inhibition of FGF/TGF-β signaling combined with promoting Wnt/LIF signaling was sufficient (Bao et al., 2009b; Greber et al., 2010). Nevertheless, the efficiency was still low, which suggests a population of mEpiSCs possess a selective advantage that facilitates their reprogramming, that other mEpiSCs are lacking.

Given the progressive nature of embryonic development, it is conceivable that pluripotent epiblast cells transition through a continuum of states. Alongside the well-characterised naïve and primed states that correspond to the pre- and post-implantation epiblast respectively, various intermediate PSC types have been generated *in-vitro* (Han et al., 2010; Morgani et al., 2017). Epi-like cells (EpiLCs) are one example that have been generated via the *in-vitro* differentiation of mESCs using short-term exposure to FGF/Activin/KSR conditions (Hayashi et al., 2011). EpiLCs exhibit an intermediate transcriptional identity between ESCs and EpiSCs, and are transcriptionally similar to postimplantation-stage E5.75 epiblast cells (Hayashi et al., 2011). Interestingly, EpiLCs are proficient in generating primordial germ cell-like cells (PGCLCs), unlike mESCs and mEpiSCs (Hayashi et al., 2011). Moreover, Schöler and colleague revealed that a sub-population of cells with characteristics associated with ESCs exists within EpiSC cultures. Comparative analysis between the different cell types and states of pluripotency has collectively revealed a number of divergent molecular characteristics, summarised in Figure 1.2. Taken together, the ability to transition between the different PSC types provides an important model to understand the molecular changes that occur during pluripotent state transitions, and how these transitions can be regulated.

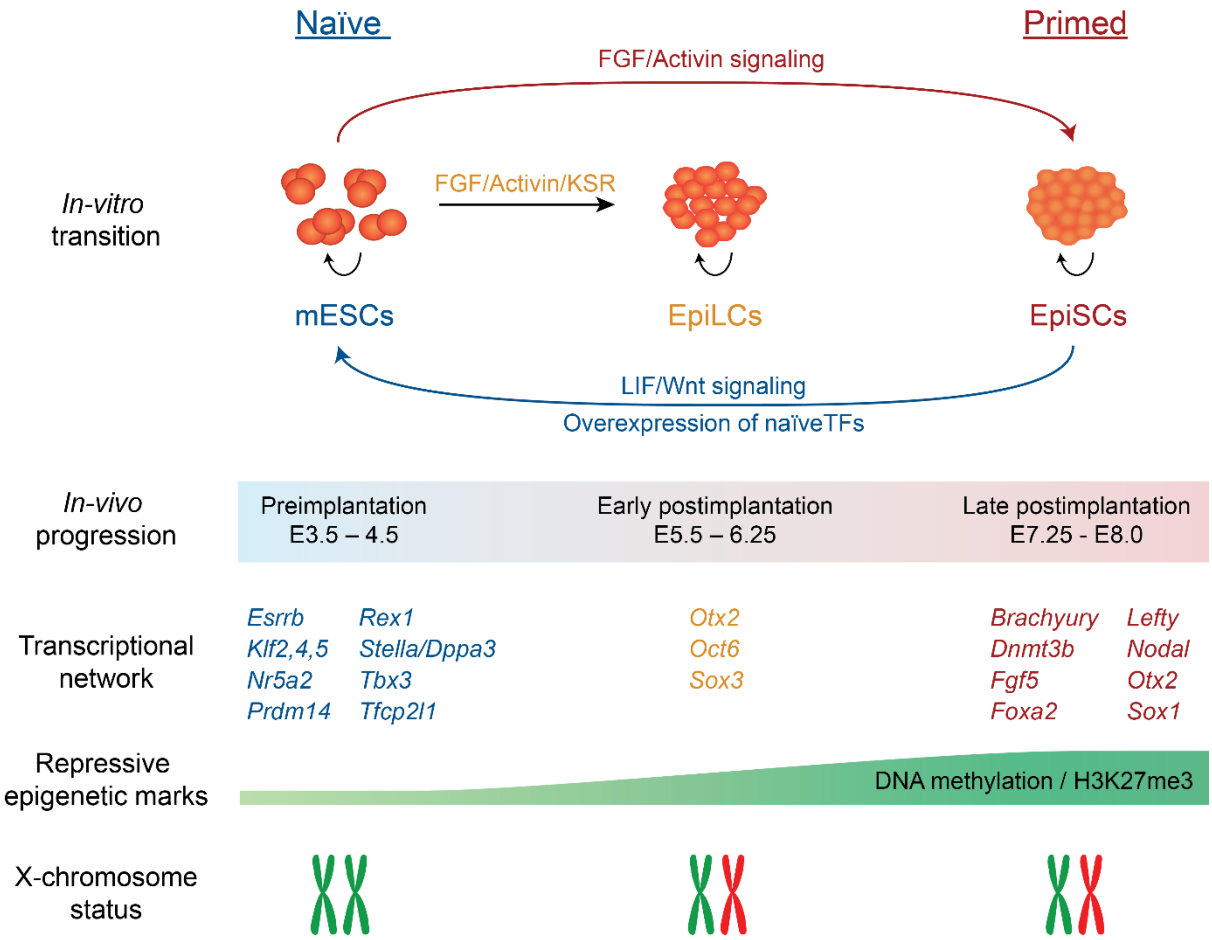


Figure 1.2. Overview of mouse pluripotent states and their associated characteristics
Alterations in the signaling environment permits the conversion between naïve, primed, and formative pluripotent states. Pluripotent state-specific transcription factors are highlighted, alongside epigenetic characteristics that vary between the cell states. Repressive DNA methylation and H3K27me3 histone modifications are acquired along the developmental progression, which is recapitulated *in-vitro*. X-chromosome status is indicated: active (green) and inactive (red).

1.2 Molecular hallmarks of mouse pluripotent states

Collectively, the experiments summarised in the previous three segments emphasise the notion that naïve mESCs and primed mEpiSCs are developmentally distinct, require different signaling cues, and possess divergent molecular features (Figure 1.2). In the subsequent sections, I will discuss in more detail several molecular features that can be used to distinguish the naïve and primed cell types.

1.2.1 Transcription factor networks and gene regulation

In line with the notion that mESCs and mEpiSCs qualify as pluripotent, both cell types express the core pluripotency transcription factors *Oct4*, *Sox2*, and *Nanog*. Yet the two cell types adopt state-specific transcriptional identities. Mouse ESCs highly express markers of pre- and peri-implantation epiblast cells. Examples include *Esrrb*, *Rex1*, *Klf4*, *Stella/Dppa3*, *Tbx3*, and *Tfc2pl1* (Boroviak et al., 2015; Ghimire et al., 2018; Ying et al., 2008); many of which are essential to maintain naïve pluripotency (Dunn et al., 2014). Conversely, EpiSCs exhibit minimal expression of these factors, and instead upregulate postimplantation epiblast markers such as *Fgf5*, *Foxa2*, *Nodal* and *Sox1* (Ghimire et al., 2018; Kojima et al., 2014; Tesar et al., 2007).

Alongside different utilisation of transcription factors, the regulatory mechanisms that govern their expression are altered depending on the pluripotent state. For instance, expression of the core pluripotency factor *Oct4* is differentially controlled by two cis-regulatory enhancer elements: the distal enhancer (DE) and the proximal enhancer (PE). The distal enhancer is preferentially engaged during naïve pluripotency (both *in-vivo* and *in-vitro*), whilst the proximal enhancer controls *Oct4* expression after implantation and in mEpiSCs (Brons et al., 2007; Choi et al., 2016; Tesar et al., 2007; Yeom et al., 1996). Further reports indicate that enhancer rewiring is a widespread phenomenon that occurs during this pluripotent state transition (Buecker et al., 2014; Factor et al., 2014; Novo et al., 2018), and is therefore a useful hallmark to distinguish the two pluripotent states.

1.2.2 Epigenetic landscape

During mammalian embryogenesis the epigenome is dramatically remodeled. This includes a genome-wide reorganisation of DNA methylation levels and histone modification occupancy, together with the silencing of an X-chromosome in female cells. Many of these events can be recapitulated *in-vitro* and largely reflect their embryonic counterparts. The first example is DNA methylation, which is a covalent chemical modification of a methyl group added to the 5’ position of cytosine residues. One of the most prevalent forms of DNA methylation occurs at CpG dinucleotides, and this has well established roles in transcriptional repression (Holliday and Pugh, 1975). The DNA methyltransferase (DNMT) family of enzymes are responsible for the deposition of DNA methylation. Conversely, the Ten Eleven Translocation (TET) enzymes are responsible for active DNA demethylation (Smith and Meissner, 2013). Throughout embryogenesis and mammalian development, the DNA methylation landscape undergoes

dramatic remodeling. The naïve state of pluripotency is characterised by a state of global hypomethylation; whilst transitioning to a primed state, either via embryo development or via the assessment of primed mEpiSCs, triggers a global increase in DNA methylation levels (Borgel et al., 2010; Hackett et al., 2013; Leitch et al., 2013; Senner et al., 2012; Smith et al., 2012; Veillard et al., 2014; Wang et al., 2014b). The differential expression of DNMT enzymes and targeting factor UHRF1 reflects this situation, such that DNMT3A, DNMT3B and UHRF1 are downregulated in naïve conditions relative to the primed state (Leitch et al., 2013; Veillard et al., 2014; von Meyenn et al., 2016). DNA methylation levels, or the factors involved in their regulation, can therefore be used to distinguish these two states of pluripotency.

The second notable example by which epigenetic features define naïve and primed states is the status of the X-chromosome in female cells. Early during mouse development, the paternal X-chromosome undergoes imprinted X-inactivation, which is maintained across extraembryonic cell types (Huynh and Lee, 2003; Mak et al., 2004; Okamoto et al., 2004). Cells of the epiblast subsequently reactivate the paternal X-chromosome during preimplantation development, before random X-inactivation occurs upon implantation (Rastan, 1982; Takagi et al., 1982). In accordance with their pre- and postimplantation status, female mESCs exhibit two active X-chromosomes, whilst one is inactivated in mEpiSCs (Bao et al., 2009a; Guo et al., 2009). X-chromosome inactivation (XCI) is mediated by the long non coding RNA (lncRNAs) *Xist*, which upon induction triggers X-chromosome silencing in cis (Penny et al., 1996). This event is followed by the exclusion of RNA pol II and active histone modifications, before the acquisition of repressive histone modifications H3K27me3 and H2A119ub, deposited by the polycomb repressive complex 2 (PRC2) and PRC1 respectively (Plath et al., 2003; Silva et al., 2003).

The third example used to distinguish alternative pluripotent states relates to the chromatin architecture and chromatin associated modifications of the histone tails. Genome-wide mapping of the chromatin landscape in 2i-cultured mESCs revealed a reduction in the number of bivalent domains, defined by the co-occurrence of active H3K4me3 and repressive H3K27me3 marks, compared to serum-cultured mESCs (Marks et al., 2012). Bivalent domains are considered to keep their associated genes in a poised state, ready for rapid activation or repression during subsequent differentiation (Azuara et al., 2006; Bernstein et al., 2006). Global H3K27me3 levels do not differ between these two states, yet feature enrichment is distinct; promoter regions in 2i-cultured mESCs are frequently devoid of H3K27me3, and instead exhibit an enrichment at satellite repeat elements (Marks et al., 2012). This could be an alternative strategy to silence repeat sequences under hypomethylated conditions.

In summary, the molecular characterisation of naïve and primed states of mouse pluripotency *in-vitro* has provided a foundation of knowledge that can be related back to *in-vivo* mouse embryonic development. The property of indefinite self-renewal in culture permits experiments to be performed, which would otherwise not be possible with the low material provided by direct embryo assessment. Moreover, our understanding of the molecular features associated with mouse pluripotency can provide a framework for studying pluripotency in other species, such as humans.

1.3 Human Pluripotency and Stem Cells

1.3.1 Deriving human pluripotent stem cells from embryos

After the successful isolation of mouse PSCs, efforts were taken to derive PSCs from a range of mammalian species, including primates. James Thomson firstly generated PSCs from two non-human primate species, the rhesus macaque (Thomson et al., 1995) and the common marmoset (Thomson et al., 1996), before deriving the first human pluripotent stem cells (hPSCs) in 1998 (Thomson et al., 1998). Five different hPSC lines were generated from donated human blastocysts acquired by *in-vitro* fertilisation (Thomson et al., 1998). Owing to their indefinite self-renewal capacity, several of these hPSCs lines are still used across the world today; most notably the female H9 line and the male H1 line (Löser et al., 2010). The pluripotent potential of hPSCs can be verified by examining the ability to generate teratomas upon injection into immunodeficient mice (Lensch et al., 2007). This can be further supported by the formation of embryoid bodies *in-vitro,* which are spherical structures that differentiate to contain cell types encompassing all three germ layers (Kurosawa, 2007). More recently, it has also been shown that hPSCs can form human-mouse interspecies chimeras, but only when introduced into postimplantation gastrula stage mouse embryos (Mascetti and Pedersen, 2016b). Conversely, hPSCs rarely contribute to the developing mouse epiblast when introduced into preimplantation-stage mouse embryos (James et al., 2006; Masaki et al., 2015). This is reminiscent of the poor chimeric contribution of mEpiSCs when transplanted into preimplantation-stage mouse embryos, yet efficient contribution is seen upon postimplantation transplantation (Huang et al., 2012). Collectively, this suggests that conventional hPSCs may have developmentally progressed compared to their mESC counterparts, and subsequently phenocopy mEpiSCs (Mascetti and Pedersen, 2016a).

1.3.2 Human induced pluripotent stem cells

The ability to capture and maintain hPSCs in an undifferentiated state indefinitely *in*-vitro, whilst retaining the capacity to differentiate, holds great promise for the study of human embryonic development, disease mechanisms and drug discovery, and ultimately the use of these cells in regenerative medicine. However, hPSCs are faced with several hurdles. Firstly, the ethical consideration that deriving hPSCs results in the destruction of human embryos, and this is procedure is not permitted in several countries. Secondly, if hPSCs are differentiated into the desired cell type and used for cell therapy, the recipient patients may elicit an immune response to the allogeneic cells. This second caveat can be overcome via somatic cell nuclear transfer (SCNT), in which patient-specific fibroblasts are injected into and fused with high-quality enucleated human oocytes. Up to 10% of SCNT embryos were able to progress to the blastocyst stage and give rise to human NT-ESCs (Tachibana et al., 2013). However, the ethical concerns surrounding the use of human oocytes persists with this method.

These concerns were subsequently addressed by the advent of induced pluripotent stem cell (iPSC) technology, which was pioneered by Shinya Yamanaka's lab. It was shown that the forced expression of four pluripotency factors (*OCT4*, *SOX2*, *KLF4* and *c-MYC* or *OCT4*, *SOX2*, *LIN28*, and *NANOG*) was sufficient for the reprogramming of adult human fibroblasts to an induced state of human pluripotency, giving rise to hiPSCs (Takahashi et al., 2007; Yu et al., 2007). Fulfilling the criteria of pluripotency, hiPSCs are able to generate teratomas *in-vivo* and embryoid bodies *in-vitro* containing cell types from all three lineages. Ensuing this discovery, a range of different adult cell types have been successfully reprogrammed to an iPSC state, including dermal fibroblasts, hepatocytes, keratinocytes and human peripheral blood cells (Aasen et al., 2008; Giorgetti et al., 2009; Liu et al., 2010; Loh et al., 2009; Maherali et al., 2008). Assessment of the transcriptional changes during iPSC reprogramming from different somatic cell types has highlighted a cell of origin effect (Nefzger et al., 2017). This is perhaps unsurprising, given the different somatic transcriptional identities that must be downregulated. However, for those cells that successfully reprogramme the transcriptional identities ultimately converge (Nefzger et al., 2017).

Importantly, there are many examples where disease phenotypes can be recapitulated *in-vitro* from patient-specific reprogrammed fibroblasts (Shi et al., 2012; The Hd iPsc Consortium, 2012; Zhang et al., 2011). This exemplifies the potential of iPSC technology for disease modelling and drug discovery (Figure 1.3). Furthermore, a variety of alternative iPSC reprogramming methods have been devised that extend beyond the original use of integrating

retroviral, or lentiviral delivery of reprogramming factors. These include safer non-integrating Sendai virus (Fusaki et al., 2009), episomal vectors (Yu et al., 2009), excisable transposons (Kaji et al., 2009; Soldner et al., 2009; Woltjen et al., 2009) and synthetic modified mRNAs (Warren et al., 2010).

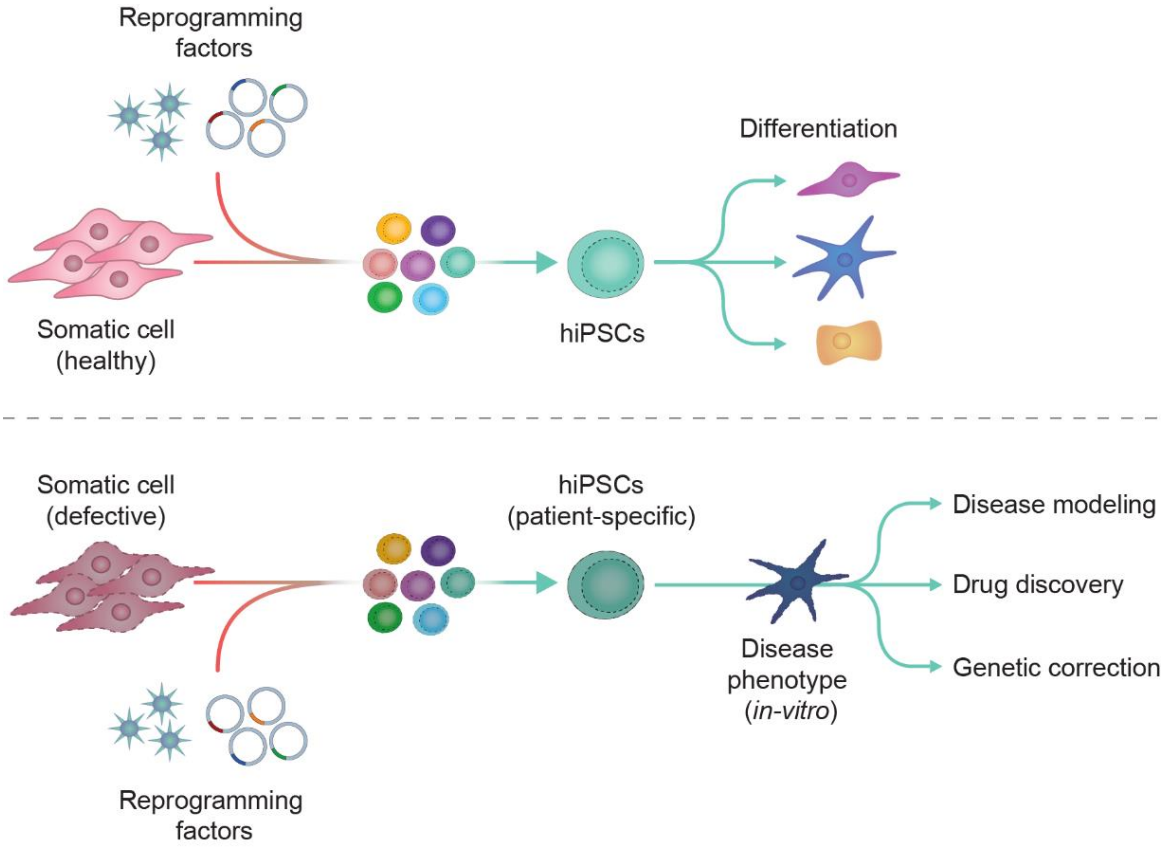


Figure 1.3. Overview of induced pluripotent stem cell reprogramming (iPSC) and subsequent applications
Reprogramming factors are delivered to the somatic cell to induce iPSC reprogramming. Human iPSCs can be subsequently differentiated into the desired cell type of interest, either from a healthy or disease-specific iPSC. Several disease-specific applications are highlighted.

The ability to generate PSCs from both embryos and via somatic cell iPSC reprogramming begs the question as to whether the resultant cell types are equivalent. Comparative transcriptional analysis of isogenic miPSCs and mESCs were indistinguishable, with the exception of several transcripts within the *Dlk1-Dio3* imprinted gene cluster (Stadtfeld et al., 2010). More recently, the same has been shown for genetically matched hiPSCs and hPSCs that cannot be distinguished based upon their transcriptional or epigenetic profiles (Choi et al., 2015). The small number of genes that were differentially expressed did not affect the developmental potential of the PSCs. Instead, biological variability between different genetic backgrounds,

length of *in-vitro* culture, and the sex of somatic cells may be responsible for the variation between hPSCs and hiPSCs observed by some groups (Bock et al., 2011; Choi et al., 2015; Kim et al., 2010; Ohi et al., 2011). Nevertheless, the advent of iPSC reprogramming hold great promise for the future study of both applied research into human disease, and as a model system to study the mechanisms that underpin cell fate reprogramming more generally.

1.3.3 Signaling requirements for conventional human pluripotent stem cells

Whilst conventional hPSCs derived from embryos qualify as pluripotent from the perspective of generating cell types from all three germ layers, there are notable differences compared to mESCs. Human PSCs have a markedly different morphology compared to their mouse counterparts. When viewed under the microscope, human PSCs grow as flat colonies of tightly packed cobblestone cells, contrary to mESCs that grow in compact dome-shaped colonies. Moreover, the optimum growth conditions for the propagation and maintenance of hPSCs are also different to those that are required for naïve mESCs. The first culture conditions used to derive hPSCs were initially determined by applying the methods developed for mESC derivation. This included the co-culture of hPSCs supported by a MEF layer in the presence of an enriched serum-containing media and LIF (Thomson et al., 1998). However, hPSCs are distinct from mESCs, such that LIF and BMP4 signalling are not sufficient in maintaining hPSCs in an undifferentiated state (Dahéron et al., 2004; Gerami-Naini et al., 2004; Humphrey et al., 2004; Xu et al., 2002). The induction of BMP signalling instead leads to trophoblast differentiation (Xu et al., 2002). Contrastingly, when the agonist noggin blocks BMP signalling, hPSCs can be maintained in an undifferentiated state (Xu et al., 2005b), collectively suggesting BMP signalling induces the differentiation of hPSCs.

The second disparity between mESC and hPSC culture is evident by the failure of 2i/LIF conditions to sustain hPSCs in a pluripotent state (Hanna et al., 2010). Instead, the optimal growth conditions for hPSCs are reminiscent of those used to culture mEpiSCs (Beattie et al., 2005; James et al., 2005a; Vallier et al., 2005; Wang et al., 2005; Xu et al., 2005a). Activation of an alternative branch of the TGF-β signalling pathway, via TGF-β/Activin/Nodal as opposed to BMP, is essential for the maintenance of hPSCs. In support of this, inhibition of Activin/Nodal signalling via the antagonist follistatin, SB43154 and Lefty, results in a loss of pluripotency (Smith et al., 2008; Vallier et al., 2005). Additionally, TGF-β signalling appears to be important for epiblast formation in humans (Blakeley et al., 2015). Upon treatment with an Activin receptor inhibitor (SB-431542), the human epiblast showed an impaired formation

assessed by the lack of NANOG and SOX17 expression. Moreover, numerous components of this pathway are highly expressed in human EPI cells, including receptor expression (*TGFBR1/ALK5* and *TDGF1*), receptor ligand expression (*LEFTY1/2*, *GDF3* and *NODAL*), and downstream signal transducers (*SMAD2* and *SMAD4*). Conversely, *SMAD1* and *SMAD5* are enriched in human trophectoderm cells, exemplifying the complex nature that the same signalling pathway can differentially regulate distinct lineages (Blakeley et al., 2015)

In addition to TGF-β signalling, hPSCs are reliant on FGF signalling via the FGF/MAPK pathway, as FGF receptor inhibition cannot be tolerated whilst maintaining an undifferentiated state. Collectively it was shown that whilst FGF signalling is necessary for maintaining hPSCs, it is not sufficient without active TGF-β/Activin/Nodal signalling, and these pathways act cooperatively to maintain hPSCs in a pluripotent state (Vallier et al., 2005).

Although human and mouse PSCs are both derived from preimplantation-stage embryos, it is evident that the culture conditions used to derive naïve mESCs are not adequate to capture the same state in humans. Instead, the conventional culture conditions used to derive hPSCs promotes a state of pluripotency that is reminiscent of primed mEpiSCs. Either this is owing to the lack of a naïve state in humans, or simply revisions must be made to the culture conditions such that this state can be captured and maintained. The discovery that multiple pluripotent states are present in the mouse embryo indicated that the same may be true for humans (Nichols and Smith, 2009). Moreover, the observation that both mESCs and EpiSCs can be derived from preimplantation mouse embryos provides evidence that *in-vivo* pluripotent cells will default to different cell types depending on the *in-vitro* culture methods used (Najm et al., 2011). Further investigation over the following years has revealed that indeed the 'naïve' state of pluripotency can be derived directly from human embryos, or more commonly via the reprogramming of primed hPSCs. Before elaborating on the culture conditions that permit the naïve hPSC state, I will discuss how the recent molecular and phenotypic studies of mouse, primate, and human embryos are providing clarity over the earliest stages of pluripotency. These studies have helped to define robust criteria by which to benchmark hPSC types against, taking into consideration species-specific differences (Boroviak and Nichols, 2017; Collier and Rugg-Gunn, 2018; Huang et al., 2014).

1.4 Human embryos provide a foundation to guide the characterisation of human pluripotent states

Assigning molecular criteria to define human pluripotent states has been hindered by the scarcity of human embryos available for scientific research, and the technical limitations of working with low-cell numbers for molecular characterisation. Over the past decade there have been great technical improvements, and consequently the molecular characteristics of human embryonic development have been studied using single-cell transcriptional profiling.

1.4.1 Transcriptional profiles of human preimplantation embryos

One of the primary revelations from sequencing human embryos indicated the temporal dynamics of lineage segregation was different in primates compared to mouse. In humans, lineage segregation occurs simultaneously in the early blastocyst, where cells of the trophectoderm (TE), the epiblast (EPI) and the primitive endoderm (PrE) are specified at the same developmental stage (Blakeley et al., 2015; Petropoulos et al., 2016; Stirparo et al., 2017). Prior to this stage (≈E5), is it not possible to clearly assign a cell fate based upon their transcriptional profile (Stirparo et al., 2017). This observation was also true in cynomolgus monkeys, whereby the three lineages (TE, EPI and PrE) could only be distinguished by ≈E8 (Nakamura et al., 2016). This is contrary to mouse lineage segregation that occurs earlier and in two stages; firstly the TE and EPI are specified at the morula stage, followed by EPI and PrE specification at the mid blastocyst stage (Figure 1.4). Interestingly, isolated and re-aggregated trophectoderm cells of human blastocyst-stage embryos are capable of generating an entirely new blastocyst, inclusive of a NANOG-expressing ICM (De Paepe et al., 2013). This further implies that lineage segregation occurs after the formation of the blastocyst, and at a later stage than during mouse development.

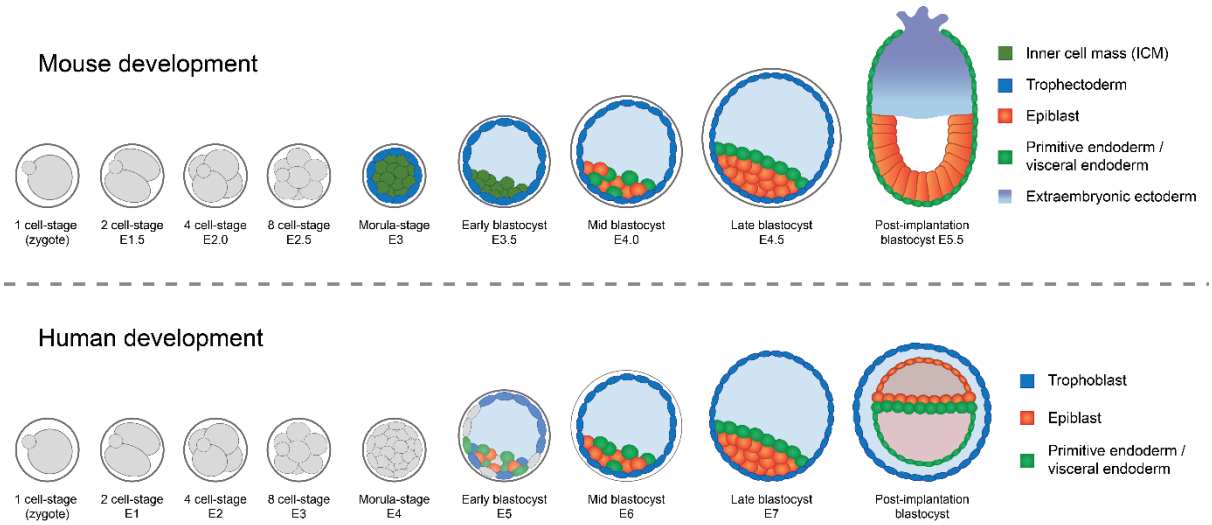


Figure 1.4. Overview of mouse and human embryonic development.
Embryonic and extraembryonic cells are specified in two waves during mouse development. Firstly, the ICM is segregated into TE and EPI cells at the morula stage, followed by EPI and PrE specification at the mid blastocyst stage. Conversely, human lineage specification occurs simultaneously from the early blastocyst-stage E5 embryo. Figure adapted from (Niakan et al., 2012).

Yan and colleagues performed a transcriptional comparison of individual human epiblast cells against conventional human PSCs (Yan et al., 2013). This highlighted a substantial difference between the two cell types, indicating that *in-vitro* hPSCs have lost their cell of origin identity. Further extensive transcriptional profiling of human embryos has reiterated this observation, whilst providing a comprehensive resource of developmental stage-specific transcriptional profiles (Blakeley et al., 2015; Petropoulos et al., 2016; Stirparo et al., 2017). This resource provides a biologically relevant benchmark for assessing hPSC types. Based on these transcriptional studies and further molecular profiling of human embryos, several characteristics of naïve human pluripotency have been proposed. This provides a helpful framework with which to evaluate human pluripotent cell types, including conventional primed hPSCs, and the derivation of 'naïve' hPSC types (Figure 1.5 and Table 1.1)

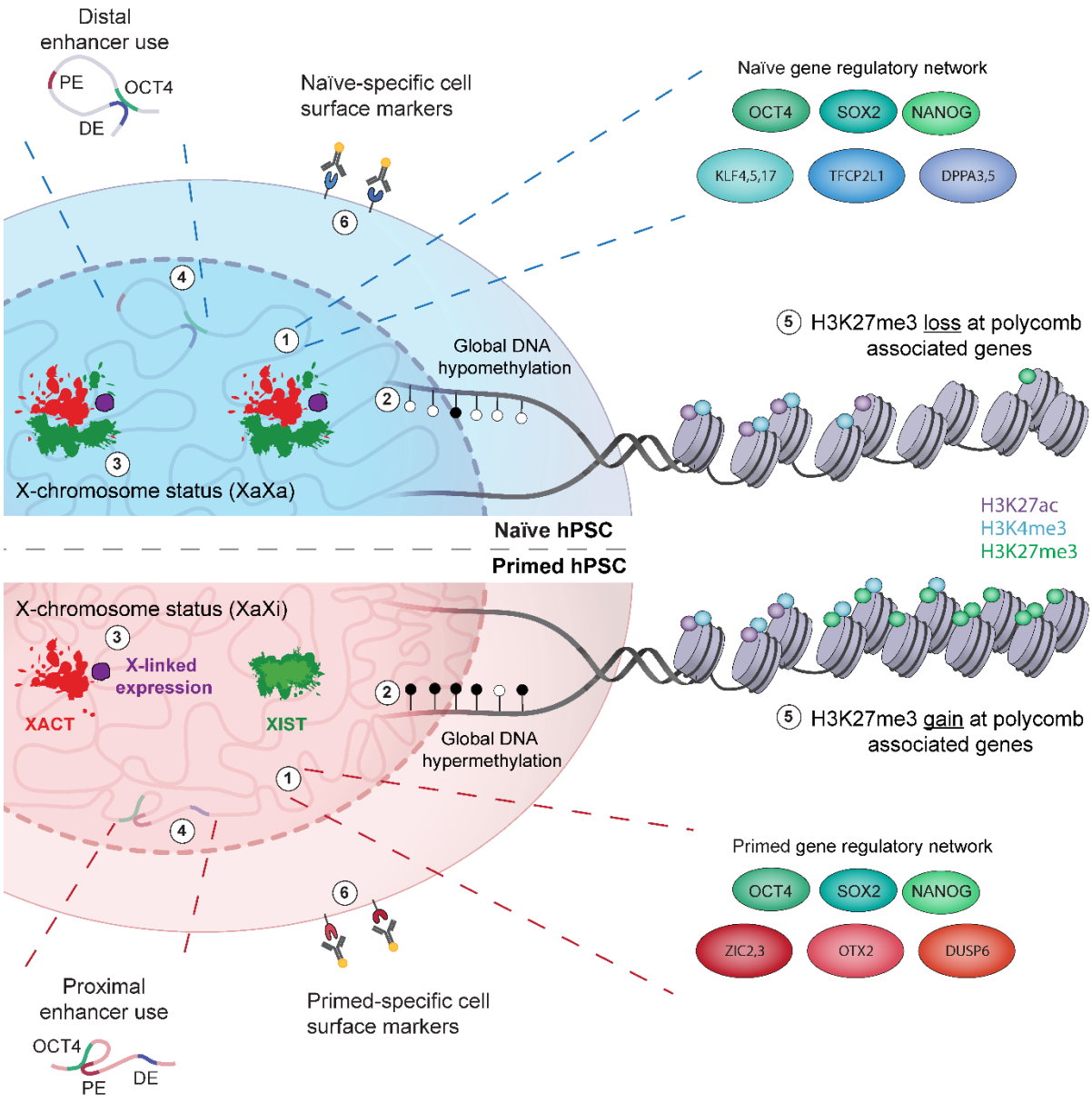


Figure 1.5. Hallmarks associated with naïve and primed human pluripotent states.
(1) The core pluripotency transcription factors OCT4, SOX2, and NANOG are expressed in both naïve and primed hPSCs, whereas several other transcription factors show state-specific enrichment. (2) Naïve hPSCs have globally hypomethylated genomes compared to primed hPSCs. (3) Naïve hPSCs have two active X-chromosomes in female cells, indicated by biallelic X-linked gene expression alongside lncRNAs expression of *XIST* and *XACT*. Conversely, primed hPSCs have one inactive X-chromosome, shown by monoallelic gene expression and the acquisition of the repressive histone mark H3K27me3. (4) Naïve hPSCs exhibit preferential activity of the *OCT4* distal enhancer, whereas primed hPSCs use the proximal enhancer. (5) H3K27me3 is absent or reduced at >3000 Polycomb-associated genes in naïve hPSCs compared to primed hPSCs. (6) Identified cell-surface markers can be used to distinguish between naïve and primed hPSCs. Figure from (Collier and Rugg-Gunn, 2018).

Table 1.1. Characteristics of naïve and primed states of pluripotency.

✓, presence of naïve state attribute; X, absence of naïve state attribute; −, attribute not examined. The brief media composition is noted alongside each publication reporting the derivation of naïve hPSCs. Conflicting results for OCT4 enhancer activity are shown, in each case the absence of DE activity was reported by: (Theunissen et al., 2014). Table from (Collier and Rugg-Gunn, 2018).

Attribute	**Naïve state**	**Primed State**	**Studies reporting the conversion of primed to naïve hPSCs**									
In vivo terminology	Preimplantation epiblast	Postimplantation epiblast	OKK+ 2iL (Hanna et al., 2010)	NHSM (Gafni et al., 2013)	3iL (Chan et al., 2013)	HDACi 2i+F (Ware et al., 2014)	5iLA(F) (Theunissen et al., 2014) (Theunissen et al., 2016)	NK2 t2iL+PKCi (Takashima et al., 2014)	2iL+ STAT3 (Chen et al., 2015)	2iL+ F,FK,AA (Duggal et al., 2015)	2iL+ FK,YAP (Qin et al., 2016)	HDACi t2iL+PKCi (Guo et al., 2017)
Appearance	Domed (✔)	Flat (✘)	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Single cell survival	High (✔)	Low (✘)	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Methylation (global CpG)	Low (≈30%)	High (≈80%)	–	High	–	–	Low	Low	Med	Med	–	Low
H3K27me3 levels	Low at promoters	High at promoters	–	Low	Low	Low	Low	–	Low	–	–	–
OCT4 enhancer	Distal (✔)	Proximal (✘)	✔	✔ / ✘	– / ✘	✔ / ✘	✔	✔	✔	–	–	✔
Metabolism	Oxidative phosphorylation (✔)	Glycolytic (✘)	–	–	–	✔	–	✔	–	–	–	–

Transcriptome profiling of human embryos has revealed both species-specific differences and similarities in the transcriptional markers used to assign cell types (Blakeley et al., 2015; Petropoulos et al., 2016; Stirparo et al., 2017). The core pluripotency transcription factors OCT4 (*POU5F1*), SOX2 and NANOG are well-conserved EPI markers expressed in mouse, primates and humans. Further conservation exists for additional mouse EPI pluripotency factors such as *Klf4*, *Sall4* and *Tfcp2l1* (Boroviak et al., 2015; Takashima et al., 2014). However, the mouse EPI factors *Klf2*, *Nr0b1* and *Esrrb* are not expressed in the human epiblast. Instead, an alternative Krueppel-Like Factor, *KLF17*, is exclusively expressed in primate EPI cells, alongside additional EPI-enriched genes such as *DPPA5*, *FGF4*, *GDF3*, *LEFTY* and *NODAL*. All of these are either absent or lowly expressed in conventional hPSCs (Stirparo et al., 2017), exemplifying the point that upon derivation these cells may have progressed.

The first studies that derived naïve-like hPSCs were published prior to the reports that revealed the molecular features of human preimplantation-stage embryos. Consequently, the initial reports may have generated cell types to match the perceived naïve properties bestowed by mESCs. Retrospective transcriptional comparisons of human EPI cells against the various naïve hPSC types has been performed (Huang et al., 2014). This revealed a considerable range in the expression levels of naïve-associated genes across the different naïve hPSC types. This suggests that some culture conditions are better than others, from the perspective of generating a cell type that aligns to the transcriptional profile of human epiblast cells (Huang et al., 2014; Pastor et al., 2016; Stirparo et al., 2017; Theunissen et al., 2016). For several of the earliest reports, a shift in transcriptional profiles towards a human preimplantation state is rather minor (Chan et al., 2013; Gafni et al., 2013; Huang et al., 2014; Ware et al., 2014). However, there are two conditions in particular, 5iLA and t2iL+PKCi, which reliably induce a global transcriptional profile that correlates with both human and monkey preimplantation-stage embryos (Guo et al., 2016b; Huang et al., 2014; Nakamura et al., 2016; Takashima et al., 2014; Theunissen et al., 2014). Moreover, these two conditions generate naïve hPSCs that express *LTR5-HERVK* and *SINE-VNTR-Alu* classes of transposable elements. Both classes are expressed in the human embryo between the 8-cell stage through to blastocyst formation, and therefore act as a useful biomarker of early human embryonic identity (Gao et al., 2018; Grow et al., 2015).

1.4.2 Naïve-specific gene regulatory elements are present in human blastocysts

Asides from the transcriptional rewiring of gene regulatory networks (GRNs), the mechanisms that govern transcription factor expression are altered depending on the pluripotent state. Gene activation frequently requires *cis*-regulatory DNA sequences, such as enhancers or promoters, to be accessed by *trans* factors. Akin to naïve mESCs, hPSCs under some naïve culture conditions utilise pluripotent state-specific enhancer elements for genes such as *POU5F1* (encoding OCT4), *KLF4*, *LEFTY1* and *NANOG* (Table 1.1) (Ji et al., 2016; Pastor et al., 2018; Theunissen et al., 2016). In the same manner that transcriptional expression can provide a useful marker of cell identity, reporter systems coupled to state-specific enhancer engagement can be engineered and used to distinguish naïve and primed hPSCs.

Perhaps owing to technical difficulties, few studies have directly examined the chromatin landscape of human embryos until recently (Gao et al., 2018; Pastor et al., 2018). This revealed that naïve hPSC-specific open chromatin regions were also accessible in human blastocyst stage embryos, whereas primed hPSC accessible regions were not conserved (Pastor et al., 2018). Moreover, naïve hPSCs mimic mESCs cultured under 2i conditions with respect to their chromatin landscape. This shows a dramatic loss of bivalent domains at >2,000 Polycomb target genes in naïve hPSCs cultured under 5iLAF conditions (Marks et al., 2012; Theunissen et al., 2014). However, there have yet to be studies that examine the genomic distribution and feature enrichment of histone modifications in human embryos. It is therefore not possible to conclude whether this phenomenon exists *in-vivo*.

1.4.3 The DNA methylation landscape of human embryos provides a hallmark characteristic of naïve pluripotency

Analogous to mouse embryo development, there are substantial genome-wide modifications to the DNA methylation landscape during human embryogenesis (Guo et al., 2014; Okae et al., 2014; Smith et al., 2014). CpG DNA methylation levels decline upon fertilisation, reaching their lowest levels in the human ICM. Upon implantation, the genome is globally remethylated and post-implantation fetal tissue samples are hypermethylated, exhibiting >70% methylation levels (Figure 1.6). DNA hypomethylation is therefore a conserved hallmark of naïve pluripotency in both mouse and humans. Comparative DNA methylation analysis of conventional human PSCs reveals a striking difference compared to the ICM cells of origin, with methylation levels of ≈80% observed for hPSCs (Okae et al., 2014). Moreover, it was shown that rapid global remethylation occurs when deriving hPSCs from human blastocysts,

indicating that the conventional culture conditions used are not sufficient to apprehend pluripotent cells in their naïve state (Smith et al., 2014).

As for the transcriptional readouts, the magnitude of DNA hypomethylation is variable between the different naïve hPSC lines (Table 1.1). Naïve-like formulations that show minimal transcriptional changes also retain global DNA methylation levels close to 70%, which is the same as primed hPSCs (Gafni et al., 2013; Pastor et al., 2016). Yet those cells maintained under 5iLA and t2iL+PKCi culture conditions show the closest resemblance to DNA methylation levels observed in the embryo (Figure 1.6) (Guo et al., 2017; Pastor et al., 2016; Theunissen et al., 2016). Despite this similarity, there are differences in the distribution of methylated CpG sites. For instance, imprinted control regions maintain methylation levels in the human embryo, but this methylation is lost in naïve hPSCs upon long-term culture (Guo et al., 2017; Pastor et al., 2016; Theunissen et al., 2016). Nevertheless, global DNA hypomethylation appears to be a species-conserved hallmark of naïve pluripotency.

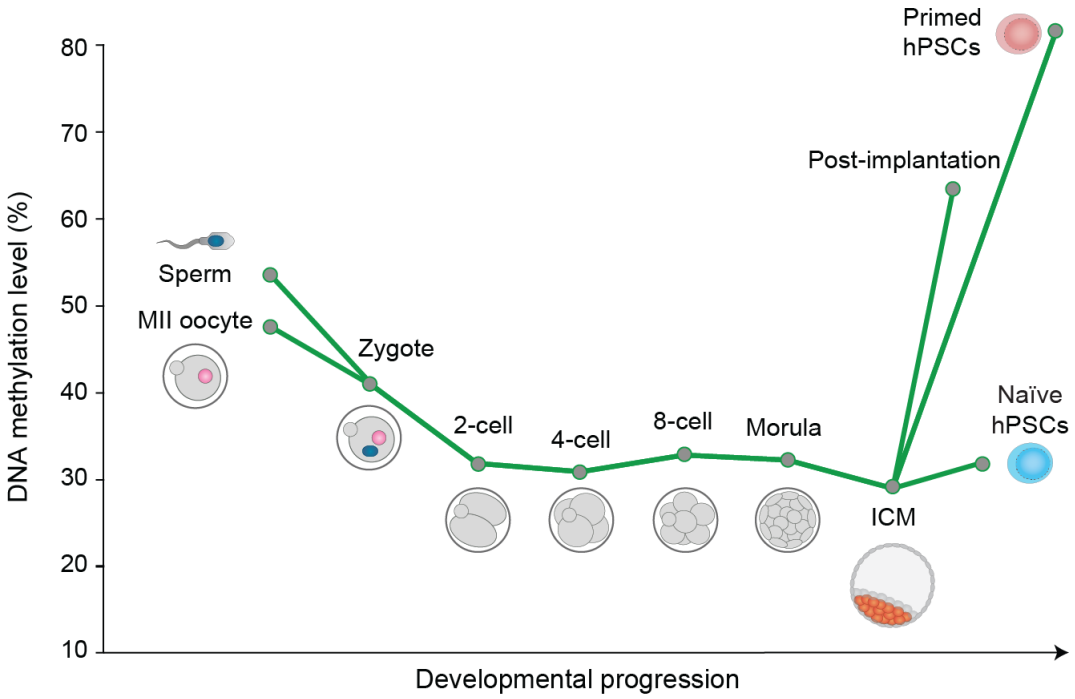


Figure 1.6. DNA methylation distribution across human embryonic development
Global CpG methylation levels are reduced upon fertilisation and the first cell division. The lowest methylation levels are observed within the ICM. Upon implantation, the genome is globally remethylated and post-implantation tissues samples exhibit >70% methylation. Conventional primed hPSCs are hypermethylated (70-80%), whereas naïve hPSCs under 5iLA(F) or t2iL+PKCi culture are hypomethylated (≈30%). Figure adapted from (Guo et al., 2014).

1.4.4 X-Chromosome activity is a hallmark of naive human pluripotency

Whilst the DNA methylation dynamics are comparable between mouse and human development, there are divergent strategies to achieve X-chromosome dosage compensation in females (Figure 1.7). Unlike the mouse system, the paternal X-chromosome is not transiently silenced in human epiblast cells. Instead, there appears to be a gradual dampening of the two X-chromosomes to halve the transcriptional output whilst remaining active. The next difference relates to the lncRNA control mechanisms behind X-chromosome inactivation. Human epiblast cells express *XIST* from both chromosomes without triggering inactivation (Petropoulos et al., 2016; Sahakyan et al., 2017; Vallot et al., 2017). Whilst speculative, the proposed mechanisms behind this is via the primate-specific lncRNA *XACT*, which antagonises *XIST* to prevent *XIST*-mediated silencing of the X-chromosome (Vallot et al., 2017). Upon embryo implantation, random X-chromosome inactivation is thought to take place, akin to mouse implantation.

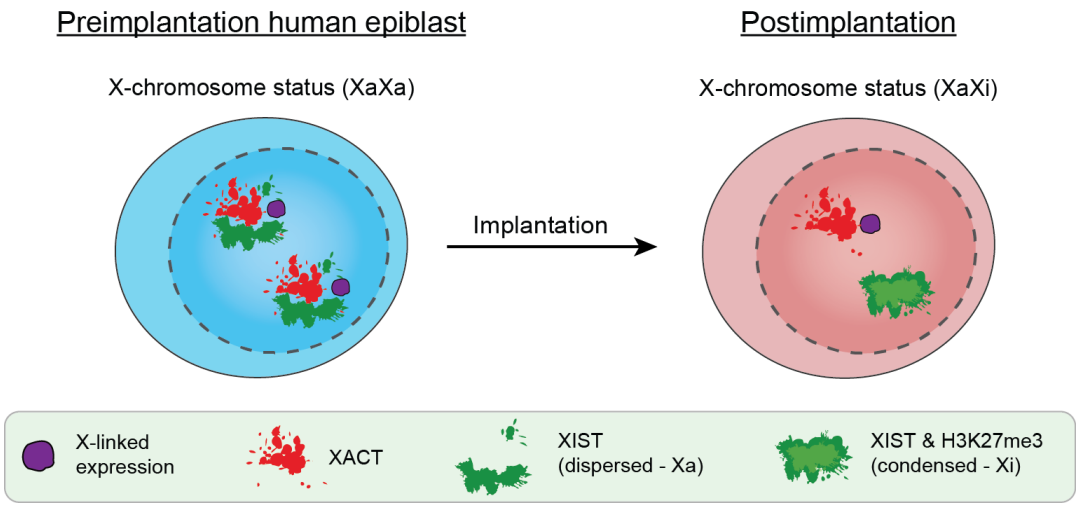


Figure 1.7. Schematic of X-chromosome status during human embryo development.
Preimplantation-stage epiblast cells contain two active X-chromosomes that exhibit biallelic expression and are associated with the lncRNAs *XIST* and *XACT*. After blastocyst implantation, random X-inactivation occurs. The silent X-chromosome condenses and is enriched with H3K27me3, whilst the active X-chromosome ceases to express *XIST*, but retains *XACT* expression.

Conventional hPSCs typically have one active and one inactive X-chromosome in female cells. This infers that upon derivation from the naïve human epiblast, a developmental progression may have occurred as cells exhibit post-implantation characteristics. Reassuringly, the inactive X-chromosome can be reactivated upon reprogramming of primed hPSCs to a naïve state, but only under certain naïve culture conditions (Table 1.2). In addition to naïve hPSCs showing biallelic X-linked gene expression, the lncRNAs *XIST* and *XACT* are expressed from both

chromosomes in a proportion of cells. This recapitulates the observed status of X-chromosome activity and regulation seen in human embryos. Consequently, naïve hPSCs will provide a valuable *in-vitro* model to study X-chromosome silencing in human cells with a more suitable transcriptional and epigenetic landscape that is akin to human epiblast cells.

Several of the initial papers where naïve hPSCs had been generated using female cells, described the reactivation of the previously silenced X-chromosome (Gafni et al., 2013; Ware et al., 2014). This conclusion was based upon the absence of *XIST* expression and a lack of H3K27me3 foci, which at the time was thought to indicate the presence of two active X-chromosomes. Subsequently the experiments that directly examined human embryos revealed that *XIST* and *XACT* are expressed from both chromosomes whilst remaining active. Therefore, the absence of *XIST* or lack of H3K27me3 foci is not a reliable readout of reactivation. Based on these recent discoveries, it is important to define X-chromosome status in hPSCs by evaluating biallelic gene expression for *XIST*, *XACT* and other genes on the X-chromosome that do not escape silencing by erosion. Following this criteria, the status of X-chromosomes in female cells can be used as a reliable hallmark to distinguish naïve pluripotent stem cells.

Table 1.2. X-chromosome status associated with naïve and primed hPSCs.
Note that X-chromosome status was not reported by: (Chan et al., 2013; Duggal et al., 2015; Qin et al., 2016). In some cases not all X-chromosome attributes were examined. Table from (Collier and Rugg-Gunn, 2018).

X-chromosome status	**Naïve state**	**Primed state**	OKK+ 2iL (Hanna et al., 2010)	NHSM (Gafni et al., 2013)	HDACi 2i+F (Ware et al., 2014)	5iLA(F) (Theunissen et al., 2014) (Theunissen et al., 2016)	NK2 t2iL+PKCi (Takashima et al., 2014)	2iL+ STAT3 (Chen et al., 2015)	HDACi t2iL+PKCi (Guo et al., 2017)
X-linked gene expression	Biallelic	Monoallelic from Xa							
XIST expression	Biallelic	Monoallelic from Xi							
XACT expression	Biallelic	Monoallelic from Xa							
H3K27me3 foci	Absence	Enrichment over Xi				on Xa			

1.5 Developing culture conditions to sustain hPSCs in a preimplantation-like state

The ability to derive and maintain naïve human PSCs *in-vitro* would provide a valuable tool to investigate pluripotent state transitions and early human development. There are now greater than ten reports that described the methodologies and growth conditions to derive human pluripotent stem cells, which share characteristics comparable to the epiblast cells of human preimplantation-stage embryos. The predominant approach used to generate naïve hPSCs is by reprogramming primed hPSCs. The cell types generated have collectively been termed 'naïve' hPSCs, even though some fulfil more of the hallmark characteristics of naïve pluripotency than others (Collier and Rugg-Gunn, 2018). Many of these attempts to derive and characterise naïve hPSCs were based upon the knowledge of mouse studies, which may account for the spectrum of pluripotent states that have been reported (Table 1.1).

1.5.1 Varying growth conditions induce a spectrum of human pluripotent states

The first studies that derived naïve-like hPSCs were published prior to the reports that define the molecular characteristics of human preimplantation-stage embryos. Consequently, cell types with features akin to naïve mESCs were initially derived (Buecker et al., 2010; Hanna et al., 2010; Li et al., 2009b). Whilst the cell types generated qualify as pluripotent, assessed by teratoma formation and *in-vitro* differentiation, only Hanna and colleagues induced a naïve-specific feature – the preferential use of the *OCT4* distal enhancer. These cells were derived by the ectopic expression of three pluripotency factors (*OCT4*, *KLF2*, and *KLF4*), in combination with 2i/LIF. However, upon transgene withdrawal the cells could not be maintained in an undifferentiated state, demonstrating that 2i/LIF alone is not sufficient to maintain hPSCs in a naïve state, contrary to mESCs (Hanna et al., 2010).

To overcome the hurdle of using transgenes to generate naïve hPSCs, multiple reports have subsequently derived stable, transgene-free naïve-like hPSCs (Chan et al., 2013; Chen et al., 2015; Gafni et al., 2013; Ware et al., 2014). The resultant cell lines exhibit several properties of naïve-like hPSCs, such as increased clonogenicity and the upregulation of several naïve-associated transcriptional markers (Chan et al., 2013; Gafni et al., 2013). However, a global shift in transcriptional profile towards a human preimplantation state has not been demonstrated; moreover, the degree of DNA demethylation is negligible where documented and biallelic X-linked expression has not been reported (Chan et al., 2013; Chen et al., 2015; Gafni et al., 2013; Ware et al., 2014).

These four studies exploit 2i/LIF as a foundation with additional supplementation (Table 1.3) (Chan et al., 2013; Chen et al., 2015; Gafni et al., 2013; Ware et al., 2014). Interestingly, the naïve-like hPSCs had transitioned to a state of LIF dependency, as withdrawal or inhibition of LIF signalling resulted in their differentiation or impaired self-renewal (Chan et al., 2013; Gafni et al., 2013; Ware et al., 2014). Chen and colleagues further demonstrate the requirement of LIF signalling by reinforcing STAT3 activation in the presence of 2i/LIF, which facilitated primed to naïve-like reprogramming (Chen et al., 2015). This is contrasting to conventional primed hPSCs that do not require exogenous LIF to maintain their pluripotent state (Dahéron et al., 2004; Humphrey et al., 2004). Naïve-like hPSCs may develop LIF-dependency in order to maintain the expression of naïve-specific transcription factors, *KLF4* and *TFCP2L1* (Martello et al., 2013; Niwa et al., 1998; Ye et al., 2013). These are known to be regulated by LIF/STAT3 signalling in naive mESCs, which appears to be a conserved feature in naïve-like hPSCs (Figure 1.8) (Chan et al., 2013).

Following on from these studies, additional signalling pathways have been targeted in order to derive hPSCs with naïve-like characteristics, including PKA and HIPPO inhibition. However these cells have not been well characterised, particularly for their epigenetic or global transcriptional profiles, and subsequently it is difficult to evaluate where they fall on the spectrum of primed to naïve pluripotency (Duggal et al., 2015; Qin et al., 2016; Zimmerlin et al., 2016). Moreover, a commercially available media has subsequently been developed (RSeT), which is based upon the 'NHSM' formulation, to generate transgene-free naïve-like hPSCs (Gafni et al., 2013). Interestingly, cell lines derived under RSeT conditions exhibit variable transcriptional profiles, with some lines clustering with primed hPSCs, whilst others are more similar to human epiblast cells. The disparity between the RSeT lines is unclear, but it appears that a spectrum of pluripotent states can be generated under these conditions (Kilens et al., 2018). To summarise the naïve-like hPSC lines describe above, it is clear that a progression from the conventional primed hPSC state has been made; yet this collection of naïve culture conditions generates cell types that do not fully satisfy the hallmark characteristics of naïve human pluripotency found within the early embryo.

Table 1.3. Media compositions associated with the induction of naïve hPSCs.

✓, addition to media; *, short-term induction / addition to the culture media; ¥, choice between two components; ✔/✕, optional addition of a component; -, not documented; (1) 1:1 ratio DMEM/F12: Neurobasal, N2B27; (2) KnockOut-DMEM, N2B27; (3) TeSR1 (4) DMEM, 20% KSR.

Chemical inhibitor, growth factor	**Target effect**	OKK+ 2iL (Hanna et al., 2010)	NHSM (Gafni et al., 2013)	3iL (Chan et al., 2013)	HDACi 2i+F (Ware et al., 2014)	5iLA(F) (Theunissen et al., 2014) (Theunissen et al., 2016)	NK2 t2iL+PKCi (Takashima et al., 2014)	2iL+ STAT3 (Chen et al., 2015)	2iL+ F,FK,AA (Duggal et al., 2015)	2iL+ FK,YAP (Qin et al., 2016)	HDACi t2iL+PKCi (Guo et al., 2017)
LIF	LIF signalling	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
PD0325901	MEK inhibition	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
CHIR99021	GSK3 inhibition	✔	✔		✔		✔	✔	✔	✔	✔
IM-12	GSK3 inhibition					✔					
BIO	GSK3 inhibition			✔							
Gö6983	PKC inhibition		✔				✔				✔
Y-27632	ROCK inhibition		✔			✔					✔
WH-4-023	SRC inhibition					✔					
SB590885	RAF inhibition					✔					
SP600125	JNK inhibition		✔								
SB203580	p38/MAPK inhibition		✔								
TGFβ	TGFβ signalling			✔							
Activin A	TGFβ signalling					✔					
FGF	FGF signalling			✔	✔	✔/✕			✔		
SAHA	HDAC inhibition				✔ *						
Sodium Butyrate	HDAC inhibition				✔ *						✔ * ¥
Valproic Acid	HDAC inhibition										✔ * ¥
Dorsomorphin	BMP inhibition			✔							
Forskolin	PKA inhibition								✔	✔	
Ascorbic Acid	Demethylation								✔		
Lysophosphatidic acid	HIPPO inhibition									✔	
Base medium		(1)	(2)	(3)	(4)	(1)	(1)	(1)(4)¥	(4)	(1)(3)	(1)
O_2 level		20%	20% or 5%	-	5%	5%	-	5%	5%	-	5%
Transgenes		*OCT4 KLF2 KLF4*					*NANOG* KLF2**	*STAT3**			

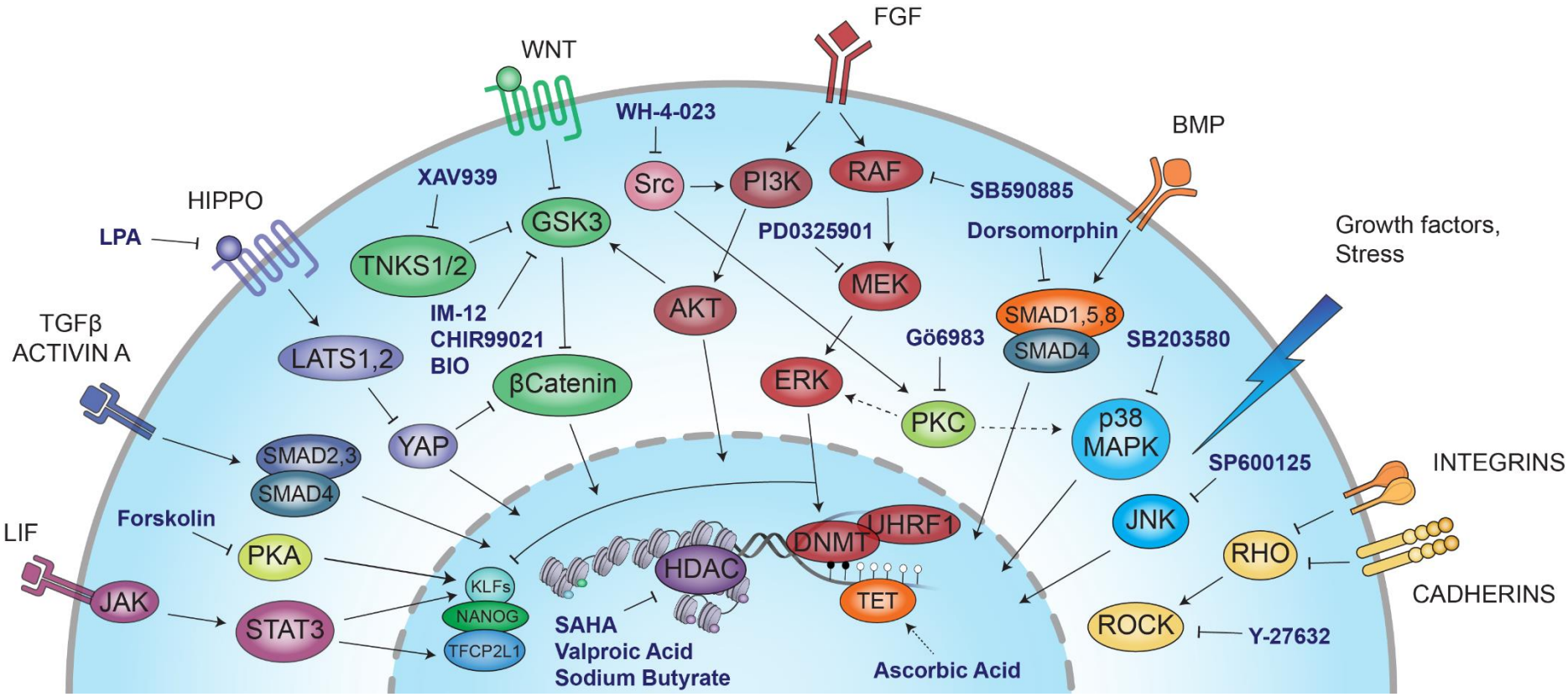


Figure 1.8. Signalling pathways in naïve and primed human pluripotent states

The different cocktails of inhibitors and growth factors used to induce naïve pluripotency are a likely cause of the reported spectrum of pluripotent states. It is important to note that additional pathways that are not highlighted here are also likely to be involved. Moreover, many of the chemical inhibitors used will non-specifically inhibit additional targets, even extending to components of different signalling pathways. Dashed lines indicate likely indirect effects. Figure from (Collier and Rugg-Gunn, 2018).

1.5.2 Identified culture conditions capture naïve hPSCs with preimplantation embryo hallmarks

Two culture conditions stand out as deriving naïve hPSCs with the closest resemblance to human epiblast cells thus far; these are the 5iLA(F) and t2iL+PKCi formulations (Table 1.1 and Table 1.2) (Guo et al., 2017; Guo et al., 2016b; Takashima et al., 2014; Theunissen et al., 2016; Theunissen et al., 2014). In the first instance, a chemical screen was performed to identify kinase inhibitors that could sustain naïve pluripotency in the presence of 2i/LIF upon transgene withdrawal (-DOX), using a doxycycline inducible *NANOG* and *KLF2* system. This identified the components that gave rise to 5iLA(F) media. Alongside classical 2i (dual inhibition of MEK and GSK3), three additional inhibitors targeting ROCK (rho-associated protein kinase), SRC and BRAF were combined with Activin A and the optional addition of FGF2 (Theunissen et al., 2016; Theunissen et al., 2014). Transgene-free naïve hPSCs derived using 5iLA(F) transcriptionally resemble cells of the late-morula and early-blastocyst stage of human and primate development (Huang et al., 2014; Nakamura et al., 2016; Theunissen et al., 2016; Wang et al., 2018).

On the other hand, naïve hPSCs generated using t2iL+PKCi may resemble more closely human ICM cells or later epiblast cells (Huang et al., 2014; Nakamura et al., 2016; Takashima et al., 2014; Theunissen et al., 2016; Wang et al., 2018). The t2iL+PKCi formulation consists of titrated levels of classical 2i, in addition to LIF and a PKC inhibitor. This combination is able to maintain naïve hPSCs that have been generated by *NANOG* and *KLF2*–induced, *KLF4*–induced, or *OCT4*, *KLF4*, *SOX2* and *MYC*-induced primed to naïve reprogramming (Kilens et al., 2018; Liu et al., 2017; Takashima et al., 2014). Alternatively, t2iL+PKCi has been used in combination with transient HDAC inhibitor treatment, termed chemical resetting, to generate transgene-free naïve hPSCs (Guo et al., 2017), and finally t2iL+PKCi with the further addition of ROCK inhibitor (t2iL+PKCi+Y) can be used to capture naïve hPSCs directly from human preimplantation blastocysts (Guo et al., 2016b).

Collectively, both 5iLA and t2iL+PKCi cultured cells also exhibit an epigenetic profile that best represents the human epiblast, including a globally hypomethylated genome and two active X-chromosomes in female cells (Table 1.1 and Table 1.2). However, there are drawbacks with both of these culture conditions. Their long-term culture (≥ ten passages) results in the irreversible loss of DNA methylation at the majority of imprinted loci (Guo et al., 2017; Pastor et al., 2016; Theunissen et al., 2016), and frequent karyotype abnormalities have been reported using 5iLA(F) and t2iL+PKCi+Y (Liu et al., 2017; Theunissen et al., 2016). Nonetheless, these

two formulations have captured cells in a state that closely resembles many aspects of human preimplantation epiblast cells. This therefore provide an expandable *in-vitro* model to functionally explore the role of naïve-specific transcription factors, the mechanisms that govern X-chromosome inactivation in a context with an appropriate epigenetic landscape, and lastly to explore the reprogramming process itself to understand pluripotent state transitions.

1.6 Methods to identify and isolate naïve human pluripotent stem cells

The predominant approach to derive naïve hPSCs is by the reprogramming of primed hPSCs. The efficiency of this method is often low and variable between protocols used, typically generating a heterogeneous population of cells during the reprogramming process. The most common approach to purify naïve hPSCs is via continued passaging of the cultures over time (Collier and Rugg-Gunn, 2018). This bulk passaging approach is dependent on reprogrammed naïve hPSCs having a competitive advantage compared to populations of cells that are refractory to reprogramming. In cases where the culture conditions permit the survival of primed and non-reprogrammed cell types, this approach is not suitable. Furthermore, this method does not easily allow the heterogeneous early stages during reprogramming to be molecularly characterised, unless single-cell approaches are used (Collier and Rugg-Gunn, 2018).

1.6.1 Transgene reporter systems for the identification and isolation of naïve hPSCs

Moving away from the bulk-passaging approach, several alternative isolation strategies have been devised to prospectively isolate naïve hPSCs, by capitalising on aforementioned knowledge of the molecular hallmarks of naïve pluripotency. A key feature of naïve mESCs is the preferential use of the *Oct4* distal enhancer, whereas primed mEpiSCs utilise a proximal enhancer (Brons et al., 2007; Choi et al., 2016; Tesar et al., 2007; Yeom et al., 1996). Several groups have subsequently devised a reporter system that exploits this pluripotent state-specific *cis*-regulatory mechanism. In cell lines that contain a deletion of the *OCT4* proximal enhancer, endogenous *OCT4* expression is coupled to *GFP* (OCT4-ΔPE-GFP), and therefore GFP expression acts as a readout of alternative enhancer activation (Gafni et al., 2013; Theunissen et al., 2014). This can be detected using flow cytometry and permits live cell isolation by fluorescence-activated cell sorting (FACS) (Figure 1.9).

Along a similar avenue, an alternative reporter system (EOS-GFP) has been used to identify naïve hPSCs (Takashima et al., 2014). EOS-GFP expression is driven from a mouse early transposon (ETn) promoter, combined with mouse *Oct4* enhancer elements that contain OCT4 and SOX2 binding sites (Hotta et al., 2009). Given that EOS-GFP expression is activated upon reprogramming of primed hPSCs to a naïve state, this reporter system could also be used to isolate naïve hPSCs. However, the timing of induction and the stringency of both EOS-GFP and OCT4-ΔPE-GFP expression have not been examined, and may therefore isolate a broader population than just naïve hPSCs.

Adopting a similar strategy, naïve-like hPSCs have been isolated using a reporter for the endogenous retrovirus *HERVH* expression, tagged with GFP (Wang et al., 2014a). Primed hPSCs expressing high levels of HERVH-GFP were isolated using FACS. These cells adopt a naïve-like morphology compared to HERVH-low/negative cells. Nonetheless, repeated re-sorting is required to maintain homogeneous naïve-like hPSCs (Wang et al., 2016). Given that endogenous *HERVH* transcription is associated with primed pluripotency, and not with naïve, this reporter system may not be the most appropriate (Theunissen et al., 2016). Moreover, HERVH-high cells only display a moderate increase in naïve associated transcripts, further questioning the ability to sort naïve hPSCs exclusively on this property (Wang et al., 2014a).

Although reporter systems can provide a simple and reliable readout of cell state, they all possess several drawbacks. Firstly, there is the requirement to genetically modify each primed hPSC line prior to the naïve reprogramming process. Secondly, in each case described above the timing and fidelity of each transgene reporter system has not been examined. These reporters may not necessarily be linked exclusively with the transition to a naïve state. Whilst the OCT4-ΔPE-GFP reporter system has been employed by multiple groups to identify naïve hPSCs, this is reliant on the assumption that *OCT4* enhancer usage is conserved from mouse to human, which may not be the case (Pastor et al., 2018).

1.6.2 Cell surface epitopes can distinguish primed and naïve hPSCs

To overcome the limitations associated with genetically engineered reporter systems, cell surface antibodies have subsequently been identified that target epitopes expressed on primed, but not naïve, hPSCs. The technique of immunophenotyping, which involves cell type identification based upon protein marker expression, has been extensively used in the field of immunology (Morilla et al., 2017). Applying this same logic, cell surface markers are routinely used to identify primed hPSCs. These markers include the glycosylated epitopes SSEA-3,

SSEA-4, TRA-1-60 and TRA-1-81 (Andrews et al., 1984; Henderson et al., 2002). Using cell surface markers to identify pluripotent cell types has a number of advantages over reporter systems. Cell surface markers enable live cells to be isolated using FACS or magnetic-activated cell sorting (MACS), without any genetic manipulation of each cell line. This straightforward, quick and routine technique is widely accessible. Moreover, antibodies against glycosylated epitopes are extensively used to isolate conventional hPSCs, which reporter systems may not be able to capture.

The first study examined the expression of CD24, which is a sialoglycoprotein (Shakiba et al., 2015b). CD24 expression was shown to be higher in both primed hPSCs and mEpiSCs, compared to NHSM cultured naïve-like hPSCs and mESCs respectively. Antibodies against primed-specific CD24 was used in combination with the pan-hPSC antigen TRA-1-60 to isolate naïve-like hPSC populations: TRA-1-60^{+}/CD24^{low} and TRA-1-60+/CD24^{high}. A transcriptional comparison between the two populations revealed that the CD24^{low} population showed an enrichment of naïve-associated transcripts compared to the CD24^{high} population. However, it took ten passages for the CD24^{low} population to arise during NHSM-induced primed to naïve reprogramming (Shakiba et al., 2015b). This may therefore limit the ability of CD24 to isolate naïve hPSCs early during the reprogramming process.

The second study again examined the expression of an individual surface marker, SSEA-4, which is routinely used to isolated primed hPSCs (Pastor et al., 2016). The author's noticed that established naïve hPSCs showed morphological heterogeneity when cultured under 5iLAF conditions, as well as variable SSEA-4 expression. Isolation of the SSEA-4 positive and negative fractions revealed that naïve cells lacking SSEA4 formed dome-shaped colonies upon their return to culture, whereas the SSEA-4 positive fraction adopt a primed morphology. A transcriptional comparison between the two fractions also revealed that human epiblast-specific genes were upregulated in the SSEA-4 negative fraction. Furthermore, SSEA4-negative cells were globally hypomethylated (≈30%), compared to the positive fraction that displayed elevated levels (≈60%). Taken together, this shows that the lack of SSEA-4 expression can be used to purify naïve hPSC cultures to enrich for cells harbouring hallmark naïve characteristics (Pastor et al., 2016). However, the dynamics of SSEA-4 expression were not examined, so it is unclear as to whether this marker could also be used to isolate naïve hPSCs during the reprogramming process.

The third report developed a cohort of new monoclonal antibodies (O'Brien et al., 2017). Those targeting CDH3, GPR64, NLGN4X and PCDH1, showed a higher reactivity against primed hPSCs compared to naïve hPSCs derived under 5iLAF conditions. Similar to the Pastor et al. study, 5iLAF cultures were heterogeneous in their surface marker expression profiles, with 30 – 50% of the naïve population expressing the four primed-enriched markers (O'Brien et al., 2017). The heterogeneity observed indicates that the 5iLAF culture conditions permit the survival of non-naïve cell types. Unfortunately, it is not known whether naïve hPSCs remain in a homogeneous state when purified using any of the surface markers described in these three studies (O'Brien et al., 2017; Pastor et al., 2016; Shakiba et al., 2015a) Moreover, the ability to unambiguously classify naïve hPSCs within a mixed population requires naïve-specific surface markers, which have yet to be identified.

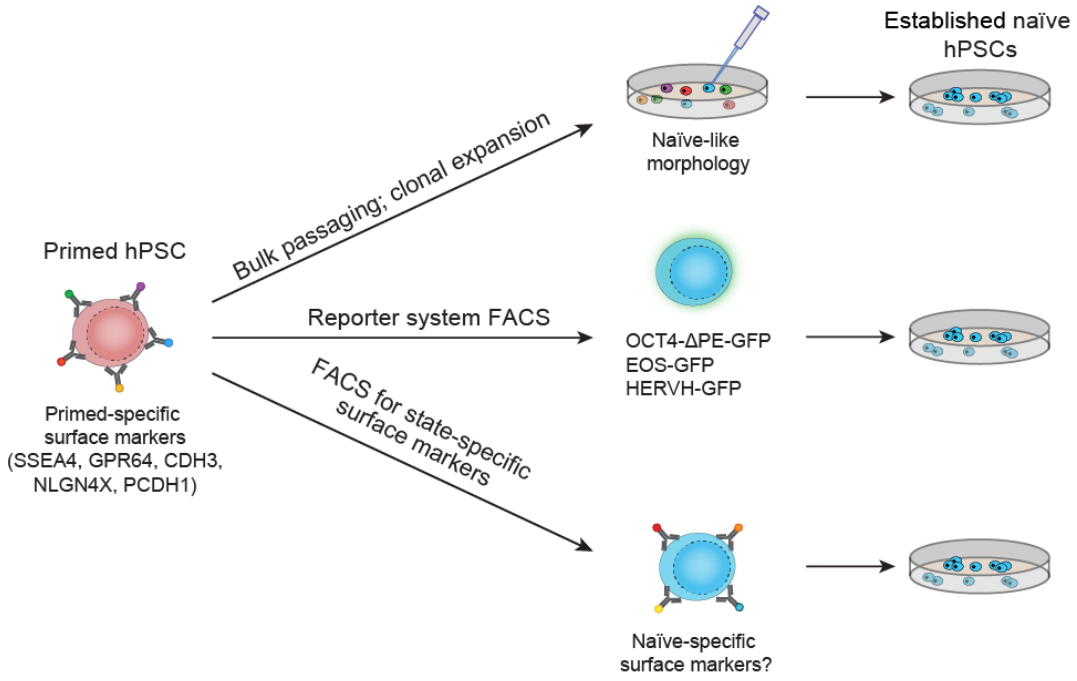


Figure 1.9. Approaches to isolate naïve hPSCs during primed to naïve reprogramming.
Primed to naïve reprogramming can be inefficient, often generating a heterogeneous population of cells during the process. Strategies to isolate and derive defined populations of naïve hPSCs include: (1) Bulk passaging and progressive selection, or clonal expansion of individual colonies that exhibit a domed naïve-like morphology. (2) Reporter systems coupled to the expression of fluorescent proteins provide a read out of characteristics that are associated with naïve pluripotency. (3) Characterisation of the cell-surface proteins expressed on naïve and primed hPSCs has identified primed state-specific markers. Using antibodies specific to these markers enables the isolation of naïve hPSCs that lack primed-specific marker expression using fluorescence-activated cell sorting (FACS). Figure from (Collier and Rugg-Gunn, 2018).

1.7 Examining the routes to pluripotency

Much of the knowledge that we currently have regarding the naïve and primed states of human pluripotency is based upon studies of the established states. Whilst this provides a valuable insight into the hallmark characteristics of the two cell types, very little is known about the transitional period or how the naïve state is acquired. We know there is a widespread remodelling of the epigenetic landscape and rewiring of the gene regulatory networks that govern pluripotency, yet the details of how these events occur and in what order were unclear at the onset of this work.

Isolating and characterising reprogramming intermediate cell types can provide a useful insight into the trajectories of cell fate changes, and the mechanisms that govern the routes taken. By studying the intermediate populations that are refractory to reprogramming, we can begin to understand why some cells fail to reprogram whilst others succeed, and subsequently use this knowledge to coerce cells down the right trajectory. The inefficient and protracted nature of iPSC reprogramming is a good example of where studying intermediate populations has helped to identify the molecular barriers to reprogramming, and subsequently how to overcome these barriers (Figure 1.10). The studies that have examined iPSC reprogramming populations typically use cell-surface markers to track and isolate the populations for subsequent molecular characterisation. Generally, the surface markers are expressed on the somatic cell of origin, such as fibroblast markers Thy1 (CD90) and CD44 (Nefzger et al., 2014; O'Malley et al., 2013). Conversely, surface markers to identify mouse iPSCs include ICAM1 and epithelial markers EpCAM and SSEA-1 (Mikkelsen et al., 2008; O'Malley et al., 2013; Stadtfeld et al., 2008), whereas SSEA-3, SSEA-4, TRA-1-60, TRA-1-81 and CD30 can be used to positively identify human iPSCs (Abujarour et al., 2013; Cacchiarelli et al., 2015; Takahashi et al., 2014; Tanabe et al., 2013). However, partially reprogrammed populations can still express iPSC markers; therefore, additional surface markers that change during the transitional period are useful to delineate cells that are prone to fully reprogram from those that are not. Using mass-cytometry, the surface markers CD73, CD49d and CD200 were identified that are absent in both fibroblasts and iPSCs, but are transiently induced in cells that are prone to fully reprogram (Lujan et al., 2015).

Using surface markers to study reprogramming populations has helped to reveal that iPSC reprogramming is a stepwise process involving the dramatic remodeling of the transcriptional and epigenetic landscapes (Buganim et al., 2012; Cacchiarelli et al., 2015; Hansson et al., 2012;

Knaupp et al., 2017; Lujan et al., 2015; Polo et al., 2012; Zunder et al., 2015). The first wave of transcriptional changes relates to the loss of somatic cell identity, which is universally accomplished to varying degrees; however, cells that are refractory to reprogramming fail to fully repress the somatic identity and do not execute the second transcriptional wave, which involves acquiring a pluripotent transcriptional identity (Mikkelsen et al., 2008; Polo et al., 2012). By identifying and modulating the roadblocks to iPSC reprogramming, refractory populations can be coerced towards an iPSC state (Ebrahimi, 2015) (Figure 1.10). This methodology and knowledge can be subsequently applied to facilitate alternative cell fate reprogramming transitions.

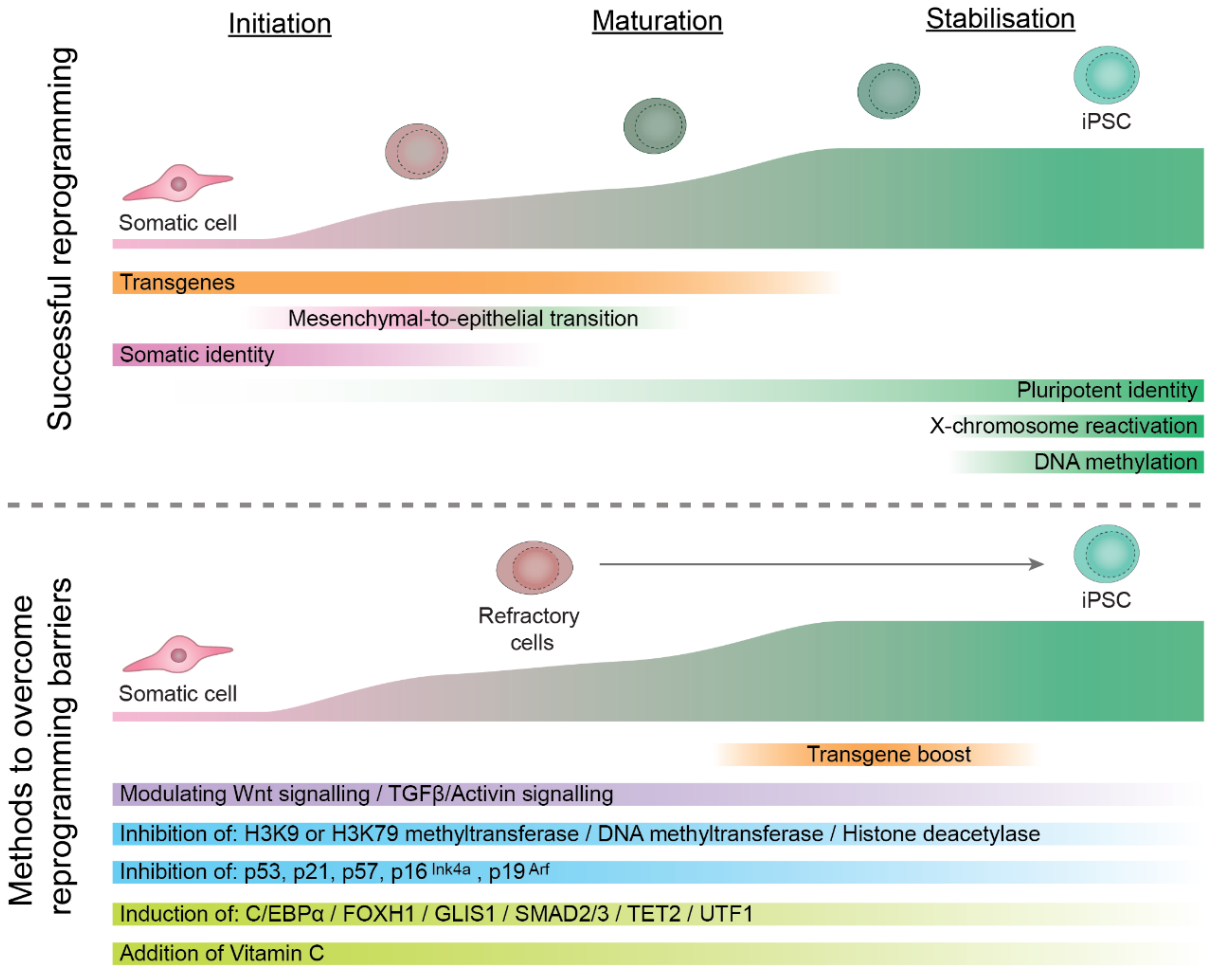


Figure 1.10. Molecular events that occur during somatic cell reprogramming.
Somatic iPSC reprogramming is accompanied by waves of transcriptional changes and remodeling of the epigenetic landscape. To enhance the reprogramming efficiency, different strategies have been devised to overcome the barriers of reprogramming and drive refractory populations to an iPSC state. Top figure panel adapted from (David and Polo, 2014).

Most of the studies that generate human iPSCs have yielded cells that exhibit primed hPSC characteristics. There are now several reports whereby naïve hPSCs have been derived using iPSC reprogramming (Kilens et al., 2018; Liu et al., 2017; Wang et al., 2018). This demonstrates that it is possible to revert human somatic cells back to a naïve preimplantation-like state. The authors in one study suggest that this may even be possible without going through a primed phase when using t2iL+PKCi conditions (Kilens et al., 2018). Further analysis would be required to confirm this observation, and to decipher whether skipping the primed state of pluripotency confers any advantage. Interestingly, naïve hiPSCs derived by *OSKM* and cultured in 5iLAF or t2iL+PKCi+Y conditions are virtually indistinguishable from their counterpart naïve hPSCs derived by primed to naïve reprogramming (Liu et al., 2017). Moreover, this study identified that *KLF4* alone could reprogram primed hPSCs to a naïve state under t2iL+PKCi+Y culture. This is reminiscent of mEpiSC to mESC conversion induced by *Klf4* under 2iL conditions, which indicates a species-conserved role of *KLF4* in the induction of naïve pluripotency (Liu et al., 2017).

As previously mentioned, mEpiSC to mESC conversion is possible at low efficiencies when naïve pluripotency factors are introduced. Testing whether the exogenous addition of specific transcription factors can facilitate mEpiSC to mESC conversion has identified essential transcription factors required for the reprogramming process. However, an in-depth analysis of the reprogramming trajectory or a systematic identification of naïve promoting factors has yet to be performed. The same is true with human primed to naive reprogramming, given that the current methods used to identify and isolate naïve hPSCs preclude the study of the early stages of the reprogramming process. Without a means to examine the earliest stages of reprogramming, it is not possible to decipher the molecular trajectories of primed to naïve reprogramming, and the mediators of this process. These are areas that I aim to address during this thesis.

1.8 Thesis aims

Naïve hPSCs are predominantly generated via the reprogramming of primed hPSCs. The efficiency of this method is often low and variable between naïve reprogramming protocols used, thereby generating a heterogeneous mix of cells at the early stages. Common approaches to purify naïve hPSCs rely on the continuous passaging of the cultures over time, reporter systems coupled to naïve state attributes, or primed-enriched surface markers. However, the current reporter systems and surface markers have not been examined for their fidelity, nor

whether the timing of induction would permit naïve hPSCs to be isolated early during reprogramming. Consequently, our current understanding of primed and naïve human pluripotent states is based upon long-term *in-vitro* cultured naïve hPSCs.

Whilst studying the established primed and naïve states has identified reliable transcriptional markers to distinguish the two cell types, we know very little about the transitional period, the transcription factors required, or the route(s) taken between the two states. This also applies to our knowledge of how the epigenetic landscape is remodelled; we know that naïve hPSCs undergo X-chromosome reactivation and a genome-wide remodelling of the DNA methylation landscape occurs, however much is unknown about the timing and regulation of these events.

Establishing a method to study cell populations as they transition from a primed to a naïve state would provide important new insights into the order of molecular changes and their possible regulatory mechanisms. Given that many of these molecular changes also occur during other cellular transitions or at the onset of disease, and may share similar pathways and targets, understanding primed to naïve transitioning may have wider implications. In order to accurately track the early stages of reprogramming, and the emergence of nascent naïve hPSCs, a new approach must be taken.

To this end, I sought to identify cell surface markers capable of positively distinguishing both naïve hPSCs and primed hPSCs, and most importantly nascent naïve hPSCs early during reprogramming. The following aims form the contents of this thesis:

I. To produce a comprehensive resource of cell surface markers expressed on naïve and primed human pluripotent stem cells. Using this resource, identify and validate pluripotent state-specific surface markers using different naïve reprogramming methods to ensure their fidelity.
II. Formulate a multiplexed antibody panel that is able to track the dynamics of primed to naïve reprogramming.
III. Using the multiplexed antibody panel, isolate and characterise nascent naïve hPSCs early during reprogramming and track their subsequent maturation.
IV. Perform transcriptional and epigenomic profiling of nascent naïve hPSCs throughout primed to naïve reprogramming to decipher the order or molecular changes.
V. Perform a comparative transcriptional analysis of nascent naïve hPSCs at an early time point using different reprogramming methods, in order to understand the route(s) to naïve pluripotency

2. Methods

2.1 Cell Culture Reagents

2.1.1 Cell Lines

All hPSC lines used in this study meet the 'ethically sourced' criterion (as defined by the National Institute for Health Registry and International Stem Cell Forum) and all work was carried out with appropriate approval from the UK Stem Cell Bank Steering Committee and the Babraham Institute Health and Safety and Human Tissue Ethics Committees.

H9 (primed)

Human pluripotent stem cell line (WA09) (WiCell Research Institute). (Thomson et al., 1998)

H9 NK2 (primed)

Derived from H9 hPSCs with the inclusion of a doxycycline inducible *NANOG* and *KLF2* construct. Obtained from Austin Smith (WT-MRC Cambridge Stem Cell Institute) with permission from WiCell. (Takashima et al., 2014).

H9 NK2 (naïve)

Derived from H9 NK2 primed hPSCs after transgene induced resetting cultured in t2iL+PKCi media. Obtained from Austin Smith (WT-MRC Cambridge Stem Cell Institute) with permission from WiCell. (Takashima et al., 2014).

FiPSC (primed)

FiPSCs were derived from fibroblasts by iPSC reprogramming and cultured in primed (KSR) hPSC media. Obtained from Austin Smith (WT-MRC Cambridge Stem Cell Institute). (Takashima et al., 2014).

FiPSC (naïve)

Naïve FiPSCs were derived from primed FiPSCs and cultured in t2iL+PKCi medium. Obtained from Austin Smith (WT-MRC Cambridge Stem Cell Institute). (Takashima et al., 2014)

WIBR3 OCT4-ΔPE-GFP (primed)

Derived from WIBR3 hPSCs containing an OCT4-GFP reporter on the same allele that has had the OCT4 proximal enhancer (PE) removed. OCT4-GFP expression is subsequently indicative of distal enhancer activity. Obtained from Rudolph Jaenisch (Whitehead Institute for Biomedical Research). (Hockemeyer et al., 2011)

WIBR3 OCT4-ΔPE-GFP (naïve)

Derived from WIBR3 OCT4-ΔPE-GFP primed hPSCs after primed-to-naïve reprogramming using 5iLAF. Obtained from Rudolph Jaenisch (Whitehead Institute for Biomedical Research). (Theunissen et al., 2014)

2.1.2 Cell Culture Media

Mouse embryonic fibroblast (MEF) media:

DMEM (ThermoFisher, 41965039)
10% Fetal Bovine Serum (FBS) (Sigma-Aldrich, F7524)
2mM L-Glutamine (ThermoFisher, 25030024)
50U/ml Penicillin-Streptomycin (ThermoFisher, 15140122)
0.1mM β-mercaptoethanol (ThermoFisher, 31350010)
1mM Sodium Pyruvate (ThermoFisher, 11360070)
1X MEM Non-Essential Amino Acids (ThermoFisher, 11140035)

Primed (KSR) hPSC medium:

Advanced DMEM (ThermoFisher, 12491015)
20% Knockout Serum Replacement (KSR) (ThermoFisher, 10828028)
2mM L-Glutamine (ThermoFisher, 25030024)
50U/ml Penicillin-Streptomycin (ThermoFisher, 15140122)
0.1mM β-mercaptoethanol (ThermoFisher, 31350010)
4ng/ml basic fibroblast growth factor (WT-MRC Cambridge Stem Cell Institute)

N2B27 basal medium for t2iL+PKCi media:

1:1 mixture of DMEM/F12 (ThermoFisher, 21331020) and Neurobasal (ThermoFisher, 21103049)

0.5x N2 supplement (ThermoFisher, 17502048)

0.5x B27 supplement (ThermoFisher, 17504044)

2mM L-Glutamine (ThermoFisher, 25030024)

50U/ml Penicillin-Streptomycin (ThermoFisher, 15140122)

0.1mM β-mercaptoethanol (ThermoFisher, 31350010)

N2B27 basal medium for 5iLA(F) media:

1:1 mixture of DMEM/F12 (ThermoFisher, 21331020) and Neurobasal (ThermoFisher, 21103049)

0.5x N2 supplement (ThermoFisher, 17502048)

0.5x B27 supplement (ThermoFisher, 17504044)

2mM L-Glutamine (ThermoFisher, 25030024)

50U/ml Penicillin-Streptomycin (ThermoFisher, 15140122)

0.1mM β-mercaptoethanol (ThermoFisher, 31350010)

50µg/ml Bovine Serum Albumin (ThermoFisher, 15260037)

0.5% KSR (ThermoFisher, 10828028)

Complete t2iL+PKCi media:

N2B27 basal medium for t2iL+PKCi media

1µM PD0325901 (WT-MRC Cambridge Stem Cell Institute)

1µM CHIR99021 (WT-MRC Cambridge Stem Cell Institute)

2µM Gö6983 (Tocris, 2285)

20ng/ml human LIF (WT-MRC Cambridge Stem Cell Institute)

Complete 5iLA(F) media:

N2B27 basal medium for 5iLA(F) media

1µM PD0325901 (WT-MRC Cambridge Stem Cell Institute)

1µM IM-12 (Cell Guidance Systems, SM04)

1µM SB590885 (Cell Guidance Systems, SM48)

1µM WH-4-023 (Cell Guidance Systems, SM16)

10µM Y-27632 (Cell Guidance Systems, SM02)

20ng/ml human LIF (WT-MRC Cambridge Stem Cell Institute)

20ng/ml Activin A (WT-MRC Cambridge Stem Cell Institute)

8ng/ml basic fibroblast growth factor (WT-MRC Cambridge Stem Cell Institute) (omitted in 5iLA)

2iL+DOX media:

N2B27 basal medium for t2iL+PKCi media

1µM PD0325901 (WT-MRC Cambridge Stem Cell Institute)

1µM CHIR99021 (WT-MRC Cambridge Stem Cell Institute)

1µg/ml Doxycycline (Sigma-Aldrich, D9891)

20ng/ml human LIF (WT-MRC Cambridge Stem Cell Institute)

2.1.3 Additional Cell Culture Reagents

PBS - Dulbecco's PBS w/o Calcium and magnesium (Sigma-Aldrich, D8537)

Reagents for passaging:

200U/ml Collagenase type IV (ThermoFisher Scientific, 17104019) in Advanced DMEM

Accutase (ThermoFisher Scientific, A1110501)

Gentle Cell Dissociation Reagent (GCDR) (StemCell Technologies, 07174) or 0.5mM EDTA (Sigma-Aldrich, 93283) in PBS

Reagents for cell culture surface treatment:

0.1% Gelatin (Sigma-Aldrich, G1890) in PBS

5µg/ml Vitronectin (ThermoFisher Scientific, A14700) in PBS

Matrigel Basement Membrane Matrix, Growth Factor Reduced (Corning, 354230) diluted 1:100 in DMEM/F12

Irradiated MEFs (30 Grays) prepared in-house from active MF1 MEFs (WT-MRC Cambridge Stem Cell Institute).

2.1.4 Human Embryos

Immunofluorescent microscopy was performed on human blastocysts by Fredrik Lanner and colleagues (Karolinska Institute) and these data are presented in Chapter 3.2.3. Human embryos were obtained from the Karolinska University Hospital, Huddinge and from the Carl von Linne´ Clinic, Uppsala, either frozen at embryonic day (E) 2 or from preimplantation genetic diagnosis testing at E4, with informed consent from donating couple and with ethical approval for these experiments to Fredrik Lanner from the Regional Ethics Board, Stockholm (2012/1765-31/1). Frozen embryos were thawed with ThawKit Cleave (Vitrolife) into G-1 Plus medium (Vitrolife) covered with Ovoil (Vitrolife) and from E3 embryos were cultured in G-2 Plus medium until E6-7 in 5% O_2, 5%CO_2 at 37°C.

2.2 Cell Culture Techniques

All cells were cultured at 37°C in humidified incubators at 5% CO_2 and 5% O_2.

2.2.1 Preparation of MEF Plates

For cells cultured on irradiated MEFs, the MEF plates were prepared as follows: Tissue culture plates were coated with 0.1% gelatin for ≥ 1 hour at 37°C. Excess gelatin was aspirated before MEFs were plated. MEFs were plated in MEF medium at a density of $2x10^6$ cells per 6-well plate or 10cm plate.

2.2.2 Preparation of Vitronectin Coated Plates

Primed hPSCs cultured under feeder-free conditions were plated on Vitronectin coated plates. Plates were coated with vitronectin for ≥ 1 hour at room temperature. Excess Vitronectin was aspirated before cells were seeded.

2.2.3 Preparation of Matrigel Coated Plates

Naïve hPSCs cultured under feeder-free conditions were plated onto Matrigel coated plates. Plates were coated with Matrigel for ≥ 1 hour at room temperature or at 4°C overnight. Excess Matrigel was aspirated before cells were seeded.

2.2.4 Primed Human Pluripotent Stem Cell Culture

Primed hPSCs were either maintained in primed hPSC medium on a MEF-layer, or feeder-free on Vitronectin coated plates in either TeSR-E8 (StemCell Technologies, 05990) or mTeSR1 (StemCell Technologies, 85850). Primed hPSCs cultured on MEFs were typically passaged every 4 days at a 1:4 – 1:6 ratio. Primed hPSCs cultured on Vitronectin were typically passaged every 4-6 days at a 1:6 – 1:18 ratio. Media was changed daily.

2.2.5 Passaging Primed Human Pluripotent Stem Cells

Passaging primed hPSCs cultured on MEFs was conducted as follows: Cell culture medium was aspirated and the culture vesseles were washed gently with PBS, followed by incubation with collagenase for 5 minutes at 37°C. Collagenase was aspirated and primed hPSC medium was added. Using a 5ml stripette, cells were manually dissociated from the culture vessels and collected for centrifugation at 200xg for 5 minutes. Medium was aspirated and cell pellets were resuspended in primed hPSC medium and passaged to freshly prepared MEF plates. Cells were passaged every 4 days at a 1:4 – 1:6 ratio.

Passaging primed hPSCs cultured on Vitronectin was conducted as follows: Cell culture medium was aspirated and the culture vesseles were washed gently with PBS, followed by incubation with GCDR for 5 minutes at room temperature. GCDR was aspirated and TeSR-E8 or mTeSR1 was added. Using a 5ml stripette, cells were manually dissociated

from the culture vessels, collected and passaged to freshly coated Vitronectin plates. Cells were passaged every 4-6 days at a 1:6 – 1:18 ratio.

2.2.6 Naïve Human Pluripotent Stem Cell Culture

Naïve hPSCs cultured in t2iL+PKCi medium or 5iLA medium were maintained as previously described (Takashima et al., 2014; Theunissen et al., 2014). Both t2iL+PKCi and 5iLA naive hPSCs were typically maintained on a MEF-layer. Complete cell culture medium was made and used within 2 days of the addition of cytokines and growth factors. Cell culture medium was replaced daily on all cells. For feeder-free culture of t2iL+PKCi naive hPSCs, cells were transfered to Matrigel coated plates for at least 2 passages before any experimental work. Both t2iL+PKCi and 5iLA naive hPSCs were passaged every 3-4 days.

2.2.7 Passaging Naïve Human Pluripotent Stem Cells

Cell culture medium was aspirated and cells were incubated with Accutase for 3-5 minutes at 37°C. Cells were collected for centrifugation and Accutase was diluted at least 1:1 with DMEM/F12 before centrifugation for 3 minutes at 300xg. Medium was aspirated and cell pellets were resuspended in complete medium. t2iL+PKCi naive hPSCs were passaged at a 1:4 – 1:5 ratio, whereas 5iLA naive hPSCs were passaged at a 1:3 ratio.

2.2.8 Freezing Cells

Naive hPSCs were typically harvested from confluent (day 3 or day 4) 6-well plates using Accutase. Approximately $3x10^5$ naive hPSCs (one confluent well of a 6-well plate) were frozen in complete t2iL+PKCi medium containing 10μM Y-27632 and 10% DMSO (Sigma-Aldrich, D2650). Primed hPSCs cultured on MEFs were harvested using Collagenase type IV and approximately $5x10^5$ primed hPSCs were frozen in complete hPSC medium containing 10μM Y-27632 and 10% DMSO. Primed hPSCs cultured under feeder-free conditions were harvested using GCDR and approximately $5x10^5$ primed hPSCs were frozen in complete TeSR-E8 or mTeSR1 medium containing 10μM Y-27632 and 10% DMSO.

2.2.9 Naïve to Primed Differentiation Protocol

For transition to primed hPSCs, t2iL+PKCi-cultured naïve hPSCs maintained on Matrigel were dissociated to single cells with Accutase for 5 minutes. Naïve hPSCs were seeded in t2iL+PKCi medium on Matrigel-coated 6-well plates at a density of $2x10^5$ cells per well. Two days later (termed day 0 of the differentiation protocol), media was changed to complete mTeSR1 medium. Cells were passaged on day 4 using Accutase for the first passage, and Collagenase type IV for subsequent passages.

2.2.10 Primed to t2iL+PKCi Naïve hPSC reprogramming

The original transgene dependent method of primed-to-naïve reprogramming (NK2 t2iL+PKCi), takes approximately 28 days (Figure 2.1A), including two weeks of NANOG and KLF2 induction (Takashima et al., 2014). Takashima and colleagues also present a condensed version of the protocol by reducing NANOG and KLF2 induction down to eight days, followed by seven days culture in t2iL+PKCi, collectively taking ≈17 days before clonal derivation and expansion of naïve hPSC lines (Figure 2.1B). I further optimised this protocol by reducing NANOG and KLF2 induction down to six days, followed by 2 days culture in t2iL+PKCi, before cells were used for downstream analysis (Figure 2.1C).

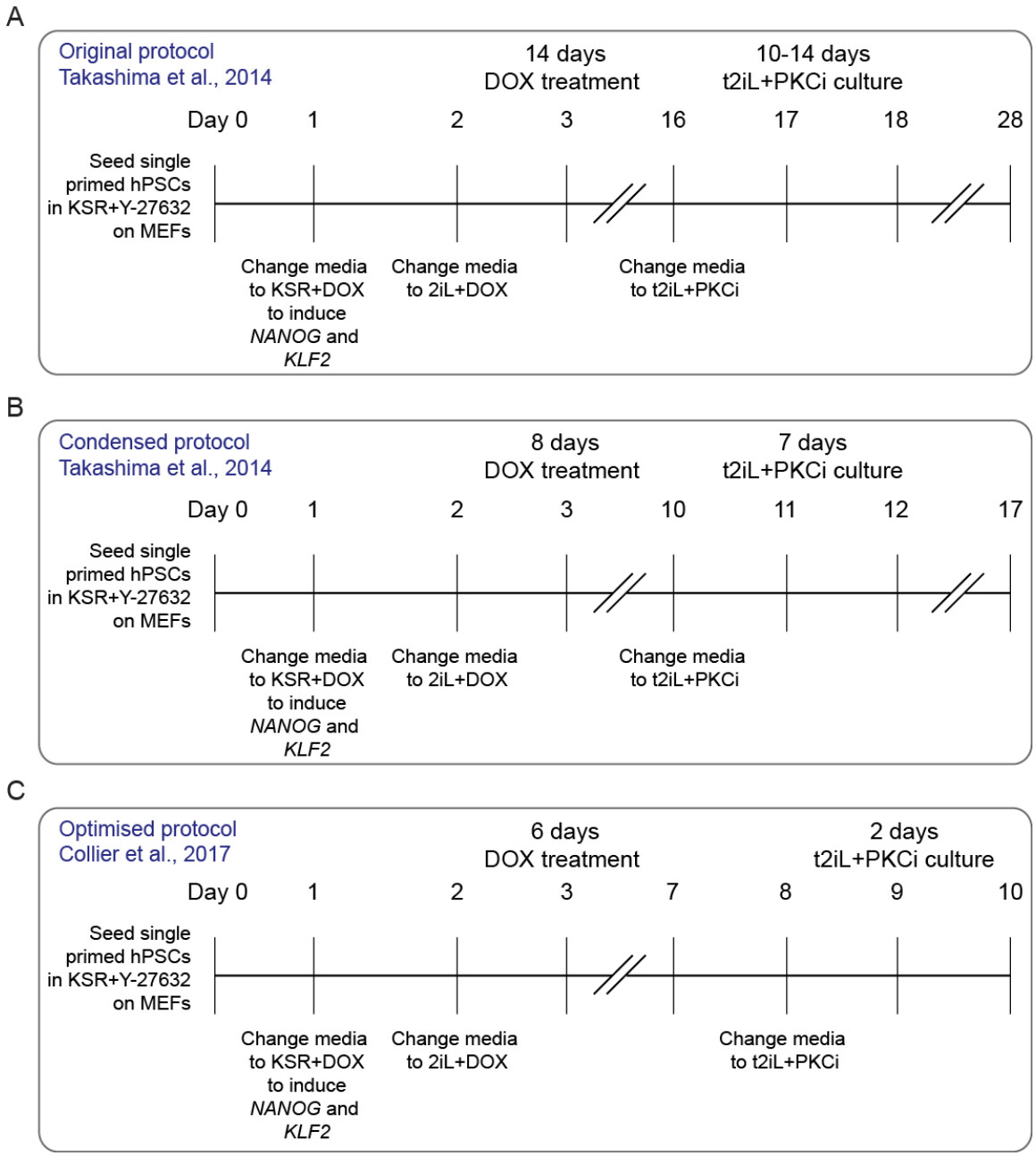


Figure 2.1. Primed to naïve reprogramming protocols for NK2 t2iL+PKCi method.
(A and B) Schematic of published primed-to-naïve NK2 t2iL+PKCi reprogramming protocols (Takashima et al., 2014). Doxycycline-inducible *NANOG* and *KLF2* transgenes activated by doxycycline treatment, followed by maintenance of the naïve state under t2iL+PKCi culture. Culture media replaced daily.
(C) Schematic of further optimised primed-to-naïve NK2 t2iL+PKCi reprogramming protocol used throughout this body of work (Collier et al., 2017).

<u>Optimised NK2 t2iL+PKCi reprogramming protocol:</u>

Primed H9 NK2 hPSCs were dissociated into single cells with Accutase. Primed hPSCs were seeded in primed hPSC medium containing 10µM Y-27632 at a density of $2x10^5$ cells per well of a MEF-coated 6-well plate. The following day (day 1), media was changed to primed PSC media containing 1 µg/ml doxycycline (DOX). On Day 2, media was changed to 2iL+DOX

media and replaced daily until day 8 where media was changed to t2iL+PKCi. Cells were passaged using Accutase on day 3 at a 1:5 ratio and again on day 7 at a 1:3 ratio. Cells were subsequently passaged every 3-5 days with Accutase.

2.2.11 Primed to 5iLA Naïve hPSC reprogramming

Primed hPSCs were dissociated into single cells with Accutase. Primed hPSCs were seeded in primed hPSC medium containing 10μM Y-27632 at a density of $2x10^5$ cells per well of a MEF-coated 6-well plate. The following day (day 1), media was changed to 5iLA media. Dome-shaped naïve colonies appeared after 4 days of culture in 5iLA. Cells were passaged with Accutase on day 5 at a 1:2.5 ratio and again on Day 10. Cells were subsequently passaged every 4 days with Accutase at a 1:2 – 1:3 ratio.

2.2.12 Colony Formation Assay

Colony formation assays were performed to assess the ability of reprogramming populations to form undifferentiated colonies from single cells. $2x10^4$ cells were sorted using Fluorescence Activated Cell Sorting into 15ml falcon tubes. Cells were centrifuged and resuspended in either t2iL+PKCi containing 10μM Y-27632 or 5iLA and transferred into a well of a MEF-coated 6-well plate. After 4-5 days of culture the colony number and morphology of each were scored. Colony formation assays were performed for at least 3 biological replicates.

2.3 Flow Cytometry

All cells used for flow cytometry were dissociated into single cells using Accutase for 5 minutes at 37°C. Cells were washed with PBS containing 2% FBS (flow buffer) and passed through 30μm cell strainers. Conjugated antibodies were mixed in 50μl Brilliant Stain Buffer (BD Biosciences, 563794) and added to 50μl cell suspension containing $1x10^6$ cells. Cells were incubated in the dark for 30 minutes at 4°C. Cells were subsequently washed with flow buffer and centrifuged at 300xg for 3 minutes. Cell pellets were resuspended in either flow buffer alone or with the inclusion of 0.2μg/ml DAPI (Sigma-Aldrich). Where DAPI was not used as a live-dead stain, cells were instead stained with a Fixable Viability Dye-eF780 according to the protocol followed for standard conjugated antibodies.

Single antibody-stained cells or OneComp eBeads (eBioScience, 01-1111-42) were used for compensation calculations. Unstained cells, live human cells and Fluorescence Minus One (FMO) controls were used in cytometer and gating set up.

Table 2.1. Antibodies used for flow cytometry and FACS

Reagent	Conjugation	Source	Volume/assay*
CD130	PE	BD Biosciences, 555756	20µl
CD24	BV650	BD Biosciences, 563720	5µl
CD24	BUV395	BD Biosciences, 563818	5µl
CD57	BV605	BD Biosciences, 563895	5µl
CD57	BV421	BD Biosciences, 563896	5µl
CD7	PE	BioLegend, 343106	1µl
CD7	PE-Cy7	BioLegend, 343114	5µl
CD75	eF660	eBioscience, 50-0759-42	5µl
CD77	PE-CF594	BD Biosciences, 563631	5µl
CD90	BUV395	BD Biosciences, 563804	5µl
CD90	PE-Cy7	BD Biosciences, 561558	5µl
CD90.2 (anti-mouse)	APC-Cy7	BioLegend, 105328	5µl
HLA-ABC	BV711	BD Biosciences, 565333	5µl
Fixable Viability Dye	eF780	eBioscience, 65-0865-18	0.6µl
Donkey anti-mouse	AF568	ThermoFisher Scientific, A10037	1:200
CD229		R&D Systems, MAB1898	2.5µl
Hoechst 33342		ThermoFisher Scientific, H3570	1:10000
DAPI		Sigma-Aldrich, D9542	1:200

*All antibodies used were first titrated. Optimised concentrations are shown for staining $1x10^6$ cells in 100µl staining buffer.

Flow Cytometry analysis was performed using a BD LSRFortessa cell analyser (BD Biosciences). The machine parameters used are shown in Table 2.2.

Table 2.2. Flow Cytometry machine settings used for analyser acquired data.

Laser(s)	Filter(s) BP	Filter(s) LP	Antibody and fluorophore
355	379/28		CD90-BUV395 or CD24-BUV395
488	530/30	505	GFP
405	450/50		CD57-BV421
	660/20	630	CD24-BV650
561	585/15		CD130-PE
	610/20	600	CD77-PE-CF594
	780/60	750	CD7-PE-Cy7
640	670/14		CD75-eF660
	780/60	750	Cd90.2-APC-Cy7 and Fixable Viability Dye eF780

Fluorescence activated cell sorting (FACS) was performed using a BD FACSAria Fusion Sorter (BD Biosciences). The machine parameters used are shown in Table 2.3.

Table 2.3. Flow Cytometry machine settings used for FACS.

Laser(s)	Filter(s) BP	Filter(s) LP	Antibody and fluorophore
355	379/28		CD90-BUV395 or CD24-BUV395
488	530/30	502	GFP
405	450/50		CD57-BV421
	670/30	630	CD24-BV650
561	582/15		CD130-PE
	610/20	600	CD77-PE-CF594
	780/60	735	CD7-PE-Cy7
640	670/14		CD75-eF660
	780/60	765	Cd90.2-APC-Cy7 and Fixable Viability Dye eF780

2.3.1 High-Throughput Flow Cytometry Antibody Screening (BioLegend LEGENDScreen)

H9 NK2 primed hPSCs and H9 NK2 t2iL+PKCi naïve hPSCs cultured under feeder-free conditions were dissociated into single cells with Accutase. Cells were washed and passed through 30µm cell strainers. Cell concentrations were adjusted to $1.8x10^6$ cells/ml in a total volume of 30ml. 75µl containing $1.5x10^5$ cells was aliquoted into each well of a 96 well LEGENDScreen plate (BioLegend, 700001), containing 25µl reconstituted PE-conjugated antibody. Cells were incubated for 20 minutes in the dark at 4°C. Plates were washed twice with 200µl Cell Staining Buffer (BioLegend) per well. Cells were resuspended in 160µl Cell Staining Buffer containing 1µg/ml DAPI used as a live-dead stain. Samples were analysed using a BD LSRFortessa cell analyser (BD Biosciences) containing a high-throughput sampler. The BioLegend LEGENDScreen was performed in biological duplicate for both primed and naïve hPSCs.

2.3.2 Fluorescence Activated Cell Sorting (FACS)

Cells cultured in t2iL+PKCi were sorted directly into 15ml falcon tubes containing t2iL+PKCi medium containing 10µM Y-27632. Cells were washed and transferred onto MEF-coated plates in ti2L+PCKi+ Y-27632. After 12-24 hours medium was replaced with t2iL+PKCi medium without Y-27632. When sorting cells cultured in t2iL+PKCi medium a strong autofluorescent signal in several channels occurs, therefore unstained H9 NK2 naïve hPSCs were used to setup the flow cytometers in order to provide an accurate autofluorescent baseline.

Cells cultured in 5iLA were sorted directly into 15ml falcon tubes containing 5iLAF medium. Cells were washed and transferred onto MEF-coated plates in 5iLAF. After two passages, cells

sorted into 5iLAF were transferred back to 5iLA. The addition of FGF2 and performing two-third media changes for the first two passages improved initial cell expansion.

2.4 Flow Cytometry Data Analysis

2.4.1 Flow Cytometry Data Quality Control Analysis (FlowAI)

FCS files were run through FlowAI V1.2.4, a quality control package that filters events affected by technical variation, such as abrupt flow rate fluctuations (Monaco et al., 2016). Gating was performed in FlowJo V10.1 (FlowJo, LLC) to examine the data and to export live, human cells for subsequent FlowSOM analysis.

2.4.2 FlowSOM Analysis

Flow cytometry data was analysed using FlowSOM (V1.2.0), an R bioconductor package that uses self-organising maps for dimensional reduction visualisation of flow data (Gassen et al., 2015). Data was scaled and logicle transformed on import. Cells were assigned to a Self-Organising Map (SOM) with a 10x10 grid, which groups similar cells into 100 different clusters. To visualise similar clusters a minimal spanning tree (MST) was constructed and cell counts were log scaled. For the time course experiments, live, human-gated cell populations from each day were exported. To account for cell number variation, an equal number of cells were exported for each time point. The same analysis was performed for the comparison of naïve and primed hPSCs. For the analysis of cell sorted samples, the N4+, N3+ and N4- populations were additionally exported and visualised on the MST constructed using day 10 live, human cells.

The FlowSOM scripts used are publically available:

https://github.com/peterch405/Collier-et-al.-2017

2.5 Molecular Biology

2.5.1 RNA extraction and cDNA synthesis

RNA for RT-qPCR analysis was extracted using the RNeasy Mini kit (QIAGEN, 74106). 1µg RNA was reverse transcribed using either SuperScript III (ThermoFisher Scientific, 11732020) or QuantiTect Reverse Transcription Kit (QIAGEN, 205311).

RNA for RNA-sequencing was extracted by lysing cells in TRIzol LS Reagent (ThermoFisher Scientific, 10296028) and performing phenol-chloroform RNA extraction according to the TRIzol LS protocol for isolation of RNA from Small Quantities.

2.5.2 Reverse Transcription Quantitative Real Time Polymerase Chain Reaction (RT-qPCR)

RT-qPCR was performed using JumpStart SYBR Green ReadyMix (Sigma-Aldrich, S4438) in the following reaction: 6µl SYBR Green, 0.24µl 10mM forward primer, 0.24µl 10mM reverse primer, 5µl DNA at 50ng/µl concentration, 0.52µl water to a total volume of 12µl. RT-qPCR was performed on Bio-Rad CFX96 RT-qPCR machine. RT-qPCR experiments were always performed using at least 3 independent biological replicates.

Table 2.4. Primers for RT-qPCR analysis

Target	**Forward Primer**	**Reverse Primer**
GAPDH	CGCTGAGTACGTCGTGGAGT	GGGCAGAGATGATGACCCTTT
HMBS	AGGAGTTCAGTGCCATCATCCT	CACAGCATACATGCATTCCTCA
POU5F1	GGATATACACAGGCCGATGTGG	ATGGTCGTTTGGCTGAATACCT
SOX2	AACCAGCGCATGGACAGTTAC	GTTCATGTAGGTCTGCGAGCTG
NANOG	TCCAGCAGATGCAAGAACTCTC	GGTTCTGGAACCAGGTCTTCAC
DUSP6	TTCCCTGAGGCCATTTCTTT	AGTGACTGAGCGGCTAATG
OTX2	CACTTCGGGTATGGACTTGC	GGTACCGGGTCTTGGCAAA
ZIC2	GATGTGCGACAAGTCCTACAC	TGGACGACTCATAGCCGGA
KLF17	CTGCAACTACGAGAACTGCG	GCAAGAATATGGCCTCTACC
KLF4	GCTGCCGAGGACCTTCTG	GCGAACGTGGAGAAAGATGG
TFCP2L1	TTTGTGGGACCCTGCGAAG	TGCTTAAACGTGTCAATCTGGA
DPPA3	AGACCAACAAACAAGGAGCCT	CCCATCCATTAGACACGCAGA
DNMT3L	CTGCTCCATCTGCTGCTCC	ATCCACACACTCGAAGCAGT

2.5.3 Western Blotting

RIPA buffer:
150 mM NaCl
1% NP-40
0.5% sodium deoxycholate
0.1% SDS
50 mM Tris pH 8

10X Tris Buffered Saline (TBS) 1L
150mM NaCl
20mM Tris-HCL

TBS-T:
1X Tris Buffered Saline
0.05% Tween-20

Whole cell lysates were extracted in RIPA buffer with 1x cOmplete Mini EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, 04693159001). Proteins were separated by electrophoresis using 12% SDS-polyacrylamide gels and transferred to 0.45 µM PVDF membranes (Amersham Hybond). Membranes were blocked for ≥ 3 hours in TBS-T 5% milk and hybridized to primary antibody overnight at 4°C. Membranes were washed three times for 10 minutes in TBS-Tween 5% milk at room temperature, followed by a 1 hour incubation with HRP-conjugated secondary antibodies at room temperature. Detection was performed using Clarity Western ECL reagent (Bio-Rad, 1705060). Precision Plus Protein Dual Colour Standards (Bio-Rad, 1610374) was used to assess protein molecular weights.

Table 2.5. Antibodies for western blotting

Reagent	Source	Concentration
TFCP2L1	R&D Systems, AF5726	1:500
KLF17	Atlas Antibodies, HPA024629	1:200
POU5F1	Santa Cruz, sc5279	1:500
β-ACTIN	Sigma-Aldrich, A5441	1:1000
HRP-conjugated goat-anti-mouse	Bio-Rad, 1721011	1:5000
HRP-conjugated goat-anti-rabbit	Bio-Rad, 1706515	1:5000
HRP-conjugated donkey-anti-goat	Jackson Immuno Research, 705-035-003	1:2500

2.5.4 Immunofluorescence

Cells were fixed with 2% formaldehyde (Agar Scientific, AGR1026) for 15 minutes and permeabilised and blocked overnight with 0.1% Triton X-100 (Sigma-Aldrich, T9284) in PBS containing 5% FBS (blocking buffer) at 4°C. Cells were subsequently incubated overnight with primary antibodies diluted in blocking buffer at 4°C. Cells were washed with 0.1% BSA and 0.1% Triton X-100 in PBS three times. Secondary antibodies diluted in blocking buffer were then applied and incubated for 4 hours at 4°C, or 1 hour at room temperature. Cells were washed

three times in 0.1% BSA and 0.1% Triton X-100 in PBS, with the final wash including 0.2µg/ml DAPI or 1µg/ml Hoechst 33342 for 10 minutes.

Images were collected on a Zeiss LSM710-NLO point scanning confocal microscope with a 20x water immersion objective, and on a NIKON A1-R confocal microscope with a 40x oil objective. Z stack images were processed with ImageJ.

Table 2.6. Antibodies used for immunofluorescence microscopy

Reagent	Source	Concentration
CD130	BD Biosciences, 555756	1:200
CD229	R&D Systems, AF1898	1:200
CD24	BD Biosciences, 555426	1:200
CD320	R&D Systems, AF1557	1:200
CD57	BD Biosciences, 555618	1:200
CD7	BD Biosciences, 555359	1:200
CD75	Abcam, ab77676	1:400
CD77	BD Biosciences, 551352	1:200
CD90	BD Biosciences, 555593	1:200
HLA-ABC	BD Biosciences, 555551	1:200
KLF17	Atlas Antibodies, HPA024629	1:300
KLF4	R&D Systems, AF3158	1:200
NANOG	Abcam, ab21624	1:500
NANOG	ReproCELL, RCAB0004P-F	1:200
POU5F1	Santa Cruz, sc5279	1:300
SOX2	R&D Systems, MAB2018	1:200
SSEA4	R&D Systems, MAB1435	1:200
TFCP2L1	R&D Systems AF5726	1:400
Donkey anti-mouse AF568	ThermoFisher Scientific, A10037	1:1000
Donkey anti-mouse AF555	ThermoFisher Scientific, A31570	1:1000
Donkey anti-mouse AF647	ThermoFisher Scientific, A31571	1:1000
Donkey anti-mouse AF594	ThermoFisher Scientific, A21203	1:1000
Donkey anti-rabbit AF647	ThermoFisher Scientific, A31573	1:1000
Donkey anti-rabbit AF488	ThermoFisher Scientific, A21206	1:1000
Donkey anti-goat AF555	ThermoFisher Scientific, A21432	1:1000
Hoechst 33342	ThermoFisher Scientific, H3570	1:10000
DAPI	Sigma-Aldrich, D9542	1:200

2.5.5 RNA-Sequencing

RNA-sequencing libraries were constructed from 500ng total RNA using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, E7530), with the Poly(A) mRNA Magnetic Isolation Module (NEB, E7490). Libraries were indexed using NEBNext Multiplex Oligos for Illumina (Index Primers Set 1 and Set2) (NEB, E7335 and E7500). Library fragment size and concentration was determined using an Agilent Bioanalyzer 2100 and KAPA Library

Quantification Kit (KAPA Biosystems, K4828). Samples were sequenced on an Illumina NextSeq500 instrument as 150bp single-end libraries at the Babraham Institute Sequencing Facility.

2.5.6 RNA-Sequencing Analysis

RNA-sequencing reads were trimmed using trim galore v0.4.2 using default parameters to remove the standard Illumina adaptor sequence. Reads were mapped to the human GRCh38 genome assembly using HISAT 2.0.5 guided by the gene models from the Ensembl v70 release (Aken et al., 2017). Quality control checks were performed using FastQC (Andrews, 2014). Samtools (Li et al., 2009a) was used to convert to BAM files that were imported to SeqMonk v36.0. Further QC of the sequencing data was performed in SeqMonk to ensure an enrichment of read counts mapping to exons, that genomic DNA contamination was not present, and more generally that the overall distribution of counts was similar between samples. Raw read counts per transcript were calculated using the RNA-sequencing quantitation pipeline on the Ensemblv70 gene set using directional counts. Differentially expressed genes were determined using DESeq2 (Love et al., 2014). Regularised log transformation was applied prior to visualization to correct for library size and variance among counts.

Principal component analysis (PCA) was performed using the top 1000 most variable genes across all samples. The first and second principal components were plotted. Gene Ontology (GO) analysis was performed using AmiGO 2 (Carbon et al., 2009) or EnrichR (Chen et al., 2013; Kuleshov et al., 2016) with default settings.

2.5.7 Transposable Element Analysis

To analyse the expression of transposable elements, probes were generated in SeqMonk over the locations of hg38 repeats and then filtered to remove those which were within 2kb of a gene. Raw counts for all of the reads that overlapped with the final probe set were exported and collated to generate counts for each class of repeat. Reads were globally normalised per million reads. Samples containing > 3% reads outside of genes were discarded due to potential DNA contamination that could mask the quantification of transposable elements. PCA was performed using the count data for repeat classes containing a minimum of 20 total reads across the samples. The first and second principal components were plotted using the top 1000 most variable transposable elements across all samples.

2.5.8 Quantification of X-linked Genes

To analyse allele-specific expression of X-linked genes, an N-masked genome was generated using the positions of heterozygous SNPs on the X chromosome of H9 or WIBR3 cells (coordinates kindly provided by Celine Vallot for H9 cells (Vallot et al., 2015). WIBR3 SNPs were identified from the GEO SNP array data file: GSM738141. R was used to identify and extract heterozygous SNPs (allele difference >-0.2 and <0.2). LiftOver was used to convert the genomic coordinates from hg19 to GRCh38. RNA-sequencing reads were trimmed using trim galore v0.4.2/v0.4.4 and aligned to the N-masked genome using HISAT2 (default settings but without soft-clipping). The mapped data was sorted into allele-specific reads using SNPsplit (v0.3.1/v0.3.3, default parameters, single-end) (Krueger and Andrews, 2016). Genome1/genome2 reads, which corresponded to reads carrying either of the two SNPs, were imported into SeqMonk. Probes were designed over informative SNP annotations and quantified in SeqMonk using linear read counts. Read counts were exported as 'Feature report' and annotated by gene name. Replicate samples were merged. Transcripts with fewer than 10 informative reads were classified as 'not expressed'. Transcripts were classified as biallelic when 25%–75% reads originated from the minor allele (i.e., allelic ratio of 3:1).

2.5.9 Comparative Transcriptional Dynamics and Clustering Analysis

Raw read counts per transcript were calculated using the RNA-sequencing quantitation pipeline in SeqMonk V.1.40.1. Biological replicates were merged and transcripts containing fewer than 5 reads were omitted. Integrated Differential Expression and Pathway analysis (iDEP) V0.71 (Ge, 2017) was used for transcriptional clustering analysis. The top 4000 most variably expressed genes were clustered based on Euclidean distance into eight lusters. Eight clusters was deemed optimal as this provided eight distinct profiles of transcriptional changes. When greater than eight clusters were selected, multiple clusters displayed the same transcriptional dynamics. Z-score normalised expression values for each gene cluster were plotted as line graphs across the reprogramming time-course.

3. Comprehensive Cell Surface Protein Profiling to identify Markers of Human Naïve and Primed Pluripotent States

3.1 Background

3.1.1 Overview of isolation methods to enrich for naïve hPSCs

The predominant method to derive naïve hPSCs is via the reprogramming of primed hPSCs to a naïve state. The efficiency of this method is often low and variable between protocols used, typically generating a heterogeneous population of cells during the reprogramming process. A common approach to purify the naïve hPSCs is via continued passaging of the cultures; however, this method relies on the assumption that naïve hPSCs have a competitive advantage compared to non-reprogrammed cells, and will therefore be enriched over time. Moreover, this approach is not suitable where the culture conditions permit the survival of primed and non-reprogrammed cell types.

As a result, several alternative approaches have been devised to prospectively isolate naïve hPSCs by capitalising on aforementioned knowledge of molecular hallmarks of naïve pluripotency. Examples include reporter systems that monitor *OCT4* distal enhancer usage, or the levels of HERVH transcription. Whilst these reporter systems can provide a simple readout of cell state, they require each cell line to be genetically modified. Furthermore, the timing of reporter induction and the fidelity of expression are not always clear, and may not necessarily report a robust change in cell state – a point that I will address in more detail in Chapter 5. Consequently, there is a need for transgene-independent isolation strategies, which had not been reported at the onset of this work.

During this time however, three reports have subsequently each identified either individual markers, or several markers that display primed-specificity (O'Brien et al., 2017; Pastor et al., 2016; Shakiba et al., 2015a). In each case, the timing of marker expression change is unknown, which restricts the utility for isolating nascent naïve hPSCs early during reprogramming. Moreover, the lack of a known positive naïve-specific marker limits the unambiguous identification of naïve hPSCs within a heterogeneous population.

3.1.2 Current strategies to identify and isolate naïve hPSCs are unsuitable for studying the early stages of primed to naïve reprogramming

As previously discussed in the main introduction, there are two media formulations that give rise to naïve hPSCs with the closest resemblance to preimplantation human blastocysts (Huang et al., 2014; Stirparo et al., 2017) (Table 1.1, Table 1.2); these are 5iLA(F) and t2iL+PKCi, and are the two methods of reprogramming that I primarily use. The naïve hPSC lines derived using 5iLA reprogramming were established by continuous passaging. (Theunissen et al., 2016; Theunissen et al., 2014). Reliance on this enrichment method means that the molecular characterisation of 5iLA(F) hPSCs, such as the DNA methylome and transcriptional analysis, has been performed on cells that have undergone ten passages. (Theunissen et al., 2016). The same is also true for the transgene dependent method of reprogramming (NK2 t2iL+PKCi), which takes approximately 28 days, before the clonal derivation and expansion of naïve hPSC lines. Therefore, much of the characterisation of NK2 t2iL+PKCi derived naïve hPSCs has also been conducted on cells that have been in culture for at least 40 days (≈10 passages) (Takashima et al., 2014). These two reprogramming methods coupled to their current isolation strategies prohibit the reprogramming process itself from being studied.

Most of what we currently know about the naïve state of human pluripotency is based upon long-term cultured naïve hPSCs. Therefore, the transcriptomics analysis will predominantly highlight the gene regulatory networks (GRNs) that are utilised in maintaining the established primed and naïve states. Whilst this has identified reliable transcriptional markers to distinguish the two cell types, we know very little about the transitional period, the route(s) taken between the two states, and the transcription factors required. The same applies to observations made regarding the epigenetic landscape; we know that naïve hPSCs undergo X-chromosome reactivation and a genome-wide remodelling of DNA methylation levels, however much is unknown about the timing and regulation of these processes.

Establishing a means to study cell populations as they transition would provide a new insight into the order of molecular changes that occur, including the identification of mediators of these processes. Many of these molecular changes, such as the rewiring of gene regulatory networks and remodelling of the DNA methylation landscape, occur during other cellular transitions or at the onset of disease, and may share similar pathways and targets. Therefore, understanding primed to naïve state transitioning will have wider implications beyond this system.

3.1.3 Aims

Accurately tracking the early stages of reprogramming, and the emergence of nascent naïve hPSCs, requires a new approach of identifying and isolating these cells that current methods do not permit. To this end, I sought to identify cell surface markers capable of distinguishing naïve hPSCs, primed hPSCs, and most importantly nascent naïve hPSCs early during reprogramming.

1. Conduct a comprehensive cell surface marker screen on naïve and primed hPSCs
2. Identify pluripotent state-specific cell surface markers
3. Validate the state-specificity of candidate surface markers
    i. Using naïve hPSCs derived using different reprogramming methods
    ii. Using preimplantation-stage human embryos
4. Examine the utility of candidate markers for tracking naïve to primed transition
5. Examine the utility of candidate markers for tracking primed to naïve reprogramming

3.2 Results

The gold standard technology for detecting and isolating live cell populations is via flow cytometry and fluorescent activated cell sorting (FACS). This technology can be used in combination with recently developed antibody library screens that contain hundreds of titrated antibodies. Individual antibodies are allotted to each well of a 96-well plate to enable high-throughput screening of your desired cell type. These antibody library screens present an opportunity to examine a large quantity of antibodies, many of which have not previously been associated with pluripotent stem cells. Even so, our knowledge of markers expressed on the cell surface of naïve hPSCs is very limited; therefore, a wider screening approach is sensible.

3.2.1 Comprehensive cell surface marker screen of naïve and primed hPSCs

Naïve and primed hPSCs were profiled using two different antibody library screens to identify pluripotent state-specific surface markers. These were the BioLegend's LEGENDScreen that I performed, and the BD-Lyoplate conducted by our collaborators at the Karolinska Institute. Both screens were used in order to increase the number of cell surface markers that could be examined, as some were unique to one screen. The LEGENDScreen contained 135 unique antibodies, whereas the BD-Lyoplate contained 45 unique antibodies. The overlapping number of antibodies covered by both screens was 197, with 109 surface markers covered by different antibody clones and 88 antibodies being the same clone. H9 primed and H9 naïve hPSCs were reprogrammed and maintained under two different naïve culture conditions (5iLA and NK2 t2iL+PKCi) to identify robust state-specific markers, and to capture any variation related to the reprogramming method and maintenance culture media used. Our collaborators at the Karolinska Institute profiled H9 primed and naïve hPSCs cultured under 5iLA conditions, whilst I profiled H9 NK2 primed and NK2 t2iL+PKCi naïve cultured cells. Primed and naïve hPSC samples were analysed by high-throughput flow cytometry, and state-specific markers were identified according to the fluorescence intensity of each marker (Figure 3.1).

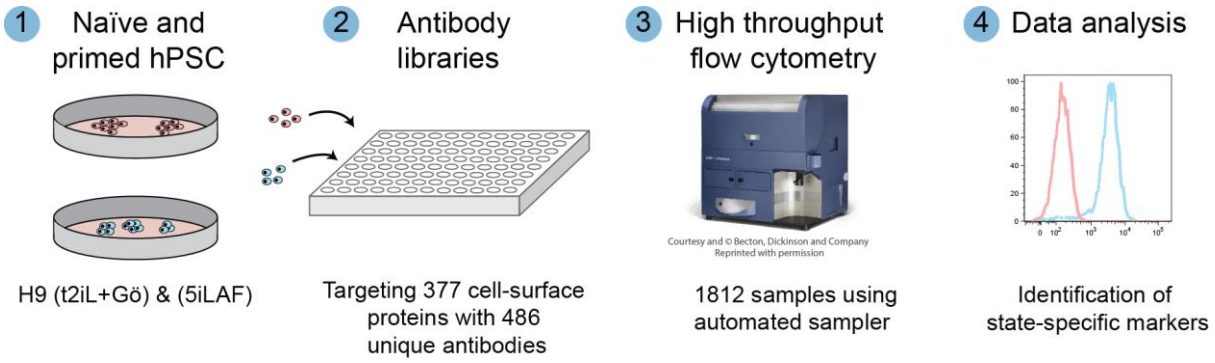


Figure 3.1. Schematic of the experimental design.
Human primed (cultured in Primed (KSR) medium on MEFs and TeSR-E8 on Vitronectin) and naïve (cultured in 5iLAF and t2iL+PKCi conditions) H9 hPSCs were profiled by two antibody library screens that targeted 377 cell surface markers. Samples were analysed by high-throughput flow cytometry. State-specific surface markers were identified according to the fluorescence intensity of each marker.

The percentage of positive cells for each cell type was determined for each surface marker after replicate values were averaged (Figure 3.2). The results have identified more than 50 primed-specific markers, 40 shared markers, and 8 naïve-specific markers (Figure 3.2). Reassuringly, the dataset includes previously reported cell surface markers that are known to be expressed in both primed and naïve hPSCs, such as TRA-1-60 and TRA-1-81 (Chen et al., 2015; Gafni et al., 2013; Pastor et al., 2016; Qin et al., 2016; Shakiba et al., 2015a; Ware et al., 2014). In line with previous reports of displaying primed-specificity, CD24 and SSEA-4 were heterogeneously or lowly expressed in naïve hPSC cultures (Pastor et al., 2016; Shakiba et al., 2015a). The results have also revealed cell surface markers that are newly identified as being expressed in human PSCs; examples include CD46, CD151, MCAM (CD146) and PDPN which were uniformly expressed in both primed and naïve hPSCs and provide a useful set of common markers. More than 50 primed state-specific markers were identified, such as THY1 (CD90), B3GAT1 (CD57), HLA-A-B-C, and members of the NOTCH family of receptors. However, far fewer naïve-specific surface markers were identified with only 8 being uniquely expressed in naïve hPSCs. These include LY9 (CD229), CD77 and CD75 that are glycosylated epitopes, and CD130 (IL6ST), also known as the LIF coreceptor.

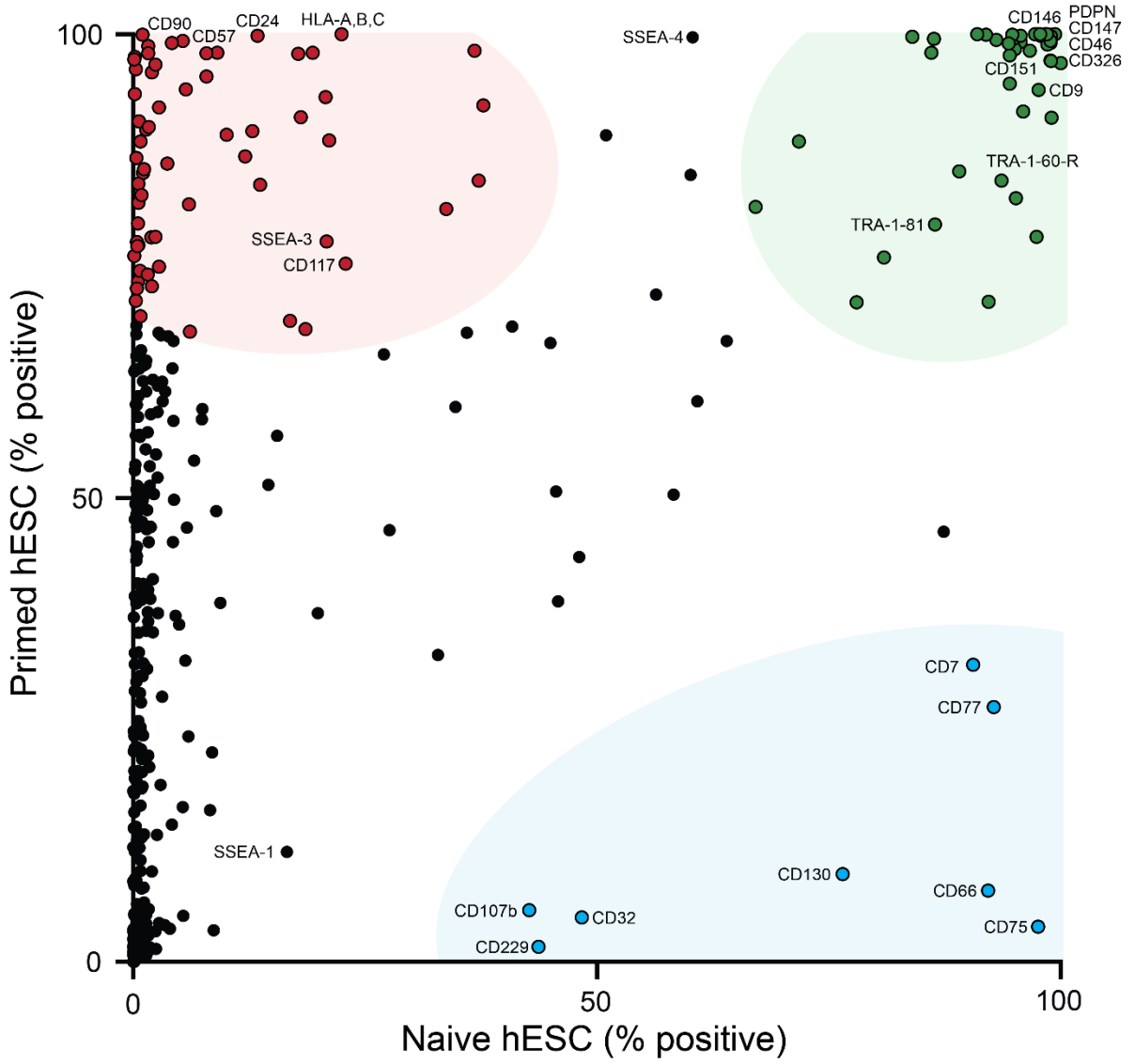


Figure 3.2. Identification of pluripotent state-specific cell surface markers

Each dot represents a different cell-surface marker and their position along the XY axis is determined by the percentage positive in naïve and primed hPSCs. Results are averaged from one to three independent assays per cell type. One BD-Lyoplate screen was conducted on 5iLAF cultured H9 naïve hPSCs and parental H9 primed hPSCs. Two biological replicates of H9 NK2 t2iL+PKCi naïve hPSCs and parental H9 NK2 primed hPSCs were independently screened using BioLegend's LEGENDScreen. Based on their position in the chart, a subset of cell surface proteins have been categorised as naïve-specific (blue), primed-specific (red), and common to both naïve and primed hPSCs (green).

The previous studies that examine surface marker expression between naïve and primed hPSCs rely exclusively on primed-specific markers (O'Brien et al., 2017; Pastor et al., 2016; Shakiba et al., 2015a). This approach is limited by the assumption that surface marker expression changes accurately reflect a change in pluripotent state; however, it is possible that surface markers may change in response to the signaling cues provided when the culture media is changed at the onset of primed to naïve reprogramming. In this instance, the primed-specific

markers would be unable to discriminate reprogramming populations, including nascent naïve hPSCs. With this in mind, I set out to validate a cohort of primed and naïve specific markers; each conjugated to different fluorophores, with the overarching aim to generate a multiplexed panel of antibodies that could be used simultaneously.

3.2.2 Validation of state-specific surface markers

One of the drawbacks to using flow cytometry or FACS to identify and isolate cell types is the requirement for high-quality antibodies. The selection of candidates that could be validated was restricted by the availability of antibodies with compatible conjugations. Many of the antibodies were only available in common conjugations, such as phycoerythrin (PE), or fluorescein (FITC). Due to the OCT4-ΔPE-GFP reporter construct integrated into hPSC lines that I commonly use during primed to naïve reprogramming, FITC antibodies would not be compatible. This therefore limited the selection of antibodies down to four compatible conjugations for naïve-specific markers targeting CD7, CD75, CD77 and CD130. The selection of primed-specific markers was based upon a newly identified hPSC marker CD57, and two additional primed-specific markers CD24 and CD90. Verification of the state-specificity of the individual markers is shown in Figure 3.3 using H9 NK2 primed and H9 NK2 Naïve hPSCs cultured under t2iL+PKCI conditions.

Consistent with the expression profiles seen from the antibody screen, all predicted primed state-specific antibodies exhibit a good separation in fluorescence intensity by flow cytometry when comparing primed and naïve hPSCs (Figure 3.3). It is clear that naïve hPSCs are uniformly negative for all primed-specific markers examined. Similarly, the four naïve-specific markers tested were consistently absent in primed hPSCs; but were all expressed in naïve hPSCs, confirming their state-specificity. The variation and magnitude of expression for some of the markers was variable. For example, CD77 was unevenly expressed across the naïve hPSC population. By comparison, CD130 was uniformly expressed, however the magnitude of difference between primed and naïve hPSCs for CD130 was lower than some other markers examined.

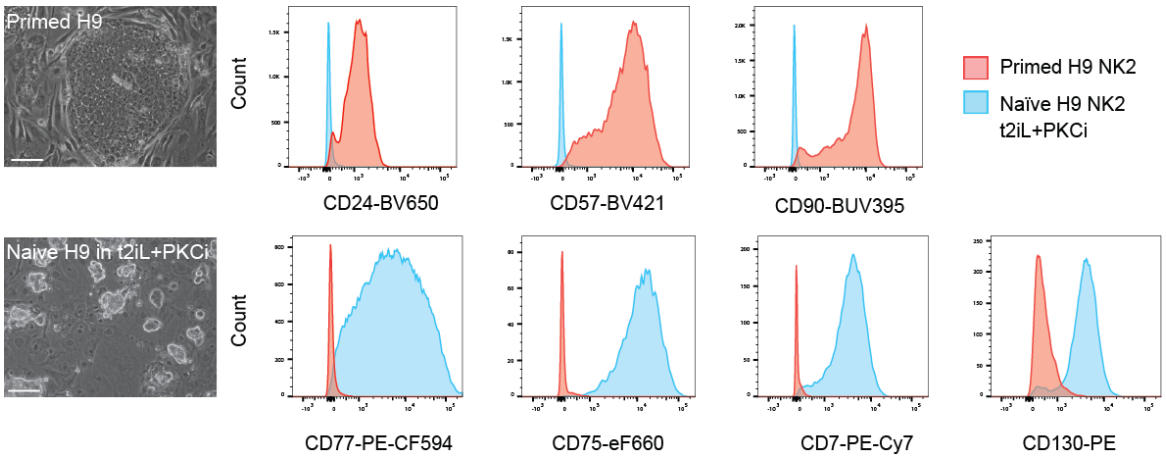


Figure 3.3. Validation of individual cell surface markers in Naïve H9 NK2 cells cultured in t2iL+PKCi and parental Primed H9 NK2 cells.
Histograms of flow cytometry data using conjugated antibodies for primed cells (red) and naïve cells (blue). Representative phase-contrast image (left); scale bars indicate 100µm.

It is possible that the state-specific markers identified could be specific to the H9 hPSC line and could also vary between naïve culture conditions used. I therefore sought to examine additional primed and naïve hPSC lines and naïve culture conditions. Shown in Figure 3.4 is the flow cytometry data for an iPSC derived line from human fibroblasts (FiPSC), with the primed cells cultured in KSR media and naïve cells cultured in t2iL+PKCi. Similar state-specific expression profiles could be seen for most of the markers; however, CD24 expression failed to clearly distinguish primed and naïve hPSCs, in contrast to the prior results using H9 hPSCs. Likewise, CD77 expression was observed in a proportion of primed FiPSCs, as well as being highly expressed in naïve FiPSCs.

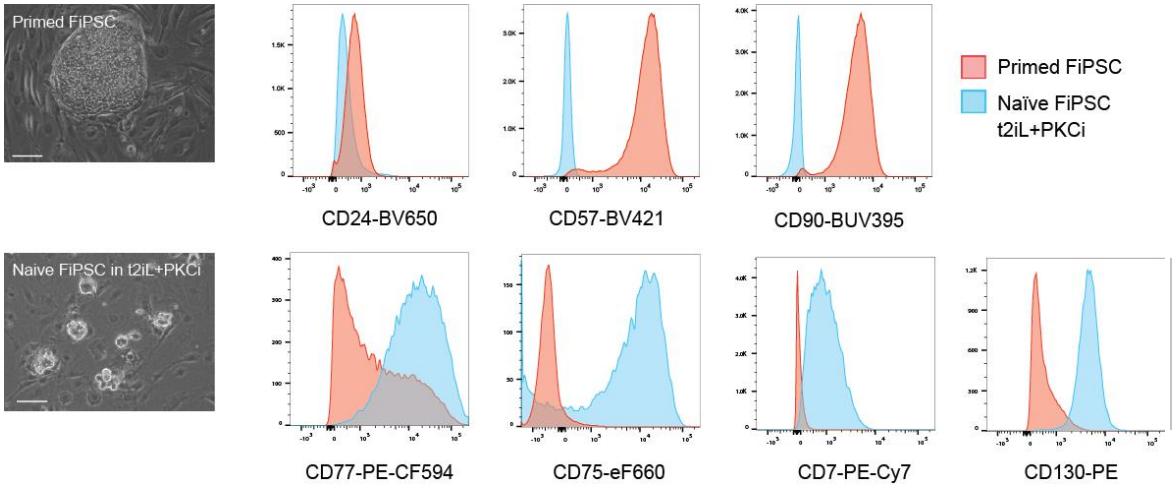


Figure 3.4. Validation of individual cell surface markers in Naïve FiPSCs cultured in t2iL+PKCi and parental Primed FiPSCs.
Histograms of flow cytometry data using conjugated antibodies for primed cells (red) and naïve cells (blue). Representative phase-contrast image (left); scale bars indicate 100µm.

To exclude the possibility of variation between culture conditions, 5iLAF cultured H9 naïve hPSCs were examined by our collaborator, Fredrik Lanner (Figure 3.5). In agreement with t2iL+PKCi cultured naïve hPSCs (Figure 3.3), all naïve-sate specific markers displayed comparable expression profiles. Primed state-specific markers also displayed concordant expression, with CD24 exclusively expressed in primed cells. An additional primed-specific marker, HLA-A,B,C, was examined and found to be enriched in H9 primed cells.

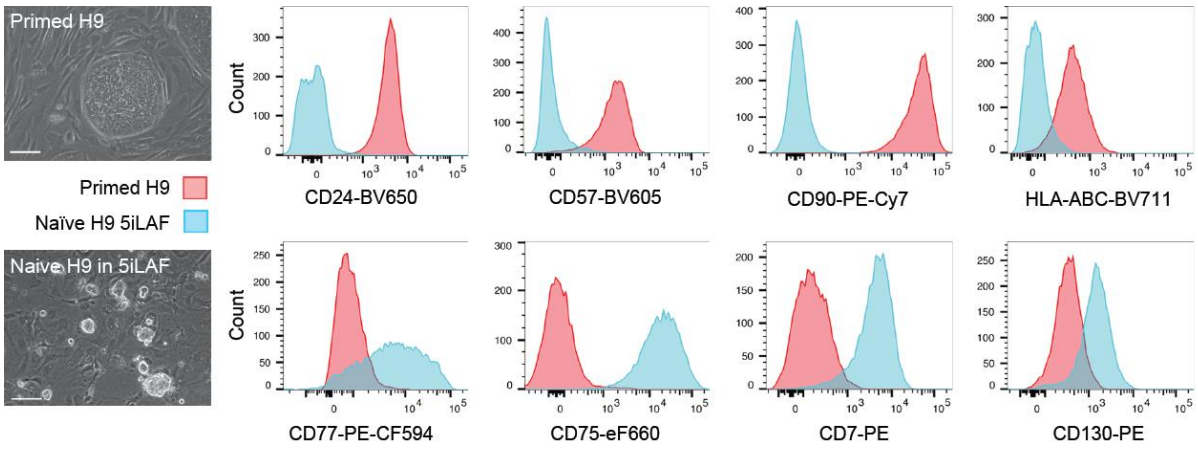


Figure 3.5. Validation of individual cell surface markers in Naïve H9 cells cultured in 5iLAF and parental Primed H9 cells.
Histograms of flow cytometry data using conjugated antibodies for primed cells (red) and naïve cells (blue); Flow cytometry data provided by Sarita Panula. Representative phase-contrast image of naïve and primed cells (left); scale bars indicate 100µm.

As previously discussed within the introduction, 5iLA(F) and t2iL+PKCi naïve hPSCs are maintained in culture media targeting similar signaling pathways, and ultimately exhibit very similar naïve hallmark characteristics. I therefore sought to examine the state-specific expression of our surface markers on naïve hPSCs generated using commercially available RSeT medium, which is based upon the NHSM formulation (Gafni et al., 2013). As noted in Table 1.1, the naïve hPSCs derived under these conditions generate an alternative type of naïve-like hPSCs with dissimilar properties. Perhaps unsurprisingly, RSeT naïve-like hPSCs fail to express any of the naïve-specific markers, but two primed-specific markers (CD24 and CD90) were downregulated (Figure 3.6).

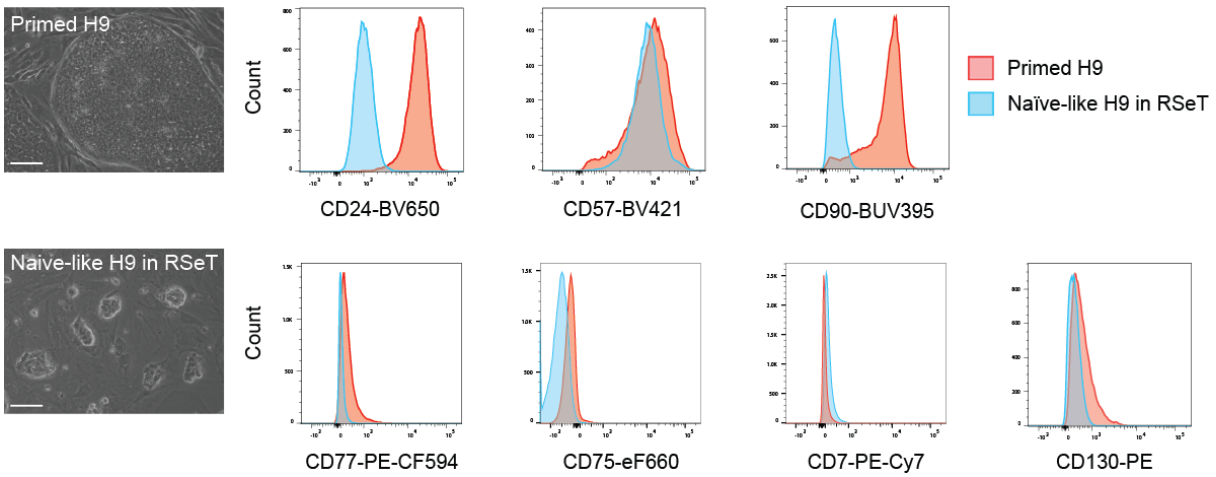


Figure 3.6. Validation of individual cell surface markers in Naïve-like H9 cells reprogrammed and maintained in RSeT medium and parental Primed H9 cells.
Histograms of flow cytometry data using conjugated antibodies for primed cells (red) and naïve-like cells (blue). Representative phase-contrast image of naïve and primed cells (left); scale bars indicate 100µm.

3.2.3 Human embryo validation of state-specific surface markers

To investigate the relevance of the identified surface markers to human embryo development, embryonic day 6-7 human embryos were stained with each antibody (Figure 3.7). This stage of embryonic development represents the preimplantation-stage, which naïve hPSCs share many characteristics with (Table 1.1). At this time point, all three lineages of the human blastocyst (the trophectoderm, epiblast and primitive endoderm) should be established (Petropoulos et al., 2016). This is confirmed by the presence of both NANOG-positive cells of the epiblast and NANOG-negative cells of the primitive endoderm found within the inner cell mass, (indicated in Figure 3.7 - ICM zoom-in for CD7 and CD130 staining). Whilst CD7 signal was not detected, CD75, CD77 and CD130 were detected at the cell surface of human blastocyst cells. Of note, CD130 was particularly enriched within the cells of the ICM; whereas CD75 and CD77 were expressed in the majority of the cells, including the ICM. This indicates that their specificity is not restricted to cells of the human epiblast. By contrast, all of the primed-specific surface markers were undetectable, except for HLA-A,B,C which was present in a small number of trophectoderm cells.

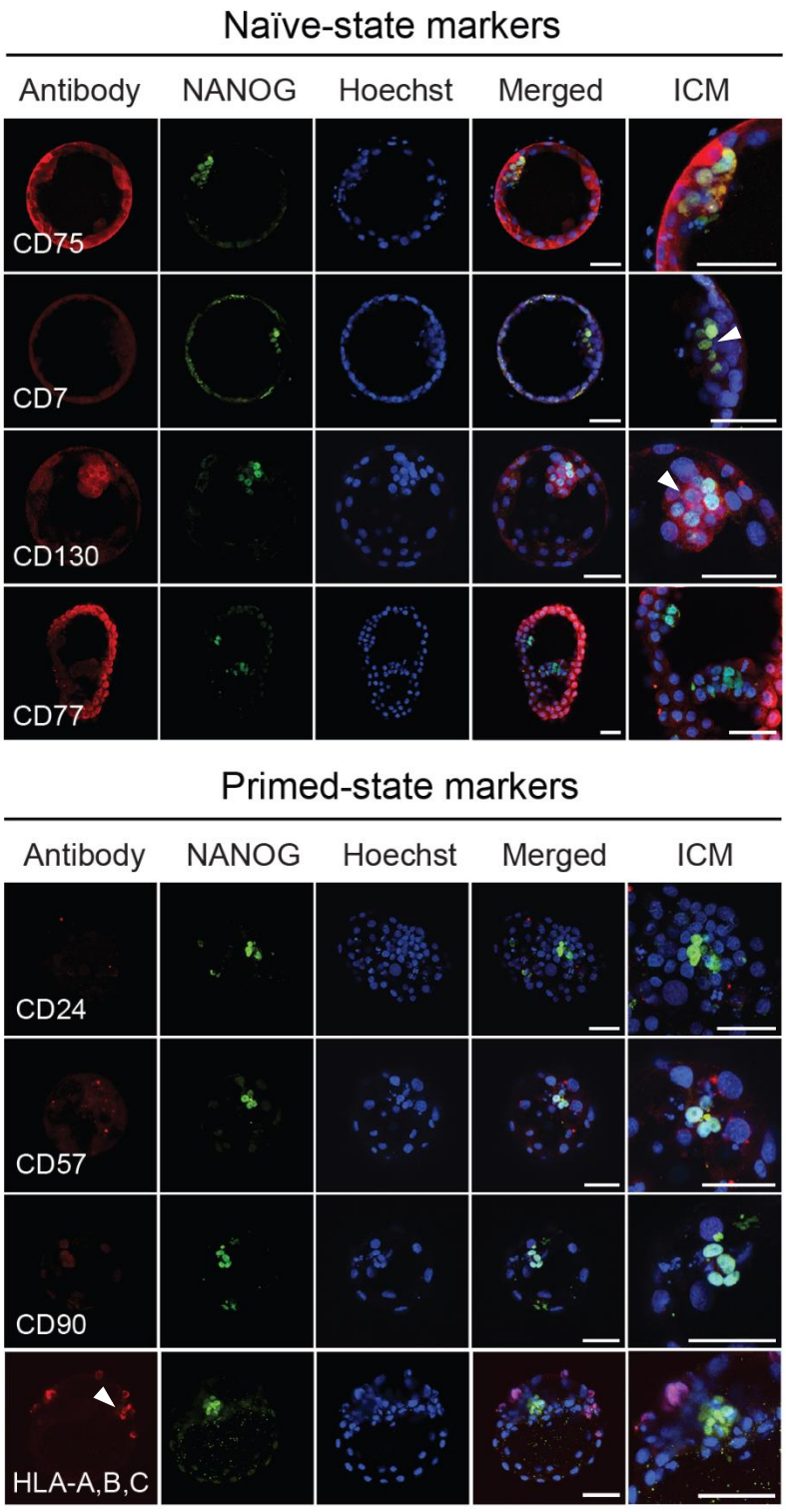


Figure 3.7. Immunofluorescent microscopy validation of candidate cell surface markers in human blastocysts.

Cross-sections of embryonic day 6 human blastocysts labelled with candidate naïve-specific and primed-specific surface markers together with NANOG, to reveal the location of the epiblast, and hoechst to stain DNA in all cells. Scale bars, 50µm. Embryo immunofluorescent microscopy data provided by members of Fredrik Lanner's lab (Sarita Panula, John Paul Schell, Alvaro Plaza Reyes and Sophie Petropoulos).

3.2.4 Multiplexed antibody panel to distinguish naïve and primed human PSCs

Given the encouraging results obtained during the individual antibody validation studies, I decided to compile a multiplexed antibody panel consisting of three primed-specific markers (CD24, CD57, CD90), and four naïve-specific markers (CD7, CD75, CD77 and CD130) (Figure 3.8A). Whilst there may be individual markers that are capable of distinguishing established primed from naïve hPSCs, the expression of each marker throughout reprogramming is unknown. Moreover, primed to naïve reprogramming generates a heterogeneous population of cells at the early stages, therefore using a multiplex antibody panel containing primed and naïve-specific markers will be more reliable for accurate cell type identification.

As reprogramming of primed hPSCs to a naïve state is conducted on a layer of mouse embryonic fibroblasts, I additionally included an APC-Cy7 conjugated anti-CD90.2 antibody that recognises MEF cells, so that these could be excluded; a viability dye was also included on the same channel. This reduced the number of channels to compensate during multi-parameter flow analysis, as cells positive for APC-Cy7 were either MEFs or dead cells (Figure 3.8B). Moreover, primed-specific marker HLA-A,B,C was not included in the final panel. The weaker HLA-A,B,C expression in combination with a BV711 conjugation could not be accurately compensated for whilst using CD24 conjugated to BV650. Nonetheless, a combination of seven state-specific markers should be capable of distinguishing naïve and primed hPSCs.

A

Antibody Panel

Naïve markers	Primed markers	Other markers
CD75 (eF660)	CD24 (BV650)	CD90.2 (APC-Cy7)
CD7 (PE-Cy7)	CD57 (BV421)	Viability dye (eF780)
CD77 (PE-CF594)	CD90 (BUV395)	
CD130 (PE)		

B

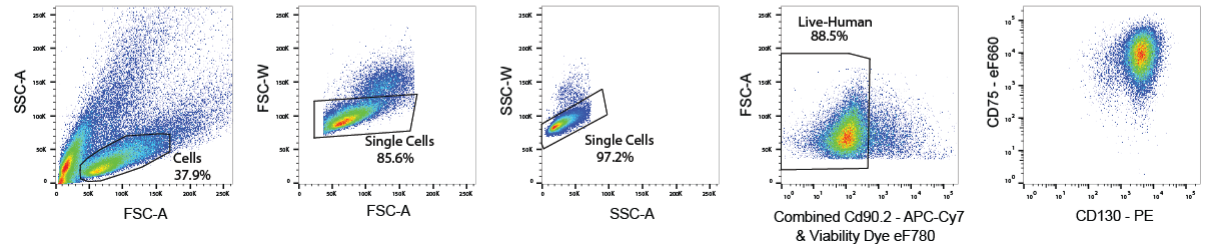


Figure 3.8. Flow cytometry gating scheme to isolate live-human single cells.
(A) A list of antibodies that are combined to form a multiplexed panel. The information in brackets shows the fluorophore conjugation of each antibody. See Table 2.1 for antibody details and Table 2.2 for flow cytometer parameters used.
(B) Flow cytometry dot plots show the gating scheme for H9 Naïve hPSCs. The first panel enables the isolation of cells versus debris, and the subsequent two panels identify single cells. The fourth panel allows live and human cells to be identified by excluding MEF cells that are positive for CD90.2 conjugated to APC-Cy7, and dead cells that are positive for an eF780 viability dye. The final panel proves an example to show the final gated population of live-human naïve PSCs that are positive for naïve-state markers CD75 and CD130.

To ensure that all conjugation were compatible when run simultaneously in a single panel, naïve and primed hPSCs were stained with all seven state-specific antibodies. Shown are the pairwise expression profiles for all seven markers in the format of contour plots (Figure 3.9). CD57, a primed specific marker, is plotted on the y axis for all combinations. The results indicate a clear state-specific expression profile for all markers, in line with the individual marker validations performed previously. Encouragingly, there is a clear separation between naïve and primed hPSCs, which will be critical when using the antibody panel to isolate naïve hPSCs at low frequency during reprogramming.

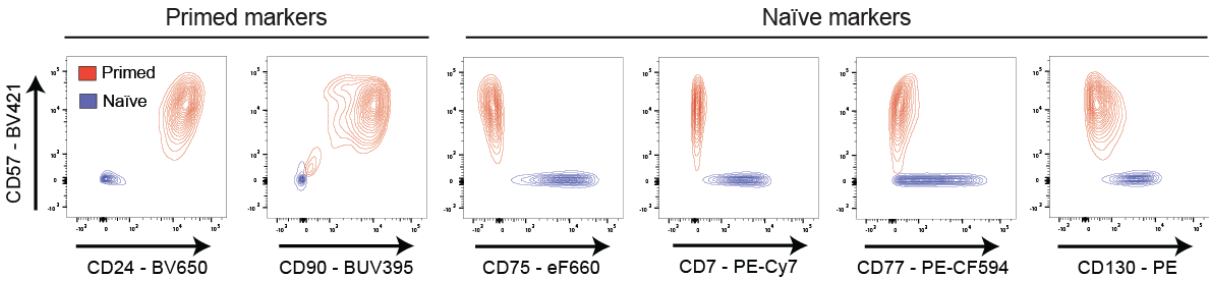


Figure 3.9. A Multiplex antibody panel to distinguish between naïve and primed hPSCs.
Flow cytometry contour plots of pairwise antibody combinations. The primed-specific marker CD57 is on the y axes, and the different primed-specific markers and naïve-specific markers are on the x axes. H9 NK2 Primed (red) and H9 NK2 t2iL+PKCi cultured naïve hPSCs (blue) are shown for each antibody combination.

A caveat with using conventional analysis methods for flow cytometry data is the inability to examine more than two markers simultaneously. For a panel with seven markers, 21 unique pairwise comparison plots would be required to examine all of the data. Even then, further gating and additional plotting would be required to examine the expression profiles of any sub-populations. Therefore, to overcome this limitation I used a dimensional reduction visualisation package called FlowSOM (Van Gassen et al., 2015). FlowSOM works by clustering the cells, typically into 100 different clusters, based on similar expression profiles obtained via flow cytometry. A self-organizing map (SOM), which is a type of artificial neural network, performs unsupervised clustering to order the different clusters; these are subsequently plotted as a minimal spanning tree (MST). This approach has the advantage of providing a clear overview of the expression levels of each marker across all cells. Moreover, subpopulations can be identified in an unsupervised manner. FlowSOM analysis was performed and is shown in Figure 3.10, using the same flow cytometry data plotted in Figure 3.9. The results again show two well-separated populations that correspond to naïve and primed hPSCs. This reiterates the finding that the antibody panel is capable of discriminating between the two cell types, (Figure 3.10B - right panel). The expression levels for each marker can be overlaid onto the MST and are shown to the left.

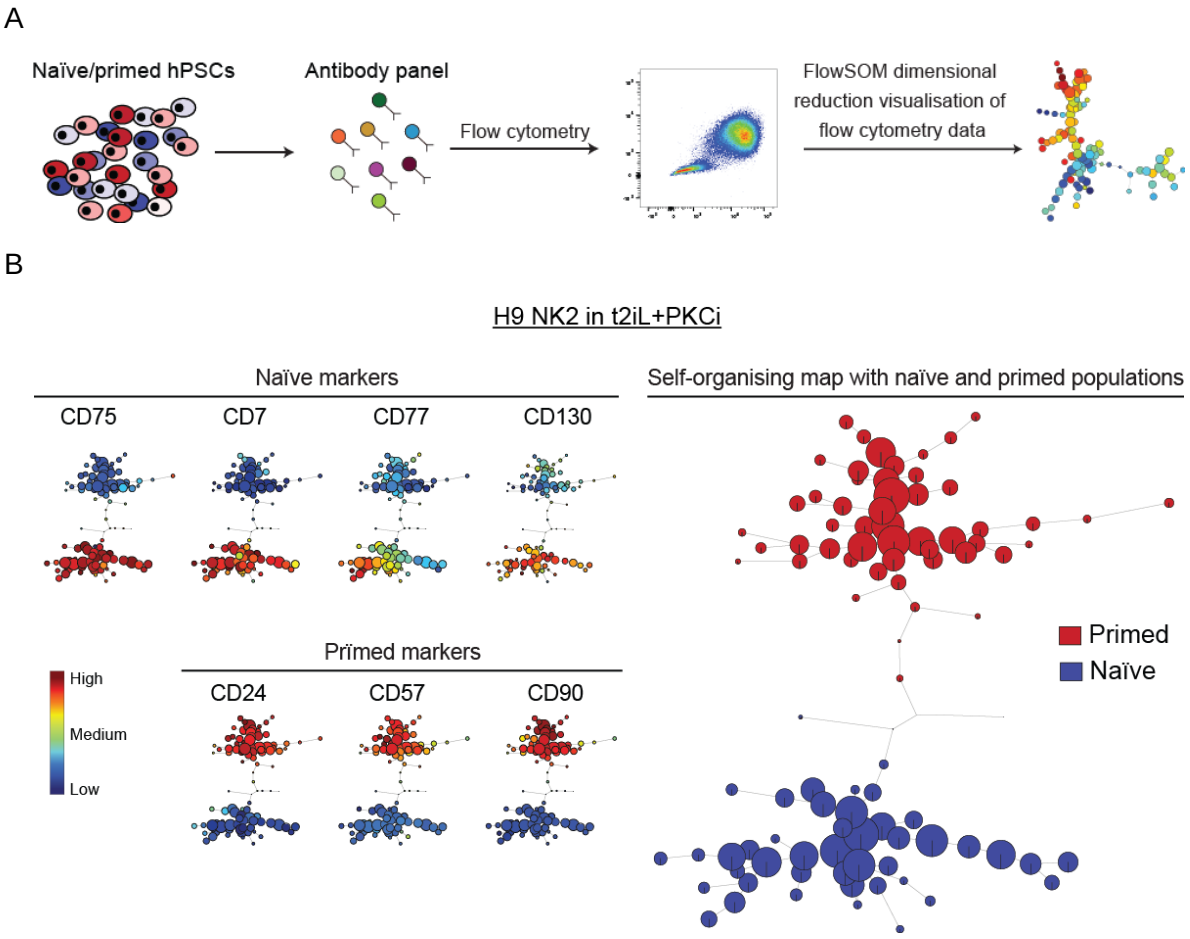


Figure 3.10. A Multiplex antibody panel to distinguish between naïve and primed hPSCs.
(A) Schematic of the experimental design and dimensional reduction visualisation of Flow cytometry data. The multiplexed panel of antibodies was applied to pluripotent cell types and flow cytometry data is obtained. FlowSOM visualisation enables the visualisation of all cells and multiple markers simultaneously by clustering cells of similar expression profiles together. (B) FlowSOM visualization of flow cytometry data for all antibodies in the panel. An unsupervised self-organising map arranges the cells into clusters (represented by circles) according to similarities in their cell surface marker expression profiles (right). Overlaying the identity of the cell type within each cluster reveals a clear separation of H9 NK2 naïve hPSCs cultured under t2iL+PKCi conditions (blue) and H9 NK2 primed (red) populations. The heatmap panels (left) show the expression level of each cell surface marker in the cell clusters. Clusters are arranged in the same position as for the minimal spanning tree of the self-organising map.

To ensure the robustness of the multiplexed antibody panel, two additional naïve and primed hPSC lines were examined; primed and naïve FiPSCs cultured in t2iL+PKCi (Figure 3.11A), and primed and naïve WIBR3 cells cultured in 5iLA (Figure 3.11B). As before, there is a good separation between the two cell types, particularly for the FiPSC line. Of note, the WIBR3 hPSC line contains an OCT4-ΔPE-GFP reporter that is active in naïve hPSCs (Theunissen et al., 2014). The FlowSOM results showed a good overlap in GFP expression with the naïve-specific surface markers, providing further authentication for the antibody panel. Curiously,

there is some overlap between the WIBR3 primed and 5iLA naïve hPSCs, with clusters containing both cell types. This is not so surprising, given that ≈40% of 5iLA cultured naïve hPSCs have been shown to express primed-specific marker SSEA-4, and display a primed or differentiated morphology (Pastor et al., 2016). These clusters could therefore correspond to a differentiated cell type, as both primed and naïve state-specific markers are lowly expressed, or absent.

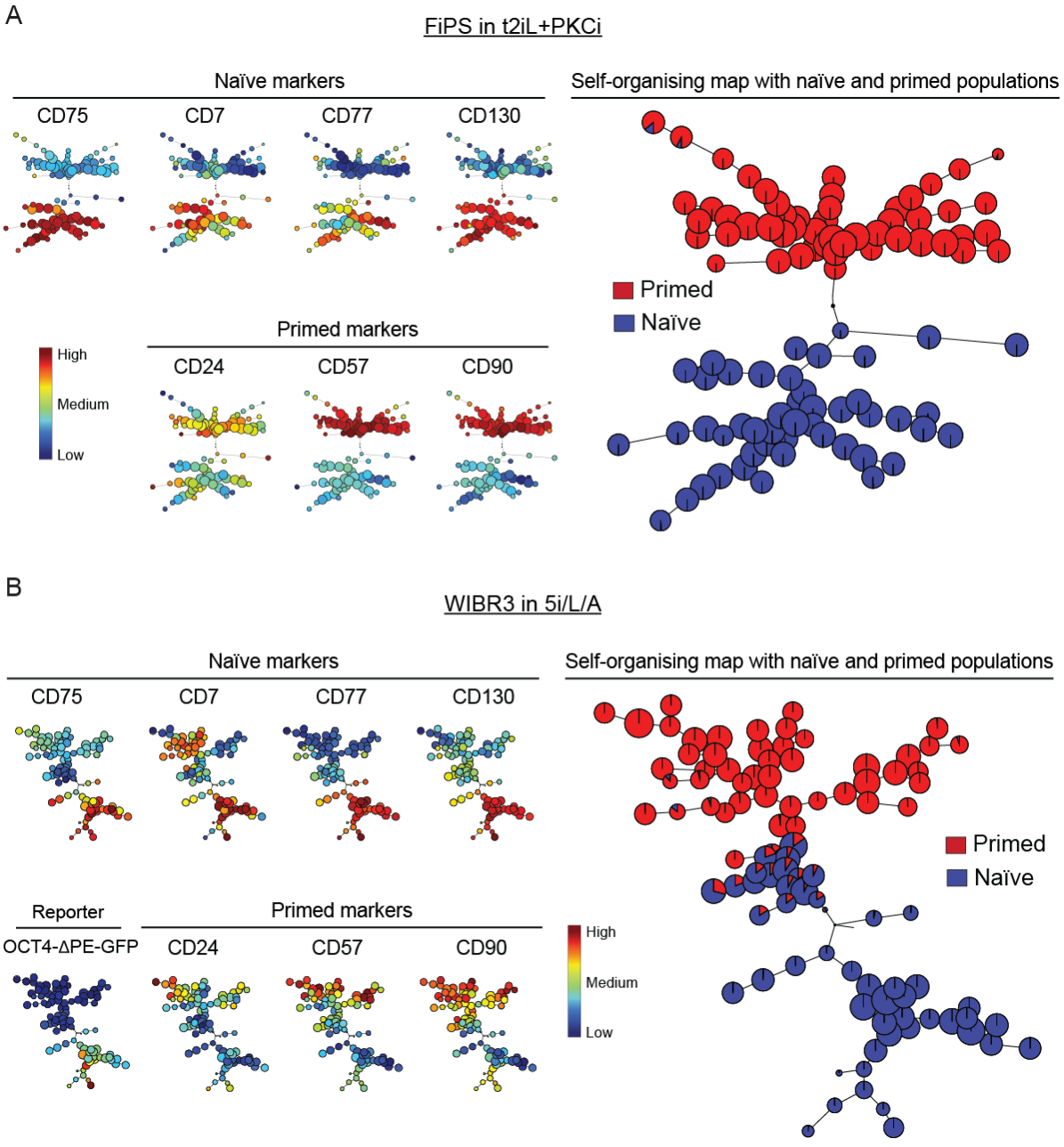


Figure 3.11. Multiplexed antibody panel enables robust distinction between naïve and primed hPSCs cultured under different naïve conditions.

(A and B) FlowSOM visualisation of flow cytometry data for all antibodies in the panel. (A) FlowSOM visualisation of FiPSC primed and FiPSC naïve hPSCs cultured under t2iL+PKCi conditions (right) exhibit clear separation according to cell type. (B) FlowSOM visualisation of WIBR3 primed and WIBR3 naïve hPSCs cultured under 5iLA conditions.

To further test the sensitivity of the antibody panel to discriminate naïve and primed hPSCs, I performed a spike-in experiment where primed and naïve hPSCs were mixed together to contain 90% primed and 10% naïve cells (Figure 3.12 – right panels). The bottom left panel shows the expression profiles of naïve hPSCs stained with two naïve-specific markers (CD75 and CD130). The same gating was used to demonstrate that primed hPSCs are not detected by this gating (top left panel), but importantly the spike-in naïve population could be identified with just these two markers. Moreover, these cells were almost exclusively negative for primed specific markers CD57 and CD24, as would be expected. This reiterates the previous findings that the antibody panel is capable of discriminating between the two cell types, and opened up the possibility of isolating emerging naïve hPSCs during primed to naïve reprogramming.

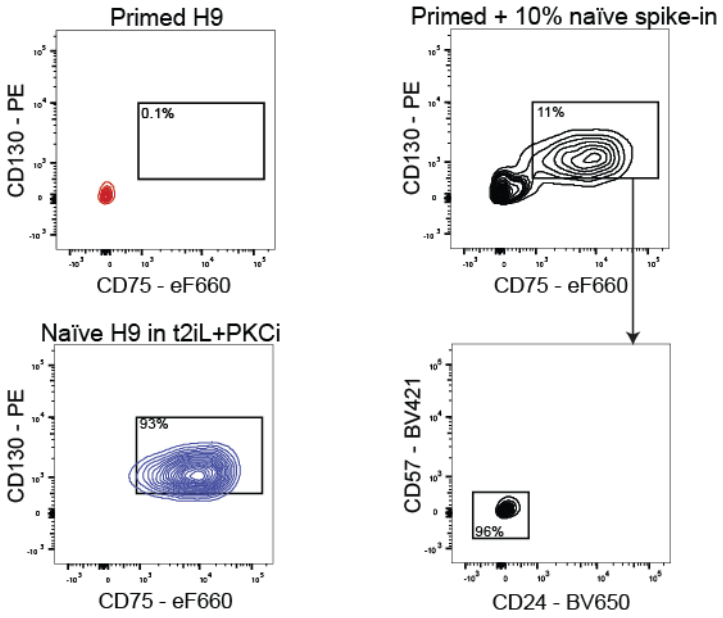


Figure 3.12. Assessing the sensitivity of identified state-specific markers to discriminate between primed and naïve hPSCs.
Flow cytometry contour plots show that the identified panel of state-specific markers can discriminate between primed and naïve hPSCs when the cells are mixed together. Left: the expression levels of two naive-specific markers (CD75 and CD130) in H9 primed (top) and naïve (bottom) hPSCs. Top right: the expression levels of the same markers in a sample of 90% primed hPSCs mixed with 10% naïve hPSCs. Bottom right: Cells positive for both CD75 and CD130 do not express the primed-specific markers CD57 and CD24. Gates were drawn based live, human PSCs which were unstained for any other marker.

3.2.5 Time course naïve to primed differentiation

To further test the utility of the antibody panel, it was important to examine whether the state-specific markers tracked the changes in cell state, rather than a direct response to changes in the signaling environment. I thereby examined the cell surface expression profiles of all markers across a state transition experiment, firstly by transitioning naïve cells to a primed state. To initiate naïve to primed conversion, naïve hPSCs cultured under feeder-free conditions in t2iL+PKCi were switched into TeSR-E8 medium, containing high levels of TGFβ and FGF. Naïve hPSCs progressively changed in their morphology from compact spherical colonies to a flattened, cobblestone primed appearance (Figure 3.13A). I monitored the cells every two days by flow cytometry using the multiplexed antibody panel. Example dotplots for two naïve-specific and two primed-specific markers are shown across the time-course (Figure 3.13B). The expression of naïve-specific markers CD75 and CD130 gradually reduced as the time-course progressed. Likewise, CD57 displayed a progressive upregulation as cells transitioned from a naïve to primed state, with the onset of expression at day four onwards. By contrast, primed-specific marker CD90 was rapidly induced two days into the conversion protocol, to similar levels observed in primed hPSCs. This strongly indicates that CD90 is responsive to the signaling cues provided by the culture media, rather than a pluripotent state change (Figure 3.13B).

FlowSOM analysis was subsequently performed to obtain a complete overview of the dynamics of all markers across the conversion experiment. The flow cytometry data collected for each time-point was merged into a single file. Importantly, FlowSOM has the ability to retain all labelling information, and expression values for each individual cell. Consequently, the FlowSOM MST output can be readily labelled to indicate the cell origin, in this case the transitional time-point. The FlowSOM clustering results show a sequential transition from the naïve to primed state (Figure 3.13C – right panel). Interestingly, the individual cell surface markers exhibit different dynamics during the ten-day time course. However, the majority of markers encouragingly track the change in cell state (Figure 3.13C – left panel). From the FlowSOM clustering results it is also notable that by day 10, CD7 and CD130 have yet to be fully downregulated, and CD24 and CD57 fully upregulated. This infers that more than 10 days are required to transition from a naïve to primed state.

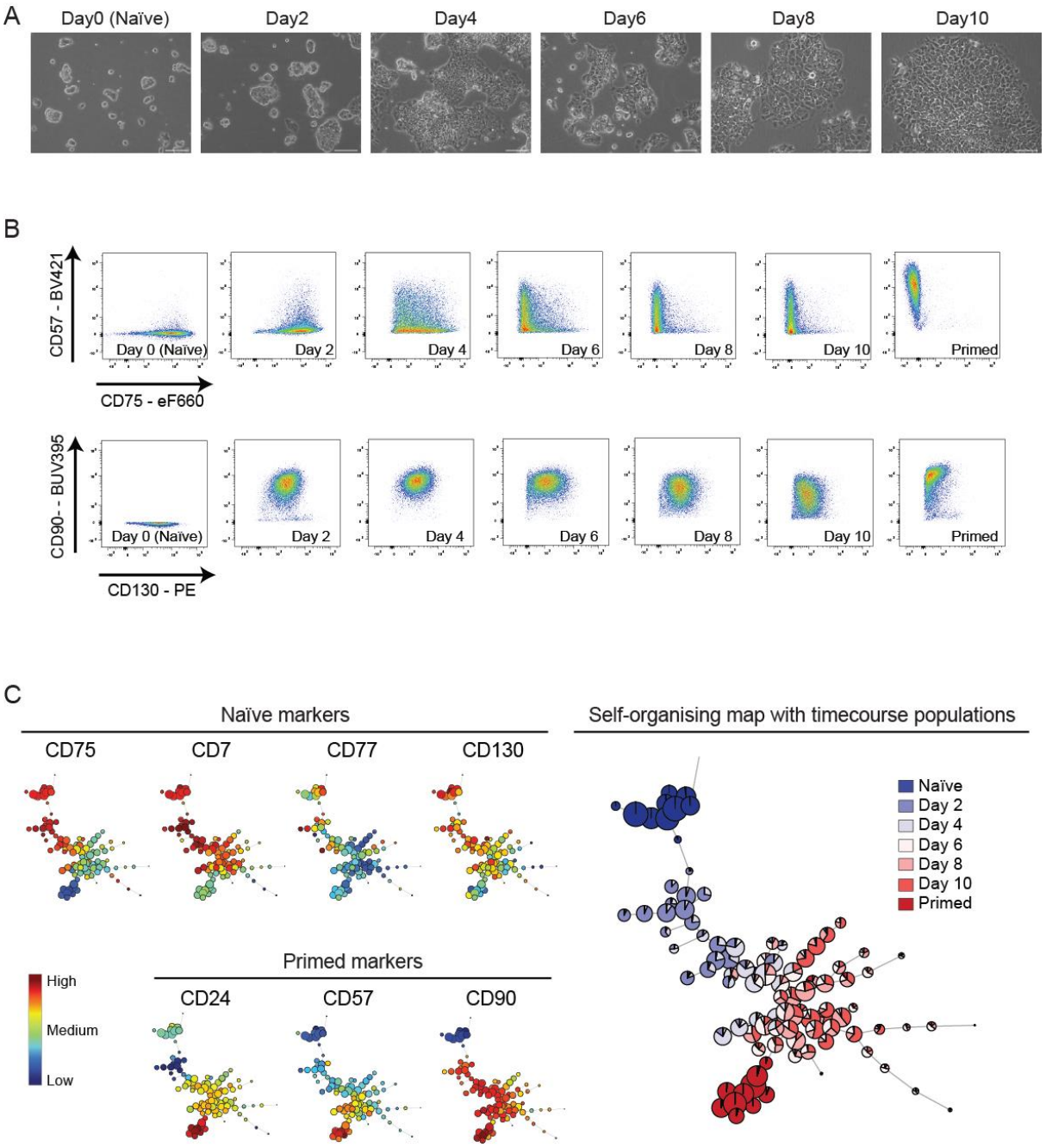


Figure 3.13. Cell Surface marker expression levels track the dynamics of naïve-to-primed hPSC transition.

(A) Phase contrast images of H9 hPSCs reveal the morphological changes that occur during naïve-to-primed state transition by switching t2iL+PKCi culture conditions to TeSR-E8. Scale bars, 100 µm.

(B) Flow cytometry dotplots of pairwise antibody combinations over the time course. Primed-specific markers are shown on the y-axis (CD57, top; CD90, bottom) and naïve-specific markers are shown on the x-axis (CD75, top; CD130, bottom).

(C) FlowSOM visualization of the flow cytometry time course data for H9 hPSCs. The minimal spanning tree of the self-organising map displays unsupervised clustering of the samples based on their cell surface marker expression levels (right). The results reveal a progressive change in cell surface protein expression during conversion from the naïve state to the primed state. The heatmap shows the expression level of each cell surface marker in the cell clusters (left).

3.2.6 Time course primed to t2iL+PKCi naïve reprogramming

The next crucial stage of the analysis was to examine the surface marker expression dynamics, this time going from a primed state to a naïve state. As discussed in the main introduction, all reprogramming protocols typically require >30 days to establish naïve hPSCs at a sufficient purity and cell number for subsequent experiments. The transgene dependent method (NK2 t2iL+PKCi) of reprogramming take approximately 28 days before clonal isolation, or ≈17 days using the condensed protocol (Figure 2.1). As mentioned in section 2.2.10, I further modified this protocol by reducing *NANOG* and *KLF2* induction down to six days, followed by two days culture in t2iL+PKCi (Figure 3.14A). Phase contrast images show the morphological changes during this transition (Figure 3.14B), with dome-shaped naïve colonies formed by day 10.

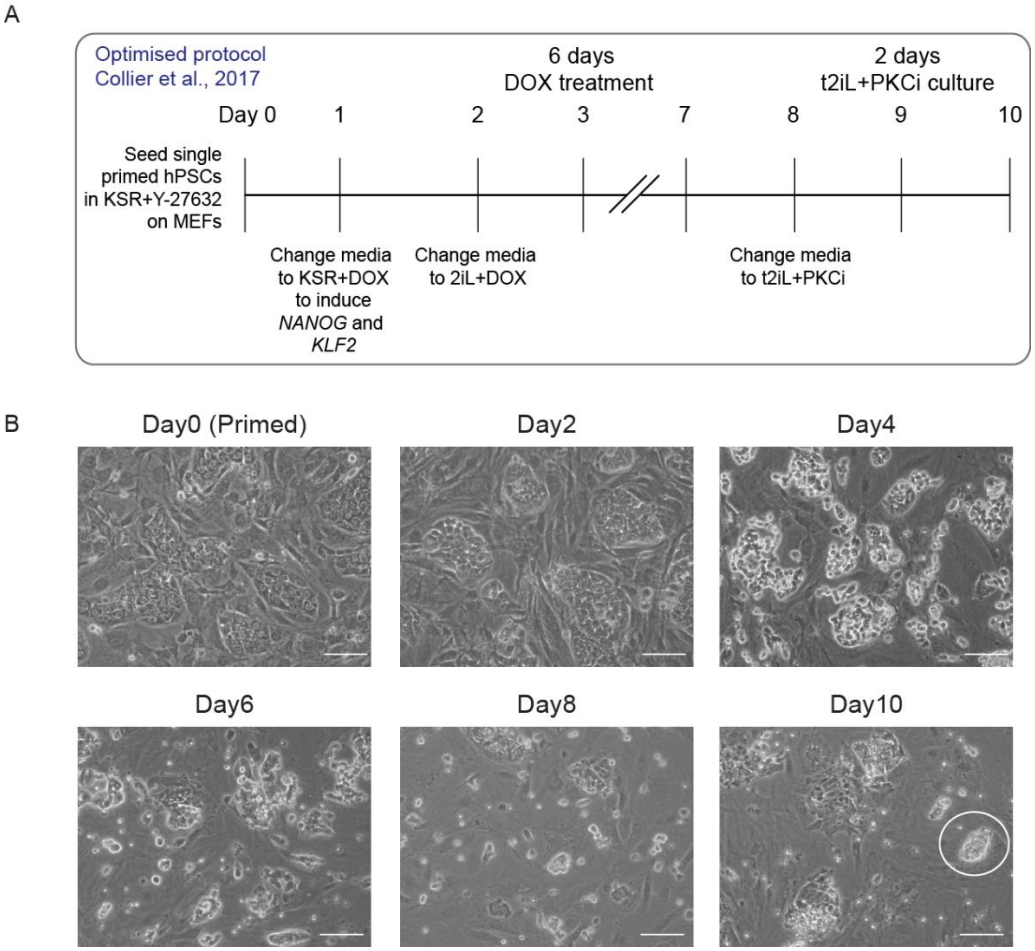


Figure 3.14. Optimisation of NK2 t2iL+PKCi primed to naïve reprogramming method.
(A) Schematic of primed-to-naïve reprogramming protocol. Doxycycline-inducible *NANOG* and *KLF2* transgenes were activated for the first 8 days, and then doxycycline was withdrawn and PKCi was added; culture media was replaced daily.
(B) Phase contrast images of H9 hPSCs reveal the morphological changes that occur during primed-to-naïve reprograming by switching primed (KSR) cultured hPSCs to t2iL+PKCi culture conditions. Scale bars, 100 µm.

To examine whether the surface markers tracked the change in cell state going from a primed to naïve state, I again performed flow cytometry every two days for a period of ten days (Figure 3.15). Primed-specific marker CD57 shows a gradual decrease in expression across the 10 days. Conversely, CD24 displays a more notable decline as early as day 2. Overall, primed-specific markers are silenced in the majority of the population (Figure 3.15A). By contrast, naïve-specific markers CD75 and CD130 are only upregulated in a sub-population of cells, with a clearly distinguishable negative and positive fraction evident by day 10 (Figure 3.15A).

The global overview provided by FlowSOM again exhibits a stepwise transition from the primed to naïve state, largely recapitulating pseudotime. However, along the trajectory are several branch points of divergence at day 8 and day 10. The majority of day 8 reprogramming cells express naïve-specific markers CD7 and CD130; just over half express CD77, but very few cells additionally express CD75 (Figure 3.15B and Figure 3.16). The profile observed for day 10 reprogramming cells is markedly different, with CD75 now detectable in 50-60% of the cells. Perhaps surprisingly, CD130 expression reduces from ≈85% on day 8, to ≈55% by day 10; similarly, CD7 expression declines from ≈90% positive down to ≈40% (Figure 3.16). It appears that the naïve-specific markers are resolved into positive and negative fractions, perhaps attributed to the removal of doxycycline and the addition of the PKC inhibitor between days 8 –10. Alternatively, the cells that a refractory to reprogramming could have a growth advantage and begin to dominant the cultures. Conversely, the naïve culture media may present a selection pressure to induce senescence in the refractory reprogramming populations. Nonetheless, the FlowSOM results highlight that day 10 cells cluster closely to established naïve hPSCs; however, individual clusters composed of both cell types occurs at a very low frequency.

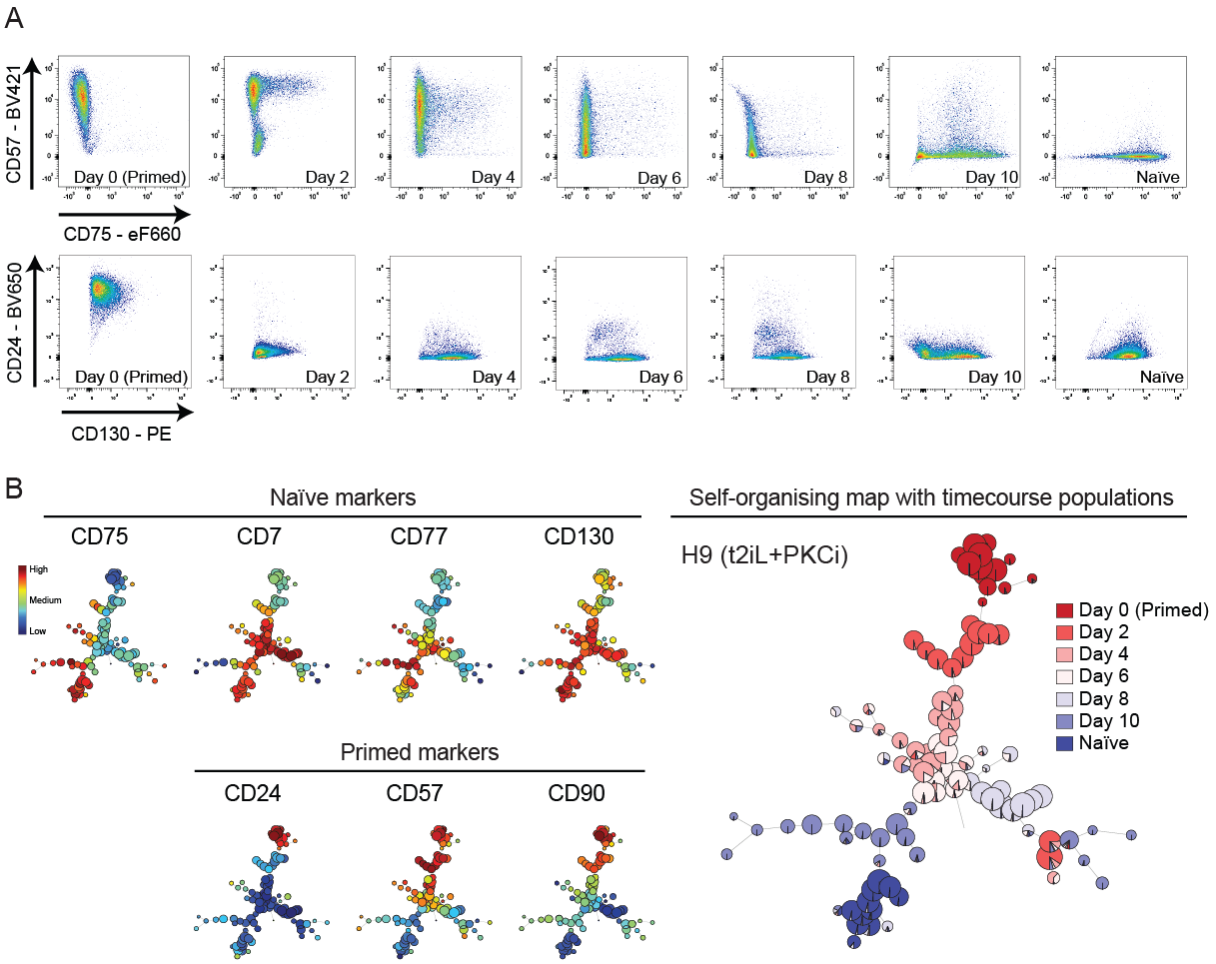


Figure 3.15. Cell Surface marker expression levels track the dynamics of primed-to-naïve hPSC reprogramming.

(A) Flow cytometry dotplots of pairwise antibody combinations over the time course. Primed-specific markers are shown on the y-axis (CD57, top; CD24, bottom) and naïve-specific markers are shown on the x-axis (CD75, top; CD130, bottom).

(B) FlowSOM visualization of the flow cytometry time course data for H9 hPSCs. The minimal spanning tree of the self-organising map displays unsupervised clustering of the samples based on their cell surface marker expression levels (right). The results reveal a progressive change in cell surface protein expression during conversion from primed to naïve pluripotency. The heatmap shows the expression level of each cell surface marker in the cell clusters (left).

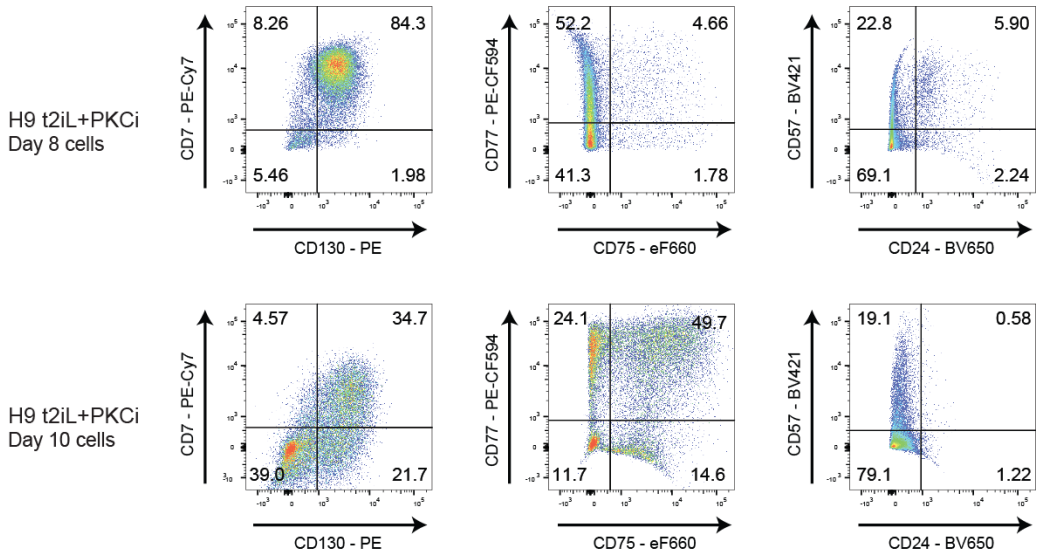


Figure 3.16. Cell Surface marker expression levels resolve upon the transition to the final stages of naïve pluripotency.
Flow cytometry dotplots of pairwise antibody combinations highlighting day 8 (top row) and day 10 samples (bottom row). Naïve-specific markers show different dynamics between the two time points.

3.2.7 Time course primed to 5iLA naïve reprogramming

Given the low efficiency of primed to naïve reprogramming observed with the NK2 t2iL+PKCi method, I wondered whether this would additionally apply to alternative methods, such as 5iLA. I therefore profiled cells in a similar manner, performing flow cytometry with the multiplex antibody panel every two days. Interestingly, the phase contrast images taken throughout reprogramming indicate that the efficiency may be higher using 5iLA. Numerous compact, dome-shaped colonies formed per field of view by day 10 – notably more frequent than observed for NK2 t2iL+PKCi reprogramming (Figure 3.17). In line with this observation, FlowSOM analysis shows that day 10 cells and established 5iLA naïve hPSCs formed well-integrated clusters (Figure 3.18B). The cell surface marker expression profiles display different dynamics between the two reprogramming methods. Markers CD7 and CD77 increase in a proportion of cells as early as day 4. Moreover, CD75 and CD130 are expressed in >50% of cells from day 6 onwards (Figure 3.18A). The OCT4-ΔPE-GFP reporter system has previously been used to identify naïve hPSCs. Providing further validation, a high overlap between the naïve-specific surface marker expression profiles and OCT4-ΔPE-GFP expression occurs (Figure 3.18B).

In summary, the state-specific surface markers collectively track the dynamics of pluripotent state transitions, in both directions. By examining the temporal dynamics of expression onset (for naïve-specific) and expression silencing (for primed-specific markers), the earliest window in which to prospectively identify and isolate putative naïve hPSCs has been identified as day 10 for NK2 t2iL+PKCi reprogramming, and as early as day 8 for 5iLA. These are both much earlier time points that previously possible.

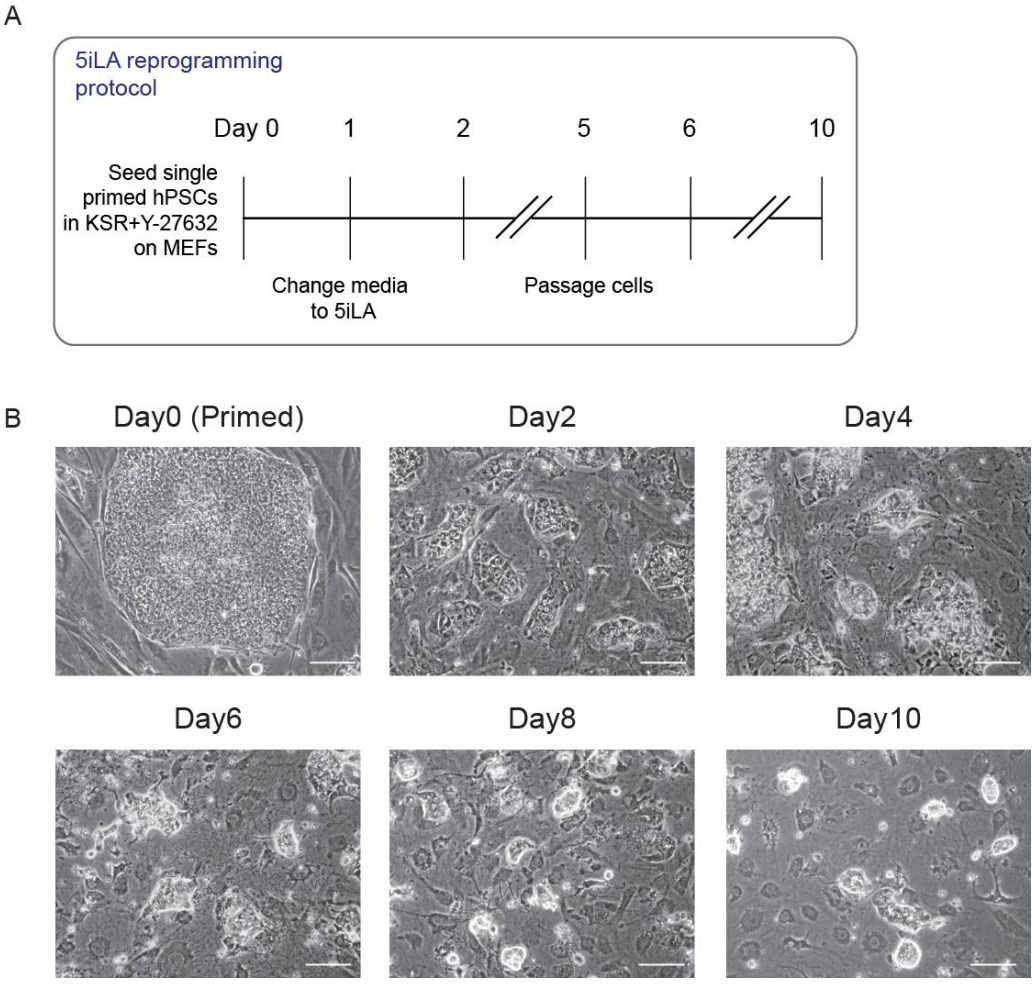


Figure 3.17. Morphological changes associated with primed to naïve reprogramming under 5iLA culture.
(A) Schematic of 5iLA primed-to-naïve reprogramming protocol. Cells were passaged on day 5; culture media was replaced daily.
(B) Phase contrast images of WIBR3 primed hPSCs reveal morphological changes that occur during primed-to-naïve reprograming by switching primed (KSR) cultured hPSCs to 5iLA conditions; media was changed to 5iLA on day 1, 24 hours after plating. Scale bars, 100 µm.

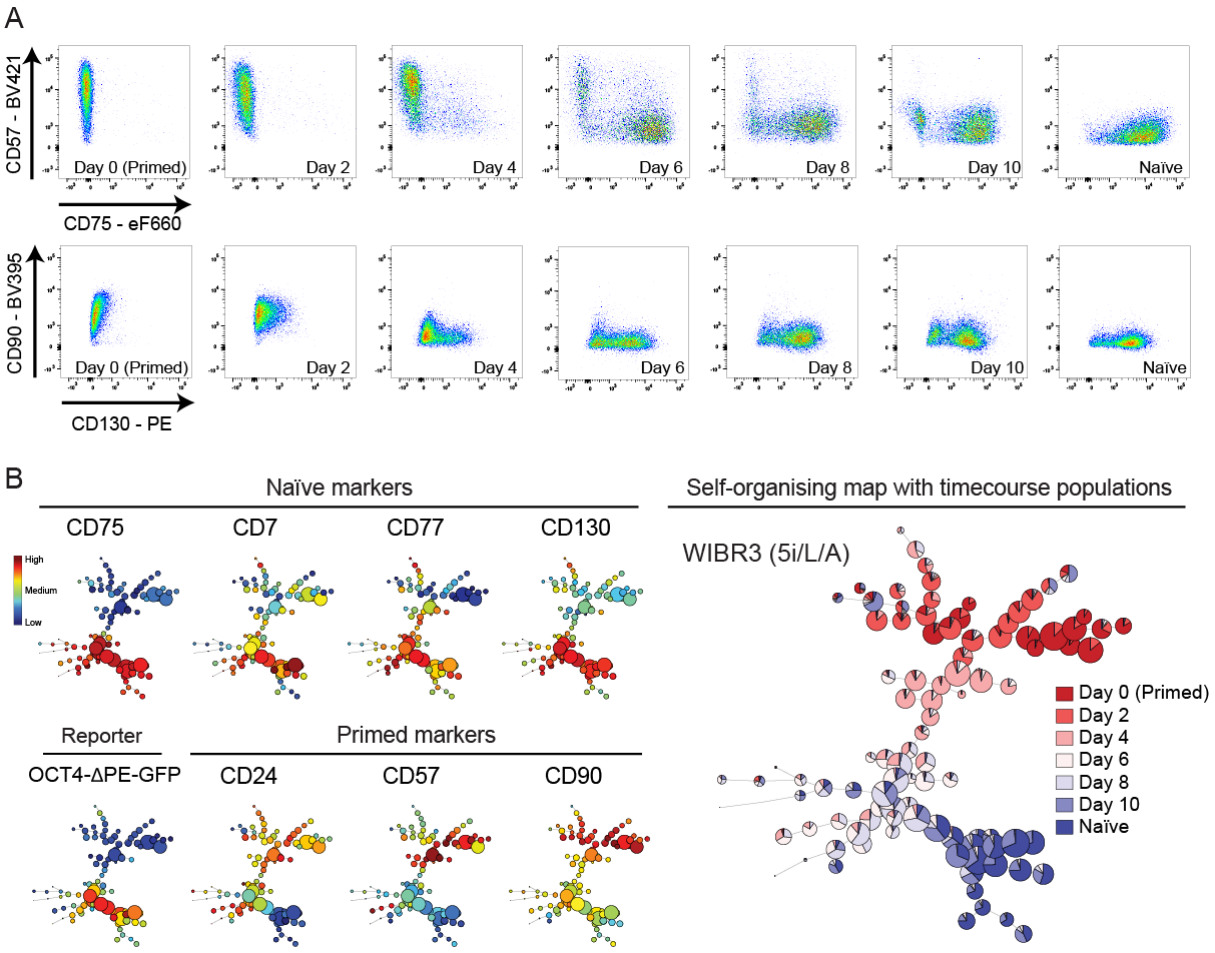


Figure 3.18. Cell Surface marker expression dynamics during primed to naïve reprogramming under 5iLA culture conditions.

(A) Flow cytometry dotplots of pairwise antibody combinations over the time course. Primed-specific markers are shown on the y-axis (CD57, top; CD90, bottom) and naïve-specific markers are shown on the x-axis (CD75, top; CD130, bottom).

(B) FlowSOM visualization of the flow cytometry time course data for WIBR3 hPSCs. The minimal spanning tree of the self-organising map displays unsupervised clustering of the samples based on their cell surface marker expression levels (right). The heatmap shows the expression level of each cell surface marker in the cell clusters (left).

3.3 Discussion

Employing antibody library screens enabled the cell surface marker expression profiles of naïve and primed hPSCs to be comprehensively studied. The approach of using two different naïve culture systems ensured that the classification of state-specific markers was robust and not biased by the culture medium used. Collectively the screens identified over 50 primed-specific markers, 8 naïve-specific markers, and 40 common markers; many of these have not been previously associated with pluripotent stem cells. One of the main advantages to using an antibody library screen was the inclusion of antibodies that detect glycosylated epitopes. Amongst the state-specific surface markers were a large number of glycosylated epitopes, such as CD77, CD75, and several SSEA family members. An alternative method of predicting surface-markers by analysing transcriptional data would have identified naïve-specific markers CD7 and CD130 (IL6ST), which both have assigned genes that are upregulated in naïve hPSCs. However, identifying modified epitopes would not be possible unless the enzymes involved in their catalysis are known. A second benefit is the quantity and diversity of antibodies that were screened, with 377 surface markers examined, a selection of which were tested by two different antibody clones to collectively cover 486 unique antibodies in total. Concordant results were obtained for different clones targeting the same surface-marker, which provided additional verification.

The screens identified eight naïve-specific surface markers that are the first described candidates that can positively distinguish naïve from primed hPSCs. The three previous reports that use surface markers to distinguish these two cell types rely only on primed-enriched surface markers (O'Brien et al., 2017; Pastor et al., 2016; Shakiba et al., 2015a). It is still unknown whether these markers accurately track the dynamics of primed to naïve conversion, by which expression is lost specifically and only once primed cells have reprogrammed. The lack of a known positive marker for human naïve PSCs has restricted the ability to unambiguously identify and isolate naïve hPSCs within a heterogeneous reprogramming population. To resolve this, I developed a multiplexed panel of antibodies consisting of three primed and four naïve-specific markers. By testing the panel on different PSC lines, I could detect line-specific differences. Similarly, I examined three different culture mediums: t2iL+PKCi, 5iLA and ReST medium (based upon NHSM). The results for t2iL+PKCi and 5iLA were comparable, and the state-specific markers were clearly able to distinguish between naïve and primed cell types. However, RSeT naïve-like cells showed a peculiar expression profile in that none of the naïve-specific markers were expressed. Interestingly, primed-specific marker CD24 had been downregulated by passage 4 under RSeT conditions (Figure 3.6). However, Shakiba and

colleagues found that it took ten passages under NHSM conditions, indicating that RSeT medium may induce naïve-like characteristics earlier than NHSM. Even so, it is clear that this naïve formulation does not derive cells with the desired naïve hallmark characteristics, such as a hypomethylated genome and the presence of two active X-chromosomes (Table 1.1, Table 1.2)(Pastor et al., 2016; Vallot et al., 2017). For this reason, I continue the focus of this body of work using t2iL+PKCi and 5iLA.

Given the encouraging results that the multiplexed antibody panel could distinguish naïve and primed hPSCs, our collaborators at the Karolinska Institute examined their expression in blastocyst stage E6 human embryos. Three of the naïve-specific markers, CD75, CD77 and CD130, were detected in the epiblast, whereas all primed-specific markers were absent. Naïve-specific markers CD75 and CD77 were also expressed in the extraembryonic cells of the embryo. This indicates that their specificity is not restricted to pluripotent cells, and that naïve hPSCs may encompass properties of both embryonic and extraembryonic cell types – a point that I will discuss further in later sections. Contrary to this, naïve-specific marker CD7 was not expressed in human embryos, inferring that naïve hPSCs may possess characteristics that do not fully equate to the epiblast cells of human preimplantation embryos.

Whilst the roles of CD7 are not well characterised, the promoter region of CD7 interestingly contains a transcription factor binding site for STAT3 (Zhang et al., 2013); it is therefore possible that JAK/STAT signaling may be utilised more in naïve hPSCs than in human blastocysts, thus the discrepancy in CD7 detection. Alternatively, CpG sites within the first exon of CD7 are methylation sensitive (Röhrs et al., 2010). Treatment with the DNA demethylating agent Aza-2'deoxycytidine induced CD7 expression in human CD7- acute myeloid leukemia and T cell lines (Röhrs et al., 2010). It is possible that this region could be hypomethylated in naïve hPSCs, but not in the human embryo. Collectively, the similarities and discrepancies observed with surface marker expression between naïve hPSCs and human embryos, provides one interesting direction for future investigation to understand their functional roles in pluripotency and signalling. Moreover, it would be interesting to examine recently developed models of postimplantation-stage embryos to validate the relevant stage-specificity of the primed-specific markers (Deglincerti et al., 2016; Shahbazi et al., 2016). Similarly, these models open up the possibility to study a time point in human embryonic development that is otherwise inaccessible. These models, alongside human preimplantation embryos, might also facilitate establishing where the different naïve and primed hPSCs fall along the human embryonic developmental trajectory.

The biological role for the naïve-specific markers CD75 and CD77 are largely uncharacterised, especially in the context of pluripotency, with no previously reported indications to their function. Little is known for each surface marker, asides for both being glycosylated epitopes that mark germinal centre B cells and B cell lymphomas (Arends et al., 1999; Billian et al., 1996). The surface of mammalian cells are decorated by complex and diverse protein and lipid linked glycan structures, which vary between different cell types (Berger et al., 2016). One of the first markers to identify mouse embryonal carcinoma and mouse embryonic stem cells was the carbohydrate-based epitope SSEA1 (Solter and Knowles, 1978). Subsequently, a range of antibodies against glycosylated epitopes such as SSEA3, SSEA4, TRA-1-60 and TRA-1-81, are all commonly used to identify human embryonic stem cells (Kannagi et al., 1983; Wright and Andrews, 2009), and were all expressed in primed cells in the antibody screen, with TRA-1-60 and TRA-1-81 also expressed in naïve hPSCs. However, the biological role for these four glycosylated epitopes are also not well understood, perhaps owing to their dispensability in the maintenance of human pluripotency *in-vitro* (Fenderson et al., 1993; Wright and Andrews, 2009). The lacking mechanistic insight into the function of glycoproteins presents an interesting, yet challenging avenue for further investigation. Possible experiments to decipher their function and requirement could include the use of neutralising antibodies against their epitopes. Alternatively, it may be possible to use antibody-mediated immunoprecipitation in combination with mass spectrometry to identify their interacting proteins, which might provide an insight into their target pathways and subsequent biological function in pluripotency.

The identified markers provide not only a means by which to distinguish primed from naïve hPSCs, but they also reveal potential differences in signaling pathways between the pluripotent states. An interesting example being CD130, otherwise known as the LIF co-receptor. The role of LIF in embryonic development was first recognised when several reports showed that LIF was required for the propagation of mESCs, through the activation of the JAK-STAT signaling pathway (Smith et al., 1988; Williams et al., 1988). However, this was not the case for the derivation and propagation of human PSCs (Reubinoff et al., 2000; Thomson et al., 1998). Moreover, hPSCs could not be maintained by the activation of LIF signaling in the absence of mouse embryonic feeder cells (Carpenter et al., 2003; Dahéron et al., 2004; Humphrey et al., 2004). Similarly, low expression levels of CD130 (*IL6ST*) indicated that LIF signaling was not active in hPSCs (Brandenberger et al., 2004). This observation is likely attributable to the derivation of hPSCs under primed culture conditions, whereas we now know that LIF/STAT3-dependent signaling is required for the maintenance of naïve hPSCs (Chan et al., 2013; Gafni et al., 2013; Hanna et al., 2010; Ware et al., 2014). Given that all the culture formulations to

derive naïve hPSCs contain 2i/LIF, perhaps it is not surprising to see enrichment of CD130. Nonetheless, future work is necessary to clarify the role of LIF signaling during human embryonic development and pluripotent state transitions.

Other interesting signaling proteins identified include the NOTCH family of receptors and NOTCH ligand JAGGED2, expressed only in primed hPSCs. Whilst I see that NOTCH1-3 are expressed in undifferentiated primed hPSCs, previous analysis of the downstream targets reveals that NOTCH signaling is inactive in this state (Yu et al., 2008). However, upon exposure to differentiation cues, NOTCH signaling is both activated and required for lineage commitment to the three germ layers (Yu et al., 2008). This suggests to me that NOTCH receptors are present on the surface of primed hPSCs to poise the cells for efficient entry into lineage specification. Perhaps it is therefore not surprising that NOTCH proteins are absent in the naïve state if their function is not imminently required.

To further test the utility of the state-specific markers, I examined whether the panel could be used to identify a small population of naïve hPSCs which had been mixed with primed cells. Reassuringly, the naïve hPSCs could be captured under gating for CD75+/CD130+ expression. However, these are both established cell states, which will not be the case during reprogramming. It was therefore necessary to examine the dynamics of each marker during primed to naïve state interconversion to check their ability in a heterogeneous setting. The results firstly reiterated their state-specificity, but secondly revealed that most could accurately track the change in pluripotent state. For example, CD57 and CD130 changed in their expression gradually during reprogramming, inferring these are coupled to a progressive change in cell state. By contrast, CD90 is rapidly downregulated at the onset of reprogramming. *THY1* (CD90) is a predicted FGF target gene and expressed accordingly in primed hPSCs cultured in media containing high levels of FGF (Choi et al., 2001; Kaufman et al., 2001). The switch to t2iL+DOX media that contains a potent FGF/MEK/ERK inhibitor (PD0325901) is likely responsible for the rapid downregulation. Consequently, CD90 is less reliable as an individual marker.

An interesting observation was made when comparing the different primed to naïve reprogramming methods, 5iLA and t2iL+PKCi; the proportion of day 8 - day10 reprogramming cells that clustered with established naïve hPSCs was greatly increased using 5iLA, compared to t2iL+PKCi (FlowSOM results: Figure 3.16 and Figure 3.18) . This suggests that reprogramming may be more efficient using 5iLA. A possible explanation for this could be the addition of inhibitors that collectively target more signalling pathways than under t2iL+PKCi

culture. Alternatively, the selection pressure to reprogram may be much stronger in 5iLA, consequently cells that do not reprogram undergo programmed cell death, which could be misinterpreted as being more efficient. Even so, it would be of interest to further examine the two reprogramming methods and the different routes to naïve human pluripotency.

Whilst the focus of the antibody library screening was to identify markers that can distinguish between naive and primed human hPSCs, the availability of an extensive catalogue of markers present on the cell surface of hPSCs should be valuable for the study of human pluripotency and differentiation more generally. Moreover, surface markers enable the straightforward isolation of live cells that can be coupled with downstream functional assays. This has several advantages over current methods that rely on an "expert eye" to handpick individual colonies for expansion, or by long-term propagation to purify the cultures. Applying the antibody panel to reprogramming populations will provide a new insight into the earliest stages of naïve cell formation; this will help to unveil the temporal dynamics of molecular changes that occur, and ultimately facilitate the discovery of what regulates these processes.

4. Cellular and Molecular Characterisation of Human Naïve Pluripotent Stem Cells at an Early Stage of Reprogramming

4.1 Background

Much of the knowledge that we currently have regarding the two states of human pluripotency is based upon studies of the established states. Whilst this provides a valuable insight into the gene regulatory networks for primed and naive cell types, we know very little about the transitional period. Isolating and characterising reprogramming intermediate cell types can provide a useful insight into the trajectories of cell fate changes and the mechanisms that govern the routes taken. Several reports have used cell surface markers to examine the transcriptional waves and alterations in the epigenetic landscape that facilitate iPSC reprogramming (Cacchiarelli et al., 2015; Knaupp et al., 2017; O'Malley et al., 2013; Polo et al., 2012). By studying the intermediate populations that are refractory to reprogramming, we can begin to understand why cells fail to reprogram, and subsequently use this knowledge to coerce cells down the right trajectory. Moreover, we can identify the order of molecular events and regulators that lead to successful reprogramming.

As previously discussed, the current methods used to isolate naïve hPSCs preclude the study of the early stages of reprogramming. Primed to naïve reprogramming generates a heterogeneous mix of cells, of which only a small proportion will be successfully transitioning. It is therefore necessary to enrich for this population of cells to faithfully track the order of molecular changes that occur during reprogramming. To address this gap in knowledge, I have identified cell-surface markers that I believe will be capable of isolating naïve hPSCs as early as day10 during primed-to-naïve reprogramming. Under t2iL+PKCi culture, a small population of day 10 cells clustered with established naïve hPSCs based upon their cell surface expression profile. I hypothesise that this population will be the nascent naïve hPSCs that will ultimately progress to form the reprogrammed population. If this is true, it would offer an opportunity to isolate naïve hPSCs at a time point far earlier than previously studied. Moreover, it would be possible to track this population at time points throughout reprogramming to examine how and when remodelling of the transcriptional and epigenetic landscape occurs.

4.1.1 Aims

1. Isolate early naïve hPSCs and reprogramming populations using FACS
2. Characterise early naïve hPSCs and reprogramming populations
    i. Assess the ability to form colonies from single cells

ii. Examine the expression of known naïve and primed state-specific markers – both transcriptionally and at the protein level
iii. Examine global transcriptional profile of early naïve hPSCs and reprogramming populations
iv. Examine epigenetic hallmarks of naïve pluripotency in early naïve hPSCs and reprogramming populations – X-chromosome status and DNA methylation levels

4.2 Results

Defining and characterising reprogramming and intermediate cell states can provide a useful insight into the mechanisms of cell fate changes. I therefore sought to use the state-specific surface markers to isolate nascent naïve hPSCs and reprogramming intermediates at an early time point during reprogramming.

4.2.1 Using cell surface marker expression profiles to define reprogramming populations under NK2 t2iL+PKCi culture

The earliest time-point at which I could hope to isolate nascent naïve hPSCs using the NK2 t2iL+PKCi method is 10 days into reprogramming (Figure 3.15). Focusing specifically on the surface marker expression profiles of day 10 cells, several distinct populations were apparent. The first, designated N4+ corresponds to the most "naïve-like" population, defined by high expression of all four naïve-specific markers, combined with low expression of primed-specific marker CD24 and CD90 (Figure 4.1). This population represents ≈2% of all live-human day 10 cells. The next population, termed N3+ expresses three naïve-specific markers: CD77, CD7 and CD130, but not CD75. Similarly, gating isolates the CD24 and CD90 low fraction. This population represents a larger proportion of cells at ≈6%. Finally, a third population with low or absent expression of naïve specific markers is apparent, designated N4- representing ≈22% of day 10 cells. I opted to isolate this population irrespective of primed-specific marker expression, as only 3% of cells lowly express CD24 and CD90. The same gate shown on all three plots is to emphasize the selection of cells that I deemed lowly expressing primed-markers (Figure 4.1 - right). This is more apparent when displaying N4- cells, as a clear CD24 positive-fraction is visible. By contrast, N4+ and N3+ populations exhibit a continuum of CD24 expression.

To gain an overview of all potential sub-populations amongst day 10 cells I performed FlowSOM analysis; this time highlighting the populations described above (N4-, N3+ and N4+) onto the MST. The three populations are distinct from each other and largely represent the predominant clusters observed using this unbiased analysis.

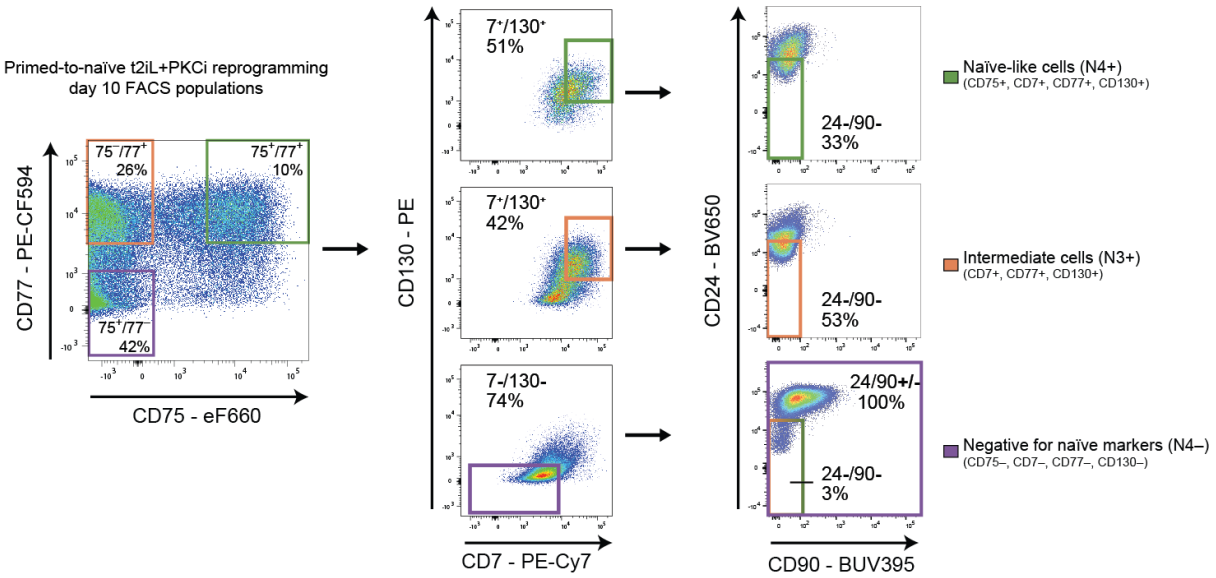


Figure 4.1. Flow cytometry gating scheme to isolate early-naïve reprogramming populations by FACS.

Flow cytometry dotplots of day 10 cells during primed to naïve-state reprogramming of H9 hPSCs under t2iL+PKCi culture conditions. Left: Expression levels of two naïve specific markers, CD75 and CD77. Gates were constructed based on control live, human stained day 10 samples. Three gates on the left correspond to CD75+/CD77+ (green box), CD75–/CD77+ (orange box), and CD75–/CD77– (purple box) cell populations. Middle: the levels of CD7 and CD130 proteins for the same three gated
cell populations. Right: Expression levels of two primed specific markers, CD24 and CD90. Boxed areas indicate the N4+ (green) and N3+ (orange) that were gated based on low expression of these markers. N4– (purple) cell populations included all cells irrespective of primed-marker expression. The cells highlighted in each gate were used for subsequent experiments. The percentage of cells within each cell sorting gate relative to all live-human cells, or the previous gate is shown.

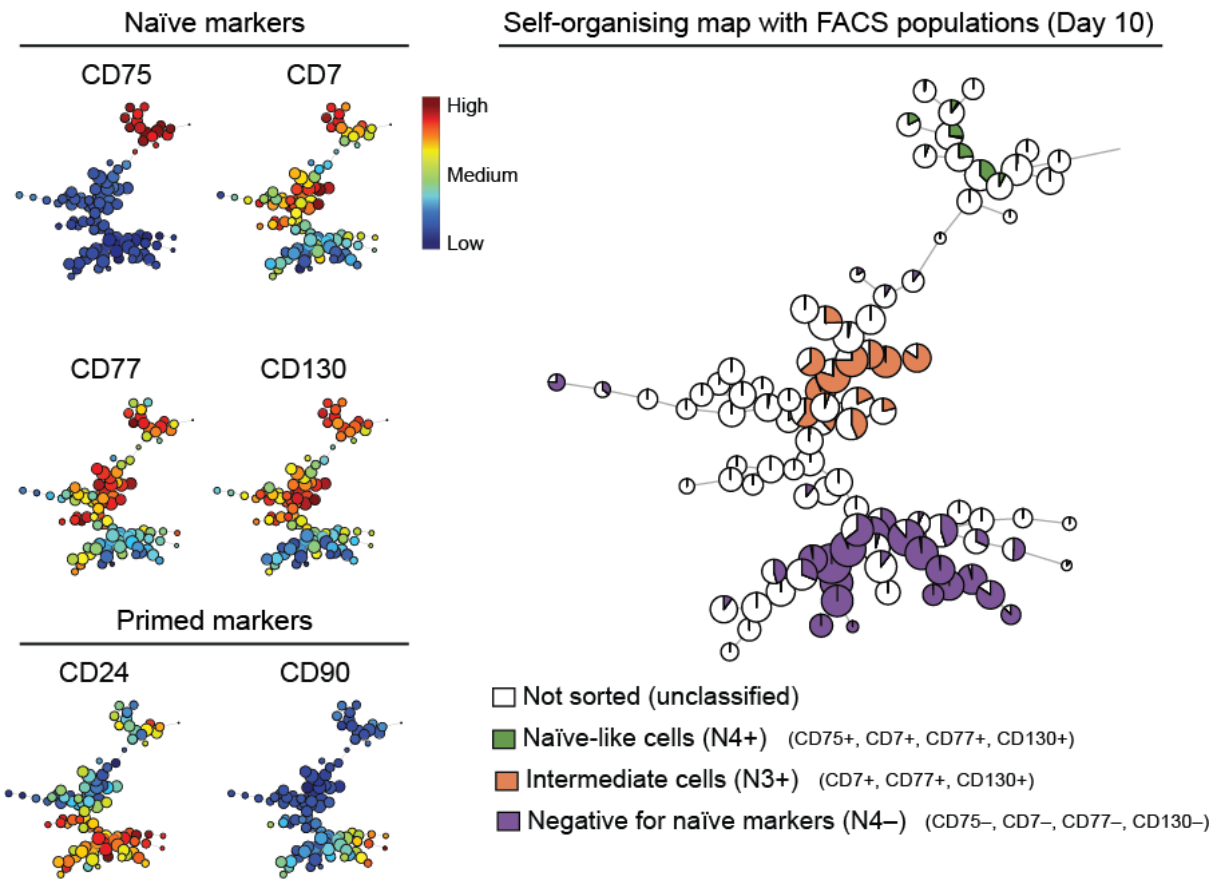


Figure 4.2. Prospective isolation of day 10 naïve reprogramming populations by FACs.
FlowSOM visualisation of the flow cytometry data for day 10 cells during primed to naïve reprogramming. The minimal spanning tree of the self-organising map displays an unsupervised clustering of the sample based on the cell surface protein expression levels (right). The cells corresponding to each cell sorted population, N4+, N3+, and N4–, are indicated. The heatmap shows the expression level of each cell surface protein marker in the cell clusters (left).

Whilst using a multiplexed panel with four naïve and two or three primed-specific markers is likely to identify a more stringent naïve-like population, there are a number of caveats to its utility. Firstly, there is the requirement for a high-end flow analyser or FACS machine with five lasers, which may impose a restriction on usability. Moreover, users must be familiar with multi-parameter flow cytometry compensation. Secondly, the degree of flexibility over compatible conjugations is restricted, as discussed in section 3.2.2. It could be problematic if users wish to examine additional surface markers or reporter expression. Finally, purchasing 8-9 antibodies required for the complete antibody panel may be financially unappealing to wider usability. With these factors in mind, I opted to compile a minimised antibody panel containing two naïve-specific and two primed-specific surface markers. The naïve-specific markers chosen were CD75 and CD130, based upon the previous panel and time-course results. CD75 is one of the most stringent naïve-specific markers, and CD130 has well-established biological importance as the LIF co-receptor. I selected CD24 and CD57 as the primed-specific markers, as they both appear to be more indicative of a pluripotency state change than CD90.

The population of interest using the minimised panel is designated N2+, a population positive for both naïve-specific markers and negative for primed markers. Shown are the flow cytometry results of day 10 cells after NK2 t2iL+PKCi reprogramming (Figure 4.3A). The N2+ population represents 4-5% of live-human cells. As anticipated, the minimal panel identifies a broader population compared to N4+. FlowSOM analysis of day 10 cells reveals N2+ cells cluster together on a distinct branch of the MST (Figure 4.3B).

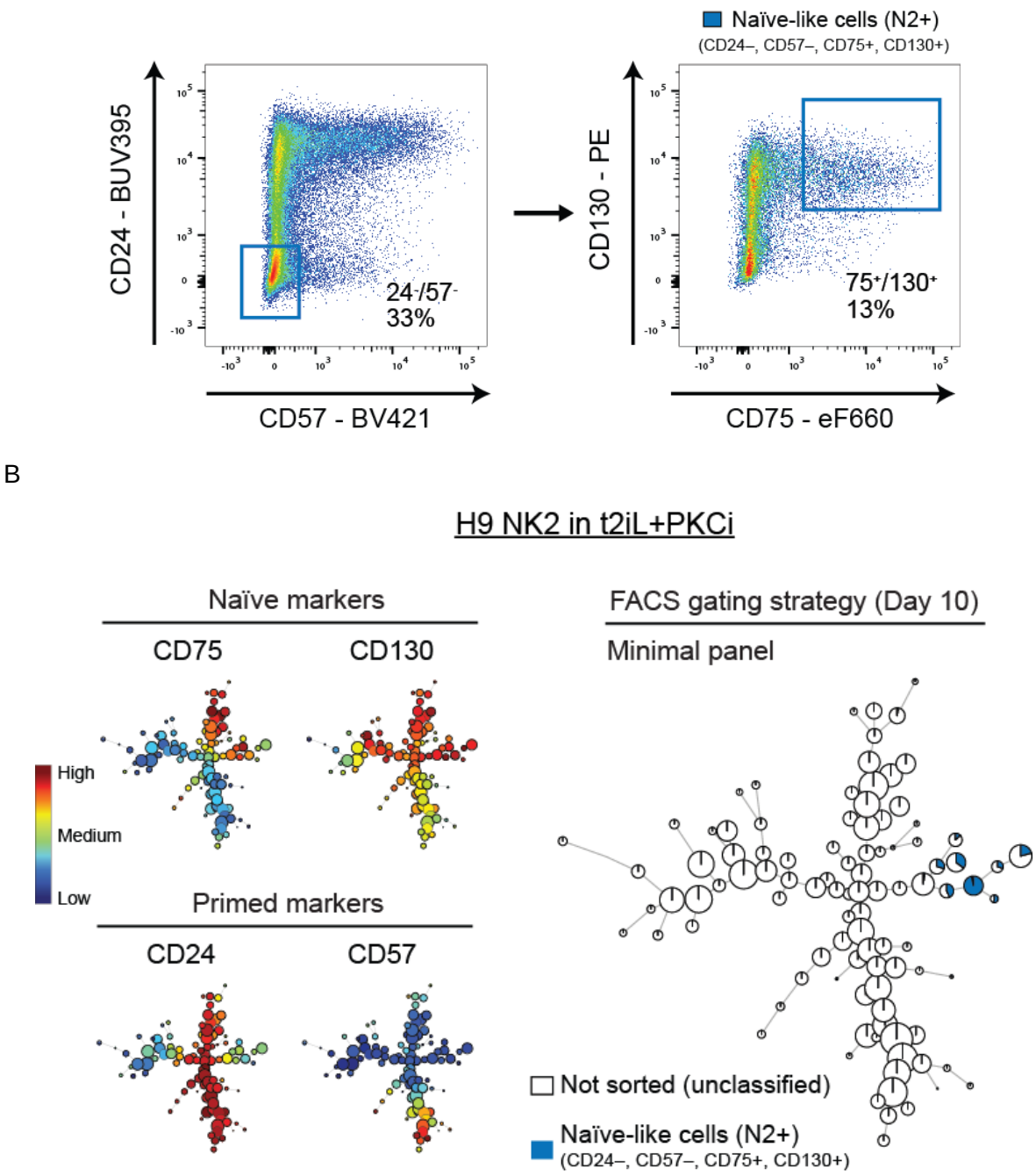


Figure 4.3. A minimised antibody panel to isolate early-naïve reprogramming cells.
(A) Flow cytometry dotplots of day 10 cells during primed to naïve-state reprogramming of H9 PSCs under t2iL+PKCi culture conditions. Left: Expression levels of two primed-specific marhkers, CD24 and CD57. The double negative population (CD24-/CD57-) was gated (blue box). Right: The expression levels of two naïve-specific markers CD75 and CD130. The double positive population (CD75+CD130+) was gated (blue box) and corresponds to N2+ cells that was used for subsequent experiments. The percentage of cells within each cell sorting gate relative to all live, human cells is shown.
(B) FlowSOM visualisation of the flow cytometry data for day 10 cells during primed to naïve reprogramming. Right: the cells corresponding to the N2+ cell sorted population is indicated on the minimal spanning tree (blue). Left: the heatmap shows the expression level of each cell surface marker in the cell clusters.

4.2.2 Characterisation of reprogramming populations

4.2.2.1 Colony formation assay to examine clonogenicity of reprogramming intermediates

One of the hallmarks of naïve pluripotency is the attribute of high clonogenicity compared to their primed counterparts. I examined this attribute by isolating each reprogramming population (N4+, N3+, N4- and N2+) and comparing their ability to form colonies from single cells against established naïve hPSCs. All populations, including established naïve hPSCs, underwent the same FACS protocol to ensure a robust comparison. Four days after isolating the various populations, I counted both the number of colonies that had formed and scored their morphology (Figure 4.4). Both N4+ and N2+ populations have the equivalent ability to form colonies from single cells when compared to established naïve hPSCs. Importantly, the colonies formed are predominantly naïve in morphology. By contrast, N3+ sorted cells form significantly fewer naïve scored colonies, indicating that the expression of CD75 is a key marker of nascent naïve hPSCs. Moreover, N4- sorted cells have an impaired ability to form colonies, with only a small number of primed and differentiated colonies observed.

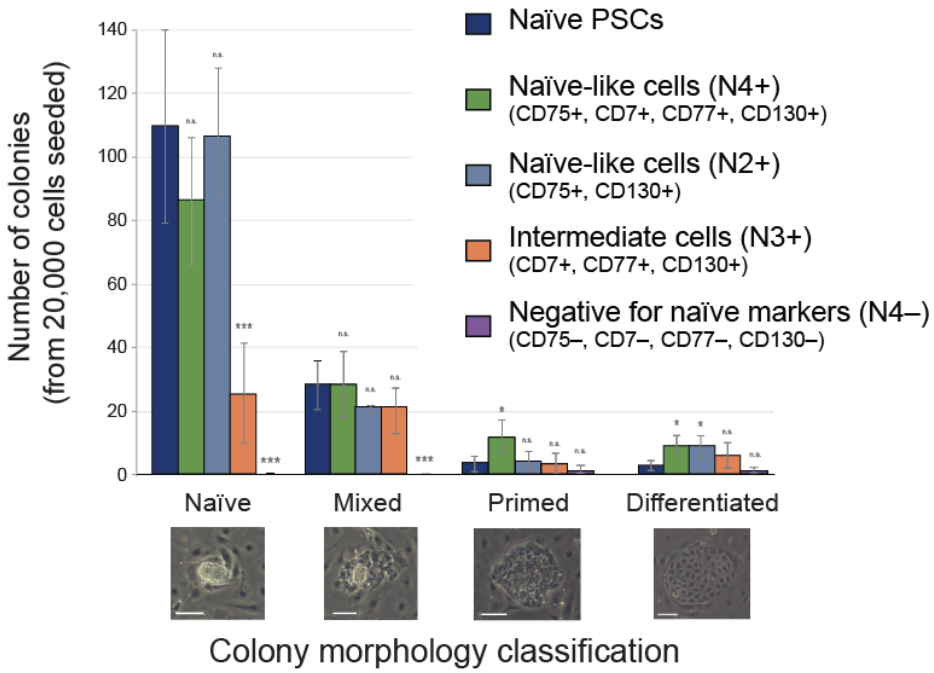


Figure 4.4. Colony formation efficiency of cell sorted populations into t2iL+PKCi naïve culture conditions.
Colonies scored as naïve, mixed, primed and differentiated: examples shown below. Data show the mean ± SD of three or four biological replicates and were compared to established naïve hPSCs using an ANOVA with Dunnett's multiple comparisons test (*$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$). Scale bars, 100 µm.

4.2.2.2 Transcriptional profiling of reprogramming intermediates for pluripotent state-specific transcription factors

One method by which to predict cell identity is to examine the transcriptional profile. This can be performed on a genome wide level using RNA-sequence, or via targeted RT-qPCR when transcriptional markers are known. To gain an insight into the fate of each reprogramming population and how each compare transcriptionally to established naïve hPSCs, I performed RT-qPCR for various genes associated with primed, naïve, and shared pluripotency (Figure 4.5). The comparison of N4+, N3+ and N2+ populations revealed no significant difference in the expression of core pluripotency factors *POU5F1*, *SOX2* or *NANOG*, when compared to established naïve cells. However, N4- cells show a decrease in *POU5F1* and *NANOG* expression, combined with an increase in *SOX2* transcription. This result combined with the inability to form colonies suggests an impairment of pluripotency. Moreover, N4- cells expressed primed-specific genes *OTX2* and *ZIC2* to higher levels than primed hPSCs. This is not the case for any other reprogramming intermediate, or established naïve hPSCs; none of which express primed-specific genes. Perhaps unsurprisingly, N4- cells also show a reduction in the transcription of naïve-specific genes. These findings suggest that N4- cells adopt a neural-like expression profile, likely caused by FGF inhibition during naïve reprogramming (Greber et al., 2011). Conversely, the expression of naïve-specific transcription factors *KLF17*, *KLF4*, *TFCP2L1*, *DPPA3* and *DNMT3L* is comparable for N2+, N3+ and N4+ cells to each other, and to established naïve hPSCs. The main exception being a ≈40 fold decrease in *KLF17* expression in N3+ cells. This is concordant with the colony formation assay highlighting a significant decrease in naïve cell formation; suggesting that N3+ cells, which lack CD75 expression, represent a partially reprogrammed cell type.

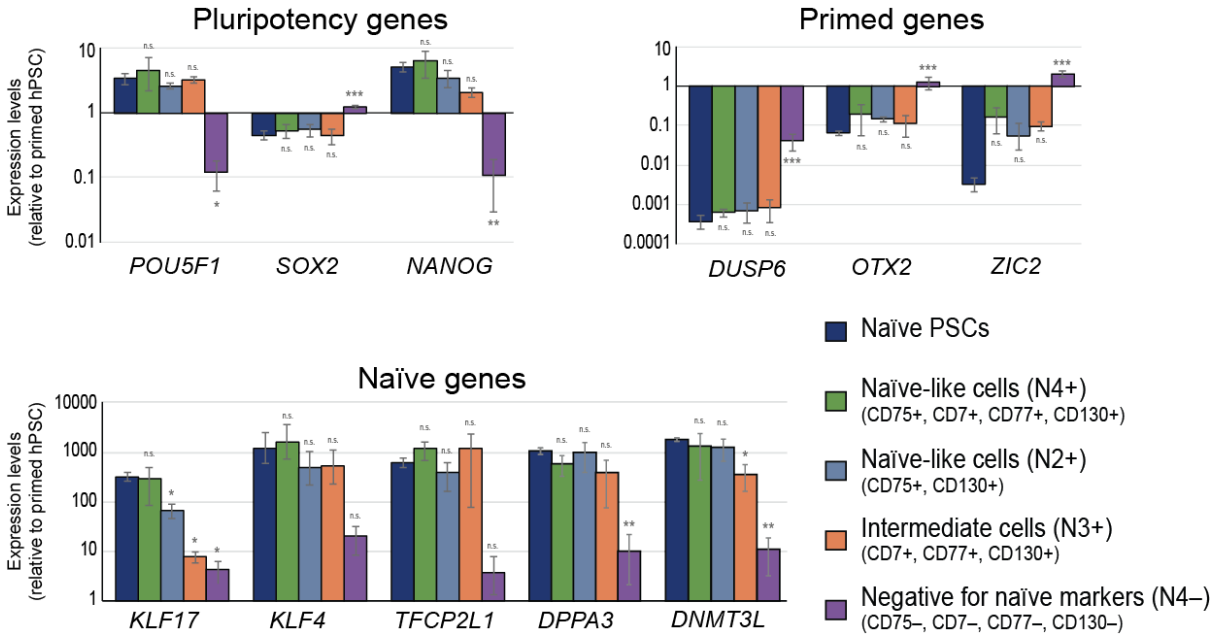


Figure 4.5. Comparative gene expression analysis for pluripotent state-specific genes across the different cell-sorted populations.
RT-qPCR analysis of gene expression levels in the different cell-sorted populations and established naïve hPSCs. Expression levels are shown on a log scale relative to primed hPSCs. Data show the mean ± SD of three or four biological replicates and were compared to established naïve hPSCs using an ANOVA with Dunnett's multiple comparisons test (*p < 0.05, **p < 0.005, ***p < 0.0005).

4.2.2.3 Isolated nascent naïve hPSCs form homogeneous naïve cell cultures

An additional readout of cell fate is to examine the protein expression and cellular localisation of key transcription factors. With this in mind, I examined the expression of core pluripotency factor OCT4 (*POU5F1*), which is common to both naïve and primed hPSCs, and naïve-specific factor KLF17. Representative immunofluorescence images of naïve hPSCs show a more compact morphology, with clear expression of both KLF17 and OCT4 (Figure 4.6). By contrast, primed hPSCs were negative for KLF17 and positive for OCT4. The results for two reprogramming intermediate populations (N4+ and N2+) are analogous to naïve hPSC results. A compact morphology is visible, combined with dual expression of OCT4 and KLF17.

To ascertain whether N4+ and N2+ isolated populations are the cells that will go on to form stable naïve hPSCs, I cultured the cells for three passages in t2iL+PKCi after FACS isolation. The cultures from both populations are homogeneous in nature and do indeed form stable cultures. I have subsequently passaged these cells greater than 20 times without deterioration in morphology.

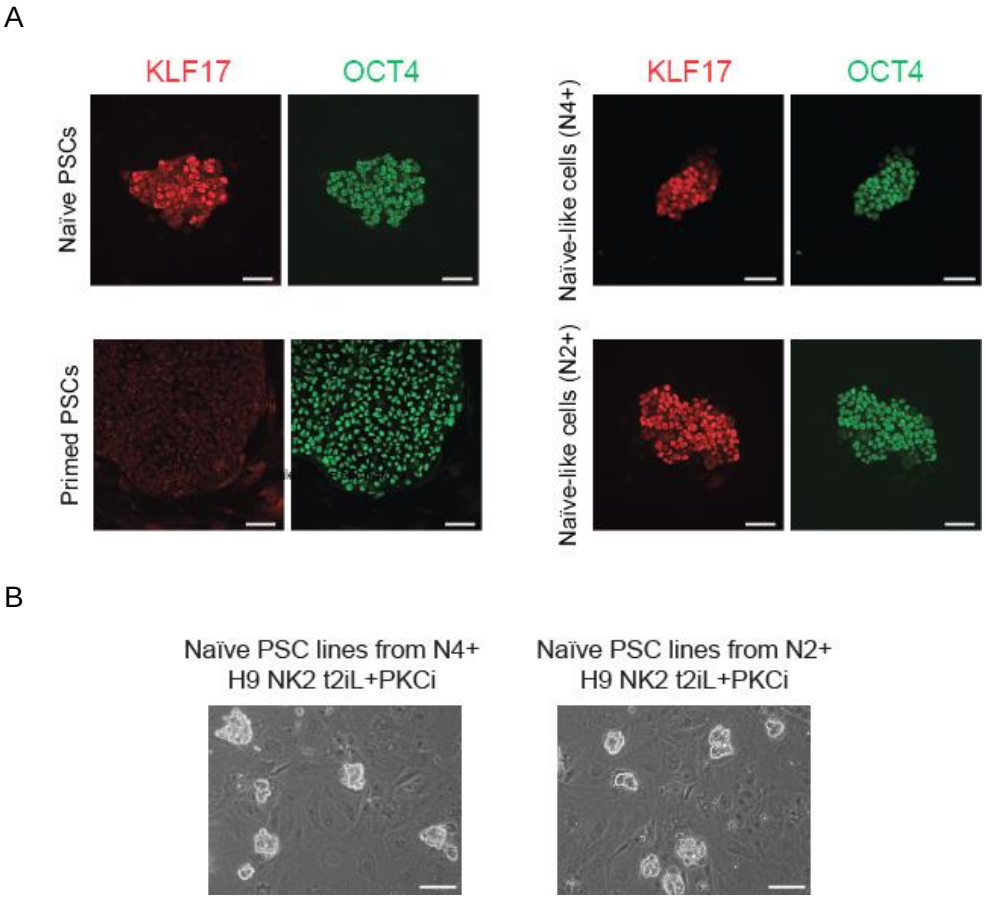


Figure 4.6. Early-naïve sorted hPSCs exhibit naïve pluripotency characteristics.
(A) Immunofluorescence microscopy for KLF17 (a naïve-specific protein) and OCT4 (a protein expressed by naïve and primed hPSCs) reveals that N4+ and N2+ cell-sorted populations can generate KLF17+/OCT4+ colonies that are similar to established naïve hPSCs. Scale bars, 100 µm.
(B) Phase contrast images showing representative fields of view of N4+ and N2+ cell-sorted populations that have been propagated under t2iL+PKCi naïve hPSC conditions for three passages. Scale bars, 100 µm.

4.2.2.4 Global transcriptional profiling to examine the transition from primed to naïve pluripotency

To determine where each reprogramming intermediate falls along the trajectory of primed to naïve pluripotency, I performed RNA sequencing of each isolated population 10 days into reprogramming to obtain a global overview of the transcriptional profiles. mRNA libraries were generated for three independent reprogramming experiments. Read counts per transcript were

quantified and differentially expressed genes were identified using DESeq2 (Love et al., 2014). Dimensional reduction visualisation of the top 1000 most variable genes across all cell types is visualised using principal component analysis (Figure 4.7). The PCA plot demonstrates that the biological replicates for each population and cell type cluster close together with minimal variation. This provides confidence in the reproducibly of the identified populations based on their cell surface marker expression profiles. The first principal component represents the majority of variation (72%), and nicely separates primed and naïve hPSCs. Along this component, N4+ cells cluster closer to naïve hPSCs. In contrast, N4- cells cluster closer to primed hPSCs and appear to take a different trajectory. The second principal component, that captures 16% of the variation, separates day 10 isolated populations from the established primed and naïve hPSCs. This suggests that day 10 samples represent early-stage cell types that have yet to acquire a mature gene expression profile. To explore this idea further, I profiled N4+ cells that were maintained for five (P5) or ten passages (P10) in t2iL+PKCi. These both clustered with established naïve hPSCs, which additionally highlights the concept that a final maturation phase occurs, but also reassuringly that the N4+ cells are the nascent naïve cells.

To further examine the variable genes that influence the clustering of samples, a loadings plot is shown (Figure 4.7). Known primed-specific (*DUSP6*, *OTX2* and *ZIC2*) and naïve-specific (*TFCP2L1*, *KLF4* and *DPPA3*) genes were amongst the most variably expressed and contribute to the separation of primed and naïve hPSCs. As alluded to previously, N4- cells exhibit a strong neural gene expression profile, with genes such as *SOX5*, *DKK1* and *MAP2* contributing to their cluster positioning. More importantly, genes that contribute to the second principal component (plotted in the top right quadrant of the loadings graph), are genes that provide a new insight into the transcriptional differences of early, nascent-naïve hPSCs compared to established naïve cells. Example genes associated with early naïve hPSCs include *TBX3*, *DPPA3*, *GDF3*, *KLF5* and *FOXC1*. These genes alongside many others that distinguish the early and late-stage naïve cell types, will be useful for establishing the temporal order of transcriptional changes between primed and naïve pluripotency. Genes associated with late stage naïve cell formation include *XIST*, *MEG3* and a collection of zinc finger proteins.

To explore the different biological processes that distinguish early and late stage naïve cells, I performed gene ontology (GO) analysis on the differentially up and down regulated genes between N4+ and naïve hPSCs (Figure 4.8). Genes that are upregulated in naïve hPSCs display enrichment for processes relating to transcriptional regulation and RNA metabolic processes. Strikingly, more than 100 zinc finger proteins are upregulated in naïve hPSCs and contribute to transcriptional regulation. This suggests an important role for this class of transcriptional

regulators in naïve cell fate. Conversely, transcripts that are downregulated in naïve hPSCs compared to N4+ cells are associated with developmental processes and cell differentiation GO terms. This indicates that N4+ cells have yet to silence lineage-priming genes that may be lowly expressed in primed hPSCs, but will be robustly silenced in late stage naïve hPSCs.

Another method by which naïve and primed hPSCs can be distinguished is based upon the expression profiles of transposable element (TE) classes (Grow et al., 2015; Theunissen et al., 2016). Different classes of TEs are upregulated during early human embryonic development and our understanding of their roles and regulation during this period remain largely unknown, but are a topic of acute interest. Theunissen and colleagues have previously classified many TE classes as being enriched in either naïve or primed hPSCs (Theunissen et al., 2016). With this in mind, I quantified the expression of TE classes accordingly and performed PCA analysis to display the results (Figure 4.9). The clustering profile closely mimics the trajectory seen when quantifying gene transcripts. N4+ cells cluster in between primed and naïve hPSCs, and these two established cell types are well separated along principal component 1. This reinforces the idea that N4+ cells represent an early stage naïve cell type, further exemplified by the finding that P5 and P10 samples have matured and again cluster with established naïve hPSCs. The loading plot highlights the transposable element classes that contribute to each sample type. *LTR5_Hs*, *LTR7Y* and *SVA* class of repeats are amongst these and have previously been labelled as naïve-specific (Theunissen et al., 2016); likewise, known primed-specific elements *HERVH-int* and *LTR7* contribute to separating these two pluripotent cell types along the first principal component. Moreover, the analysis identifies additional TE families that contribute not only to the separation of naïve and primed hPSCs, but additionally influence the separation of early versus late naïve stage formation. *LTR7B*, *MER11A* and *LTR5B* elements contribute to principal component two and are associated with early-stage naïve hPSCs, whereas *MER47C*, *MER57E3* and *BSR/Beta* elements distinguish late stage naïve hPSCs. Taken together, the transcriptional comparison of populations at an early-stage of primed to naïve reprogramming reveals the temporal order of transcriptional changes associated with naïve cell formation and maturation.

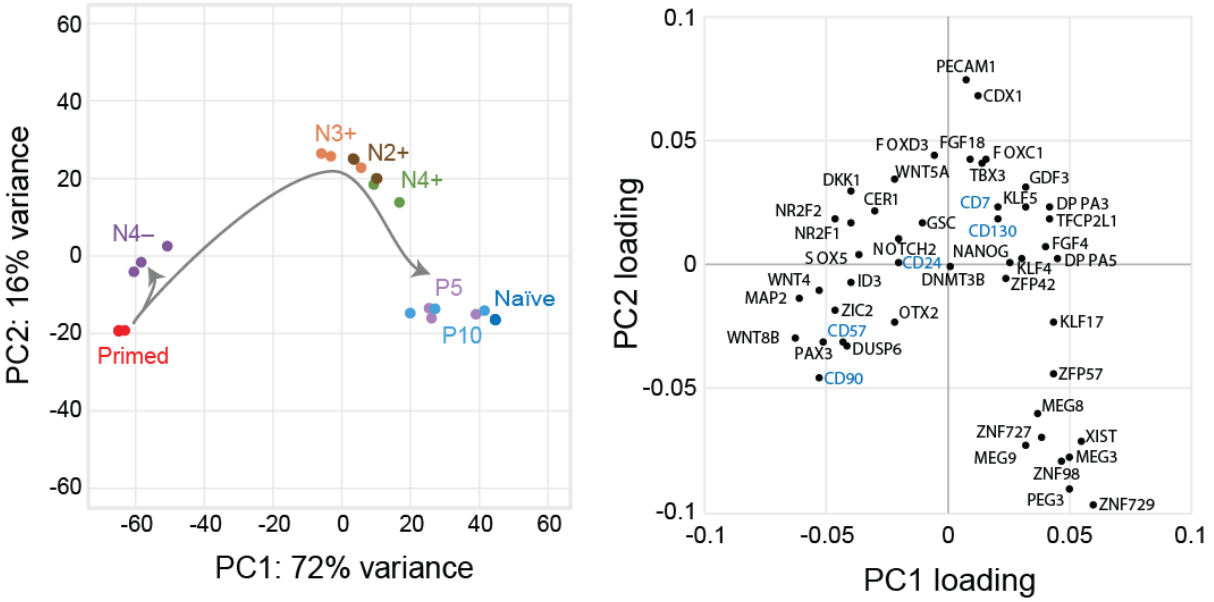


Figure 4.7. Transcriptional analysis of primed to naïve reprogramming populations.
Principal component analysis of RNA-sequencing gene expression data from the different cell-sorted populations, primed, and established naïve hPSCs (left). Right: the contribution of selected genes to the first and second principal components.

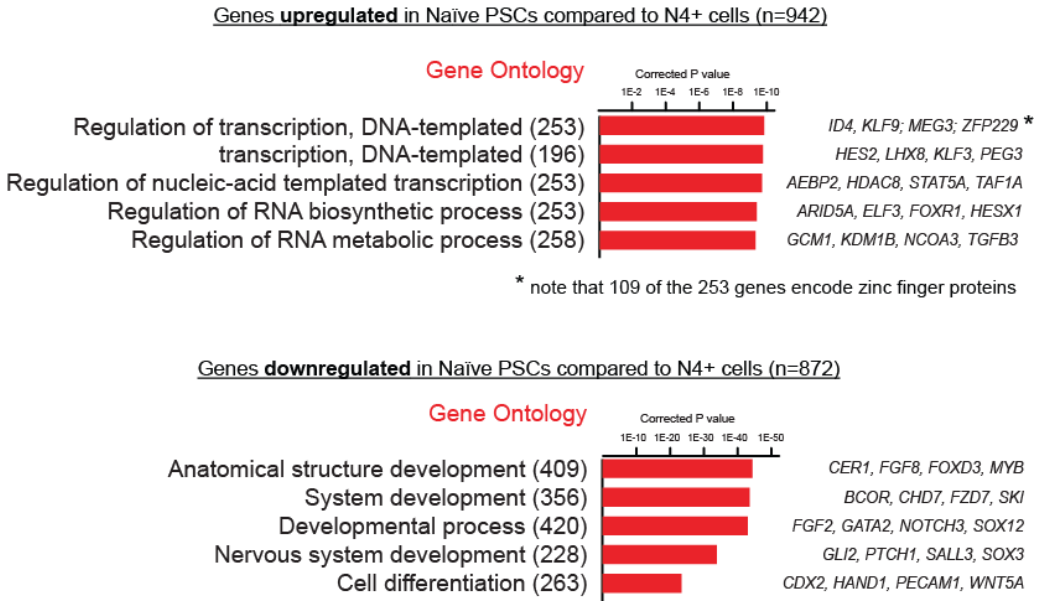


Figure 4.8. Gene Ontology analysis of differentially expressed genes between early-naïve and established-naïve hPSCs.
Top GO terms of genes that were differentially expressed between N4+ and established naïve hPSCs. The numbers of genes contributing to each GO term are shown; example genes within each GO category are listed (right). Corrected p values were calculated using a modified Fisher's exact test followed by Bonferroni's multiple comparisons test.

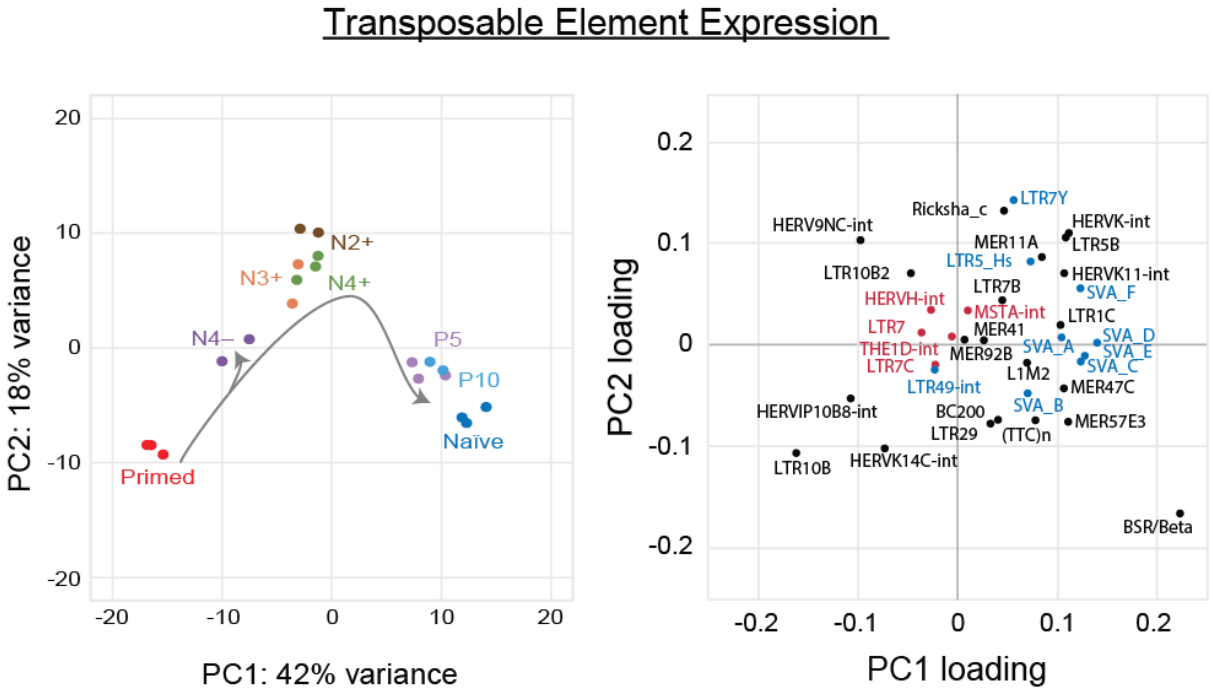


Figure 4.9. Transposable element expression can recapitulate primed to naïve reprogramming dynamics.
PCA of TE classes from the different cell-sorted populations (left). Right: the contribution of TEs to the first and second principal component. Selected TEs are labelled as having a previously defined naïve (blue) or primed (red) TE signature (Theunissen et al., 2016).

4.2.2.5 X-chromosome status

One of the hallmark characteristics of naïve pluripotency is the presence of two active X-chromosomes in female cells. As discussed in the main introduction, the methodologies of inferring X-chromosome activity are controversial. To examine the status of X-chromosome activity, I utilised the RNA-sequencing datasets generated throughout the primed to naïve reprogramming time-course. Using SNP information for the H9 line enabled the quantification of X-linked gene expression from each X-chromosome. Genes classed as biallelic infer the presence of two active X-chromosomes (XaXa), whereas mono-allelic expression is indicative of XaXi status. This analysis reveals that X-chromosome reactivation occurs predominantly during the late stage of naïve cell formation (Figure 4.10). All day 10 reprogramming intermediates have yet to reactivate the X-chromosome, even though N4+ and N2+ cells already exhibit a naïve-like transcriptional profile. However, the culturing of N4+ cells for an additional for five or ten passages results in X-chromosome reactivation. This narrows down the window whereby the inactive X-chromosome will reactivate to somewhere between day 10 to day 30 of reprogramming. A peculiar observation from this analysis reveals a handful of genes on the p-arm of the X-chromosome were monoallelically expressed in established naïve hPSCs

(Takashima et al., 2014). The cause of this is unclear, but it could indicate an erosion of X-chromosome activation during long-term maintenance of naïve hPSCs. Nevertheless, N4+ cells cultured for five (P5) and ten (P10) passages undergo X-chromosome reactivation and display uniform biallelic expression. I believe this analysis supports the notion that X-chromosome reactivation is a robust molecular marker, and perhaps one of the most faithful indicators of bona-fide mature naïve hPSCs.

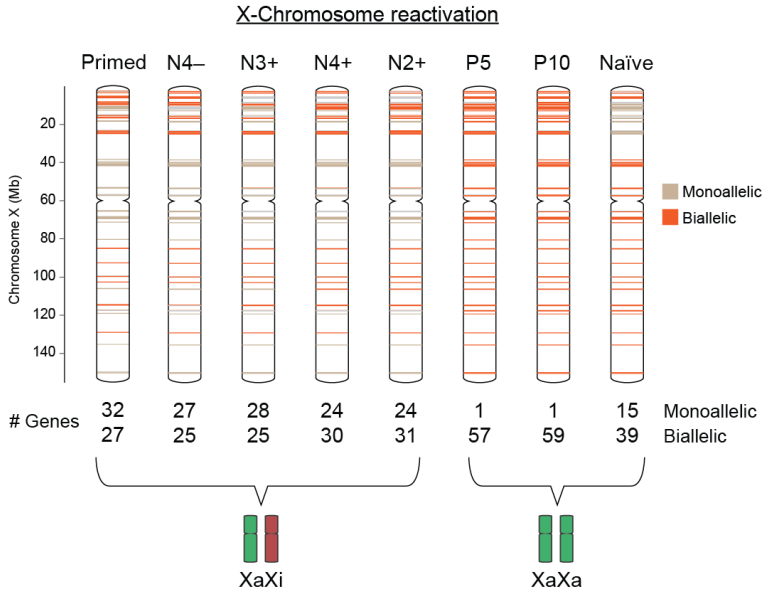


Figure 4.10. X-chromosome reactivation is a late event during primed to naïve reprogramming. Schematic of X chromosomes that summarise the results from an allelic analysis of RNA-seq data for the indicated cell types. Informative SNPs within X-linked genes of the H9 PSC line (Vallot et al., 2017) were used to classify expression as monoallelic (brown, <25% from minor allele), biallelic (orange, 25%–75% from minor allele), or not expressed (gray, <10 reads/sample). The number of monoallelic and biallelic genes is shown below.

4.2.2.6 DNA methylation analysis

Global DNA hypomethylation is an additional characteristic of human ICM cells that naïve hPSCs also share. Typically, primed hPSCs exhibit uniformly high CpG methylation levels (≈80% globally), whereas long-term cultured naïve hPSCs reduce to 30% globally (Guo et al., 2017; Pastor et al., 2016). The DNA methylation dynamics have yet to be examined at intervals during this state transition. Therefore, we examined the methylome status at several time-points: N4+ cells at day 10, and passage 10 (day 50) (Figure 4.11). There is a notable decrease (30%) in global CpG DNA methylation as early as 10 days into reprogramming compared with parental H9 primed cells. A further decrease occurs by passage 10, with global levels dropping

to ≈30%. This is the equivalent level to established naïve hPSCs that have been cultured for >20 passages. Unfortunately, samples at an intermediate time-point (passage 5) did not provide sufficient coverage for reliable quantification. However, the rapid loss of DNA methylation between primed cells and day 10 precedes the reactivation of the X-chromosome. This infers that the removal of DNA methylation may be required for X-chromosome reactivation.

We further examined the DNA methylation levels at imprinted loci where sufficient coverage was obtained (n=36). Of these, 30/36 showed retention of DNA methylation 10 days into reprogramming; However, by passage 10 only 10/36 were still methylated. This phenomenon has been previously highlighted, also in naïve hPSCs cultured for >10 passages (Guo et al., 2017; Pastor et al., 2016).

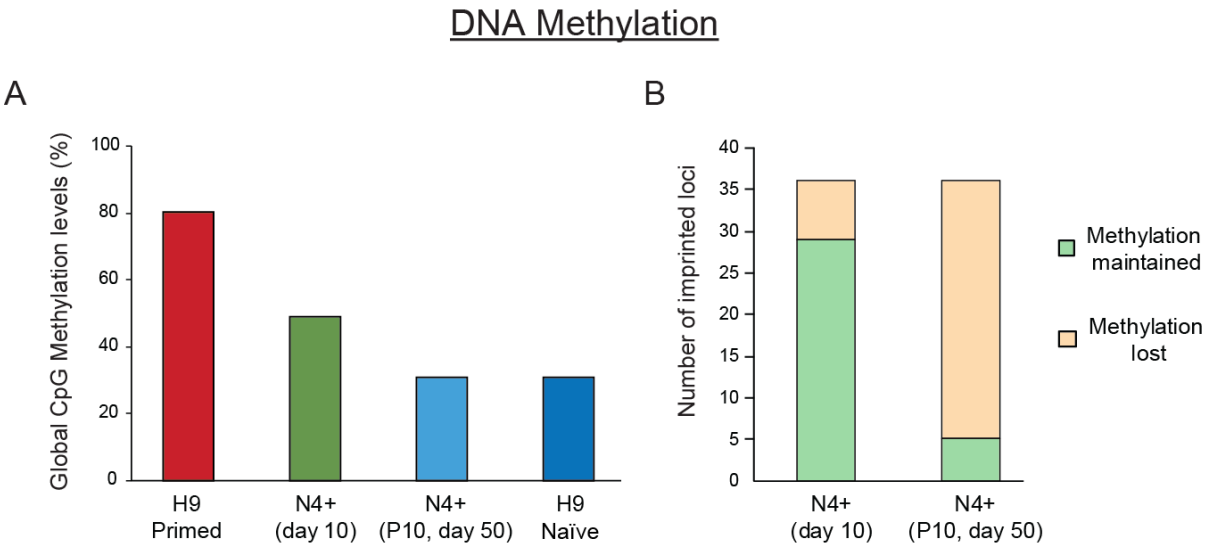


Figure 4.11. Global CpG DNA Methylation levels across primed to naïve reprogramming populations.
(A) Global CpG methylation levels (%) from three independent biological replicates which were pooled after sequencing.
(B) Classification of imprinted loci methylation status between day 10 and passage 10 samples. Methylation was either maintained or lost (classified according to a ≥50% loss in CpG methylation compared to parental primed hPSC levels).

4.3 Discussion

Flow cytometry analysis of day 10 t2iL+PKCi reprogrammed cells revealed several distinct populations. The most "naïve-like" population, designated N4+, is defined by the high-expression of all four naïve-specific markers, and low expression of primed-specific markers. After 10 days of reprogramming, between ≈2% of all live-human cells are classed as N4+, highlighting the inefficient nature of t2iL+PKCi reprogramming. Alongside N4+ cells, I additionally sorted two other reprogramming intermediates, N3+ that expressed CD7, CD77, CD130, but not CD75, and N4- that failed to express any naïve-specific markers. N3+ and N4- populations represented 6% and 22% of the live-human population respectively. Collectively, the populations represent 30% of cells at this time-point; this highlights one of the caveats to the gating strategy used, as many cells were unable to be characterised. Nonetheless, the FlowSOM results reveal three predominant cell clusters which N4+, N3+ and N4- are embedded within (Figure 4.2). These populations should therefore provide a good overview and capture the variation in reprogramming populations.

Whilst multiplexing a large panel of antibodies provides a high-resolution insight into different cell populations, it is not without challenges. Users will require a good knowledge of antibody panel design to ensure appropriate compensation is applied. Moreover, the current panel design requires a high-end FACS machine with five lasers. Finally, the availability of high-quality antibodies in a range of conjugations is necessary. For commonly used markers this is normally not an issue; however, the choice of infrequently studied naïve-specific markers CD77, CD75, and CD130, was restricted to common conjugations. To improve the usability of our approach, I refined the set of antibodies down to a combination of two naïve-specific (CD75 and CD130) and two primed-specific markers (CD24 and CD57). These could largely recapitulate the full antibody panel, isolating a slightly wider population that corresponded to 4-5%, and was designated as N2+.

An increase in clonal efficiency is one of the hallmark characteristics of naïve pluripotency compared to their primed counterparts. I was pleasantly surprised to see that both N4+ and N2+ populations possess the ability to form colonies with a naïve-morphology that equates to the standard of established naïve hPSCs. Contrary to this, the N3+ population performed poorly by comparison and generated far fewer naïve colonies. This suggests that N3+ cells represent a partially reprogrammed population, and that CD75 is a crucial marker of naïve hPSCs. CD75 is a glycosylated epitope purportedly catalysed by the sialytransferase ST6GAL1 (Munro et al., 1992). Sialyation is known to be involved in a variety of cellular processes, such as cell

adhesion, cell signalling, and the regulation of glycoprotein stability (Pshezhetsky and Ashmarina, 2013; Schauer, 2009). Whilst the role of CD75 and sialyation has not been examined in the context of naïve pluripotency, the perturbation of the sialytransferase ST6GAL1 caused a reduction in iPSC reprogramming efficiency and led to the impaired self-renewal of primed hPSCs (Wang et al., 2015). Moreover, the deletion of an enzyme responsible for sialic acid biosynthesis (UDP-GlcNAc 2-epimerase) resulted in embryonic lethality, supporting the notion that sialyation is important for embryonic development in mice (Schwarzkopf et al., 2002). It would be interesting in the future to isolate the N3+ population at day 10 of reprogramming and subsequently re-evaluate whether subpopulations can mature to an N4+ state – or whether a failure to upregulate CD75 by day 10 is a defining criterion to distinguish nascent naïve hPSCs from refractory populations.

To further characterise the different reprogramming populations, I examined their transcriptional profiles, and several epigenetic hallmarks that distinguish primed and naïve pluripotent states. Firstly, the transcriptional analysis revealed N4+ and N2+ populations represent an early-stage naïve cell type and have yet to acquire a mature naïve gene expression profile. To determine whether these populations represent nascent naïve hPSCs, I continued to culture N4+ isolated cells and reassessed five or ten passages later. By passage five, the cells had acquired a fully mature naïve expression profile, providing evidence that N4+ cells are indeed nascent naïve hPSCs. Moreover, N4+ cells can be maintained long-term for more than 20 passages whilst retaining a homogeneous naïve morphology. Collectively this shows the antibody panel provides an opportunity to study naïve hPSC formation at a time point far earlier than previously possible.

Importantly, analysis of early-stage naïve hPSCs provided a new insight into the temporal order of gene expression changes. Of note, transcription factors *TBX3, DPPA3, GDF3* and *KLF5* are strongly upregulated by day 10 of reprogramming, to near maximum levels observed, suggesting these play a crucial role during reprogramming. Whilst TBX3 and DPPA3 have not been studied during human pluripotent state transitions, several groups have emphasised the connection between the two in mouse pluripotency. TBX3appears to maintain steady levels of DPPA3 to prevent the premature exit from pluripotency, and is strongly downregulated upon the transition from naïve mESCs to primed mEpiSCs states (Buecker et al., 2014; Russell et al., 2015; Waghray et al., 2015). The co-expressed nature of *TBX3* and *DPPA3* in this context infers a conserved relationship in human pluripotency. Moreover, *TBX3, DPPA3, GDF3* and *KLF5* are all highly expressed within the ICM cells of human embryos assessed by single-cell RNA-

sequencing (Stirparo et al., 2017). This provides additional support to their biological relevance in naïve human pluripotency.

The transcriptional analysis of the reprogramming populations additionally highlighted a set of genes that are associated with established naïve hPSCs and N4+ cells that had been passaged five or ten times. Several interesting examples include *KLF17*, *DPPA5*, and a large number of zinc finger proteins. Both *KLF17* and *DPPA5* are transcriptionally present in the epiblast cells of both primate and human embryos. However, the function and targets of KLF17 in human pluripotency are largely unknown, and provides an interesting avenue for future studies. DPPA5 - another member of the developmental pluripotency-associated protein family - has been shown to post-transcriptionally regulate NANOG stability in primed hPSCs (Xu et al., 2016). Moreover, DPPA5 (Esg1 in mouse) is a small RNA binding protein that has been shown to bind over 900 transcripts in mouse ESCs. Notable validated examples include the chromatin remodelling factor *Ezh2*, and transcriptional activator *Nr5a2* (Tanaka et al., 2006). Future studies employing enhanced sequencing technologies for the identification of RNA-binding protein target RNAs, such as iCLIP, will help to reveal the function of DPPA5 in human pluripotency. Collectively, KLF17, DPPA5 and zinc finer proteins - particularly KRAB-ZFPs - have all been implicated in transcriptional repression, amongst other roles (Gumireddy et al., 2009; Tanaka et al., 2006; Yang et al., 2017a). It is possible that given the predominantly blank epigenetic landscape of naïve hPSCs, as far as repressive marks at promoter regions are concerned (Ji et al., 2016; Theunissen et al., 2014), ZFPs and other transcription factors may act as transcriptional repressors to thwart precarious gene activation and silence transposable elements (TEs).

Different transposable element classes exhibit developmental stage-specific expression profiles during embryogenesis. The mode of action for transposable elements are also variable; examples include cis-regulation of nearby gene expression, chimeric transcript formation by developing splice donor sites, or acting as long-range enhancer elements (Peaston et al., 2004; Thompson et al., 2016; Xie et al., 2010). Based upon the expression of various TE classes, it is possible to distinguish naïve and primed hPSCs (Theunissen et al., 2016). PCA analysis of TE expression for the various day 10 reprogramming intermediates, alongside established primed and naïve hPSCs (Figure 4.9), was able to recapitulate the PCA clustering observed by assessing mRNA transcript levels (Figure 4.7).

Whilst I observe concordant classification of TE classes previously implicated in human naïve and primed pluripotency, I detected additional TE classes with state-specific enrichment. For example, β-satellite repeats which are highly enriched in naïve hPSCs, but were not described

in naïve hPSCs under 5iLA culture (Theunissen et al., 2016); possibly owing to PSC line-specific differences, or the different naïve reprogramming methods. Moreover, I found *LTR5_Hs* elements to be enriched in N2+, N3+ and N4+ day 10 reprogramming cells, as well as being highly expressed in naïve hPSCs. Both *LTR5_Hs* and *HERVK* are expressed in 8-cell stage human embryos through to blastocyst formation, and exert their function to modulate mRNA levels via the accessory protein Rec (Grow et al., 2015). Wysocka and colleagues show the hypomethylated state of *LTR5_Hs* in human embryonic carcinoma cells and naïve-like hPSCs results in *LTR5_Hs* upregulation (Grow et al., 2015). We also examined the DNA methylation profile of N4+ nascent naïve hPSCs at day 10 and passage 10 (Figure 4.11); from this, we see a global decrease in CpG methylation of 30% by day 10. Whilst further examination is necessary to decipher loci-specific DNA demethylation dynamics, the transcriptional upregulation of *LTR5_Hs* correlates with the loss of DNA methylation. Future work will seek to address the entwined relationship between the regulation of transposable elements, and how in turn TEs regulate gene expression in the context of human pluripotency.

Finally, I sought to examine the status of X-chromosome reactivation (XCR) – a vital hallmark of naïve pluripotency in female cells. The results revealed X-chromosome reactivation had not occurred in any of the day 10 reprogramming intermediates; however subsequent passaging of N4+ nascent naïve hPSCs displayed a convincing XaXa status by passage 5. Utilising RNA-sequencing data to examine X-chromosome status has the advantage of being matched to a global transcriptional profile for cell type identification; however, there are a few caveats to classifying the X-Chromosome status by this method. Firstly, it is only possible to examine allelic-specific expression for PSC lines where SNP information is available. Secondly, the results display a population average. This means that the details of how XCR is mediated, for instance by lncRNAs, cannot be observed in detail at a single cell resolution.

Analysis of female human preimplantation-stage embryos has revealed the two active X-chromosomes are associated with expression of both *XIST* and *XACT* lncRNAs (Petropoulos et al., 2016; Vallot et al., 2017). Whilst this biallelic expression is observed in established naïve hPSCs cultured under t2iL+PKCi and 5iLA, substantial heterogeneity of *XIST* expression by RNA-FISH is seen (Kilens et al., 2018; Sahakyan et al., 2017; Vallot et al., 2017). This observation would be missed by the quantification of X-linked gene expression alone. Regardless, this method can still serve as a reliable population status of X-chromosome activity. One correlation from our data that can be drawn is the upregulation of *XIST* that occurs at a late-stage of naïve cell formation, correlating with XCR. Interestingly, the rapid DNA demethylation observed by day 10 of naïve reprogramming occurs before XCR. This implies

that further DNA demethylation may be necessary for robust reactivation of the silent X-chromosome. In line with this observation, XCR does not take place during human iPSC reprogramming to the primed state, perhaps unsurprisingly given their hypermethylated status (Mekhoubad et al., 2012). Moreover, DNA methylation persists on the inactive X-chromosome throughout mouse iPSC reprogramming. It is not until the very latest stages that sufficient DNA demethylation occurs to allow faithful XCR (Pasque et al., 2014).

In summary, by identifying the earliest responders during reprogramming, inference towards an interconnected transcription factor network or hierarchy can be made, and subsequently tested experimentally in future studies. Our understanding of pluripotent gene regulatory networks and their importance has been shown extensively for mouse pluripotency (Boyer et al., 2005; Chen et al., 2008; Kim et al., 2008; van den Berg et al., 2010; Wang et al., 2006). Whilst these studies provide a valuable foundation for mammalian pluripotency, extrapolating to human may not always be accurate, and human or primate-specific pluripotency factors may be overlooked. More recent reports which examine the molecular and phenotypic characteristics of human, primate, and mouse embryos, will facilitate our understanding of human pluripotency and highlight species-specific differences (Nakamura et al., 2016; Stirparo et al., 2017). Nonetheless, due to the ethical restrictions and technical challenges associated with studying early human embryos, *in-vitro* models that recapitulate the developmental progression of human embryogenesis would be highly desirable. Accordingly, I have shown that the identified state-specific surface markers reliably track the dynamics of primed to naïve reprogramming, and enable nascent naïve hPSCs to be isolated. For the first time, this enables the timing and order of molecular changes that occur during reprogramming to be examined.

5. Comparative analysis of different primed to naïve reprogramming methods

5.1 Background

Until now, the inability to isolate nascent naïve hPSCs at an early time point during primed-to-naïve reprogramming has precluded the study of the reprogramming process. This gap in knowledge means that little is known about the transitional period and consequently there are many unanswered questions. Primarily, are there different molecular routes from primed to naïve pluripotency? If so, are there benefits of one reprogramming route over another – such as efficiency, timing, or stability of the end naïve hPSC population? To date, all comparative studies that examine the molecular characteristics of naïve hPSCs derived using different reprogramming methods, were restricted to long-term cultured naïve hPSCs. Without a means to examine the earliest stages of reprogramming, it is not possible to decipher if there are alternative routes to naïve pluripotency. The ability to answer these questions relies upon a method to study the reprogramming process with frequent temporal precision, in order to determine with accuracy the dynamics and potential coordination of molecular changes that occur. A better understanding of this process may present an opportunity to coerce refractory reprogramming populations down the right route.

From the comparative molecular characterisation of naïve hPSCs derived by 5iLA and NK2 t2iL+PKCi reprogramming, we know that the resulting cell types are very similar and share a close transcriptional and epigenetic resemblance to preimplantation epiblast cells (Huang et al., 2014; Stirparo et al., 2017). However, we know almost nothing about the routes taken from primed pluripotency to a naïve state. As previously mentioned, the cell type of origin can influence the molecular trajectory taken during iPSC reprogramming (Nefzger et al., 2017). However, in this instance both the starting (primed) and end cell type (naïve) are hypothetically the same; instead the reprogramming method is what differs. 5iLA reprogramming is transgene independent, with the addition of several chemical inhibitors that are not present in t2iL+PKCi conditions. Conversely, t2iL+PKCi reprogramming is reliant on the transient overexpression of *NANOG* and *KLF2* transgenes, followed by continued culture in t2iL+PKCi; yet remarkably the resulting naïve cell types generated by each method are very similar. This infers that convergent signalling pathways and gene regulatory networks are utilised, at least towards the later stages of reprogramming, or for the maintenance of naïve pluripotency. Yet given the divergent methodologies employed at the onset of reprogramming, I propose that the early stages of primed-to-naïve reprogramming are likely to be transcriptionally and phenotypically dissimilar.

In support of this idea, I observed a substantial difference in the proportion of day 10 reprogramming cells that cluster with established naïve hPSCs, based upon their cell surface expression profile, between the two reprogramming methods (Figure 3.15and Figure 3.18). During 5iLA reprogramming the majority of day 10 cells clustered with established naïve hPSCs; by contrast only ≈2% of t2iL+PKCi day 10 day cells were cluster with their corresponding naïve hPSCs, which is approximately 40 fold lower than 5iLA. This allows me to infer that the reprogramming efficiency of 5iLA may be higher, or that reprogramming may be faster. To address this, I propose using the multiplexed antibody panel to isolate nascent naïve hPSCs at early matched time-points for 5iLA reprogramming, to be compared against the t2iL+PKCi transition data discussed throughout chapter 4. Comparative analysis of the transcriptional profiles will help to reveal the similarities and differences in gene regulatory networks utilised when transitioning from a primed to naïve state using the different reprogramming methods. Moreover, it may be possible to associate the transcriptional routes taken to specific chemical inhibitors used in one culture medium, and not the other. Alternatively, common mediators that facilitate reprogramming may be identified, which would subsequently open up the possibility of these being experimentally tested for their requirement during primed to naïve reprogramming. If the identified mediators or repressors of reprogramming are susceptible to chemical induction or inhibition respectively, future targeted efforts can be taken towards improving current naïve hPSC induction media. Taken together, the isolation of nascent naïve hPSCs using an additional reprogramming method will help to address unresolved questions, whilst also highlighting the robust and versatile capability of the multiplexed antibody panel.

5.1.1 Aims

1. Isolate early naïve hPSCs using different primed to naïve reprogramming methods
2. Characterise early naïve hPSCs
    i. Assess the ability to form colonies from single cells
    ii. Examine global transcriptional dynamics throughout primed to naïve reprogramming at frequent time points
    iii. Examine the dynamics of X-chromosome reactivation throughout primed to naïve reprogramming
    iv. Perform comparative transcriptional profile analysis of early naïve hPSCs to decipher the routes taken from primed to naïve pluripotency

5.2 Results

To gain a better understanding of the cell fate transition from primed to naïve pluripotency, I wished to examine different reprogramming methods in more detail. As alluded to earlier (in section 3.2.7) the efficiency of reprogramming is variable among the reprogramming protocols. Whilst NK2 t2iL+PKCi reprogramming forms a nascent naïve population of ≈2% by day 10 (Figure 4.2), 5iLA reprogramming revealed that the majority of day 10 cells were clustered amongst established naïve hPSCs (Figure 3.18) – indicating that the reprogramming efficiency may be higher using this method.

5.2.1 Using cell surface marker expression profiles to define reprogramming populations under 5iLA culture

I first examined what intermediate reprogramming populations were present by day 10 using the complete multiplexed antibody panel. The results visualised using FlowSOM reveal a far greater percentage of N4+ cells (≈11% on average) (Figure 5.1). Curiously, the N3+ population (lacking CD75 induction) is considerably smaller, perhaps indicating that this population has progressed, or that this trajectory is infrequent with 5iLA reprogramming. As previously observed in naïve cells, there is a good overlap between the naive-specific markers and OCT4-ΔPE-GFP expression. Interestingly, the vast majority of day 10 cells are positive for OCT4-ΔPE-GFP expression, even though there is considerable heterogeneity in the morphology at this time point (Figure 5.2). Moreover, OCT4-ΔPE-GFP positive cells are apparent that are lowly expressing several naïve-specific markers, and are still positive for primed-specific marker CD57 (designated with black arrows, Figure 5.1). This indicates that OCT4-ΔPE-GFP expression may precede the expression of naïve-specific markers. Alternatively, this could indicate that OCT4-ΔPE-GFP is not adequately stringent for prospective isolation of homogeneous naïve hPSCs, and may identify a broader cell population. – a point I shall address in subsequent sections.

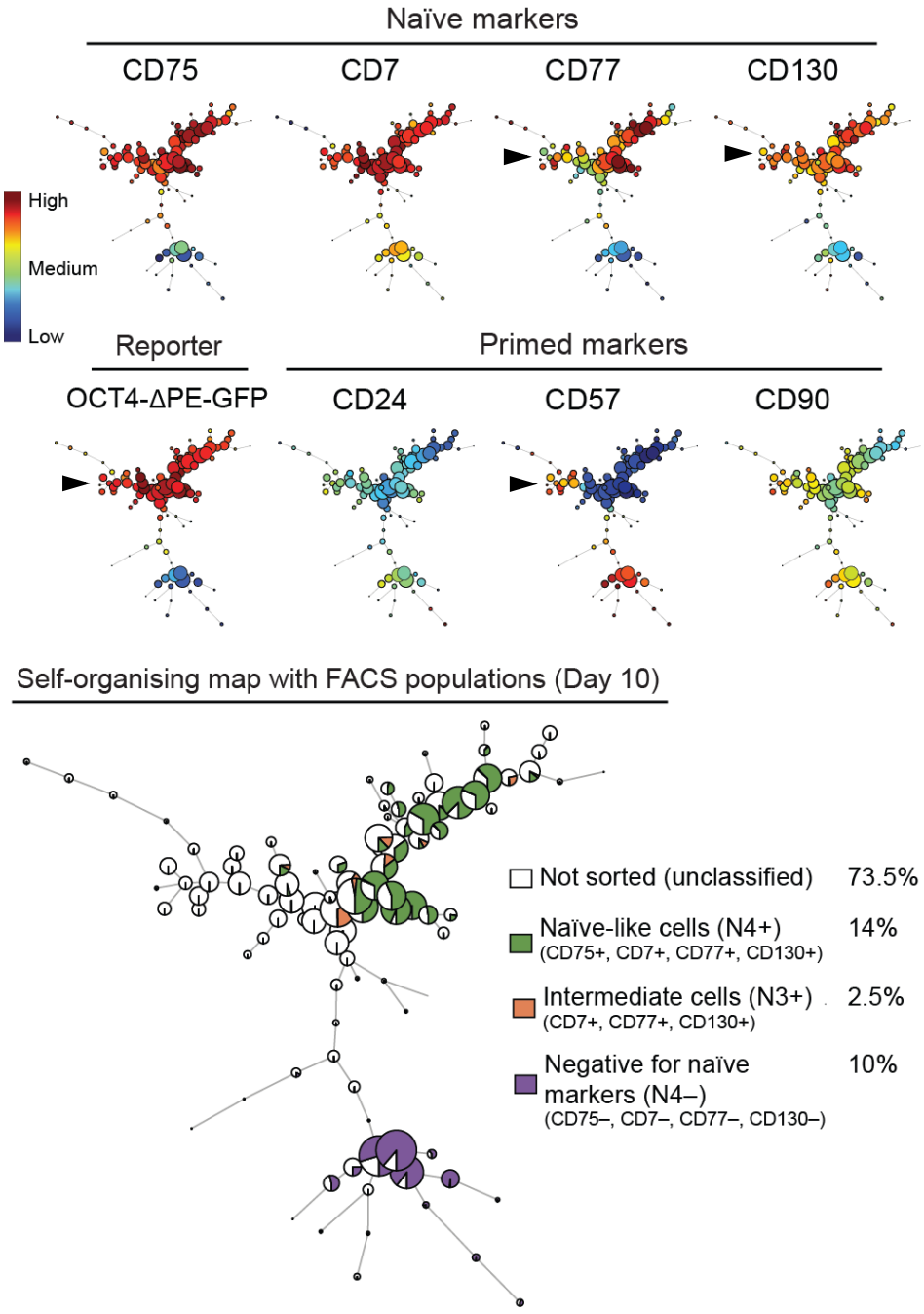


Figure 5.1. Prospective isolation of 5iLA day 10 reprogramming populations by FACs.
FlowSOM visualisation of the flow cytometry data for day 10 cells during primed to 5iLA naïve reprogramming. The minimal spanning tree of the self-organising map displays an unsupervised clustering of the sample based on the cell surface protein expression levels (bottom). The cells corresponding to each cell sorted population, N4+, N3+, and N4–, are indicated. The heatmap shows the expression level of each cell surface protein marker in the cell clusters (top).

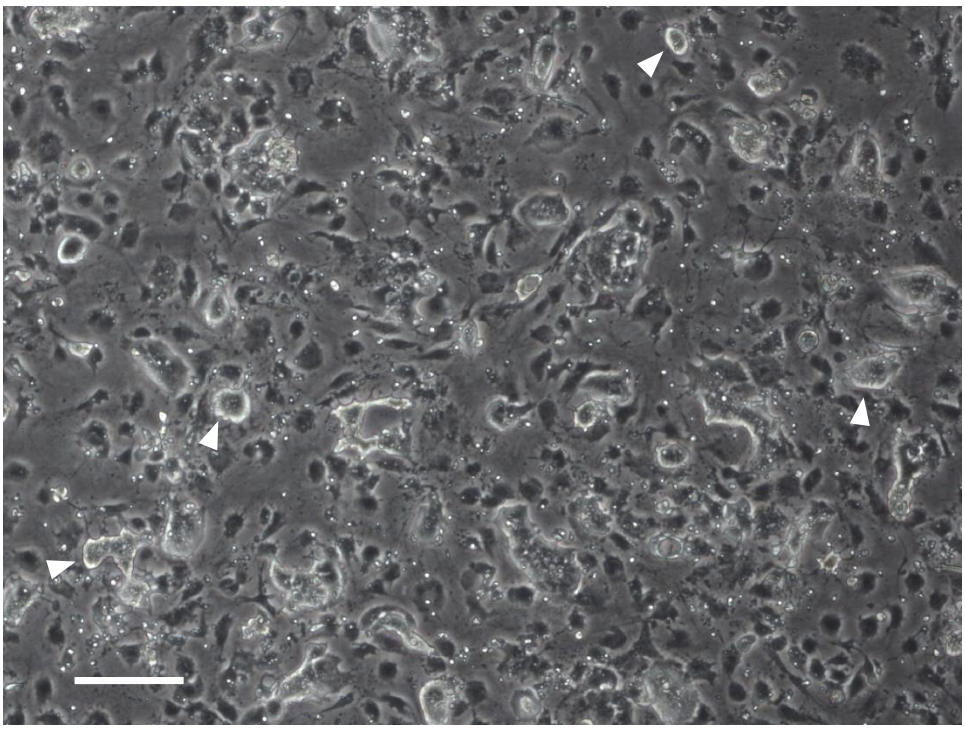


Figure 5.2. Representative morphology of day 10 5iLA reprogramming hPSCs
Phase contrast image showing a representative field of view of WIBR3 hPSCs under 5iLA reprogramming conditions (day 10). Colonies with a naïve morphology are beginning to emerge (indicated by white arrows), yet considerable heterogeneity is apparent. Scale bars, 200 µm; 4x objective.

5.2.2 Isolating nascent naïve hPSCs early during 5iLA reprogramming using a refined multiplexed antibody panel

I next decided to further optimise the multiplexed antibody panel, by removing the naïve specific CD7 and the primed-specific CD90 antibodies. These two markers contribute the least to the panel when the objective is to identify naïve hPSCs. The majority of the cells that are triple positive for CD75, CD77 and CD130 are positive for CD7 (≈98%). Moreover, CD7 expression is absent in human embryos, unlike CD75, CD77 and CD130. Likewise, CD24 and CD57 low populations correspond to the lowest CD90 expressing cells. Removal of these two markers additionally served to simplify cytometry compensation and improve the reliability, as is the case when fewer compensation calculations are required. I additionally considered removing CD130 from the antibody panel, as 97% of cells that are double positive for CD75 and CD77 are also CD130 positive under 5iLA reprogramming; however, CD130 is a useful marker during NK2 t2iL+PKCi reprogramming to delineate successfully reprogrammed cells upon DOX-induced transgene withdrawal. I therefore decided to retain CD130 to safeguard future continuity in the case where additional reprogramming methods are examined.

As discussed previously, there is a potential caveat with the isolation strategy performed previously for NK2 t2iL+PKCi samples. It is possible that additional nascent naïve hPSC populations exist, but may not have been captured within the stringent gating strategies of N3+ and N4+. Moving forward with the condensed antibody panel (CD75, CD77, CD130, CD24 and CD57), I sought to address this concern by broadening the gated populations into "Reset" and "Not Reset". Gating for "Reset" cells is comparable to N4+ in that the CD75/CD77/CD130 high and CD24/CD57 low expressing cells are isolated (Figure 5.3). However, the gating of "Not Reset" cells has been broadened considerably to include all cells that do not fall into the initial "Reset" category of CD75/CD77 high. This includes cells that will be CD77 or CD75 high, but not those that express both markers highly. With this modified gating strategy, the Reset population corresponds to ≈10%, whilst the Not Reset cells corresponds to ≈55% on average at 10 days into reprogramming with 5iLA (Figure 5.3). If I were to apply this modified gating strategy to NK2 t2iL+PKCi day 10 reprogramming cells analysed previously, the Reset population would correspond to ≈2.5%, a slight increase from the ≈2% N4+ gating.

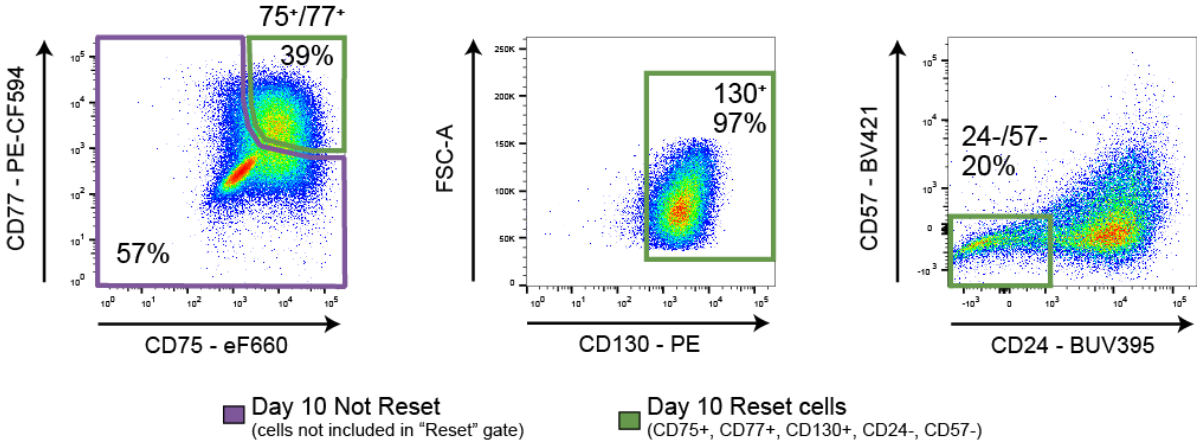


Figure 5.3. Refined gating strategy for the prospective isolation of day 10 reprogramming populations by FACS.

Flow cytometry dotplots of day 10 cells during primed to naïve-state reprogramming of WIBR3 hPSCs under 5iLA culture conditions. Left: Expression levels of two naïve-specific markers, CD75 and CD77. The double positive population (CD75+/CD77+) was gated (green box), and cells not included were gated separately and correspond to day 10 "Not Reset" cells. These cells can express CD77 and CD75 individually to high levels, but not together. Middle: The expression levels of a third naïve-specific marker CD130 was gated (CD75+/CD77+/CD130+). Right: Expression levels of two primed-specific markers, CD24 and CD57. The double negative fraction was gated and corresponds to day 10 "Reset" cells that was used for subsequent experiments. The percentage of cells within each cell sorting gate relative to all live, human cells or the previous gated population is shown.

The FlowSOM projection for 5iLA reprogramming highlights the both the Reset and Not Reset populations, and again exemplifies the previous finding that OCT4-ΔPE-GFP is expressed in the majority of day 10 cells (≈75%) (Figure 5.4), even though most of these cells do not display the appropriate surface marker expression profile indicative of naïve cells. I have further highlighted the OCT4-ΔPE-GFP expression across three day 10 populations: Reset, Not Reset, and all live-human cells (Figure 5.5). Whilst the Reset population is almost exclusively positive for the OCT4-ΔPE-GFP reporter, more than 50% of the Not Reset population is also positive. This infers that either the surface markers are too stringent and will therefore fail to capture additional nascent naïve hPSCs, or the OCT4-ΔPE-GFP reporter is too permissive.

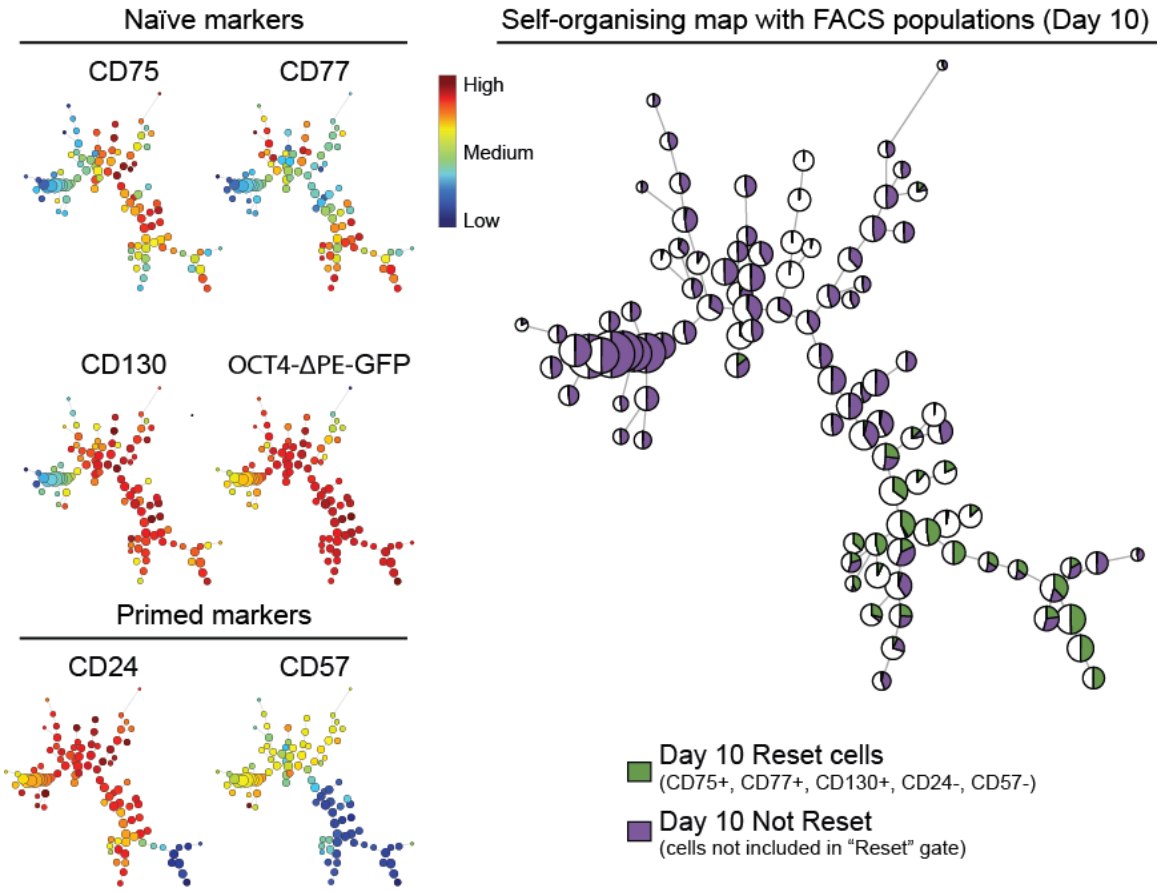


Figure 5.4. Prospective isolation of 5iLA day 10 reprogramming populations by FACS.
FlowSOM visualisation of flow cytometry data for day 10 cells during primed to naïve reprogramming under 5iLA conditions. Right: cells corresponding to day 10 "Reset" (green), and day 10 "Not Reset" (purple) cell sorted populations indicated on the minimal spanning tree. Left: heatmap showing the expression levels of each surface marker in the cell clusters.

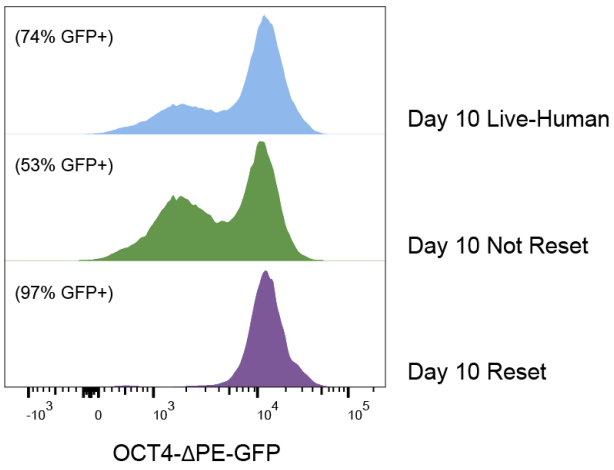


Figure 5.5. OCT4-ΔPE-GFP expression levels across day 10 5iLA reprogramming populations
Representative histograms of day 10 GFP expression levels for three populations: all live-human cells, Not Reset cells and Reset cells. Noted are the percentage of cells classified as GFP+ for each population.

5.2.3 Colony formation assay to examine clonogenicity of reprogramming populations

To assess whether the modified gating strategy faithfully distinguishes between nascent naïve hPSCs and the cells that have not reprogrammed, I performed a colony formation assay. Strikingly, the Reset population isolates almost all cells with the capability of giving rise to colonies with a naïve morphology (Figure 5.6A). On contrast, the Not Reset population fails to generate colonies with a naïve or mixed morphology, even though this population includes cells that express two out of the three naïve-specific markers, and OCT4-ΔPE-GFP positive cells. This result reassuringly addresses the potential concern of missing important nascent naïve hPSC populations, and exemplifies that the state-specific surface markers can isolate naïve hPSCs generated by different reprogramming methods. Moreover, 5iLA day 10 Reset cells isolated by FACS give rise to homogeneous naïve hPSC cultures (Figure 5.6). Taken together, the modified antibody panel robustly identifies nascent naïve hPSCs early during 5iLA reprogramming.

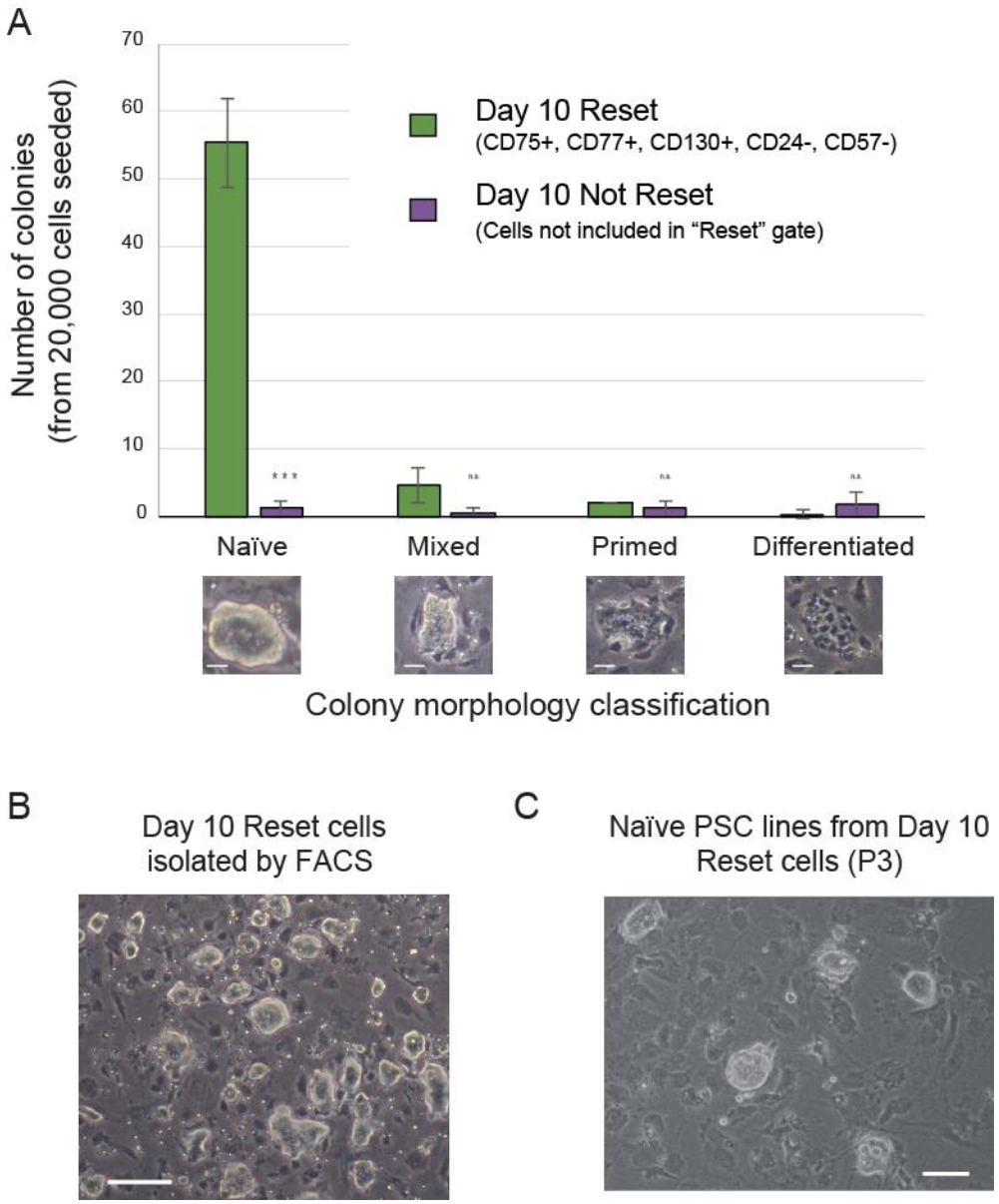


Figure 5.6. Colony formation efficiency of cell sorted populations into 5iLA naïve hPSC conditions.

(A) Colonies were scored as naïve, mixed, primed and differentiated: examples shown below. Data show the mean ± SD of three biological replicates. Day 10 Reset cells were compared to Day 10 Not Reset cells for each morphology classification using a two-tailed t-test (***$p < 0.005$). Scale bars, 100 µm.

(B) Representative phase contrast image of day 10 Reset sorted population, cultured for 4 days after sorting for colony morphology Scale bars, 200 µm, 4x objective.

(C) Representative phase contrast image naïve hPSC line derived from day 10 Reset cells, cultured for 3 passages under 5iLA culture conditions. Scale bar, 100 µm, 10x objective.

5.2.4 Transcriptional dynamics of 5iLA primed to naïve reprogramming

5.2.4.1 Global overview of the transcriptional dynamics during 5iLA reprogramming

Examining the transcriptional dynamics between primed and naïve pluripotency using NK2 t2iL+PKCi reprogramming provided an initial insight into the potential regulators of this process. However, I wished to elaborate upon this work in a number of ways. Firstly, the assessment of an earlier time point, given that emerging naïve cells appear on day 8 during 5iLA reprogramming based upon FlowSOM clustering. Secondly, more frequent time-points to enhance the resolution of transcriptional dynamics and order of molecular changes. The transcriptional analysis performed using NK2 t2iL+PKCi reprogramming revealed that considerable transcriptional and epigenetic remodeling continues between day 10 and passage 5. I therefore decided to examine each passage between these time points. Finally, I wished to use an alternative method of reprogramming which would enable me to decipher the similarities and differences in routes taken between pluripotent states at matched time points. To this end, I sorted "Reset" cells at the following time points: day 8, day 10, passage 1 (day 14), passage 2 (day 18), passage 3 (day 22), passage 4 (day 26) and passage 5 (day 30). In addition, "Not Reset" cells were isolated at early time points of 8 and 10 days into reprogramming for a comparison against the "Reset" population. Cells were isolated at each time point by FACS and mRNA-seq libraries were generated from three independent reprogramming experiments.

PCA was performed using the top 1000 most variable genes across all cell types. Time-point replicates cluster closely together and exhibit minimal variation, providing assurance in the reproducibility of the antibody panel to isolate nascent naïve hPSCs (Figure 5.7). Similar to the PCA generated for NK2 t2iL+PKCi reprogramming, PC1 separates primed and naïve hPSCs, representing 74% of the variation. Along this same component, there is a progressive trajectory from day 8 through to passage 5. Interestingly, cells appear to have acquired a mature naïve identity by passage 2, with little variation between P2, P3, P4, P5 and established naïve hPSCs. Analogous to the trajectory observed for NK2 t2iL+PKCi reprogramming, the "Not Reset" population takes a divergent path that mimics the N4- population. This population similarly expresses neural associated genes such as *MAP2*, *CER1* and *SOX5*. Likewise, genes associated with early-naïve hPSCs include *DPPA3*, *GDF3*, *TFCP2L1* and *CD7*, all of which are common with N4+ early-naïve hPSCs; this indicates similarities in the gene expression changes and trajectory taken from primed to naïve pluripotency using both reprogramming methods.

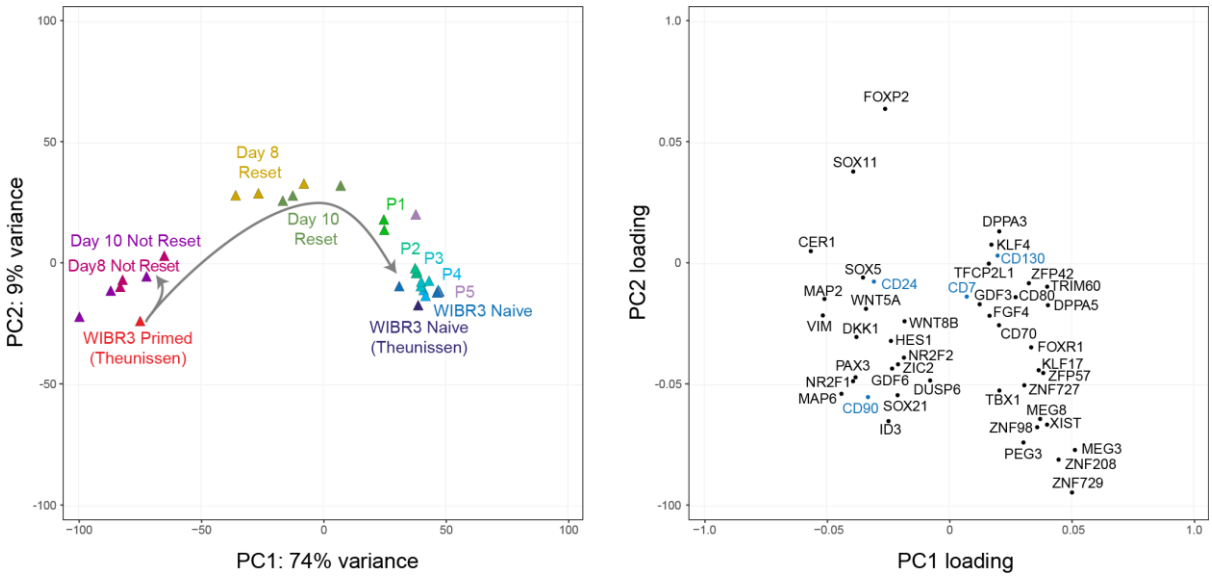


Figure 5.7. Transcriptional analysis of primed to naïve 5iLA reprogramming populations.
Principal component analysis of RNA-sequencing gene expression data from the different cell-sorted populations throughout reprogramming, primed, and established naïve hPSCs (left). Right: the contribution of selected genes to the first and second principal components.

Whilst PCA analysis is useful for deciphering the broad similarities and differences between sample types, the majority of variation in gene expression occurs between primed and later naïve hPSC samples; therefore, the loading plot predominantly highlights genes that contribute to their separation, as opposed to emphasising the transcriptional changes that occur early during the reprogramming transition.

5.2.4.2 Comparative transcriptional analysis of naïve hPSCs throughout 5iLA reprogramming

To examine the reprogramming transition in more detail, I broaden the analysis by extending the calling of differentially expressed genes across all sample types to the top 4,000 genes, followed by clustering analysis (Figure 5.8). Reassuringly, the samples (columns) largely cluster in the time-point order of reprogramming. Each row corresponds to a cluster of genes that exhibit the same expression pattern. Using eight clusters for the expression dynamics allowed genes to be classified into distinct clusters of co-expressed genes that describe the major stages I would expect during a state transition; including up and downregulation of genes at early, middle and late stages of reprogramming.

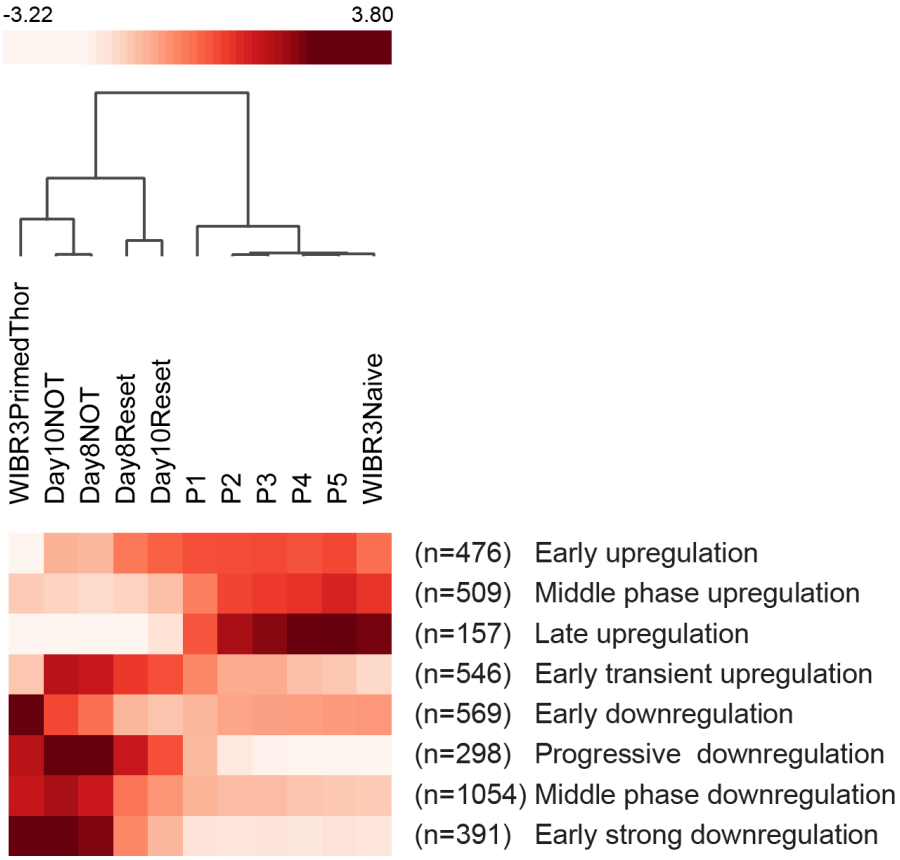


Figure 5.8. Clustering of differentially expressed genes during primed to naïve reprogramming reveals distinct clusters of co-expressed genes

Heatmap showing the enrichment or depletion of the identified gene clusters during primed to naïve reprogramming at frequent time points. Gene clusters were generated from the top 4,000 most differentially expressed genes across all time-points. Z-score normalized expression data is shown for each cluster of genes at each time-point. Co-expressed gene clusters have been given a short description of the transcriptional trajectory.

I subsequently plotted the expression values for all genes within each cluster along the reprogramming time course (Figure 5.9). Presented are the number of genes within each cluster, a selection of representative genes, and gene ontologies associated with each gene cluster.

The early upregulated gene cluster, which shows upregulation in day 8 Reset cells through to established naïve cells, contains genes that I have previously identified as being associated with early stage naïve hPSCs, such as *CD7*, *IL6ST* (*CD130*), *DPPA3* and *KLF4*, along with >450 additional genes, many of which are unknown in this context. The identification of genes within this cluster that have previous links with naïve pluripotency provides confidence in those that have no prior association. Moving forward more generally, this will provide new insights into transcriptional dynamics during primed-to-naïve reprogramming. This gene cluster provides several significant gene ontology categories. Examples include immune response and cellular

response to chemical stimuli – likely due to a number of cell surface receptor changes, including multiple interleukin and chemokine receptors. Moreover, several growth factors are upregulated during this early stage of reprogramming, such as *FGF4* and *PGF* (placental growth factor). PGF is known to be expressed from trophoblast cells during human embryo development (Athanassiades and Lala, 1998). Similarly, FGF4 is an important growth factor produced throughout mouse embryonic development (Niswander and Martin, 1992; Rappolee et al., 1994), and found to be necessary for the derivation and maintenance of trophoblast stem (TS) cells (Tanaka et al., 1998). This cluster also contains notable epigenetic modifiers such as *TET2* and *DNMT3L* that both have key roles in regulating DNA methylation levels that is an important hallmark to distinguish pluripotent states.

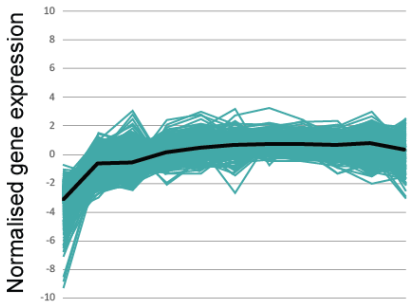
Whilst the next cluster termed "middle phase upregulation" did not return any significant GO terms, there are a number of interesting genes with relevance to primed-naïve pluripotency. *GDF3*, a known TGFβ superfamily member and naïve-pluripotency-associated marker, follows this expression trajectory. Likewise, *CD320* shown to be expressed in E6 blastocyst stage human embryos (Collier et al., 2017) is upregulated during the middle phase of reprogramming.

The late upregulated cluster contains many genes which are common to both methods of reprogramming, such as *KLF17*, *XIST*, *DPPA5* and many zinc finger proteins. In addition to these, several genes in this cluster contribute to DNA methylation and piRNA related GO terms, examples include *ZFP57*, *TDRD9* and *PIWIL2*. Both TDRD9 and PIWIL2 mediate the repression of transposable elements during germline development (Aravin et al., 2009; Shoji et al., 2009), and may therefore be required to moderate transposable element expression in naïve hPSCs.

Interestingly, >500 genes were clustered as being transiently upregulated early during primed to naïve reprogramming, with many significant GO terms associated with these. A few examples include cell-cell adhesion, driven by a collection of cadherin/protocadherin genes, and terms associated with the regulation of hormone levels including genes such as *GATA3*, *IL6*, *TFAP2B* and *CGB3*. *CGB3* encodes for a beta subunit of chorionic gonadotropin, a glycoprotein hormone released by trophoblast cells between implantation and gestation stages of pregnancy (Cole, 2009). Identification of candidate small molecules or growth factors, such as IL6 and CGB amongst many others, can be examined for their ability to facilitate primed to naïve reprogramming via media supplementation. Moreover, new signalling pathways previously not implicated in primed to naïve reprogramming may be identified with further analysis of the transcriptional dynamics, and can subsequently be experimentally tested.

The next four clusters all exhibit a downregulated trajectory. The early downregulated gene cluster includes previously identified primed-specific markers *DUSP6, OTX2* and *ETV5*, which are known regulators of, and/or regulated by FGF/ERK signaling (Hoch et al., 2015; Li et al., 2007). During primed to naïve reprogramming there is silencing of lineage-priming genes that are expressed in subpopulations of primed hPSCs, even though these cells are still pluripotent (Hough et al., 2014; Takashima et al., 2014). The GO terms associated with the four downregulated clusters include systems development, embryo development and fate specification. Within the gene cluster of progressively downregulated genes include key mesoderm lineage specification genes such as *T* (*brachyury*), *BMP7* and *MESP1*. Similarly, the middle phase and early strongly downregulated gene clusters include important lineage specification and stem cell differentiation genes such as *LEFTY1, ZIC2/3* and *ID2/3*. This infers that silencing of the primed pluripotent state-specific and lineage-priming GRNs is required for conversion back to naïve pluripotency (Kalkan et al., 2017; Luo et al., 2015; Smith, 2017). More than 1000 genes are downregulated during the middle phase and are silenced by passage 3, which is longer than I would have anticipated given the potent inhibitor treatment from the onset of reprogramming. This could mean that priming is a robust phenomenon, acting as a barrier to naïve conversion, or alternatively that complete silencing of the priming GRNs is not required before adopting naïve hallmark characteristics.

Early upregulation

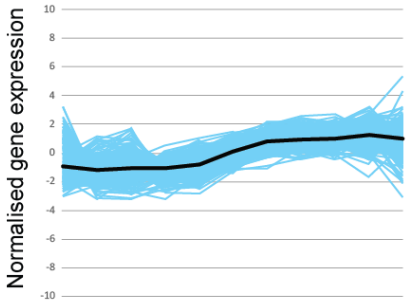


n=476
CD7
DPPA2, 3
DNMT3L
FGF4
KLF2,4,5
IL6R
IL6ST
TET2
UTF1

Gene Ontology Analysis

regeneration	6.03E-05
immune response	6.34E-05
cellular response to chemical stimulus	7.22E-05
regulation of cell proliferation	7.73E-05
organic anion transport	8.54E-05

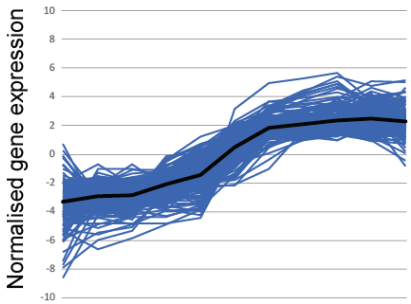
Middle phase upregulation



n=509
BMI1
CD320
FGFR4
FOXC1
GDF3
PRDM13
ZSCAN1

No significant GO terms

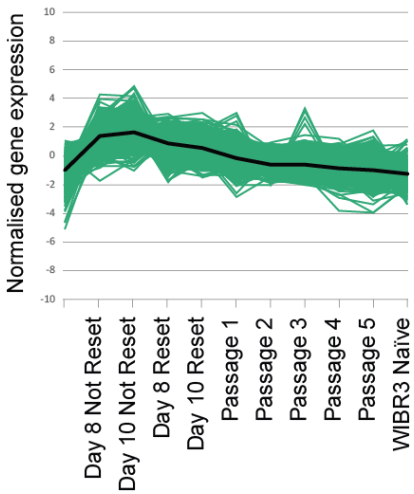
Late upregulation



n=157
CD70, CD80
DPPA5
DDX4
FOXR1
KLF17
MEG3,8,9
TBX1
XIST
ZNF729

regulation of transcription by RNA polymerase II	3.56E-07
transcription, DNA-templated	1.35E-06
piRNA metabolic process	5.81E-06
DNA methylation involved in gamete generation	8.37E-06
meiotic cell cycle	4.13E-05
DNA methylation or demethylation	4.74E-05

Early transient upregulation



n=546
FGF13
FOXP2
GATA3,6
H19
HAND1
IL6
PCAM1
PCGF5
TFAP2B

cytokine-mediated signaling pathway	3.22E-06
ovulation	1.90E-05
organelle organization	3.50E-05
nerve development	7.02E-05
cell fate determination	7.02E-05
regulation of hormone levels	7.89E-05
stem cell differentiation	8.71E-05
cellular protein metabolic process	1.73E-04
regulation of actin cytoskeleton reorganization	1.81E-04
cell-cell adhesion	1.83E-04
long-chain fatty acid metabolic process	2.21E-04
positive regulation of protein secretion	2.98E-04

Figure legend on the next page.

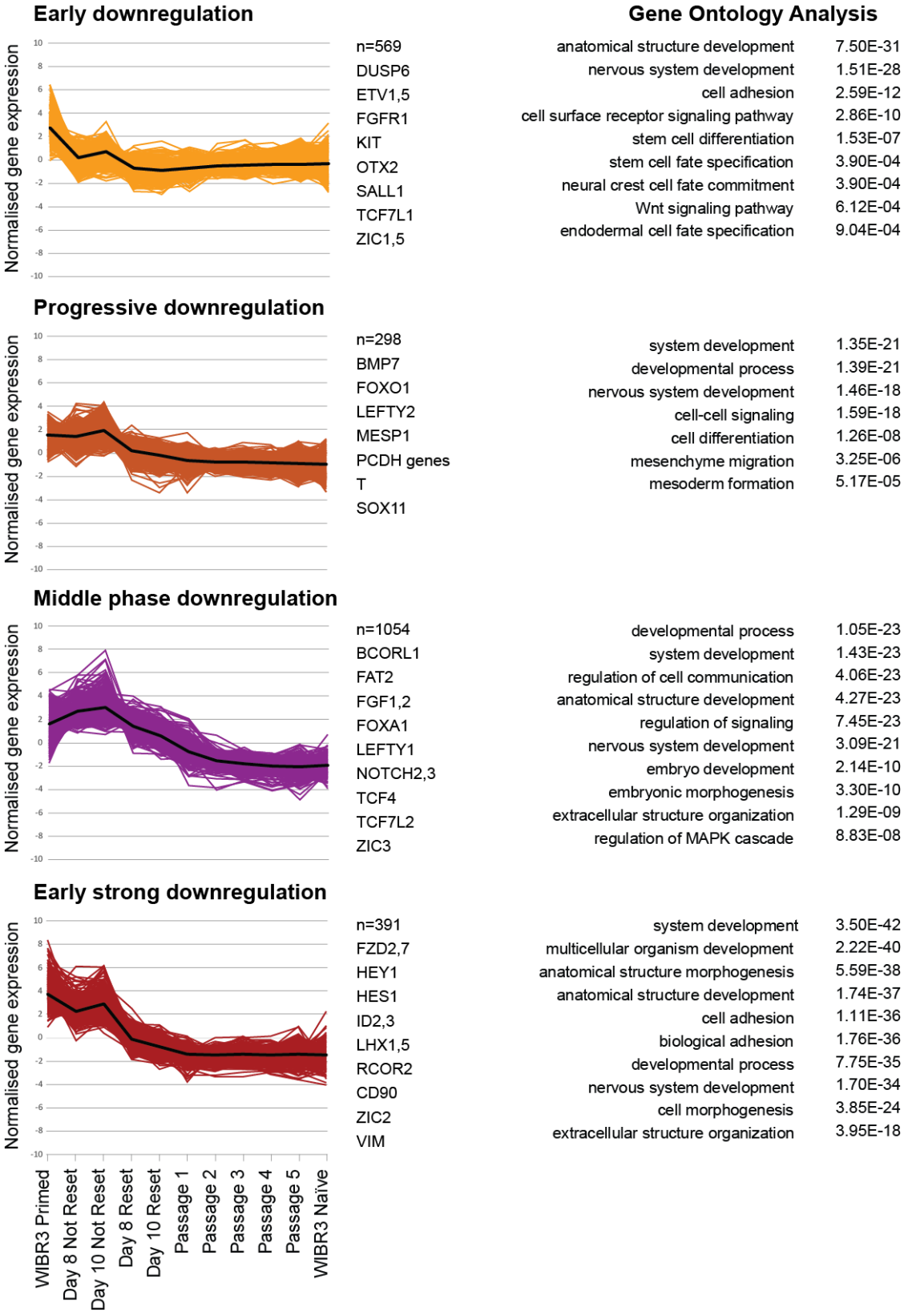


Figure 5.9. Transcriptional dynamics of co-expressed gene clusters during reprogramming. The expression (z-score normalised FPKM) of all genes within each cluster are shown for each reprogramming time-point. Alongside each plot are the number of genes in the cluster and representative genes. A selection of gene ontologies associated with each cluster are shown. Corrected p values were calculated using a modified Fisher's exact test followed by Bonferroni's multiple comparisons test.

One use of the transcriptional data could be the inference of upstream regulators that modulate the expression dynamics and subsequently naïve reprogramming. There are many possible options for further data mining in the long-term, but in the first instance I sought to identify candidate compounds that are known to regulate the genes within the four-upregulated clusters by performing DSigDB analysis (drug signatures database for gene set analysis) (Yoo et al., 2015). Highlighted in Table 5.1 are a selection of compounds, their target effect, and example genes that each compound is known to regulate. This has identified a number of conceivable compounds, such as the DNA methyltransferase inhibitor Decitabine (5-aza-2'-deoxycytidine), and the histone deacetylase inhibitor Trichostatin A. DNA demethylation occurs rapidly during primed to naïve reprogramming, likely via the downregulation of *DNMT1*, *DNMT3B*, and *UHRF1*, and the concurrent upregulation of *TET1* and *TET2*. Even so, 5-aza-2'-deoxycytidine has been shown to promote a variety of cell fate transitions (Kim et al., 2010; Mikkelsen et al., 2008; Yoon et al., 2006; Zhao et al., 2015), and the addition during primed to naïve reprogramming could enhance this process. Additionally, the compound Trichostatin A was identified, that is a histone deacetylase (HDAC) inhibitor. Several reports of naïve cell derivation by alternative reprogramming methods have used HDAC inhibitors (Guo et al., 2017; Ware et al., 2014). This provides some reassurance for the other compounds identified, all of which have yet to be examined for their ability to promote primed to naïve reprogramming, and this offers an avenue for further experimental exploration.

Table 5.1. DSigDB results identifying compounds that target genes which are upregulated during primed to naïve reprogramming.

Compound	Predicted target effect	Adjusted P-value	Genes
Prenylamine	Calcium channel blocker	9.07E-07	*MT2A, MX1, MX2, MT1M, MT1X, MT1G, MT1H, MT1E, MT1HL1, SGK1*
Progesterone	Steroid hormone	1.09E-05	*GATA6, CGA, CGB, DEPTOR, KLF4, KLF5, GDF15, IL6, IL6R, IL6ST, TGFB3*
Decitabine	DNA methyltransferase inhibition	2.32E-05	*BMI1, CD70, CD80, KLF2, KLF5, KLF9, IL6, IL6R, IL6ST, SOX15, SOX18,*
Dexamethasone	Glucocorticosteroid	8.62E-04	*BMP6, IL6, IL6R, IL32, NFKB2, PCAM1, POMC, TGFB3, VEGFA, WNT5B,*
Etynodiol	Synthetic progestational hormone	2.48E-03	*CEBPD, IL1R1, KLF9, MT1X, MT2A, OLAH, SPTLC2*
Prenylamine	Calcium channel blocker	2.61E-03	*CDKN1A, GDF15, GPRC5A, IFIT3, IFIH1, IL6, KLF5, SAT1*
Mometasone	Glucocorticoid	3.21E-03	*CCNA1, FKBP5, KLF9, MT1G, MT1X, MT2A, OLAH, SMARCA2*
Trichostatin A	HDAC inhibitor	6.25E-03	*BMI1, DEPTOR, DPPA2, DPPA5, GATA2, GATA3, KLF5, KLF9, TET2, TFAP2B, XIST*

Gonadorelin	Gonadotropin-releasing hormone receptor agonist	1.35E-02	*CGB5, CYGB, ESRRB, FGF13, IGF1, IL6, NFKB2, PGF, VEGFA,*
Phenoxy-benzamine	Alpha-adrenergic antagonist	1.61E-02	*GATA6, GDF15, IFI16, IL6ST, KLF4, STYK1*

5.2.5 X-chromosome status

As previously discussed, one of the most faithful indicators of naïve hPSCs is the reactivation of the silenced X-chromosome in female cells. Naïve and primed hPSCs have a close resemblance to the transcriptional and epigenetic landscape of primate pre- and post-implantation epiblast respectively (Boroviak and Nichols, 2017). Given that epigenetic mechanisms regulate X-chromosome activity, it is important to study this event in a model system that best resembles the *in-vivo* cellular context. For this reason, naïve and primed hPSCs offer an opportunity to obtain a better mechanistic understanding of how X-chromosome status may be regulated in humans. Analysis of the time-course NK2 t2iL+PKCI reprogramming revealed X-chromosome reactivation occurred between day 10 and day 30 (passage 5). To examine the X-chromosome status during 5iLA reprogramming a similar approach was taken. RNA-sequencing datasets were examined and using SNP information genes could be classed as mono- or bi-allelically expressed. This analysis again reveals that X-chromosome reactivation occurs predominantly during the later maturation phase of naïve cell formation, with reactivation occurring between passage 1 (day 14) and passage 2 (day 18) (Figure 5.10). It was not possible to examine as many genes for the WIBR3 line compared to H9 NK2 cells, because WIBR3 cells have only been genotyped by SNP array and not by sequencing. Therefore, the total number of informative SNPs in coding regions is reduced. Nonetheless, a consistent trend of reactivation during P1-P2 is observed for those genes that could be assessed. As observed under NK2 t2iL+PKCi reprogramming, *XIST* is also upregulated during the later stages of reprogramming concomitant with XCR. As previously discussed, the upregulation of *XIST* expression is counterintuitively indicative of X-chromosome reactivation (Sahakyan et al., 2017; Vallot et al., 2017). This result is therefore consistent with X-chromosome reactivation occurring between day 14 and day 18, and the conclusion that X-chromosome status is a robust hallmark of naïve pluripotency.

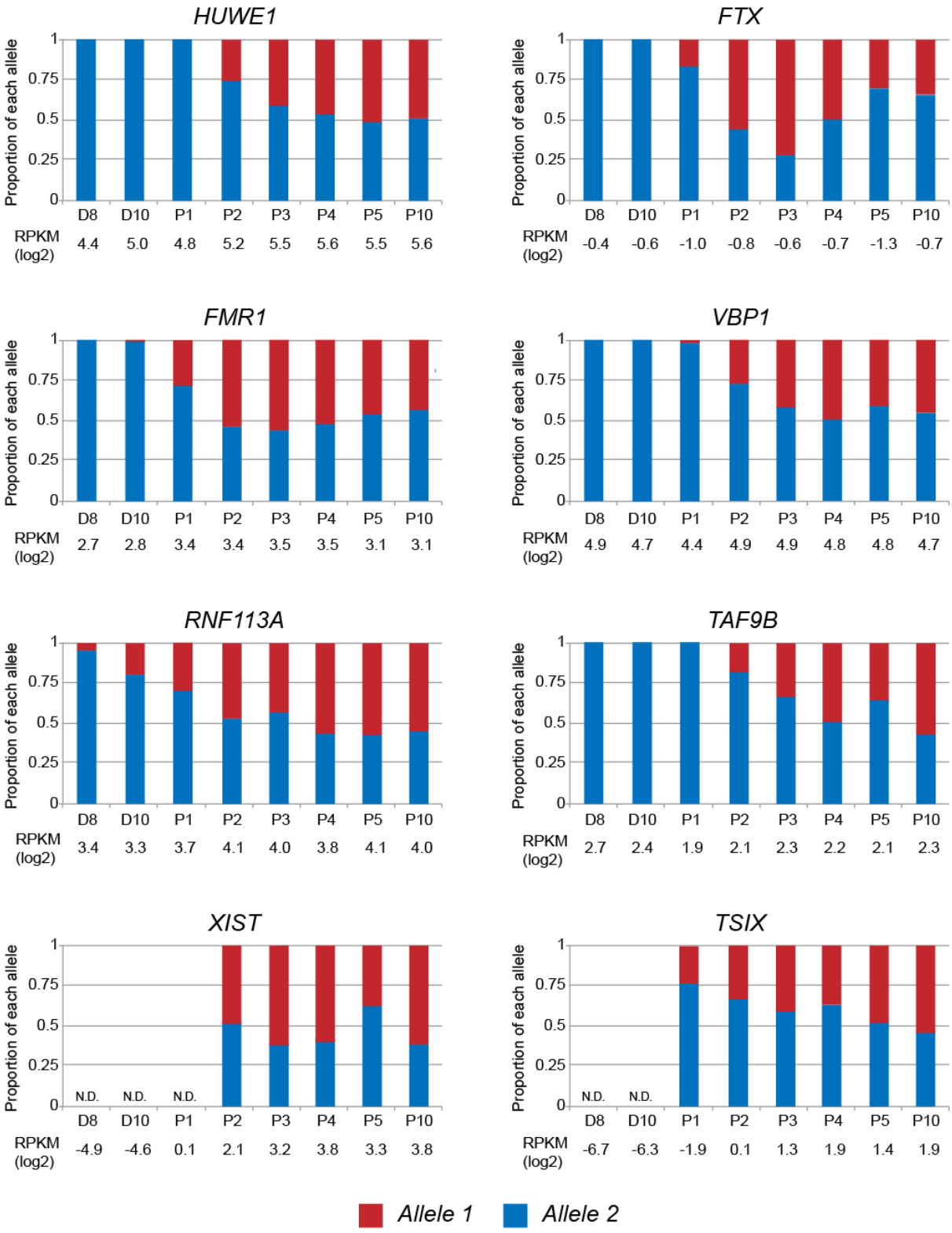


Figure 5.10. X-chromosome reactivation dynamics during primed to naïve reprogramming. Allelic analysis of RNA-seq data for X-linked genes at time-points throughout reprogramming. Informative SNPs within X-linked genes of the WIBR3 hPSC line were used to classify expression from each allele: allele 1 (red), allele 2 (blue); Log2 RPKM values for each time point are displayed below each transcript graph.

5.2.6 Comparative transcriptional analysis of the routes from primed to naïve pluripotency by different reprogramming methods

To better understand the route from primed to naïve pluripotency I not only wanted to examined more intricate time points, but also to compare different reprogramming protocols to establish the similarities and differences in paths taken. To this end, I firstly performed a transcriptional comparison between the two reprogramming methods NK2 t2iL+PKCi and 5iLA (Figure 5.11). The PCA plot shows that the established naïve hPSCs derived by both reprogramming methods exhibit a remarkable transcriptional similarity; however, there are notable differences in the trajectory taken from primed pluripotency to get there. PC1 clearly separates primed from naïve hPSCs. Along this component, there is a good alignment between the matched cell types: Primed, Not Reset, Day 10 Reset, and Naïve hPSCs. However, the two reprogramming methods exhibit divergence in their trajectories from primed to naïve pluripotency shown along PC2.

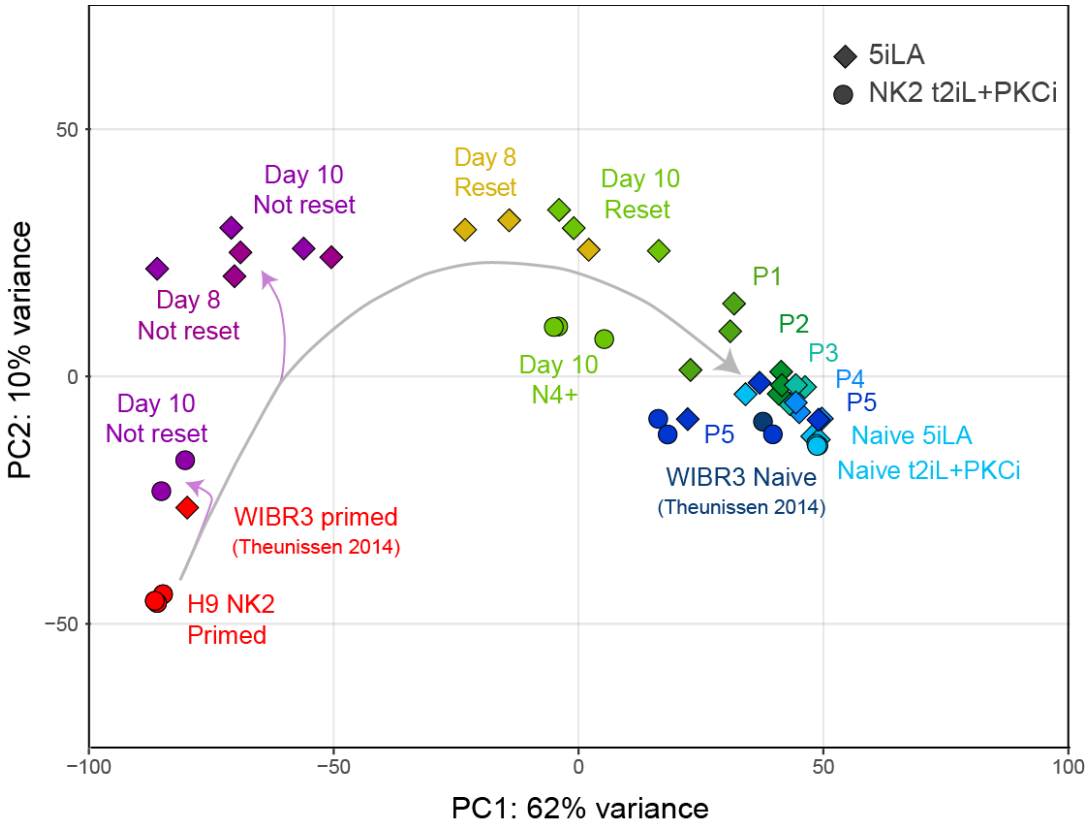


Figure 5.11. Comparative transcriptional analysis of 5iLA and NK2 t2iL+PKCi reprogramming.
Principal component analysis of RNA-sequencing gene expression data from time points throughout primed to naïve reprogramming. Circles represent H9 NK2 t2iL+PKCi reprogramming time-series, diamonds denote 5iLA.

Next I chose to focus in on the differences between day 10 time points of the two reprogramming methods, as this is the first matched time-point that also displays some of the greatest divergence. PCA results show the samples cluster by reprogramming method, as would be expected, and the majority of variation is captured in PC1 (91%) (Figure 5.12). To examine the most variable genes that influence the clustering of samples, a loadings plot is shown with a selection of these genes. This immediately highlights some curious differences between the two methods. Both *TET2* and *TET3*, which are capable of active DNA demethylation, are upregulated in day 10 5iLA reset cells compared to day 10 cells under NK2 t2iL+PKCi reprogramming. This may indicate that DNA demethylation could occur earlier, or more rapidly, under 5iLA reprogramming. Another gene of considerable interest and which is also upregulated in 5iLA samples is *TFAP2C*. This transcription factor is essential for naïve human pluripotency using both 5iLA and t2iL+PKCi reprogramming (Pastor et al., 2018). TFAP2C is able to bind the proximal enhancers of pluripotency related genes, including a naïve-specific enhancer that regulates *POU5F1* (Pastor et al., 2018). Of note, day 10 5iLA cells also express *GATA2, GATA3* and *GATA6* to higher levels. All three of these GATA factors are expressed in both the mouse and human trophectoderm (Home et al., 2009; Koutsourakis et al., 1999; Ma et al., 1997; Ralston et al., 2010; Roode et al., 2012). This raises the speculative possibility that early reprogramming cells could transition through a TE-like state under 5iLA reprogramming. Alternatively, the differential expression of GATA factors may be due to cell line differences, and GATAs 3 and 6 are lowly expressed in WIBR3 primed cells, but not in the H9 NK2 line. Whilst in the first instance I have focused on these PSC lines, in the future it will be important to perform the reciprocal experiments to distinguish line-specific from reprogramming-specific differences.

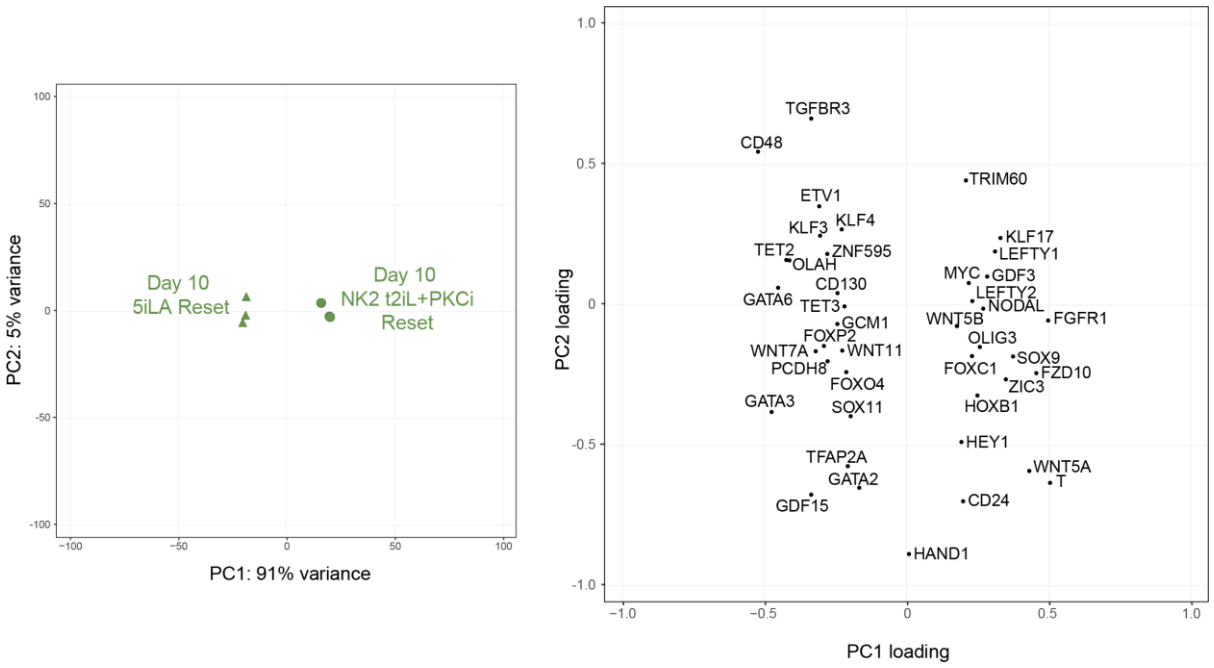


Figure 5.12. Comparative transcriptional analysis of early-naïve reset cells via different reprogramming methods.
Principal component analysis of RNA-sequencing gene expression data from Day 10 Reset cells generated using 5iLA and NK2 t2iL+PKCi reprogramming methods (top). Bottom: the contribution of selected genes to the first and second principal components.

To examine the top 1,000 most variably expressed genes between the two reprogramming methods in more detail, I performed gene-clustering analysis. Genes were firstly classified as up regulated in 5iLA cells (n=603), or upregulated in t2iL+PKCi cells (n=397) (Figure 5.13). A number of distinct clusters were apparent within the two gene sets: "primed genes", defined as being upregulated in primed and day 10 cells (red), "transient genes", upregulated in day 10 cells only (green), "naïve genes", upregulated in day 10 and naïve cells (blue), and "unaffected genes" with minor expression changes between the three cell types (orange). One of the first notable observations is the continued expression of "primed genes" in both 5iLA and t2iL+PKCi cells at the day 10 time point, even though many genes associated with naïve pluripotency have already been induced. This indicates that it may not be necessary to silence lineage-priming and primed-pluripotency-specific genes before converting back to a naïve state. Another notable observation is a cluster of genes that are transiently upregulated in day 10 cells under 5iLA reprogramming conditions (Figure 5.13). Within this cluster are four genes (*GCA*, *CGB*, *CGB5*, and *CGB8*) that encode for the alpha and beta subunits of chorionic gonadotropin, a hormone released by trophoblast cells (Zhuang and Li, 1991). Intriguingly neither *CGA* nor *CGB* genes are expressed at any stage during t2iL+PKCi reprogramming, and are only expressed transiently in early-naïve 5iLA cells. This could mean that chorionic

gonadotropin is produced during 5iLA reprogramming and may facilitate this process, suggesting one possible divergent pathway between the two reprogramming methods.

Finally, a cluster of "naïve" genes that are upregulated in day 10 and naïve hPSCs exists for both reprogramming methods. A few interesting examples that are differentially upregulated in 5iLA cells compared to t2iL+PKCi include *GCM1*, *KLF3* and *KLF4*. GCM1 has been shown to bind the promoter region of all six *CGB* genes to promote gene expression (Cheong et al., 2016). Moreover, a TFAP2C binding motif is also present at these promoter regions (Cheong et al., 2016), which may be another method by which *CGB* genes are upregulated, given that TFAP2C is also an early responder during 5iLA reprogramming. Another notable gene within the "naïve" gene cluster is *NODAL*, which encodes for a secreted protein that binds to type I and type II TGF-β receptors (James et al., 2005b; Massague, 1998). *NODAL* is not expressed in H9 NK2 primed cells, but is strongly upregulated by day 10 and remains expressed in naïve hPSCs. Related to this, *GDF3*, *LEFTY1* and *LEFTY2* exhibit this same expression pattern in H9 NK2 cells – low in primed, but strongly upregulated in early and established naïve hPSCs, suggesting a potential role for Activin/Nodal signalling in naïve pluripotency. In line with this observation, Blakeley and colleagues found that human embryos treated with a potent TGF-β inhibitor resulted in the lack of NANOG expression and impaired epiblast formation (Blakeley et al., 2015). Interestingly, *NODAL* is expressed at similar levels throughout the entire 5iLA transition, extending to WIBR3 primed cells and established 5iLA naïve hPSCs. This again highlights a potential PSC line specific difference and could infer that WIBR3 primed cells have a head start towards the acquisition of naïve pluripotency. The PCA plot displaying both reprogramming trajectories additionally suggests that this may be the case.

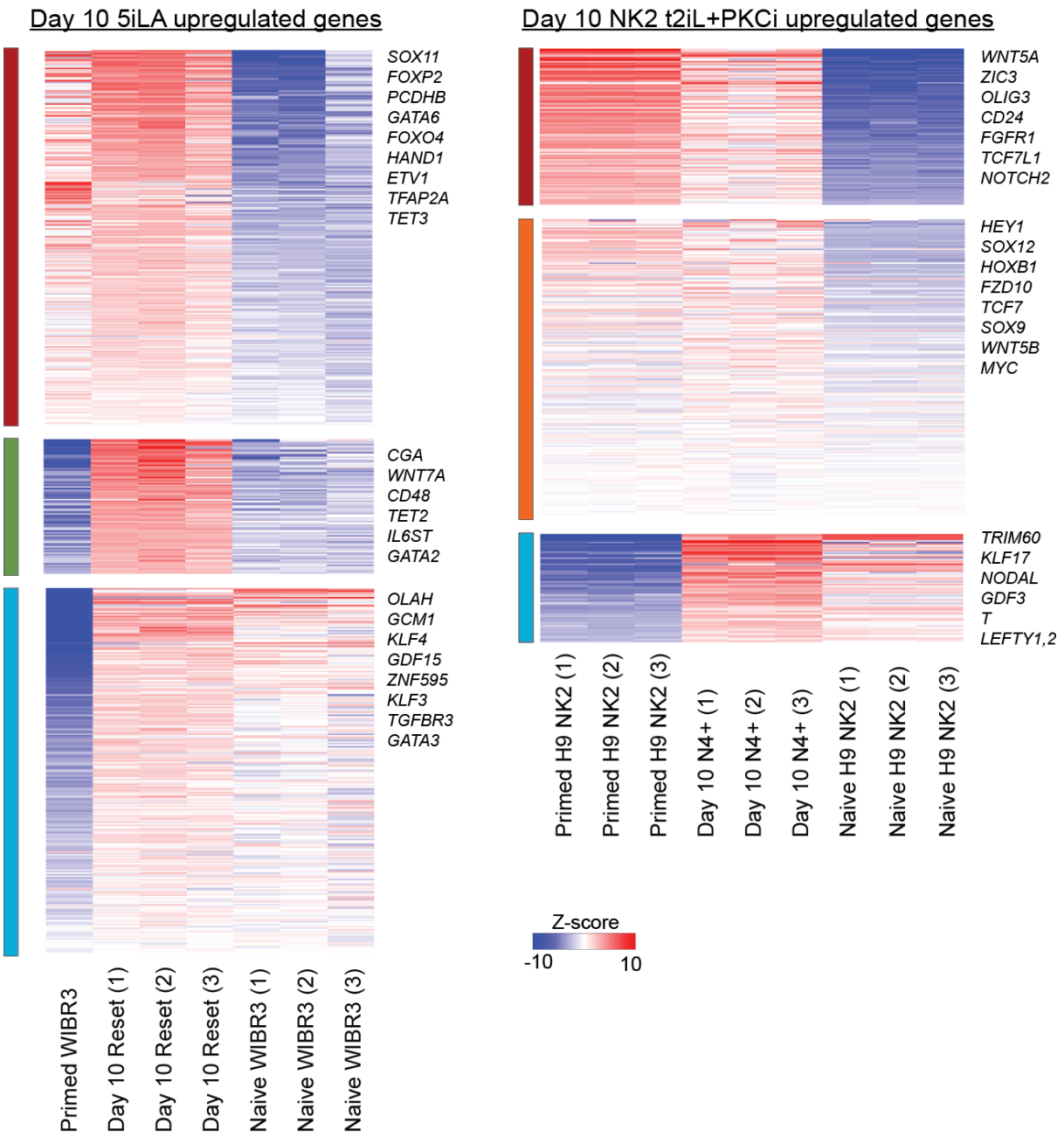


Figure 5.13. Clustering of differentially expressed genes between 5iLA and NK2 t2iL+PKCi Day 10 Reset cell populations reveals distinct clusters of co-expressed genes.

The top 1,000 differentially expressed genes between 5iLA and t2iL+PKCi Day 10 Reset cells were used for clustering analysis. Data from the parental primed hPSCs and established naïve hPSCs plotted for comparison. Left: Genes that are upregulated in day 10 5iLA compared to NK2 t2iL+PKCi reprogramming form three distinct clusters: genes upregulated in primed and day 10 cells (red), genes transiently upregulated in day 10 cells (green), and genes upregulated in day 10 and naïve cells (blue). Right: Genes that are upregulated in day 10 NK2 t2iL+PKCi compared to 5iLA reprogramming separate into three clusters: genes upregulated in primed and day 10 cells (red), genes with minor expression changes between the three cell types (orange), and genes upregulated in day 10 and naïve cells (blue). Z-score normalised FPKM expression and example genes within each cluster are shown.

In light of the unexpected observation that several trophectoderm related markers are upregulated transiently during 5iLA reprogramming, I subsequently assessed a broader panel of 500 TE-enriched genes. These 500 genes are the most differentially expressed genes between the trophectoderm and epiblast, determined by scRNA-seq of human embryos (Blakeley et al., 2015). To my surprise, the transient upregulation of TE-related genes appears to be more widely applicable, with a significant upregulation in day 10 reprogramming samples compared to primed hPSCs. Example genes that exhibit some of the greatest expression changes include *GATA2*, *LRP2*, *CDX2*, *SLC7A2*, *TEAD3* and *HOOK1*, which are all recognised human TE-markers (Assou et al., 2012; Blakeley et al., 2015; Petropoulos et al., 2016). Further to this, I have observed colonies with a trophoblast-like morphology arising during primed to naïve reprogramming (Figure 5.15); however, to validate this observation it will be necessary to assess the protein expression of TS-markers. Even so, these results suggests it may be possible to derive trophoblast-like stem cells; by amending the culture conditions to enrich for their survival and expansion, as opposed to selecting for naïve hPSCs, it could be possible to preserve this population in the future.

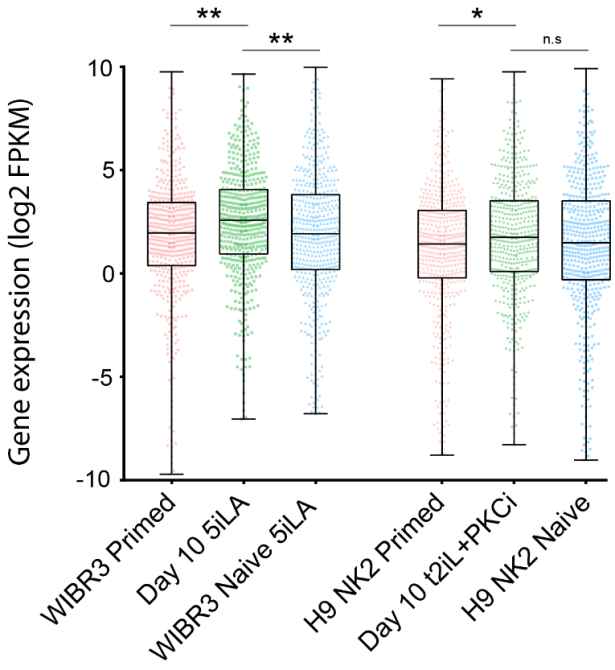


Figure 5.14. Comparative transcriptional analysis of trophectoderm-enriched genes across primed to naïve reprogramming time-points.
Assessment of the top 500 differentially expressed genes (TE compared to EPI, acquired from scRNA-seq of human embryos (Blakeley et al., 2015), across the reprogramming time-points for 5iLA and NK2 t2iL+PKCi reprogramming methods. Log2 FPKM expression shown. Average FPKM calculated from 3 biological replicates, with the exception of WIBR3 primed (n=1, (Theunissen et al., 2014)). Boxplots show the median, min and max values. ANOVA with Tukey's multiple comparisons test performed (*$p < 0.05$, **$p < 0.005$).

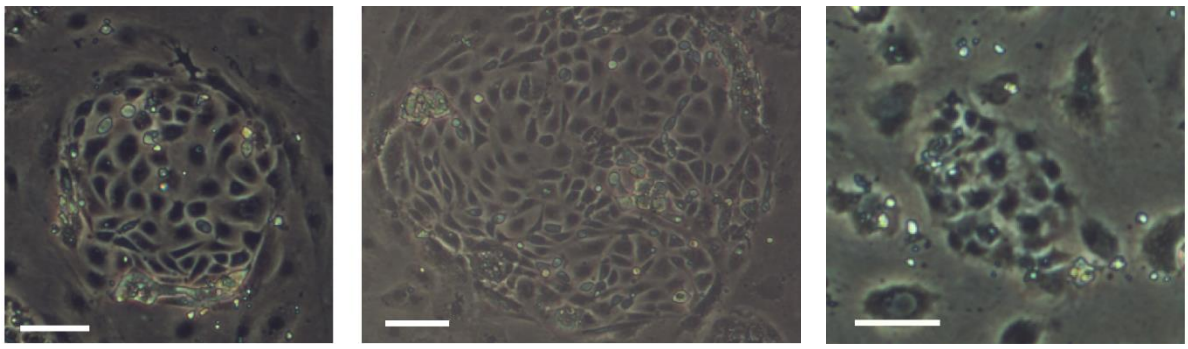


Figure 5.15. A subset of human PSCs adopt a trophoblast-like stem cell morphology during primed to naïve reprogramming.
Representative phase contrast image of early reprogramming cells which adopt a trophoblast-like morphology. Left and centre: NK2 t2iL+PKCi reprogramming; Right 5iLA reprogramming. Scale bars, 100 µm, 4x objective.

KLF transcription factors are known to play an important role in regulating pluripotency (Bialkowska et al., 2017; Pastor et al., 2018; Takashima et al., 2014). To my surprise, most of the different KLF factors appear to be utilised more so during early 5iLA reprogramming, with the exception of *KLF17* that is expressed to significantly higher levels in NK2 t2iL+PKCi day 10 cells (Figure 5.16). Under both reprogramming methods, *KLF17* is ultimately upregulated to similar levels in mature naïve hPSCs; however, it appears that during the early stages of reprogramming, there is differential utilisation of the KLF family members. Interestingly, *KLF2* was expressed to higher levels during early 5iLA reprogramming, even though NK2 t2iL+PKCi reprogramming requires the transient over-expression of *NANOG* and *KLF2*. This infers that by day 10 the transgenes are no longer induced and reprogramming resumes independently of transgene expression. Given the importance of KLF factors in human pluripotency, little is known about their target specificity and probable redundancy, and this therefore provides an interesting avenue for future exploration.

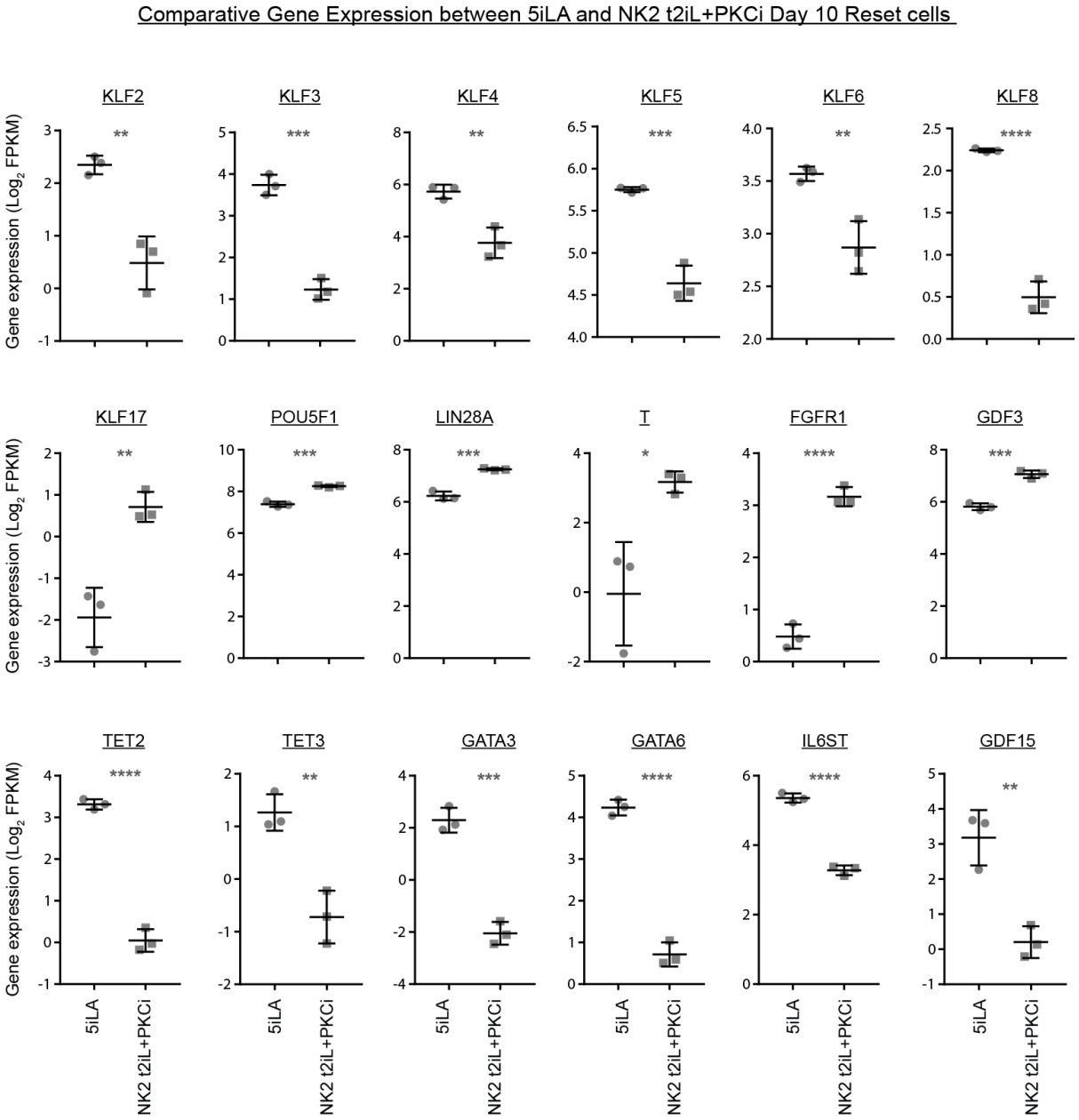


Figure 5.16. Important regulatory genes exhibit divergent expression dynamics between 5iLA and NK2 t2iL+PKCi Day 10 Reset cell populations.

Plotted are the gene expression profiles of example genes that exhibit differential expression between the two reprogramming methods. Data show the mean log2 FPKM ± SD of three biological replicates and were compared using a two-tailed t-test (*$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$, ****$p < 0.00005$).

5.3 Discussion

One of the first indicators that 5iLA reprogramming may be more efficient than NK2 t2iL+PKCi came from the morphological observation that dome-shaped colonies appeared more frequently by day 10. In support of this idea, the flow cytometry results of day 10 reprogramming cells revealed a population of ≈10% corresponding to Reset cells, which is far greater than ≈2% observed under t2iL+PKCi culture. Although this implies the efficiency of reprogramming may be higher, without performing lineage tracing experiments it would not be possible to conclude this. There is the possibility that the same efficiency is observed by both methods, yet expansion of naïve hPSCs under 5iLA culture is faster. Alternatively, the selection pressure of 5iLA culture may be greater, and subsequently the cells that are refractory to reprogramming do not survive, giving the illusion of enhanced efficiency. I believe this is a more likely outcome given the initial reports of "wide-spread cell death" in the original publication (Theunissen et al., 2014), and from what I experience early during 5iLA reprogramming.

In a similar manner observed for NK2 t2iL+PKCi reprogramming, the day 10 Reset population under 5iLA culture gave rise to colonies with a predominantly naïve morphology. Furthermore, the Not Reset population that corresponds to >50% of all day 10 cells, struggled to form colonies of any morphology. This result instils confidence in the multiplexed antibody panel for the reliable and versatile identification of the nascent naïve hPSC population using a different reprogramming method. Moreover, I have shown the antibody panel is effective for the tracking and subsequent isolation of naïve hPSCs throughout the reprogramming process.

The examination of early reprogramming populations under 5iLA has inadvertently highlighted a potential caveat to the OCT4-ΔPE-GFP reporter system. By day 10 of reprogramming, ≈75% of the population are GFP positive; however, only 10% of cells are classed as Reset according to the surface marker panel. Likewise, ≈50% of the Not Reset population – which does not form naïve colonies – are GFP positive. This strongly indicates that the expression of OCT4-ΔPE-GFP is not restricted to colony forming naïve hPSCs, and is therefore not a robust method to isolate naïve cells under 5iLA reprogramming. The OCT4-ΔPE-GFP reporter system was originally devised based upon enhancer utilisation during mouse preimplantation development and the conserved activity seen in naïve mouse ESCs (Choi et al., 2016; Tesar et al., 2007; Yeom et al., 1996). However, the enhancer elements driving *OCT4* expression in humans had not been examined in detail until recently (Pastor et al., 2018). Pastor and colleagues discovered a naïve-specific intronic enhancer element that is essential for *OCT4* expression in human naïve PSCs (Pastor et al., 2018). Amending the previously used reporter system, such that GFP

expression is coupled to the newly identified intronic-enhancer element usage, may provide a more faithful readout of naïve state reprogramming, and subsequently enable using this reporter system to accurately isolate naïve hPSCs.

The chemical screening approach used by Theunissen and colleagues was reliant on the induction of OCT4-ΔPE-GFP reporter expression. Compounds that promoted GFP expression were ultimately combined to form the 5iLA media formulation. However, given the finding that the OCT4-ΔPE-GFP reporter is not restricted to naïve hPSCs, it raises questions over their specificity for inducing naïve hPSCs. Adopting a similar compound library screen using a more reliable report system, or the surface marker panel, could prove useful for the future identification of compounds that induce naïve pluripotency.

Using the antibody panel to isolate nascent naïve cells as they emerge and progress throughout reprogramming has provided the first detailed insight into the transcriptional dynamics across human pluripotent states. The clustering analysis of variably expressed genes across the 5iLA reprogramming time course revealed distinct clusters of co-expressed genes, many of which have thought-provoking biological functions. Of note, transcription factors *KLF2*, *KLF4* and *KLF5* are all upregulated early during reprogramming, along with >400 other genes. It is well established that KLFs 2, 4 and 5 work in cooperation with Nanog to maintain mESC pluripotency and self-renewal (Jiang et al., 2008). Furthermore, *in-vitro* studies of mESCs have highlighted MEK/ERK signalling as a negative regulator of *Klf2*, and upon treatment with PD0325901 *Klf2* expression is quickly elevated (Yeo et al., 2014). Similarly, *Klf4* and *Klf5* are targets of Jak/Stat3 signalling and are upregulated in the presence of LIF (Hall et al., 2009). It is therefore not all that surprising that specifically these three KLF factors are induced at the very earliest stages of naïve reprogramming. Moreover, this infers a conserved signalling pathway regulation of these factors in humans.

The importance of KLF transcription factors in relation to naïve human pluripotency has been touched upon in several studies. Takashima and colleagues show that shRNA knockdown of *KLF4* in t2iL+PKCi naïve hPSCs (but not in primed hPSCs) caused a vast reduction in AP+ pluripotent colony formation (Takashima et al., 2014). More recently, KLF binding motifs were the most enriched factor found at open chromatin sites in human blastocysts and naïve hPSCs, signifying an overlapping and important role in both *in-vitro* and *in-vivo* naïve cell types (Pastor et al., 2018). Nonetheless, the target specificity and redundancy of the different KLF factors in human primed and naïve pluripotency are currently unknown, and this provides an interesting avenue for further investigation.

There is still an incomplete understanding of the transcriptional dynamics that drive primed and naïve pluripotency. Intricate analysis throughout 5iLA reprogramming has provided a comprehensive insight into the order of transcriptional changes. A few examples of the top most variably expressed genes across the 5iLA transition include *DPPA5*, *KHDC1L*, *TRIM60*, *OLAH*, *KHDC3L*, *VIM*, *ALPP* and *GTSF1*; alongside many other examples, these will provide a useful resource of transcriptional markers for the future assessment of cell identity along the primed to naïve axis. Furthermore, it will be interesting to utilise the time-series expression data for impending gene regulatory network modelling. Due to time constraints it has not yet been possible to perform this analysis, however recently devised tools such as dynamical GENIE3, that are specifically designed for GRN inference from time course transcriptional data, will help to reveal the GRNs utilised throughout reprogramming in the future (Huynh-Thu and Geurts, 2018). A recent study has similarly performed time-series analysis of transcriptional profiling across 5iLAF reprogramming, although starting from immortalised secondary human fibroblasts (Wang et al., 2018). Bulk RNA-sequencing was performed at regular intervals throughout, revealing a transient reactivation of transcripts associated with 8-cell stage embryos during the later stages of reprogramming. Given that lineage segregation has not occurred at this stage of human embryonic development (Petropoulos et al., 2016), this again infers naïve hPSCs may adopt a transcriptional identity encompassing both embryonic and extraembryonic cell types. Further experiments are warranted to confirm whether this is the case.

Transcriptional data is frequently used as a method to infer cell identity. However, there can be a disconnect between transcript levels and protein expression, particularly in the case of state transitions (Liu et al., 2016). There is often a time lag between transcription, transcript export and subsequent translation, and the mRNA sequence and a multitude of additional factors can influence translation rates (Liu et al., 2016). For this reason, it would be nice to perform additional experiments in the future employing mass cytometry (CyTOF) technology, with the formulation of antibody panels to study primed to naïve reprogramming at the protein level. Similar to flow cytometry, mass cytometry utilises antibodies conjugated to metal isotopes and enables the quantification of >50 different protein markers to be assessed at a single-cell resolution (Bendall et al., 2012). Moreover, the concurrent examination of phospho-specific antibodies, and antibodies for epigenetic modifiers, will provide a comprehensive overview of the reprogramming trajectory, and the cross talk between signalling, epigenetic, and gene regulatory networks.

An unexpected outcome of the transcriptional dynamics of 5iLA reprogramming was the observation that multiple trophectoderm related genes were transiently upregulated. I firstly

noticed that *GATAs 2, 3* and 6 were all upregulated by day 10 of reprogramming. All three of these GATA factors are strongly enriched in the trophectoderm cells of both mouse and human embryos compared to the epiblast (Blakeley et al., 2015; Home et al., 2009; Petropoulos et al., 2016; Ralston et al., 2010; Roode et al., 2012). Further analysis of reported human TE enriched genes revealed that many were transiently upregulated under both reprogramming conditions. However, there were some notable exceptions which displayed 5iLA specificity, such as a collection of genes encoding for the alpha and beta subunits of human chorionic gonadotropin (hCG). As mentioned previously, both GCM1 and TFAP2C are also upregulated early under 5iLA reprogramming, and are known regulators of CGB gene expression (Cheong et al., 2016). However, this may not be the mechanisms for CGB upregulation, as both GCM1 and TFAP2C continue to be expressed throughout both reprogramming methods, whilst CGB genes are silenced at later stages in 5iLA culture, and CGB genes are not expressed at any stage under t2iL+PKCi culture. Collectively, this infers hCG could be transiently produced during 5iLA reprogramming and might promote naïve state transitioning.

Alternatively, early naïve hPSCs may possess the ability to become both naïve hPSCs with embryonic potential, and cell types with extraembryonic potential such as trophoblast-like stem cells; but with prolonged culture in 5iLA the route to naïve pluripotency may be enforced, and thus the trophectoderm related genes are largely downregulated at later stages. Conversely, it is possible that multiple cell types are isolated using the antibody panel, including a population with embryonic potential and a separate population with an extraembryonic identity. A similar phenomenon has been reported during chemical mediated iPSC reprogramming (CiPSC), where populations of cells can pass through an extra-embryonic endoderm (XEN)-like state before transitioning to form mouse CiPSCs (Li et al., 2017; Zhao et al., 2018; Zhao et al., 2015). It has yet to be shown whether this is also true during human iPSC reprogramming. As it currently stands there are no protocols to establish and maintain human XEN cells *in-vitro*. However, there has been a recent report describing human trophoblast stem cell derivation and maintenance (Okae et al., 2018). A chemical cocktail that promotes Wnt and EGF signalling, in combination with TGF-β, HDAC, and ROCK inhibition permitted the expansion of several trophoblast cell types. These culture conditions provide a framework to guide the amendment of current naïve culture conditions, which will be required to fully explore the possibility of deriving trophoblast-like stem cells during primed to naïve reprogramming. Furthermore, the surface marker KRT7 was used to positively assess the human TS cells, in combination with a lack of HLA-A,B,C expression (Okae et al., 2018). The addition of these two markers to the

current multiplexed antibody panel offers an alternative method to enrich for potential TS-like cells during primed to naïve reprogramming.

An additional aim of this chapter was to characterise the transcriptional dynamics of primed to naïve reprogramming, with the intent to identify candidates that could be added to the culture media to facilitate naïve cell formation. The first example of this is human chorionic gonadotropin, as mentioned previously. The second candidate is interleukin-6 (*IL-6*) that is highly expressed in 5iLA day 8 and day 10 Reset cells, and day 10 N4+ t2iL+PKCi cells, but subsequently downregulated in established naïve hPSCs. IL-6 exerts its biological activities by binding to the IL-6R and inducing homodimerization of gp130 (*IL6ST/*CD130) (Heinrich et al., 1998). Unsurprisingly *IL6ST* is highly expressed throughout reprogramming, as CD130 is one of the naïve-specific markers used for isolating reprogramming cells. Many other cytokines can also signal via gp130, including LIF (Gearing et al., 1992; Heinrich et al., 1998); however, the additional upregulation of both *IL-6* and the *IL-6 receptor* indicates an important function for this cytokine and signal transduction pathway. However, experiments to examine IL6-R protein expression and IL-6 secretion would be necessary to both validate and distinguish their importance from LIF signalling in human PSCs. Finally, by performing DSigDB analysis I was able to identify candidate compounds to be added via media supplementation during reprogramming, with the overarching aim to enhance naïve cell formation. Amongst these were compounds with conceivable modes of action, such as promotion of DNA demethylation (Decitabine) and histone deacetylase inhibition (Trichostatin A). Additionally, several steroid hormones and hormone receptor agonists were identified, none of which have previously been examined for their ability to promote naïve pluripotency. This therefore provides an interesting and unexplored direction for future studies.

The analysis of X-chromosome reactivation (XCR) under NK2 t2iL+PKCi reprogramming revealed that between day 10 and passage 5, the silent X-chromosome undergoes XCR. By performing the equivalent analysis using the 5iLA reprogramming transition with more frequent time-points, the window of XCR has been narrowed to passage 1 (day 14) – passage 2 (day 18). In the scope of this work, is has not been possible to elaborate on this observation; however, deciphering the time point of interest to study will greatly facilitate the identification of candidates that regulate XCR. Future experiments can now be taken to mechanistically understand the function of key lncRNAs that purportedly govern X-chromosome activity in humans such as *XIST* and *XACT* (Sahakyan et al., 2017; Vallot et al., 2017), at a single-cell resolution throughout this period. More broadly, GRN analysis could be used to identify candidates that correlate with XCR; either in an unbiased or targeted manner by supplying

information of pre-existing regulators of X-chromosome activity. Identified candidates could subsequently formulate an enriched CRISPR library screen to be applied at the onset of reprogramming. Using a primed hPSC reporter line containing two different fluorescent proteins coupled to an X-linked gene on each X-chromosome would provide a readout of bi- or mono-allelic gene expression. The surface marker panel in combination with this type of reporter system would allow nascent naïve hPSCs with two active X-chromosomes, and populations that do not reprogram or do not display XaXa status, to be isolated using FACS. By comparing the different populations to see which CRISPR candidates are enriched or depleted, it would be possible to validate the predicted regulators of human X-chromosome activity.

To summarise, I have shown that the antibody panel provides an opportunity to study naïve hPSC formation at a time point far earlier than previously possible, and the panel can be used reliably across different reprogramming methods. By studying the transcriptional dynamics across reprogramming, I have identified clusters of co-expressed genes that will provide a framework for the future assessment of cell identity along the spectrum of primed to naïve states. Similarly, a large number of candidates have been identified that will be experimentally tested in the future to establish the mechanisms by which they promote or inhibit reprogramming to a naïve state. These include chemical compounds and growth factors for cell culture media supplementation, and candidates for transcription factor mediated reprogramming. Moreover, the cross-comparison of early reprogramming populations has revealed that there are likely divergent routes from primed to naïve pluripotency. Further exploration of additional reprogramming methods, and reciprocal experiments to rule out PSC line differences, will help to elucidate these divergent routes and their significance. Moreover, it will be interesting to assess the transcriptional similarity between the time points of reprogramming and the different embryonic lineages present within the human preimplantation blastocyst. This will help to provide clarity over their developmental staging and potential applications in stem cell biology.

6. Summary and Conclusions

The molecular and phenotypic studies of mouse, primate, and human embryos are providing clarity over the earliest stages of pluripotency, and are helping to define robust criteria by which to benchmark the different hPSC types against (Boroviak and Nichols, 2017; Huang et al., 2014). There are now greater than ten reports that described the methodologies and growth conditions to derive human pluripotent stem cells with characteristic akin to those of the ICM of preimplantation-stage embryos. These publications have generated a spectrum of different hPSC types that have collectively been termed 'naïve', even though some fulfil more of the hallmark characteristics of naïve pluripotency than others (Collier and Rugg-Gunn, 2018). Nonetheless, the transcriptional and epigenetic profiling of these cells has provided a foundation of knowledge to distinguish primed and naïve human pluripotent states. This knowledge is providing new insights into the regulation of transcription, epigenetic states, and signalling requirements, which we can relate back to the human embryo for experimental validation. However, this information is restricted to the established states with no previous reports assessing how primed cells enter the naïve state of pluripotency, until now. The foremost reason for this being the absence of an isolation strategy capable of unambiguously identifying nascent naïve hPSCs early during reprogramming.

The first outcome of my work addressed this problem through the comprehensive profiling of cell surface markers expressed by naïve and primed hPSCs. The results of the antibody library screen have identified eight naïve-specific surface markers, which are the first described candidates that can positively identify naïve from primed hPSCs; alongside a large number of primed-specific and shared markers. Collectively screening over 400 antibodies with diverse biological functions and inclusion of glycosylated epitopes was critical to identifying several naïve-specific markers, which would not have been identified by transcriptional analysis. Using the screen results, I subsequently validated a cohort of pluripotent state-specific markers across several different PSC lines and naïve culture conditions. Consistent state-specific results were obtained across PSC lines cultured under 5iLA and NK2 t2iL+PKCi conditions; however, I observed a mixed immunophenotype for RSeT conditions, which inferred RSeT cells fall somewhere in between primed and naïve states of pluripotency. Taken together, the screen results provide a large resource of markers present on the cell surface of primed and naive hPSCs, which should be valuable for future functional studies to interrogate the mechanisms that underlie the different states of human pluripotency. Nonetheless, there are bound to be many more state-specific surface markers that were not included across the two antibody library screens used during this work. This presents an opportunity in the future to examine more broadly the surface proteome and glycome of naïve and primed hPSCs to capture additional

markers, which in turn may provide new biological insights into additional signalling pathways regulating human pluripotent states.

To fully demonstrate the pluripotent state-specific nature of the candidate surface markers, preimplantation-stage human embryos were examined for their expression. Three of the naïve-specific markers, CD75, CD77, and CD130, were detected at the cell surface. CD130 showed an enrichment within the NANOG-positive cells of the ICM, whereas CD75 and CD77 were additionally expressed within the trophectoderm. This highlights that naïve hPSCs can be used to reveal new information relevant to human embryos. Moreover, this result indicates that CD75 and CD77 are not restricted to pluripotent epiblast cells at this stage of human embryonic development, and suggests naïve hPSCs may possess some features of extraembryonic cell types. By contrast, all of the primed-specific surface markers were not expressed at this stage. It would be interesting in the future to examine their developmental expression using human post-implantation *in-vitro* models (Deglincerti et al., 2016; Shahbazi et al., 2016). Taken together, these results confirm the PSC state-specific markers generally correspond to their respective developmental staging.

Whilst I have shown the state-specificity for a handful of identified markers, their biological roles in the context of pluripotency is unknown for many of these. Aforementioned, CD77 and CD75 are largely uncharacterised glycosylated epitopes. The enzymes purportedly responsible for the catalysis of CD75 and CD77, ST6GAL1 and A4GALT respectively, are transcriptionally upregulated in both 5iLA and NK2 t2iL+PKCi reprogramming cells and established naïve hPSCs. Additionally, both are highly expressed in human epiblast cells by scRNA-sequencing (Stirparo et al., 2017). Based upon the previous report of impaired hiPSC reprogramming and maintenance of primed pluripotency by *ST6GAL1* perturbation, appropriate sialyation is required for pluripotency more generally (Schwarzkopf et al., 2002; Wang et al., 2015). Similarly, reports show that numerous glycosyltransferase enzymes are required to maintain pluripotency and self-renewal of mESCs (comprehensively reviewed in: (Nishihara, 2018)). Glycans participate in a variety of important cellular processes, including cell adhesion and ligand-receptor binding that helps to elicit signalling cascades. Several glycans that are more well characterised include heparan sulfate and chondroitin sulfate, which are known to bind to Wnt, FGF, TGF-β, and BMP ligands in *Drosophila* and mice (reviewed in (Bernfield et al., 1999)). Considering their significance in cell fate regulation by modulating cell signalling, surprisingly little is known about the specific downstream pathways beyond many glycosylated surface receptors in humans (Pshezhetsky and Ashmarina, 2013; Schauer, 2009). As mentioned previously, a handful of glycosylated epitopes (SSEA3, SSEA4, TRA-1-60 and TRA-1-8) are

routinely used to identify primed hPSCs (Kannagi et al., 1983; Wright and Andrews, 2009). However, the biological role for these four glycosylated epitopes are also not well understood. Collectively, these results prompt further exploration of sialyltransferase and galactosyltransferase enzymes in the context of human pluripotency. Future experiments that exploit CRISPR-deletion/inhibition, or an auxin-based degron system for targeted degradation of candidate enzymes, will help to confirm their involvement in the biosynthesis of specific surface receptors and subsequent requirement for human pluripotency. Future combinatorial approaches to perform antibody-mediated immunoprecipitation coupled to mass spectrometry of specific receptors, may also help to identify interacting protein partners to elucidate the target signalling pathways associated with each receptor.

The second notable outcome from this work was the development of a multiplexed panel of state-specific antibodies capable of distinguishing naïve from primed cells by flow cytometry. I confirmed the sensitivity of the antibody panel by mixing together 10% naïve and 90% primed hPSCs. By gating for naïve surface marker expression and the absence of primed marker expression, all of the naïve hPSCs could be subsequently identified. This was the first critical step to ensuring the antibody panel could be used to isolate live naïve cells during reprogramming. The second critical step was to show that the antibody panel tracked the dynamics of heterogeneous pluripotent state conversions. The antibody panel collectively reflects the change from a primed to a naïve state, with the surface markers CD57, CD75, and CD130 being most reliable. By contrast, the predicted FGF target gene CD90 responded rapidly to the potent FGF/MEK/ERK inhibitor (PD0325901) added at the onset of reprogramming (Choi et al., 2001; Kaufman et al., 2001). Nonetheless, I have shown that the antibody panel is capable of robustly isolating nascent naïve hPSCs early during reprogramming for the very first time, alongside reprogramming intermediates that are refractory to reprogramming. Whilst there are a number of reported alternative isolation strategies (reviewed: (Collier and Rugg-Gunn, 2018; Trusler et al., 2018)), each one has been unable to demonstrate the ability to isolate early-naïve hPSCs. Consequently, our knowledge of the molecular events and their order during the reprogramming process has been limited. By using the antibody panel to isolate and characterise reprogramming populations and intermediate cell types, I have begun to map the molecular trajectory of cell fate changes during reprogramming. This has revealed new insights into the mechanisms which promote primed to naïve reprogramming, and highlighted possible methods to enhance this process.

Using the complete antibody panel comprising of four naïve-specific and two primed-specific markers, several populations were isolated by FACS after only ten days of reprogramming. The

most “naïve-like” population designated N4+, and two reprogramming intermediates N3+ and N4- were sorted for downstream analysis. With an outlook on future usability, I also developed a minimised antibody panel combining two naïve- and two primed-specific markers. This was used to isolated N2+ cells to resemble the N4+ population. A variety of molecular assays were performed to evaluate how each reprogramming population compared to established naïve hPSCs. Encouragingly, N4+ and N2+ populations were indistinguishable from established hPSCs in terms of colony formation efficiency, KLF17 expression, and initial transcriptional analysis by RT-qPCR. Moreover, both N2+ and N4+ populations formed homogeneous naïve PSC lines that could be maintained for >20 passages, emphasising that the antibody panel is able to identify and isolate nascent naïve hPSCs. This analysis also underscored the importance of CD75 expression, as N3+ cells that lack CD75 generated significantly fewer naïve colonies, and expressed *KLF17* to significantly lower levels.

Upon inspection of the global transcriptome by RNA-sequencing, it became apparent that day 10 nascent naïve hPSC populations (N2+ and N4+) had not yet acquired a mature naïve gene expression profile. However, by passage five and ten the N4+ population clustered with established naïve hPSCs, providing evidence that the nascent naïve population was isolated using the antibody panel. Consequently, assessment of the transcriptional profile of day 10 N4+ cells has provided the first insight into the order of gene expression changes that occur during reprogramming. Collectively, this revealed both early and late transcriptional responders under NK2 t2iL+PKCi reprogramming. This information will be useful to guide future experiments that will functionally examine the requirement of these factors to induce and sustain a naïve pluripotent state, and to understand their mode of action. Furthermore, I used the transcriptional data to examine the expression of transposable element classes, which Theunissen and colleagues propose as a means to distinguish naïve from primed hPSCs (Theunissen et al., 2016). In agreement with their findings, transposable element classes were able to recapitulate the spatial separation of samples by PCA. This finding prompts further investigation into the intertwined relationship of the pluripotent state-specific roles and regulation of transposable element expression.

Current reports that cross-compare established naïve and primed hPSCs generated using different reprogramming methods, have begun to highlight considerable variation in both the transcriptional and epigenetic profiles (Collier and Rugg-Gunn, 2018; Guo et al., 2017; Pastor et al., 2016; Takahashi et al., 2018). To examine the epigenetic profiles across the time-series, I firstly determined the X-chromosome status of each population. This revealed that X-chromosome reactivation occurred at a late time-point during reprogramming, somewhere

between day 10 and passage 5 (day 30), and is therefore a marker for mature naïve hPSCs. In collaboration with Ferdinand von Meyenn, assessment of the global CpG DNA methylation levels revealed a substantial decrease from 80% (NK2 primed) to 50% in day 10 N4+ cells. By passage 10, there was a further decrease to the equivalent level observed for established naïve hPSCs. This was accompanied by a widespread loss of imprint methylation, which was not the case at day 10.

One of the theoretical benefits to naïve hPSCs was the hope that akin to naïve mESCs cultured in 2i/LIF, the epigenetic memory and lineage bias of primed hPSCs would be erased upon reprogramming (Bock et al., 2011; Guo et al., 2016a; Weinberger et al., 2016). The epigenetic memory in established naïve hPSCs is erased to some extent, exemplified by global DNA demethylation and reactivation of the silent X-chromosome (Guo et al., 2017; Pastor et al., 2016; Sahakyan et al., 2017; Theunissen et al., 2016). However, there is a failure to faithfully reinstate these events after long-term culture in 5iLA(F) or t2iL+PKCi. Upon re-priming of naïve hPSCs, few imprinted loci regain DNA methylation, and X-chromosome silencing is biased towards the parental primed silenced X-chromosome (Pastor et al., 2016; Sahakyan et al., 2017; Theunissen et al., 2016). Moreover, the ability to erase any lineage bias present in the parental primed lines upon reprogramming and subsequent differentiation has yet to be examined. Nevertheless, the isolation of N4+ cells at early stages of reprogramming may harbour the desired characteristics to overcome lineage bias, whilst retaining genome stability. Future work will seek to address the differentiation potential of early naïve hPSCs.

One of the major caveats to current human naïve PSCs is the loss of genome stability, such as the erasure of imprint methylation and frequent karyotype abnormalities that arise with long-term culture. Efforts are required to solve these problems to fully harness the theoretical advantages that naïve hPSCs could offer. From the transcriptional comparison between day 10 N4+ cells and established naïve hPSCs, gene ontology analysis revealed that >100 zinc finger proteins had yet to be upregulated by day 10. The failure to upregulate this class of transcriptional repressors before the onset of global DNA hypomethylation could be one cause for the loss of imprint methylation that cannot be faithfully reinstated. Alternatively, prolonged exposure to PD0325901 (FGF/MEK/ERK inhibitor) is reportedly responsible for the genome instability that arises with mESCs cultured in 2i/LIF (Choi et al., 2017). The authors infer the same is true for naïve hPSCs. However, this may not be the case as PD0325901 is a common component to all human naïve culture conditions, even those that do not exhibit global DNA hypomethylation (Gafni et al., 2013; Pastor et al., 2016). It is likely that a combination of factors are responsible and future efforts are necessary to establish these and to determine a solution.

To obtain a broader understanding of the cell fate transition from primed to naïve pluripotency, I used the antibody panel to examine 5iLA reprogramming in addition to NK2 t2iL+PKCi. The "Reset" population that closely mimics the N4+ gating strategy isolated ≈10% of day 10 reprogramming cells under 5iLA culture, alongside a population that is refractory to reprogramming, "Not Reset" representing ≈55%. Assessing the colony formation efficiency revealed the day 10 Reset population captured virtually all reprogramming cells with the ability to generate naïve colonies, contrary to the Not Reset population that struggled to form colonies of any type. The isolation of nascent naïve hPSCs using different reprogramming methods highlights that the multiplexed antibody panel is a versatile and robust tool for the isolation of naïve hPSCs. Moreover, a cross-comparison of antibody panel populations and OCT4-ΔPE-GFP reporter expression inadvertently revealed that this reporter is not specific to naïve hPSCs.

Using the antibody panel to isolate nascent naïve cells as they emerge and progress throughout NK2 t2iL+PKCi reprogramming provided the first detailed insight into the transcriptional dynamics across the transitional period. From this it was apparent that substantial molecular changes occurred between day 10 and passage five. To address this, I used the antibody panel to track and subsequently isolate Reset cells at frequent intervals throughout 5iLA reprogramming. This provided an enhanced resolution to decipher the temporal order of molecular changes. Notably, X-chromosome reactivation occurred between P1 (day 14) and P2 (day 18). This helps to narrow down the window to determine how this process is mechanistically regulated. Furthermore, frequent transcriptional profiling revealed extensive changes occur between primed hPSCs and day 8 and day 10 Reset cells; conversely, by P2 (day 18) a naïve transcriptional profile had been established. Gene expression clustering analysis of the 5iLA reprogramming transition has revealed a large number of candidates that may be required at different stages of primed to naïve reprogramming. Future work will seek to experimentally establish the mechanisms by which they promote or inhibit reprogramming to a naïve state. As discussed previously, these include chemical compounds and growth factors for cell culture media supplementation, and transcriptional regulators that were common to both 5iLA and NK2 t2iL+PKCi reprogramming.

The similarity in gating strategies between the antibody panels used, and the matched time-points examined, permitted a direct comparison between the two reprogramming methods. The global transcriptional overview by PCA, in combination with gene clustering analysis, revealed that there are likely divergent routes from primed to naïve pluripotency. Whilst some transcriptional changes can be attributed to inherited PSC line-specific differences, there was

still considerable variation in the utilisation of different transcription factors during the early stages of reprogramming; most notably the KLF and GATA family of transcription factors. To robustly exclude PSC line specific differences from those attributable to divergent pathways during reprogramming, the reciprocal experiments will be required. Moreover, it is important to consider the limitations to the bulk RNA-sequencing approach taken, even though an explicit and seemingly homogeneous population of cells was isolated. To reliably conclude that there are multiple routes from primed to naïve pluripotency, future experiments that utilise lineage tracing technologies coupled with single-cell RNA sequencing would be desirable. Numerous CRISPR "scarring" technologies such as LINNAEUS (Spanjaard et al., 2018), GESTALT (McKenna et al., 2016), or MEMOIR coupled with seqFISH (Frieda et al., 2017) could be employed to address this question moving forward. Similarly, this would provide an insight into the alternative cell types generated during the initial phases of reprogramming that might be masked by bulk sequencing. The expression of many trophoblast stem cell related transcription factors, hormones and growth factors during primed to naïve reprogramming led me to speculate that it may be possible to derive human TS-like cells during this process. Single cell analysis would reveal whether this is either driven by a subpopulation of TS-like cells, or if these genes are more generally involved in primed to naïve reprogramming, or whether cells with both an embryonic and extraembryonic transcriptional profile are generated.

The idea of deriving a cell type with embryonic and extraembryonic capacity has been recently reported (Yang et al., 2017b). The human cells, termed extended pluripotent stem (EPS) cells, were shown to contribute to both the embryo and the placenta by performing an interspecies chimeric assay. However, the functionality and viability of the human cells was not assessed due to the limited integration. Moreover, the authors did not present data to suggest human EPS cells could differentiate into trophoblast cell types. Nonetheless, the ability to derive isogenic stem cells with both embryonic and extraembryonic capabilities would be highly desirable for the study of early human development, and pathogenesis of developmental disorders attributed to trophoblast defects. The recently reported culture conditions for the maintenance of human TS-cells will provide a basis to amend current naïve hPSC culture conditions in the future (Okae et al., 2018), which will allow us to explore the possibility of deriving TS-like cells during primed to naïve reprogramming.

Along a similar avenue, new and exciting *in-vitro* models have been developed that mimic *in-vivo* self-organisation events that occur during mouse embryogenesis. By combining mouse ESC and TS cells, these reports have derived blastocyst like structures (termed blastoids) that resemble preimplantation (Rivron et al., 2018), and peri- to post- implantation-stage mouse

embryo structures (Harrison et al., 2017). In contrast to blastocysts, blastoids can be generated in large quantities that allows sufficient material to be collected for downstream assays that would be restricted using blastocysts, such as high-throughput biochemical and genetic engineering screens. Consequently, blastoids offer a scalable model to study the principals underlying mammalian embryogenesis. It will be exciting to see in the future whether human synthetic blastoids can be produced to mimic human embryogenesis, and what pluripotent cell types will be required. On a related note, human embryos have been cultured *in-vitro* for up to 14 days to mimic elements of postimplantation development, a time point that is otherwise inaccessible (Deglincerti et al., 2016; Shahbazi et al., 2016). In the future, a combination of these models will help to validate transcriptionally where the different primed, naïve, and reprogramming intermediate cell types fall along the human embryonic developmental trajectory.

7. Bibliography

Aasen, T., Raya, A., Barrero, M.J., Garreta, E., Consiglio, A., Gonzalez, F., Vassena, R., Bilic, J., Pekarik, V., Tiscornia, G., *et al.* (2008). Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat Biotechnol *26*, 1276-1284.

Abujarour, R., Valamehr, B., Robinson, M., Rezner, B., Vranceanu, F., and Flynn, P. (2013). Optimized Surface Markers for the Prospective Isolation of High-Quality hiPSCs using Flow Cytometry Selection. Scientific Reports *3*, 1179.

Aken, B.L., Achuthan, P., Akanni, W., Amode, M.R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., *et al.* (2017). Ensembl 2017. Nucleic acids research *45*, D635-d642.

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

Andrews, S. (2014). FastQC A Quality Control tool for High Throughput Sequence Data. http://wwwbioinformaticsbabrahamacuk/projects/fastqc/.

Aravin, A.A., van der Heijden, G.W., Castañeda, J., Vagin, V.V., Hannon, G.J., and Bortvin, A. (2009). Cytoplasmic Compartmentalization of the Fetal piRNA Pathway in Mice. PLOS Genetics *5*, e1000764.

Arends, J.E., Bot, F.J., Gisbertz, I.A., and Schouten, H.C. (1999). Expression of CD10, CD75 and CD43 in MALT lymphoma and their usefulness in discriminating MALT lymphoma from follicular lymphoma and chronic gastritis. Histopathology *35*, 209-215.

Assou, S., Boumela, I., Haouzi, D., Monzo, C., Dechaud, H., Kadoch, I.-J., and Hamamah, S. (2012). Transcriptome Analysis during Human Trophectoderm Specification Suggests New Roles of Metabolic and Epigenetic Genes. PLoS ONE *7*, e39306.

Athanassiades, A., and Lala, P.K. (1998). Role of placenta growth factor (PIGF) in human extravillous trophoblast proliferation, migration and invasiveness. Placenta *19*, 465-473.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., *et al.* (2006). Chromatin signatures of pluripotent cell lines. Nature cell biology *8*, 532.

Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., and Surani, M.A. (2009a). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature *461*, 1292.

Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., and Surani, M.A. (2009b). Epigenetic reversion of postimplantation epiblast cells to pluripotent embryonic stem cells. Nature *461*, 10.1038/nature08534.

Beattie, G.M., Lopez, A.D., Bucay, N., Hinton, A., Firpo, M.T., King, C.C., and Hayek, A. (2005). Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. Stem Cells *23*, 489-495.

Beddington, R.S., and Robertson, E.J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. Development (Cambridge, England) *105*, 733-737.

Bendall, S.C., Nolan, G.P., Roederer, M., and Chattopadhyay, P.K. (2012). A deep profiler's guide to cytometry. Trends in immunology *33*, 323-332.

Berger, R.P., Dookwah, M., Steet, R., and Dalton, S. (2016). Glycosylation and stem cells: Regulatory roles and application of iPSCs in the study of glycosylation-related disorders. BioEssays : news and reviews in molecular, cellular and developmental biology *38*, 1255-1265.

Bernfield, M., Götte, M., Park, P.W., Reizes, O., Fitzgerald, M.L., Lincecum, J., and Zako, M. (1999). Functions of Cell Surface Heparan Sulfate Proteoglycans. Annual review of biochemistry *68*, 729-777.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315-326.

Bialkowska, A.B., Yang, V.W., and Mallipattu, S.K. (2017). Krüppel-like factors in mammalian stem cells and development. Development (Cambridge, England) *144*, 737-754.

Billian, G., Bella, C., Mondiere, P., and Defrance, T. (1996). Identification of a tonsil IgD+ B cell subset with phenotypical and functional characteristics of germinal center B cells. European journal of immunology *26*, 1712-1719.

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 (Cambridge, England) *142*, 3151-3165.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., *et al.* (2011). Reference Maps of Human ES and iPS Cell Variation Enable High-Throughput Characterization of Pluripotent Cell Lines. Cell *144*, 439-452.

Boeuf, H., Hauss, C., Graeve, F.D., Baran, N., and Kedinger, C. (1997). Leukemia Inhibitory Factor–dependent Transcriptional Activation in Embryonic Stem Cells. The Journal of Cell Biology *138*, 1207-1217.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nature genetics *42*, 1093.

Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. Developmental Cell *35*, 366-382.

Boroviak, T., and Nichols, J. (2017). Primate embryogenesis predicts the hallmarks of human naive pluripotency. Development (Cambridge, England) *144*, 175-186.

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.

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K., *et al.* (2004). Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. Nature Biotechnology *22*, 707.

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.

Buecker, C., Chen, H.H., Polo, J.M., Daheron, L., Bu, L., Barakat, T.S., Okwieka, P., Porter, A., Gribnau, J., Hochedlinger, K., *et al.* (2010). A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. Cell stem cell *6*, 535-546.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency. Cell stem cell *14*, 838-853.

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.

Buehr, M., and Smith, A. (2003). Genesis of embryonic stem cells. Philosophical transactions of the Royal Society of London Series B, Biological sciences *358*, 1397-1402; discussion 1402.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell gene expression analyses of cellular reprogramming reveal a stochastic early and hierarchic late phase. Cell *150*, 1209-1222.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Developmental biology *210*, 30-43.

Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X., *et al.* (2015). Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. Cell *162*, 412-424.

Canham, M.A., Sharov, A.A., Ko, M.S.H., and Brickman, J.M. (2010). Functional Heterogeneity of Embryonic Stem Cells Revealed through Translational Amplification of an Early Endodermal Transcript. PLOS Biology *8*, e1000379.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288-289.

Carpenter, M.K., Rosler Elen, S., Fisk Gregory, J., Brandenberger, R., Ares, X., Miura, T., Lucero, M., and Rao Mahendra, S. (2003). Properties of four human embryonic stem cell lines maintained in a feeder-free culture system. Developmental Dynamics *229*, 243-258.

Chan, Y.-S., Göke, J., Ng, J.-H., Lu, X., Gonzales, Kevin Andrew U., Tan, C.-P., Tng, W.-Q., Hong, Z.-Z., Lim, Y.-S., and Ng, H.-H. (2013). Induction of a Human Pluripotent State with Distinct Regulatory Circuitry that Resembles Preimplantation Epiblast. Cell stem cell *13*, 663-675.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC bioinformatics *14*, 128.

Chen, H., Aksoy, I., Gonnot, F., Osteil, P., Aubry, M., Hamela, C., Rognard, C., Hochard, A., Voisin, S., Fontaine, E., *et al.* (2015). Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency. *6*, 7095.

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.

Cheong, M.-L., Wang, L.-J., Chuang, P.-Y., Chang, C.-W., Lee, Y.-S., Lo, H.-F., Tsai, M.-S., and Chen, H. (2016). A Positive Feedback Loop between Glial Cells Missing 1 and Human Chorionic Gonadotropin (hCG) Regulates Placental hCGβ Expression and Cell Differentiation. Molecular and Cellular Biology *36*, 197-209.

Choi, D.Y., Toledo-Aral, J.J., Lin, H.Y., Ischenko, I., Medina, L., Safo, P., Mandel, G., Levinson, S.R., Halegoua, S., and Hayman, M.J. (2001). Fibroblast growth factor receptor 3 induces gene expression primarily through Ras-independent signal transduction pathways. The Journal of biological chemistry *276*, 5116-5122.

Choi, H.W., Joo, J.Y., Hong, Y.J., Kim, J.S., Song, H., Lee, J.W., Wu, G., Scholer, H.R., and Do, J.T. (2016). Distinct Enhancer Activity of Oct4 in Naive and Primed Mouse Pluripotency. Stem Cell Reports *7*, 911-926.

Choi, J., Huebner, A.J., Clement, K., Walsh, R.M., Savol, A., Lin, K., Gu, H., Di Stefano, B., Brumbaugh, J., Kim, S.-Y., *et al.* (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. Nature *548*, 219.

Choi, J., Lee, S., Mallard, W., Clement, K., Tagliazucchi, G.M., Lim, H., Choi, I.Y., Ferrari, F., Tsankov, A.M., Pop, R., *et al.* (2015). A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. Nature Biotechnology *33*, 1173.

Cole, L.A. (2009). New discoveries on the biology and detection of human chorionic gonadotropin. Reproductive Biology and Endocrinology : RB&E *7*, 8-8.

Collier, A.J., Panula, S.P., Schell, J.P., Chovanec, P., Plaza Reyes, A., Petropoulos, S., Corcoran, A.E., Walker, R., Douagi, I., Lanner, F., *et al.* (2017). Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. Cell stem cell *20*, 874-310065408.

Collier, A.J., and Rugg-Gunn, P.J. (2018). Identifying Human Naïve Pluripotent Stem Cells − Evaluating State-Specific Reporter Lines and Cell-Surface Markers. BioEssays *40*, 1700239.

Czechanski, A., Byers, C., Greenstein, I., Schrode, N., Donahue, L.R., Hadjantonakis, A.-K., and Reinholdt, L. (2014). Derivation and characterization of mouse embryonic stem cells (mESCs) from permissive and non-permissive strains. Nature protocols *9*, 559-574.

Dahéron, L.O., Sarah L., Zaehres, H., Lensch, W.M., Andrews, P.W., Itskovitz-Eldor, J., and Daley, G.Q. (2004). LIF/STAT3 Signaling Fails to Maintain Self-Renewal of Human Embryonic Stem Cells. STEM CELLS *22*, 770-778.

David, L., and Polo, J.M. (2014). Phases of reprogramming. Stem cell research *12*, 754-761.

De Paepe, C., Cauffman, G., Verloes, A., Sterckx, J., Devroey, P., Tournaye, H., Liebaers, I., and Van de Velde, H. (2013). Human trophectoderm cells are not yet committed. Human reproduction (Oxford, England) *28*, 740-749.

Deglincerti, A., Croft, G.F., Pietila, L.N., Zernicka-Goetz, M., Siggia, E.D., and Brivanlou, A.H. (2016). Self-organization of the in vitro attached human embryo. Nature *533*, 251.

Dhaliwal, N.K., Miri, K., Davidson, S., Tamim El Jarkass, H., and Mitchell, J.A. (2018). KLF4 Nuclear Export Requires ERK Activation and Initiates Exit from Naive Pluripotency. Stem Cell Reports *10*, 1308-1323.

Duggal, G., Warrier, S., Ghimire, S., Broekaert, D., Van der Jeught, M., Lierman, S., Deroo, T., Peelman, L., Van Soom, A., Cornelissen, R., *et al.* (2015). Alternative Routes to Induce Naive Pluripotency in Human Embryonic Stem Cells. Stem Cells *33*, 2686-2698.

Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science *344*, 1156-1160.

Ebrahimi, B. (2015). Reprogramming barriers and enhancers: strategies to enhance the efficiency and kinetics of induced pluripotency. Cell Regeneration *4*, 10.

Evans, M.J. (1972). The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. Journal of embryology and experimental morphology *28*, 163-176.

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

Factor, Daniel C., Corradin, O., Zentner, Gabriel E., Saiakhova, A., Song, L., Chenoweth, Josh G., McKay, Ronald D., Crawford, Gregory E., Scacheri, Peter C., and Tesar, Paul J. (2014). Epigenomic Comparison Reveals Activation of "Seed" Enhancers during Transition from Naive to Primed Pluripotency. Cell stem cell *14*, 854-863.

Fenderson, B.A., Radin, N., and Andrews, P.W. (1993). Differentiation Antigens of Human Germ Cell Tumours: Distribution of Carbohydrate Epitopes on Glycolipids and Glycoproteins Analyzed Using PDMP, an Inhibitor of Glycolipid Synthesis. European Urology *23*, 30-37.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, Simon 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.

Frieda, K.L., Linton, J.M., Hormoz, S., Choi, J., Chow, K.K., Singer, Z.S., Budde, M.W., Elowitz, M.B., and Cai, L. (2017). Synthetic recording and in situ readout of lineage information in single cells. Nature *541*, 107-111.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proceedings of the Japan Academy Series B, Physical and biological sciences *85*, 348-362.

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., *et al.* (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282-286.

Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., Yi, L., Yu, G., Hou, Z., Fan, D., *et al.* (2018). Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. Cell *173*, 248-259.e215.

Gassen, S.V., Callebaut, B., Helden, M.J.V., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytometry Part A *87*, 636-645.

Ge, S.X. (2017). iDEP: An integrated web application for differential expression and pathway analysis. bioRxiv.

Gearing, D.P., Comeau, M.R., Friend, D.J., Gimpel, S.D., Thut, C.J., McGourty, J., Brasher, K.K., King, J.A., Gillis, S., Mosley, B., *et al.* (1992). The IL-6 signal transducer, gp130: an oncostatin M receptor and affinity converter for the LIF receptor. Science *255*, 1434-1437.

Gerami-Naini, B., Dovzhenko, O.V., Durning, M., Wegner, F.H., Thomson, J.A., and Golos, T.G. (2004). Trophoblast differentiation in embryoid bodies derived from human embryonic stem cells. Endocrinology *145*, 1517-1524.

Ghimire, S., Van der Jeught, M., Neupane, J., Roost, M.S., Anckaert, J., Popovic, M., Van Nieuwerburgh, F., Mestdagh, P., Vandesompele, J., Deforce, D., *et al.* (2018). Comparative analysis of naive, primed and ground state pluripotency in mouse embryonic stem cells originating from the same genetic background. Scientific Reports *8*, 5884.

Giorgetti, A., Montserrat, N., Aasen, T., Gonzalez, F., Rodriguez-Piza, I., Vassena, R., Raya, A., Boue, S., Barrero, M.J., Corbella, B.A., *et al.* (2009). Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2. Cell stem cell *5*, 353-357.

Greber, B., Coulon, P., Zhang, M., Moritz, S., Frank, S., Müller-Molina, A.J., Araúzo-Bravo, M.J., Han, D.W., Pape, H.-C., and Schöler, H.R. (2011). FGF signalling inhibits neural induction in human embryonic stem cells. The EMBO journal *30*, 4874-4884.

Greber, B., Wu, G., Bernemann, C., Joo, J.Y., Han, D.W., Ko, K., Tapia, N., Sabour, D., Sterneckert, J., Tesar, P., *et al.* (2010). Conserved and Divergent Roles of FGF Signaling in Mouse Epiblast Stem Cells and Human Embryonic Stem Cells. Cell stem cell *6*, 215-226.

Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D., Martin, L., Ware, C., Blish, C.A., Chang, H.Y., *et al.* (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature *522*, 221-225.

Gumireddy, K., Li, A., Gimotty, P.A., Klein-Szanto, A.J., Showe, L.C., Katsaros, D., Coukos, G., Zhang, L., and Huang, Q. (2009). KLF17 is a negative regulator of epithelial–mesenchymal transition and metastasis in breast cancer. Nature cell biology *11*, 1297.

Guo, G., Pinello, L., Han, X., Lai, S., Shen, L., Lin, T.-W., Zou, K., Yuan, G.-C., and Orkin, S.H. (2016a). Serum-based culture conditions provoke gene expression variability in mouse embryonic stem cells as revealed by single cell analysis. Cell reports *14*, 956-965.

Guo, G., and Smith, A. (2010). A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. Development (Cambridge, England) *137*, 3185-3192.

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 (Cambridge, England) *144*, 2748.

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., and Nichols, J. (2016b). Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. Stem Cell Reports *6*, 437-446.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development (Cambridge, England) *136*, 1063-1069.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., *et al.* (2014). The DNA methylation landscape of human early embryos. Nature *511*, 606.

Hackett, Jamie A., Dietmann, S., Murakami, K., Down, Thomas A., Leitch, Harry G., and Surani, M.A. (2013). Synergistic Mechanisms of DNA Demethylation during Transition to Ground-State Pluripotency. Stem Cell Reports *1*, 518-531.

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 Krüppel Factors to Sustain Embryonic Stem Cell Self-Renewal. Cell stem cell *5*, 597-609.

Han, D.W., Tapia, N., Joo, J.Y., Greber, B., Araúzo-Bravo, M.J., Bernemann, C., Ko, K., Wu, G., Stehling, M., Do, J.T., *et al.* (2010). Epiblast Stem Cell Subpopulations Represent Mouse Embryos of Distinct Pregastrulation Stages. Cell *143*, 617-627.

Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proceedings of the National Academy of Sciences *107*, 9222-9227.

Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A.W., Cassady, J.P., Staerk, J., Carey, B.W., Lengner, C.J., Foreman, R., Love, J., *et al.* (2009). Metastable Pluripotent States in NOD-Mouse-Derived ESCs. Cell stem cell *4*, 513-524.

Hansson, J., Rafiee, Mahmoud R., Reiland, S., Polo, Jose M., Gehring, J., Okawa, S., Huber, W., Hochedlinger, K., and Krijgsveld, J. (2012). Highly Coordinated Proteome Dynamics during Reprogramming of Somatic Cells to Pluripotency. Cell Reports *2*, 1579-1592.

Harrison, S.E., Sozen, B., Christodoulou, N., Kyprianou, C., and Zernicka-Goetz, M. (2017). Assembly of embryonic and extra-embryonic stem cells to mimic embryogenesis in vitro. Science.

Hayashi, K., Lopes, S.M.C.d.S., Tang, F., and Surani, M.A. (2008). Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. Cell stem cell *3*, 391-401.

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.

Heinrich, P.C., Behrmann, I., Müller-Newen, G., Schaper, F., and Graeve, L. (1998). Interleukin-6-type cytokine signalling through the gp130/Jak/STAT pathway. Biochemical Journal *334*, 297-314.

Henderson, J.K., S., D.J., S., B.H., S., F., A., T.J., H., M., and W., A.P. (2002). Preimplantation Human Embryos and Embryonic Stem Cells Show Comparable Expression of Stage-Specific Embryonic Antigens. STEM CELLS *20*, 329-337.

Hoch, R.V., Lindtner, S., Price, J.D., and Rubenstein, J.L.R. (2015). OTX2 Transcription Factor Controls Regional Patterning Within The Medial Ganglionic Eminence (MGE) And Regional Identity Of The Septum. Cell reports *12*, 482-494.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., *et al.* (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol *29*, 731-734.

Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226-232.

Home, P., Ray, S., Dutta, D., Bronshteyn, I., Larson, M., and Paul, S. (2009). GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. The Journal of biological chemistry *284*, 28729-28737.

Hotta, A., Cheung, A.Y.L., Farra, N., Vijayaragavan, K., Seguin, C.A., Draper, J.S., Pasceri, P., Maksakova, I.A., Mager, D.L., Rossant, J., *et al.* (2009). Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. Nat Meth *6*, 370-376.

Hough, Shelley R., Thornton, M., Mason, E., Mar, Jessica C., Wells, Christine A., and Pera, Martin F. (2014). Single-Cell Gene Expression Profiles Define Self-Renewing, Pluripotent, and Lineage Primed States of Human Pluripotent Stem Cells. Stem Cell Reports *2*, 881-895.

Huang, K., Maruyama, T., and Fan, G. (2014). The Naive State of Human Pluripotent Stem Cells: A Synthesis of Stem Cell and Preimplantation Embryo Transcriptome Analyses. Cell stem cell *15*, 410-415.

Huang, Y., Osorno, R., Tsakiridis, A., and Wilson, V. (2012). In Vivo Differentiation Potential of Epiblast Stem Cells Revealed by Chimeric Embryo Formation. Cell Reports *2*, 1571-1578.

Humphrey, R.K., Beattie, G.M., Lopez, A.D., Bucay, N., King, C.C., Firpo, M.T., Rose-John, S., and Hayek, A. (2004). Maintenance of Pluripotency in Human Embryonic Stem Cells Is STAT3 Independent. STEM CELLS *22*, 522-530.

Huynh-Thu, V.A., and Geurts, P. (2018). dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. Scientific Reports *8*, 3384.

Huynh, K.D., and Lee, J.T. (2003). Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. Nature *426*, 857.

James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005a). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development (Cambridge, England) *132*, 1273-1282.

James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005b). TGFβ/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development (Cambridge, England) *132*, 1273-1282.

James, D., Noggle, S.A., Swigut, T., and Brivanlou, A.H. (2006). Contribution of human embryonic stem cells to mouse blastocysts. Developmental biology *295*, 90-102.

Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I., *et al.* (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell stem cell *18*, 262-275.

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. Nature cell biology *10*, 353.

Kahan, B.W., and Ephrussi, B. (1970). Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. Journal of the National Cancer Institute *44*, 1015-1036.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature *458*, 771-775.

Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Marks, H., Walker, R., Stunnenberg, H.G., Lilley, K.S., *et al.* (2017). Tracking the embryonic stem cell transition from ground state pluripotency. Development (Cambridge, England).

Kannagi, R., Cochran, N.A., Ishigami, F., Hakomori, S., Andrews, P.W., Knowles, B.B., and Solter, D. (1983). Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells. The EMBO journal *2*, 2355-2361.

Kaufman, D.S., Hanson, E.T., Lewis, R.L., Auerbach, R., and Thomson, J.A. (2001). Hematopoietic colony-forming cells derived from human embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *98*, 10716-10721.

Kilens, S., Meistermann, D., Moreno, D., Chariau, C., Gaignerie, A., Reignier, A., Lelievre, Y., Casanova, M., Vallot, C., Nedellec, S., *et al.* (2018). Parallel derivation of isogenic human primed and naive induced pluripotent stem cells. Nat Commun *9*, 360.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. Cell *132*, 1049-1061.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., *et al.* (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285.

Kleinsmith, L.J., and Pierce, G.B. (1964). Multipotentiality of Single Embryonal Carcinoma Cells. Cancer Research *24*, 1544-1551.

Knaupp, A.S., Buckberry, S., Pflueger, J., Lim, S.M., Ford, E., Larcombe, M.R., Rossello, F.J., de Mendoza, A., Alaei, S., Firas, J., *et al.* (2017). Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. Cell stem cell *21*, 834-845.e836.

Kojima, Y., Kaufman-Francis, K., Studdert, Joshua B., Steiner, Kirsten A., Power, Melinda D., Loebel, David A.F., Jones, V., Hor, A., de Alencastro, G., Logan, Grant J., *et al.* (2014). The Transcriptional and Functional Properties of Mouse Epiblast Stem Cells Resemble the Anterior Primitive Streak. Cell stem cell *14*, 107-120.

Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R., and Grosveld, F. (1999). The transcription factor GATA6 is essential for early extraembryonic development. Development (Cambridge, England) *126*, 723-732.

Krueger, F., and Andrews, S. (2016). SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes [version 1; referees: 3 approved], Vol 5.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., *et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research *44*, W90-97.

Kunath, T. (2011). Primed for Pluripotency. Cell stem cell *8*, 241-242.

Kurosawa, H. (2007). Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells. Journal of Bioscience and Bioengineering *103*, 389-398.

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.* (2013). Naive pluripotency is associated with global DNA hypomethylation. Nature Structural &Amp; Molecular Biology *20*, 311.

Lensch, M.W., Schlaeger, T.M., Zon, L.I., and Daley, G.Q. (2007). Teratoma Formation Assays with Human Embryonic Stem Cells: A Rationale for One Type of Human-Animal Chimera. Cell stem cell *1*, 253-258.

Li, C., Scott, D.A., Hatch, E., Tian, X., and Mansour, S.L. (2007). Dusp6 (Mkp3) is a negative feedback regulator of FGF-stimulated ERK signaling during mouse development. Development (Cambridge, England) *134*, 167-176.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009a). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

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.

Li, W., Wei, W., Zhu, S., Zhu, J., Shi, Y., Lin, T., Hao, E., Hayek, A., Deng, H., and Ding, S. (2009b). Generation of Rat and Human Induced Pluripotent Stem Cells by Combining Genetic Reprogramming and Chemical Inhibitors. Cell stem cell *4*, 16-19.

Li, X., Liu, D., Ma, Y., Du, X., Jing, J., Wang, L., Xie, B., Sun, D., Sun, S., Jin, X., *et al.* (2017). Direct Reprogramming of Fibroblasts via a Chemically Induced XEN-like State. Cell stem cell *21*, 264-273.e267.

Liu, H., Ye, Z., Kim, Y., Sharkis, S., and Jang, Y.Y. (2010). Generation of endoderm-derived human induced pluripotent stem cells from primary hepatocytes. Hepatology (Baltimore, Md) *51*, 1810-1819.

Liu, X., Nefzger, C.M., Rossello, F.J., Chen, J., Knaupp, A.S., Firas, J., Ford, E., Pflueger, J., Paynter, J.M., Chy, H.S., *et al.* (2017). Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. Nat Methods *14*, 1055-1062.

Liu, Y., Beyer, A., and Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell *165*, 535-550.

Loh, Y.H., Agarwal, S., Park, I.H., Urbach, A., Huo, H., Heffner, G.C., Kim, K., Miller, J.D., Ng, K., and Daley, G.Q. (2009). Generation of induced pluripotent stem cells from human blood. Blood *113*, 5476-5479.

Löser, P., Schirm, J., Guhr, A., Wobus Anna, M., and Kurtz, A. (2010). Human Embryonic Stem Cell Lines and Their Use in International Research. STEM CELLS *28*, 240-246.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology *15*, 550.

Lujan, E., Zunder, E.R., Ng, Y.H., Goronzy, I.N., Nolan, G.P., and Wernig, M. (2015). Early reprogramming regulators identified by prospective isolation and mass cytometry. Nature *521*, 352.

Luo, Z., Gao, X., Lin, C., Smith, E.R., Marshall, S.A., Swanson, S.K., Florens, L., Washburn, M.P., and Shilatifard, A. (2015). Zic2 is an enhancer-binding factor required for embryonic stem cell specification. Molecular cell *57*, 685-694.

Ma, G.T., Roth, M.E., Groskopf, J.C., Tsai, F.Y., Orkin, S.H., Grosveld, F., Engel, J.D., and Linzer, D.I. (1997). GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo. Development (Cambridge, England) *124*, 907-914.

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., and Hochedlinger, K. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell stem cell *3*, 340-345.

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., and Brockdorff, N. (2004). Reactivation of the Paternal X Chromosome in Early Mouse Embryos. Science *303*, 666-669.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A., *et al.* (2012). The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. Cell *149*, 590-604.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. The EMBO journal *32*, 2561-2574.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Gottgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell stem cell *11*, 491-504.

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

Masaki, H., Kato-Itoh, M., Umino, A., Sato, H., Hamanaka, S., Kobayashi, T., Yamaguchi, T., Nishimura, K., Ohtaka, M., Nakanishi, M., *et al.* (2015). Interspecific in

vitro assay for the chimera-forming ability of human pluripotent stem cells. Development (Cambridge, England) *142*, 3222.

Mascetti, Victoria L., and Pedersen, Roger A. (2016a). Contributions of Mammalian Chimeras to Pluripotent Stem Cell Research. Cell stem cell *19*, 163-175.

Mascetti, Victoria L., and Pedersen, Roger A. (2016b). Human-Mouse Chimerism Validates Human Stem Cell Pluripotency. Cell stem cell *18*, 67-72.

Massague, J. (1998). TGF-beta signal transduction. Annual review of biochemistry *67*, 753-791.

McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science *353*.

Mekhoubad, S., Bock, C., de Boer, A.S., Kiskinis, E., Meissner, A., and Eggan, K. (2012). Erosion of dosage compensation impacts human iPSC disease modeling. Cell stem cell *10*, 595-609.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49.

Mintz, B., and Illmensee, K. (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. Proceedings of the National Academy of Sciences of the United States of America *72*, 3585-3589.

Monaco, G., Chen, H., Poidinger, M., Chen, J., de Magalhães, J.P., and Larbi, A. (2016). flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. Bioinformatics *32*, 2473-2480.

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 Developmental Biology *17*, 7.

Morilla, R., Morilla, A.M., and Nadal-Melsió, E. (2017). 16 - Immunophenotyping by Flow Cytometry. In Dacie and Lewis Practical Haematology (Twelfth Edition), B.J. Bain, I. Bates, and M.A. Laffan, eds. (Elsevier), pp. 330-349.

Munro, S., Bast, B.J.E.G., Colley, K.J., and Tedder, T.F. (1992). The B lymphocyte surface antigen CD75 is not an α-2,6-sialyltransferase but is a carbohydrate antigen, the production of which requires the enzyme. Cell *68*, 1003.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. Proceedings of the National Academy of Sciences *90*, 8424-8428.

Najm, F.J., Chenoweth, J.G., Anderson, P.D., Nadeau, J.H., Redline, R.W., McKay, R.D., and Tesar, P.J. (2011). Isolation of epiblast stem cells from preimplantation mouse embryos. Cell stem cell *8*, 318-325.

Nakamura, T., Okamoto, I., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Yamamoto, T., and Saitou, M. (2016). A developmental coordinate of pluripotency among mice, monkeys and humans. Nature *537*, 57.

Nefzger, C.M., Alaei, S., Knaupp, A.S., Holmes, M.L., and Polo, J.M. (2014). Cell Surface Marker Mediated Purification of iPS Cell Intermediates from a Reprogrammable Mouse Model. JoVE, e51728.

Nefzger, C.M., Rossello, F.J., Chen, J., Liu, X., Knaupp, A.S., Firas, J., Paynter, J.M., Pflueger, J., Buckberry, S., Lim, S.M., *et al.* (2017). Cell Type of Origin Dictates the Route to Pluripotency. Cell Reports *21*, 2649-2660.

Niakan, K.K., Han, J., Pedersen, R.A., Simon, C., and Pera, R.A.R. (2012). Human pre-implantation embryo development. Development (Cambridge, England) *139*, 829-841.

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. Nature Medicine *15*, 814.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell stem cell *4*, 487-492.

Nishihara, S. (2018). Glycans in stem cell regulation: from Drosophila tissue stem cells to mammalian pluripotent stem cells. FEBS letters.

Niswander, L., and Martin, G.R. (1992). Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. Development (Cambridge, England) *114*, 755-768.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes & Development *12*, 2048-2060.

Novo, C.L., Javierre, B.-M., Cairns, J., Segonds-Pichon, A., Wingett, S.W., Freire-Pritchett, P., Furlan-Magaril, M., Schoenfelder, S., Fraser, P., and Rugg-Gunn, P.J. (2018). Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. Cell Reports *22*, 2615-2627.

O'Brien, C.M., Chy, H.S., Zhou, Q., Blumenfeld, S., Lambshead, J.W., Liu, X., Kie, J., Capaldo, B.D., Chung, T.L., Adams, T.E., *et al.* (2017). New Monoclonal Antibodies to Defined Cell Surface Proteins on Human Pluripotent Stem Cells. Stem Cells *35*, 626-640.

O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., and Kaji, K. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature *499*, 88.

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J.M., Guo, T., Qi, Z., Downey, S.L., Manos, P.D., Rossi, D.J., *et al.* (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nature cell biology *13*, 541.

Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M., *et al.* (2014). Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development. PLOS Genetics *10*, e1004868.

Okae, H., Toh, H., Sato, T., Hiura, H., Takahashi, S., Shirane, K., Kabayama, Y., Suyama, M., Sasaki, H., and Arima, T. (2018). Derivation of Human Trophoblast Stem Cells. Cell stem cell *22*, 50-63.e56.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic Dynamics of Imprinted X Inactivation During Early Mouse Development. Science *303*, 644-649.

Papaioannou, V.E., McBurney, M.W., Gardner, R.L., and Evans, M.J. (1975). Fate of teratocarcinoma cells injected into early mouse embryos. Nature *258*, 70.

Pasque, V., Tchieu, J., Karnik, R., Uyeda, M., Sadhu Dimashkie, A., Case, D., Papp, B., Bonora, G., Patel, S., Ho, R., *et al.* (2014). X Chromosome Reactivation Dynamics Reveal Stages of Reprogramming to Pluripotency. Cell *159*, 1681-1697.

Pastor, William A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, Steven E., and Clark, Amander T. (2016). Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. Cell stem cell *18*, 323-329.

Pastor, W.A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T.J., Lukianchikov, A., Liu, X., Polo, J.M., Jacobsen, S.E., *et al.* (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. Nature cell biology *20*, 553-564.

Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell *7*, 597-606.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131-137.

Pereira, L., Yi, F., and Merrill, B.J. (2006). Repression of Nanog Gene Transcription by Tcf3 Limits Embryonic Stem Cell Self-Renewal. Molecular and Cellular Biology *26*, 7479-7491.

Petropoulos, S., Edsgard, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell *165*, 1012-1026.

Pierce, G.B., Jr., and Verney, E.L. (1961). An in vitro and in vivo study of differentiation in teratocarcinomas. Cancer *14*, 1017-1029.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Polo, Jose M., Anderssen, E., Walsh, Ryan M., Schwarz, Benjamin A., Nefzger, Christian M., Lim, Sue M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., *et al.* (2012). A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. Cell *151*, 1617-1632.

Pshezhetsky, A.V., and Ashmarina, L.I. (2013). Desialylation of surface receptors as a new dimension in cell signaling. Biochemistry (Moscow) *78*, 736-745.

Qin, H., Hejna, M., Liu, Y., Percharde, M., Wossidlo, M., Blouin, L., Durruthy-Durruthy, J., Wong, P., Qi, Z., Yu, J., *et al.* (2016). YAP Induces Human Naive Pluripotency. Cell Rep *14*, 2301-2312.

Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development (Cambridge, England) *137*, 395-403.

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

Rastan, S. (1982). Timing of X-chromosome inactivation in postimplantation mouse embryos. Journal of embryology and experimental morphology *71*, 11-24.

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

Rivron, N.C., Frias-Aldeguer, J., Vrij, E.J., Boisset, J.-C., Korving, J., Vivié, J., Truckenmüller, R.K., van Oudenaarden, A., van Blitterswijk, C.A., and Geijsen, N. (2018). Blastocyst-like structures generated solely from stem cells. Nature *557*, 106-111.

Röhrs, S., Scherr, M., Romani, J., Zaborski, M., Drexler, H.G., and Quentmeier, H. (2010). CD7 in acute myeloid leukemia: correlation with loss of wild-type CEBPA, consequence of epigenetic regulation. Journal of Hematology & Oncology *3*, 15-15.

Roode, M., Blair, K., Snell, P., Elder, K., Marchant, S., Smith, A., and Nichols, J. (2012). Human hypoblast formation is not dependent on FGF signalling. Developmental biology *361*, 358-363.

Rosenthal, M.D., Wishnow, R.M., and Sato, G.H. (1970). In vitro growth and differetiation of clonal populations of multipotential mouse clls derived from a transplantable testicular teratocarcinoma. Journal of the National Cancer Institute *44*, 1001-1014.

Russell, R., Ilg, M., Lin, Q., Wu, G., Lechel, A., Bergmann, W., Eiseler, T., Linta, L., Kumar P, P., Klingenstein, M., *et al.* (2015). A Dynamic Role of TBX3 in the Pluripotency Circuitry. Stem Cell Reports *5*, 1155-1170.

Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T.W., Kuoy, E., Langerman, J., Clark, A.T., Jaenisch, R., *et al.* (2017). Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. Cell stem cell *20*, 87-101.

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

Schauer, R. (2009). Sialic acids as regulators of molecular and cellular interactions. Current Opinion in Structural Biology *19*, 507-514.

Schwarzkopf, M., Knobeloch, K.-P., Rohde, E., Hinderlich, S., Wiechens, N., Lucka, L., Horak, I., Reutter, W., and Horstkorte, R. (2002). Sialylation is essential for early development in mice. Proceedings of the National Academy of Sciences *99*, 5267-5270.

Senner, C.E., Krueger, F., Oxley, D., Andrews, S., and Hemberger, M. (2012). DNA methylation profiles define stem cell identity and reveal a tight embryonic-extraembryonic lineage boundary. Stem Cells *30*, 2732-2745.

Shahbazi, M.N., Jedrusik, A., Vuoristo, S., Recher, G., Hupalowska, A., Bolton, V., Fogarty, N.M.E., Campbell, A., Devito, L.G., Ilic, D., *et al.* (2016). Self-organization of the human embryo in the absence of maternal tissues. Nature cell biology *18*, 700.

Shakiba, N., White, C.A., Lipsitz, Y.Y., Yachie-Kinoshita, A., Tonge, P.D., Hussein, S.M.I., Puri, M.C., Elbaz, J., Morrissey-Scoot, J., Li, M., *et al.* (2015a). CD24 tracks divergent pluripotent states in mouse and human cells. *6*, 7329.

Shakiba, N., White, C.A., Lipsitz, Y.Y., Yachie-Kinoshita, A., Tonge, P.D., Hussein, S.M.I., Puri, M.C., Elbaz, J., Morrissey-Scoot, J., Li, M., *et al.* (2015b). CD24 tracks divergent pluripotent states in mouse and human cells. Nature Communications *6*, 7329.

Shi, Y., Kirwan, P., Smith, J., MacLean, G., Orkin, S.H., and Livesey, F.J. (2012). A human stem cell model of early Alzheimer's disease pathology in Down syndrome. Science translational medicine *4*, 124ra129.

Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., Kondoh, G., Okawa, K., Chujo, T., Suzuki, T., *et al.* (2009). The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. Dev Cell *17*, 775-787.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. PLOS Biology *6*, e253.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell *4*, 481-495.

Smith, A. (2017). Formative pluripotency: the executive phase in a developmental continuum. Development (Cambridge, England) *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, J.R., Vallier, L., Lupo, G., Alexander, M., Harris, W.A., and Pedersen, R.A. (2008). Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. Developmental biology *313*, 107-117.

Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature Reviews Genetics *14*, 204.

Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G.W., Cook, E.G., Hargus, G., Blak, A., Cooper, O., Mitalipova, M., *et al.* (2009). Parkinson's Disease Patient-Derived Induced Pluripotent Stem Cells Free of Viral Reprogramming Factors. Cell *136*, 964-977.

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

Solter, D., ŠKreb, N., and Damjanov, I. (1970). Extrauterine Growth of Mouse Egg-cylinders results in Malignant Teratoma. Nature *227*, 503.

Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N., and Junker, J.P. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. Nature Biotechnology *36*, 469.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature *465*, 175.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell stem cell *2*, 230-240.

Stevens, L.C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. Developmental biology *21*, 364-382.

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

Stirparo, G.G., Boroviak, T., Guo, G., Nichols, J., Smith, A., and Bertone, P. (2017). Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human preimplantation epiblast. bioRxiv.

Tachibana, M., Amato, P., Sparman, M., Gutierrez, Nuria M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.-S., Sritanaudomchai, H., *et al.* (2013). Human Embryonic Stem Cells Derived by Somatic Cell Nuclear Transfer. Cell *153*, 1228-1238.

Takagi, N., Sugawara, O., and Sasaki, M. (1982). Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. Chromosoma *85*, 275-286.

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., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K., and Yamanaka, S. (2014). Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nature Communications *5*, 3678.

Takahashi, S., Kobayashi, S., and Hiratani, I. (2018). Epigenetic differences between naïve and primed pluripotent stem cells. Cellular and Molecular Life Sciences *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.

Tanabe, K., Nakamura, M., Narita, M., Takahashi, K., and Yamanaka, S. (2013). Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. Proceedings of the National Academy of Sciences of the United States of America *110*, 12172-12179.

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.

Tanaka, T.S., Isabel, L.d.S., V., S.L., Hidenori, A., Toshiyuki, Y., Hisayuki, A., Shinya, Y., Myriam, G., and H., K.M.S. (2006). Esg1, expressed exclusively in preimplantation embryos, germline, and embryonic stem cells, is a putative RNA-binding protein with broad RNA targets. Development, Growth & Differentiation *48*, 381-390.

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 Hd iPsc Consortium (2012). Induced Pluripotent Stem Cells from Patients with Huntington's Disease Show CAG-Repeat-Expansion-Associated Phenotypes. Cell stem cell *11*, 264-278.

Theunissen, Thorold W., Friedli, M., He, Y., Planet, E., O'Neil, Ryan C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., *et al.* (2016). Molecular Criteria for Defining the Naive Human Pluripotent State. Cell stem cell *19*, 502-515.

Theunissen, Thorold W., Powell, Benjamin E., Wang, H., Mitalipova, M., Faddah, Dina A., Reddy, J., Fan, Zi P., Maetzel, D., Ganz, K., Shi, L., *et al.* (2014). Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. Cell stem cell *15*, 471-487.

Thompson, P.J., Macfarlan, T.S., and Lorincz, M.C. (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. Molecular cell *62*, 766-776.

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, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., Becker, R.A., and Hearn, J.P. (1995). Isolation of a primate embryonic stem cell line. Proceedings of the National Academy of Sciences of the United States of America *92*, 7844-7848.

Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., and Hearn, J.P. (1996). Pluripotent cell lines derived from common marmoset (Callithrix jacchus) blastocysts. Biol Reprod *55*, 254-259.

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

Trusler, O., Huang, Z., Goodwin, J., and Laslett, A.L. (2018). Cell surface markers for the identification and study of human naive pluripotent stem cells. Stem cell research *26*, 36-43.

Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. Journal of cell science *118*, 4495-4509.

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 (Cambridge, England) *136*, 1339-1349.

Vallot, C., Ouimette, J.F., Makhlouf, M., Feraud, O., Pontis, J., Come, J., Martinat, C., Bennaceur-Griscelli, A., Lalande, M., and Rougeulle, C. (2015). Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. Cell stem cell *16*, 533-546.

Vallot, C., Patrat, C., Collier, A.J., Huret, C., Casanova, M., Liyakat Ali, T.M., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P.J., *et al.* (2017). XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. Cell stem cell *20*, 102-111.

van den Berg, D.L., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. Cell stem cell *6*, 369-381.

Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytometry Part A : the journal of the International Society for Analytical Cytology *87*, 636-645.

Veillard, A.-C., Marks, H., Bernardo, A.S., Jouneau, L., Laloë, D., Boulanger, L., Kaan, A., Brochard, V., Tosolini, M., Pedersen, R., *et al.* (2014). Stable Methylation at Promoters Distinguishes Epiblast Stem Cells from Embryonic Stem Cells and the In Vivo Epiblasts. Stem Cells and Development *23*, 2014-2029.

von Meyenn, F., Iurlaro, M., Habibi, E., Liu, Ning Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., *et al.* (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. Molecular cell *62*, 848-861.

Waghray, A., Saiz, N., Jayaprakash, Anitha D., Freire, Ana G., Papatsenko, D., Pereira, C.-F., Lee, D.-F., Brosh, R., Chang, B., Darr, H., *et al.* (2015). Tbx3 Controls Dppa3 Levels and Exit from Pluripotency toward Mesoderm. Stem Cell Reports *5*, 97-110.

Wang, G., Zhang, H., Zhao, Y., Li, J., Cai, J., Wang, P., Meng, S., Feng, J., Miao, C., Ding, M., *et al.* (2005). Noggin and bFGF cooperate to maintain the pluripotency of human embryonic stem cells in the absence of feeder layers. Biochemical and biophysical research communications *330*, 934-942.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364.

Wang, J., Singh, M., Sun, C., Besser, D., Prigione, A., Ivics, Z., Hurst, L.D., and Izsvak, Z. (2016). Isolation and cultivation of naive-like human pluripotent stem cells based on HERVH expression. Nat Protocols *11*, 327-346.

Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Rasko, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., *et al.* (2014a). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature *516*, 405-409.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., *et al.* (2014b). Programming and inheritance of parental DNA methylomes in mammals. Cell *157*, 979-991.

Wang, Y.-C., Stein, J.W., Lynch, C.L., Tran, H.T., Lee, C.-Y., Coleman, R., Hatch, A., Antontsev, V.G., Chy, H.S., O'Brien, C.M., *et al.* (2015). Glycosyltransferase ST6GAL1 contributes to the regulation of pluripotency in human pluripotent stem cells. Scientific Reports *5*, 13317.

Wang, Y., Zhao, C., Hou, Z., Yang, Y., Bi, Y., Wang, H., Zhang, Y., and Gao, S. (2018). Unique molecular events during reprogramming of human somatic cells to induced pluripotent stem cells (iPSCs) at naïve state. eLife *7*, e29518.

Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., *et al.* (2014). Derivation of naïve human embryonic stem cells. Proceedings of the National Academy of Sciences *111*, 4484-4489.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., *et al.* (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell stem cell *7*, 618-630.

Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nature Reviews Molecular Cell Biology *17*, 155.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hamalainen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., *et al.* (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766-770.

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. Nature cell biology *13*, 838-845.

Wright, A.J., and Andrews, P.W. (2009). Surface marker antigens in the characterization of human embryonic stem cells. Stem cell research *3*, 3-11.

Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C., and Zhong, S. (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. Genome research *20*, 804-815.

Xu, C., Rosler, E., Jiang, J., Lebkowski, J.S., Gold, J.D., O'Sullivan, C., Delavan-Boorsma, K., Mok, M., Bronstein, A., and Carpenter, M.K. (2005a). Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium. Stem Cells *23*, 315-323.

Xu, Q., Koo, K.J., Wilbur, T., G., V.D.L., and H., K.P. (2016). DPPA5 Supports Pluripotency and Reprogramming by Regulating NANOG Turnover. STEM CELLS *34*, 588-600.

Xu, R.-H., Chen, X., Li, D.S., Li, R., Addicks, G.C., Glennon, C., Zwaka, T.P., and Thomson, J.A. (2002). BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nature Biotechnology *20*, 1261.

Xu, R.H., Peck, R.M., Li, D.S., Feng, X., Ludwig, T., and Thomson, J.A. (2005b). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nat Methods *2*, 185-190.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., *et al.* (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature Structural &Amp; Molecular Biology *20*, 1131.

Yang, P., Wang, Y., and Macfarlan, T.S. (2017a). The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. Trends in Genetics *33*, 871-881.

Yang, Y., Liu, B., Xu, J., Wang, J., Wu, J., Shi, C., Xu, Y., Dong, J., Wang, C., Lai, W., *et al.* (2017b). Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. Cell *169*, 243-257.e225.

Ye, S., Li, P., Tong, C., and Ying, Q.L. (2013). Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. The EMBO journal *32*, 2548-2560.

Yeo, J.-C., Jiang, J., Tan, Z.-Y., Yim, G.-R., Ng, J.-H., Göke, J., Kraus, P., Liang, H., Gonzales, Kevin Andrew U., 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 (Cambridge, England) *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.

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

Yoo, M., Shin, J., Kim, J., Ryall, K.A., Lee, K., Lee, S., Jeon, M., Kang, J., and Tan, A.C. (2015). DSigDB: drug signatures database for gene set analysis. Bioinformatics *31*, 3069-3071.

Yoon, B.S., Yoo, S.J., Lee, J.E., You, S., Lee, H.T., and Yoon, H.S. (2006). Enhanced differentiation of human embryonic stem cells into cardiomyocytes by combining hanging drop culture and 5-azacytidine treatment. Differentiation *74*, 149-159.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I.I., and Thomson, J.A. (2009). Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences. Science (New York, NY) *324*, 797-801.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Yu, X., Zou, J., Ye, Z., Hammond, H., Chen, G., Tokunaga, A., Mali, P., Li, Y.-M., Civin, C., Gaiano, N., *et al.* (2008). Notch Signaling Activation in Human Embryonic Stem Cells Is Required for Embryonic, but Not Trophoblastic, Lineage Commitment. Cell stem cell *2*, 461-471.

Zhang, J., Lian, Q., Zhu, G., Zhou, F., Sui, L., Tan, C., Mutalif, R.A., Navasankari, R., Zhang, Y., Tse, H.F., *et al.* (2011). A human iPSC model of Hutchinson Gilford Progeria reveals vascular smooth muscle and mesenchymal stem cell defects. Cell stem cell *8*, 31-45.

Zhang, J.X., Zhang, J., Yan, W., Wang, Y.Y., Han, L., Yue, X., Liu, N., You, Y.P., Jiang, T., Pu, P.Y., *et al.* (2013). Unique genome-wide map of TCF4 and STAT3 targets using ChIP-seq reveals their association with new molecular subtypes of glioblastoma. Neuro-oncology *15*, 279-289.

Zhao, T., Fu, Y., Zhu, J., Liu, Y., Zhang, Q., Yi, Z., Chen, S., Jiao, Z., Xu, X., Xu, J., *et al.* (2018). Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming. Cell stem cell.

Zhao, Y., Zhao, T., Guan, J., Zhang, X., Fu, Y., Ye, J., Zhu, J., Meng, G., Ge, J., Yang, S., *et al.* (2015). A XEN-like State Bridges Somatic Cells to Pluripotency during Chemical Reprogramming. Cell *163*, 1678-1691.

Zhou, H., Li, W., Zhu, S., Joo, J.Y., Do, J.T., Xiong, W., Kim, J.B., Zhang, K., Schöler, H.R., and Ding, S. (2010). Conversion of Mouse Epiblast Stem Cells to an Earlier Pluripotency State by Small Molecules. Journal of Biological Chemistry *285*, 29676-29680.

Zhuang, L.Z., and Li, R.H. (1991). Study on reproductive endocrinology of human placenta (II)--Hormone secreting activity of cytotrophoblast cells. Science in China Series B, Chemistry, life sciences & earth sciences *34*, 1092-1097.

Zimmerlin, L., Park, T.S., Huo, J.S., Verma, K., Pather, S.R., Talbot, C.C., Agarwal, J., Steppan, D., Zhang, Y.W., Considine, M., *et al.* (2016). Tankyrase inhibition promotes a stable human naïve pluripotent state with improved functionality. Development (Cambridge, England) *143*, 4368-4380.

Zunder, Eli R., Lujan, E., Goltsev, Y., Wernig, M., and Nolan, Garry P. (2015). A Continuous Molecular Roadmap to iPSC Reprogramming through Progression Analysis of Single-Cell Mass Cytometry. Cell stem cell *16*, 323-337.