From Department of Clinical Science, Intervention and Technology,
Division of Obstetrics and Gynecology
Karolinska Institutet, Stockholm, Sweden

EXPLORING MAMMALIAN PREIMPLANTATION DEVELOPMENT & PLURIPOTENCY

John Paul Schell

Karolinska Institutet

Stockholm 2020

Published by Karolinska Institutet.
Printed by E-Print AB 2020
Cover: Pluripotent gene OCT4 (red) expression in human blastocyst and two stem cell states.

ISBN 977-91-7831 -842-1

Exploring Mammalian Preimplantation Development & Pluripotency

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

John Paul Schell

Principal Supervisor:
Assoc. Prof. Fredrik Lanner, Ph.D
Karolinska Institutet
Department of Clinical Science, Intervention and Technology
Division of Obstetrics and Gynecology

Co-supervisor(s):
Assoc. Prof. Anna Falk
Karolinska Institutet
Department of Neuroscience

Prof. Juha Kere
Karolinska Institutet
Department of Bioscience and Nutrition

Academic Mentor:
Prof. Urban Lendahl
Karolinska Institutet
Department of Cell and Molecular Biology

Opponent:
Thorsten Edwin Boroviak, Ph.D
University of Cambridge
Department of Physiology, Development and Neuroscience

Examination Board:
Prof. Anneli Stavreus-Evers
Uppsala University
Department of Women and Children's Health
Biology of Reproduction

Prof. Sergey Rodin
Uppsala University
Department of Chemistry/Surgical Science
Polymer Chemistry/Thoracic Surgery

Prof. Thomas Perlmann
Karolinska Institutet
Department of Cell and Molecular Biology

I dedicate this thesis to my Mom and Dad, for instilling a balance of curiosity, caution, and confidence, that continues to guide me in all walks of life.

"Science is the poetry of reality."

-Richard Dawkins

ABSTRACT

Our current understanding of how stem cells arise and transition during embryonic development has been limited by analysis tools that have lacked single-cell, whole-genome resolution. This thesis emphasizes the use of novel techniques and cutting-edge technology to better evaluate the biological underpinning of stem cell dynamics in both mouse and human.

Development relies on stem cells to establish different lineage potentials from the same starting material. Stem cell populations are produced during embryogenesis at multiple stages, existing in various cellular states. These cells have unique self-renewal properties that allow them to divide without differentiating. Stem cell plasticity becomes more restricted as development progresses. A totipotent stem cell state arises after fertilization once embryo cells can generate exact copies of themselves. Totipotent cells maintain the competency for specification into both embryonic (organism) and extraembryonic (placenta and yolk sack) lineages. Once the mammalian blastocyst is formed, the embryonic lineage is maintained exclusively in epiblast (EPI) cells. Both pre- and postimplantation EPI cells are considered pluripotent stem cells, which lack the capacity for generating extraembryonic tissues but maintain full competency to develop into all embryonic germ lineages. During embryogenesis EPI cells transition through several definable pluripotent states, several of which can be maintained *in vitro*. This thesis focuses on utilizing better methods for evaluating how well *in vitro* stem cell culture systems recapitulate endogenous developmental cell types.

In **Paper I** we assessed pre- and postimplantation mouse embryonic stem cells and compared their allelic and transcriptional profiles with developing *in vivo* cell types. We were able to make unprecedented observations of X chromosome inactivation (XCI) dynamics, elucidating evidence that *in vitro* mouse XCI does not follow the perceived dogma that preimplantation stem cells express two fully active X chromosomes. By assessing the full length of each X chromosome with allelic resolution we found that XCI is initiated heterogeneously in preimplantation female stem cells with an observable elongated transition between stem cell states.

In **Paper II** we screen two states of human pluripotent stem cells and preimplantation human embryos to define cell surface markers that attempts to effectively separate preimplantation from postimplantation epiblast. The markers provide a sorting method for state conversions.

Paper III and **IV** complement one another in their intent to define the limits of mouse totipotency using transcriptomics and implementation of functional aggregation assays that effectively evaluate lineage specification and commitment. We determined when the first lineage segregation is defined and used an assortment of molecular tools to evaluate embryonic and extraembryonic contribution. This establishes a benchmark for defining totipotency.

Together the findings presented in this thesis add significant contribution toward an improved understanding of mammalian embryonic development and stem cell biology.

LIST OF SCIENTIFIC PAPERS INCLUDED IN THE THESIS

I. Chen G, **SCHELL JP****, Benitez JA, Petropoulos S,Yilmaz M, Reinius B, Alekseenko Z, Shi L, Hedlund E, Lanner F, Sandberg R, Deng Q. (2016). Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. **Genome Research**, 10.11.1/gr.201954.115

II. Collier AJ*, Panula SP*, **SCHELL JP****, Chovanec P, Plaza Reyes A, Petropoulos S, Corcoran AE, Walker R, Douagi I, Lanner F, Rugg-Gunn PJ. (2017). Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. **Cell Stem Cell**, Jun 1;20(6):874-890.e7

III. Posfai E, Petropoulos S, de Barros F, **SCHELL JP**, Jurisica I, Sandberg R, Lanner F, Rossant J. (2017). Position- and Hippo signalling-dependent plasticity during lineage segregation in the early mouse embryo. **Elife**, Feb 22;6. pii: e22906

IV. Posfai E*, **SCHELL JP***, Adrian Janiszewski*,Isidora R, Murray A, Bradshaw B, Pardon T, Bakkali M, Talon I, Geest N, Kumar P, To S, Petropoulos S, Jurisicova A, Pasque V, Lanner F, Rossant J. (2020) Defining Totipotency Using Criteria of Increasing Stringency. **bioRxiv**, 2020.03.02;972893

LIST OF SCIENTIFIC PAPERS NOT INCLUDED IN THE THESIS

I. Faridani OR, Abdullayev I, Hagemann-Jensen M, **SCHELL JP**, Lanner F, Sandberg R. (2016). Single-cell sequencing of the small-RNA transcriptome. **Nature Biotechnology**. 10.1038/nbt.3701

II. Gelali E, Girelli G, Matsumoto M, Wernersson E, Custodio J, Mota A, Schweitzer M, Ferenc K, Li X, Mirzazadeh R, Agostini F, **SCHELL JP**, Lanner F, Crosetto N, Bienko M. (2019). iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. **Nature Communications**, 10.1038/s41467-019-09616-w

*First author
**Second author

CONTENTS

LIST OF ABBREVIATIONS

BSA	bovine serum albumin
DiM	(S)-(+)- Dimethindene maleate
DNA	deoxyribonucleic acid
EB	embryoid body
ECs	embryonic carcinoma
EDTA	ethylenediaminetetraacetic acid
EPI	epiblast
EpiSCs	epiblast stem cells
FACs	fluorescence activated cell sorting
FISH	fluorescent in situ hybridisation
hESCs	human embryonic stem cells
hPSCs	human pluripotent stem cells
ICM	inner cell mass
iPSCs	induced pluripotent stem cells
IVF	*in vitro* fertilization
KOSR	knock out serum replacement
LIF	leukaemia inhibitory factor
MEFs	mouse embryonic feeders
mESCs	mouse embryonic stem cells
mPSCs	mouse pluripotent stem cells
mRNA	messenger RNA
NHSM	naïve human stem cell medium
OSKM	Oct4, Sox2, Klf4, c-Myc
PBS	phosphate buffered saline
PGC	primordial germ cells
PGCLC	primordial germ like cells
Pol III	polymerase III
PSCs	pluripotent stem cells
RNA	ribonucleic acid
SCENIC	single cell regulatory network inference and clustering
SNPs	single nucleotide polymorphism
TADs	topological associating domains
TE	trophectoderm

tRNA	transfer RNA
UMI	unique molecular identifier
XCI	X chromosome inactivation

1 INTRODUCTION

As sperm meets egg, a cellular cascade of remodeling events is initiated, and life is set in motion. Upon fertilization the mammalian embryo undergoes an elaborate transformation of size and shape during the first rounds of division. These divisions give rise to seemingly identical cells that compose the totipotent morula, which compacts and expands into a blastocyst structure that physically segregates embryonic from extraembryonic lineages. Pluripotent stem cells are a unique cell type with dynamic properties, including self-renewal and the ability to differentiate into all germ lineages. Several pluripotent states can be captured from mouse and human embryos using culture conditions containing different combinations of inhibitors, growth factors, and small molecules. These culture conditions modulate state specific gene expression programs responsible for cell survival and self-renewal. The developmental significance of different stem cell states is key to unraveling their biological potential for future research and biomedical application. To understand pluripotency we first look at the molecular mechanisms driving plasticity from within the cell itself.

1.1 A Simple Code Under Complex Control

DNA can be described as a unit of information, conserved in all multicellular forms of life. It is a sequence of self-perpetuating chemistry that reacts, replicates, and evolves to endure vastly different environments with changing physiological demands. By gaining selective advantages with each surviving copy, molecular code can govern an entirely more complex existence than the simple four-nucleotide base variants (Adenine, Thymine, Guanine, and Cytosine) of which DNA is composed. These sequences can be arranged into a cosmic number of combinations that maintain the integrity of a species by allowing each and every cell a blueprint for the entire organism. DNA orchestrates how a cell identifies itself in response to its local community of cells in tissues and organs. In order to react to biochemical or physical demands that influence life essential processes, cells are constantly evaluating and reacting to their local environments. While the DNA is found encapsulated in the nucleus of mammalian cells, it is regulated on multiple levels and responsible for coordinating how each cell reacts to the outside world. The sequence achieves this by having evolved programmed responses to exogenous situations. Using complex intercellular protein machinery, the cell transcribes a close relative of DNA, a single-stranded chemical code known as messenger RNA (mRNA). RNA is the connection between blueprint DNA and

functional protein machinery. Our ability to obtain whole-genome mRNA profiles from individual cells has revolutionized our understanding of cellular biology.

Every Human cell contains roughly 3 billion base pairs of DNA (A,T,G,C nucleotides in a sequence). Our cells maintain these 3 billion bases on 46 different strings of sequence known as chromosomes. As humans we have 22 pairs (2 copies) of each chromosome, plus sex chromosomes (XX female, or XY male), each spanning different lengths and coding for various genes. Chromosomes are made of strings of double-stranded DNA, in a very specific sequence. Single nucleotide polymorphisms (SNPs) are genetically traceable mutations that are responsible for some of the observable individual-to-individual variations in our species. The vast majority of sequence is conserved human-to-human.

DNA's chromosome pairs are being held in the nucleus of each cell. The approximate 2 meters of nucleotide sequence that make up our combined chromosomes can be condensed into a regulated form of DNA known as chromatin. Through the implementation of protein complexes known as histones, the chromatin is wound and un-wound based on epigenetic regulation of histone modification. Topological associating domains (TADs) also govern nuclear dynamics, which regulate access to particular gene sequences (Yu et al., 2017). Epigenetic control over chromatin state also determines which genes a cell can and can not transcribe. Access to specific genes, along with their promoter and enhancer sequence regions, is what dictates how a cell can react.

Cells need access to genetic sequences for a protein complex known as polymerase III (Pol III) to read that gene sequence and generate a complementary single-stranded RNA chain (Reyes-Lamothe et al., 2010). Generating mRNA through this process is called transcription. The mRNA is recognized by complementing tRNAs which associate tri-nucleotide arrangements with specific amino acid assembly in the process of translation. Amino acid chains are the sub-units of the protein machines that govern the cellular environment. While our understanding of the molecular kinetics controlling this vast complexity remains limited, basic research is constantly gaining new insights into how cells communicate and react.

1.2 A Breif History Of Embryonic Stem Cells

The term 'pluripotent embryonic stem cells' was first coined by Leroy Stevens, who conducted studies in mouse strain 129, which he recognized for it's propensity to form teratocarcinomas in the testis (Stevens, 1958). Stevens found that when he excised cells from

these teratoma tumor growths, he could transplant portions of them into other mice and continuously give rise to cells that resembled early developmental tissues (Stevens, 1959). His landmark discovery that led to the generation of stem cells was in 1970 when he observed that primordial germ cells which were competent to form teratomas, also gave rise to embryoid bodies and cell types that morphologically resemble the early embryonic niche (Stevens, 1970). These cells were later termed 'embryonic carcinoma' (ECs).

Beatrice Mintz at the Institute for Cancer research in Philadelphia collaborated with Stevens and showed that these embryonic like cells not only had the capacity to form teratocarcenomas, but also could integrate into developing organisms (Mintz & Illmensee, 1975). She observed the capacity for chimeric contribution when ECs were injected into developing blastocysts. This was the first real indication that stem cells could functionally reintegrated into a developing organism, slowly elucidating the full potential of pluripotency. Gale Martin at the University of California, San Francisco, and subsequently Mathew Kaufman and Martin Evans at the University of Cambridge, went on to derive the first mouse embryonic stem cells (mESCs) from embryos (Evans & Kaufman, 1981; Martin, 1981). These were maintained *in vitro* on a substrate of mouse embryonic feeders (MEFs). In these first landmark papers, the three scientists showed that they could maintain stem cells *in vitro* for a prolonged number of passages and sufficiently validated pluripotency by chimeric integration with germline contribution. Germline contribution can be assessed by integrating either teratocarcenoma cells or embryonic stem cells into a developing blastocyst, and when introduced back into the uterus of a pseudo pregnant mother, will develop into a viable transgenic contributing chimera (Martin, 1981). Thus the genetic information contained in the *in vitro* system can be expressed and conserved in future generations of a mouse strain. Now for the first time, stem cells became an *in vitro* tool, used to edit and introduce changes to the germline of well-studied mouse breeds. Stem cell contribution is the cornerstone of chimeric transgenics and has been utilized as an intermediate in a large portion of mutant/knockout strains produced to date. Germline contribution is the absolute gold standard to validate if putative pluripotent cells retain the genomic capacity to generate the entire organism (De Los Angeles et al., 2015).

1.3 Chimeric Integration and Germ Line Contribution

Contribution to the germline depends on the developmental potential of integrating pluripotent stem cells (PSCs). When injected or aggregated with morula or blastocysts, PSCs must make contact amongst the other inner cell mass (ICM) cells and assume the identity and

behavior of the endogenous epiblast (Nagy et al.,1993). PSCs integrated with epiblast gives rise to primordial germ cells still harboring the genetic diversity of the *in vitro* system and successfully introducing the capacity to pass this genetic information onto their progeny.

1.4 Primordial Germ Cells

The ability to generate viable primordial germ cells (PGCs) is an evolutionarily conserved key element in terms of multi-generational survival (Extavour & Akam, 2003). Since this is the only cell type that goes on to contribute to the germline, PGCs must accomplish genome-wide demethylation and epigenetic resetting in order to give rise to haploid gamete cells (Tang, Kobayashi, Irie, Dietmann, & Surani, 2016). If conception occurs, these gametes will once again achieve *in vivo* totipotency, and propagate functional progeny.

Primordial germ cells begin their existence after emerging from a population of the developing epiblast (McLaren, 2000). While it still remains uncertain how this transition occurs in human, it is speculated that PGCs are likely differentiated from a pool of competent epiblast or selected based on a spatiotemporal influence (McLaren, 1999). PGCs divide as the neural tube elongates, migrating toward the site of the developing gonad (Molyneaux, Stallock, Schaible, & Wylie, 2001). During this migration the PGCs respond to the strong BMP gradient differently than the patterning somatic lineages, and are propagated along the body axis (Nakamura & Extavour, 2016). BLIMP1 expression inhibits neuronal and endoderm commitment, while upstream SOX17 expression sustains this population of gonad determinants (Irie et al., 2015). Mavericks in reproductive biology such as Anne McLaren and Azim Surani have pioneered this field by highly contributing to our current understanding of early embryogenesis, epigenetics, and PGC specification (A. Surani & Smith, 2007; M. A. Surani, 2016).

1.5 Harnessing and Validating Pluripotency

In vitro validation of pluripotency consists of an assortment of recognized features indicative of state specific cell physiology. Pluripotent stem cells are known to express developmental genes responsible for self-renewal and maintenance of an undifferentiated state. Core transcription factors such as OCT4 (POU5F1), SOX2, and NANOG are known to be key regulators of pluripotent circuitry (De Los Angeles et al., 2015). Initially Leukemia inhibitory factor (LIF), in combination with serum, was recognized to promote expression of these genes in a heterogeneous cell population known as LIF/serum culture (Nagy et al., 1993). mESCs under LIF/serum conditions maintain a metastable expression of pluripotency,

meaning individual cells express variable levels of pluripotent factors which inevitably influences their efficiency for embryonic integration (Alexandrova et al., 2016). Stem cells *in vitro* are generally characterized by immunocytochemistry of known pluripotent factors and state specific surface markers. *In vitro* ESCs are also assayed through embryoid body (EB) formation and differentiation into the three germ layers (Peterson & Loring, 2012). mESC EBs have the capacity to differentiate into teratomas when implanted into immune compromised mice (Iles, 1977). The first reported human embryonic stem cells (hESCs) were derived from blastocyst stage embryos by Jamie Thomson in 1998 (Thomson et al., 1998). Due to obvious ethical restraints the gold standard for validation of pluripotency resorted to the teratoma assay. Although the culture conditions, morphology, and gene expression profile of these conventional hESCs dramatically differed from mESCs, their inequalities with respect to developmental potential went unrecognized for some time.

Conventional human embryonic stem cells grow in flat epithelial like colonies that differ greatly from mESCs. These stem cells are more prone to differentiation and display low clonal efficiency when seeded as single cells (Pera & Trounson, 2004). These conventional hESCs depend primarily on bFGF/FGF2 and ActivinA/TGFβ in their medium to support and maintain undifferentiated self-renewal (Vallier, Alexander, & Pedersen, 2005). This requirement for FGF differs significantly from maintenance of mouse pluripotency, where it has been observed in mouse embryos that FGF exposure pushes all ICM cells toward a primitive endoderm fate rather than epiblast (Lanner & Rossant, 2010; Yamanaka, Lanner, & Rossant, 2010).

1.6 The Influence of X Inactivation

X chromosome inactivation is another developmental property where differentiated female cells epigenetically inactivate one of their X chromosomes in order to equalize the gender discrepancy in transcriptional output (Bermejo-Alvarez, Ramos-Ibeas, & Gutierrez-Adan, 2012; Lyon, 1961). XCI during differentiation has been observed to coincide with the loss of pluripotency as well as responsible for specific lineage segregations in mouse blastocyst formation (Okamoto et al., 2011). High integrating mouse pluripotent mESCs maintain active X chromosomes due to being hypomethylated, yet conventional human PSCs exhibit heterogeneous expression of different conflicting XCI states (Vallot et al., 2015). Prior to efforts such as Paper II, conclusions regarding XCI have been extrapolated from observed expression of only a handful of genes and generally a single methylation mark. Because the

X chromosome contains over 200 genes, current approaches using single-cell transcriptomics are more equipped for resolving this type of biological phenomenon.

1.7 Induced Pluripotentcy

The landmark work of Shinya Yamanaka's lab in their advent of reprograming differentiated cell types into induced pluripotent stem cells (iPSCs) in 2006 brought about an observation that supported the possibility that hESCs were developmentally equivalent to mESCs. This reprogramming phenomenon dedifferentiated terminal somatic cell types into a pluripotent state through over expression of key pluripotency genes (Takahashi & Yamanaka, 2006). The fact that mouse iPSCs and human iPSCs can both be reprogrammed using the same combination of factors: OCT4, SOX2, KLF4, cMYC (OSKM), suggested their developmental equivalence. This phenomenon somewhat corroborated how such vast variability between mouse and human pluripotent networks could still maintain similar nodes of transcriptional similarity. While representative stem cell lines in human and mouse both express key pluripotency factors *in vitro*, their endogenous developmental equivalents were only recognized after the isolation of postimplantation epiblast in mouse, as mentioned later. Several years later it was shown that OSKM has the capacity to reprogram to totipotency *in vivo*; allowing the dedifferentiation of somatic tissues into both embryonic and extraembryonic resembling cell types (Abad et al., 2013). This potential is somewhat brought into question by Paper IV, and is reviewed further in the discussion section of this thesis.

1.8 State Contradictions

For many years the stem cell field assumed that mouse and human embryogenesis was so developmentally different that indeed their representative pre-implantation schemas were established and maintained by contrasting signaling pathways. This was also somewhat further self-validated by a survival phenomenon evolved in mouse known as diapause, which allows mouse embryos to maintain quiescent in a pregnant mother under distress; a phenomenon not conserved in primates (Ptak et al., 2012). It has been hypothesized that this diapause 'induced stasis' could be why mESCs were more amenable to *in vitro* culture. Later the oncogene MYC was shown to be responsible for this transition between mitotic and non-mitotic pluripotency, thus proposing a mechanistic separation of self-renewal from the maintenance of an undifferentiated state (Scognamiglio et al., 2016). As a growing body of evidence separated mouse PSCs (mPSCs) and human PSCs (hPSCs), one critical developmental similarity went unrecognized until 2007 that completely revamped our understanding of the undifferentiated state, and sub-divided pluripotency forever.

Nearly ten years after the derivation of human pluripotent stem cells, Paul Tesar made the groundbreaking realization that indeed if you derived mouse embryonic stem cells from postimplantation epiblast and cultured in FGF/ActivinA, you in fact get outgrowths that can be propagated and maintained with a phenotype identical to that of conventional hPSCs (Tesar et al., 2007, Brons et al). Tesar termed these postimplantation ESCs, Epiblast Stem Cells (EpiSCs). This was the first definitive evidence that even though human embryonic stem cells were derived from blastocyst stage ICM, their developmental significance was likely more representative of a post-implantation schema. This distinction has led to a much more thorough investigation and classification of 'naïve' PSCs: representing a pre-implantation epiblast, and 'primed' PSCs: representing post-implantation epiblast (Nichols & Smith, 2009).

Naïve and primed stem cells represent two generalized states of the pluripotent spectrum. During the initial transition from pre-implantation to post-implantation epiblast, evidence from *in vivo* mouse post-implantation studies suggests a dramatic shift in the underlying signaling pathways controlling self-renewal and pluripotency (Nichols & Smith, 2009). This transition elicits changes in the transcriptional profile, metabolic capacity, X inactivation status, epigenetic identity, and morphology (Takashima et al., 2014). Naive cells exhibit a domed morphology, retain a hypomethylated genome, and utilize a metabolic state that employs high glycolytic activity as well as aerobic respiration for energy production (Gu et al., 2016). Primed cells have a flat epithelial-like morphology, rely solely on a glycolytic metabolism, and are epigenetically primed for differentiation (Nichols & Smith, 2009). The naïve state maintains pluripotency through expression of key transcriptional circuitry that utilizes LIF/STAT3 and TFCP2L1 signaling (Qin et al., 2016). As previously described, primed stem cells employ FGF/TGFβ signaling for maintenance of pluripotency (Tesar, 2016, Brons et al).

The distinction between naïve and primed states of pluripotency pushed the field to identify and homogenize culture conditions. In searching for a solution to this variability, the group of Austin Smith and Jennifer Nichols discovered the 2-inhibitor (2i) culture system. 2i is composed of a MEKi and GSK3i, that when added to serum free medium, is able to maintain much more transcriptionally homogenous cells that express a pluripotent profile more similar to the epiblast than conventional mESCs (Ying et al., 2008, Boroviak et al., 2014). 2i cells exhibit much higher integration efficiency than conventional pluripotent cultures, can be

cultured with less heterogeneity, and were physiologically more relevant to the developmental time point from when they were derived. The term 'ground state' was coined to describe the homogenous hypomethylated state of being impervious to extrinsic cues, although much of this definition now overlaps with the description of 'naivety' (Davidson, Mason, & Pera, 2015). 2i maintains a more homogenous naïve state of pluripotency that conserves high REX/NANOG populations and thus supports more efficient contribution in comparison to conventional LIF/Serum cultured cells (Alexandrova et al., 2016).

1.9 Human Naive Pluripotency

Once it was accepted that conventional human embryonic stem cells in fact did not reflect the naïve pluripotent characteristics indicative of preimplantation, there have been many attempts to convert, derive, and reprogram human naïve embryonic stem cell lines (Chan et al., 2013; Gafni et al., 2013; Valamehr et al., 2014). This would soon become a contentious area of debate as several labs stepped up to the task of aiming for a representative human ground state. The first big claim for the derivation of naïve hESCs was from the lab of Jacob Hanna at the Weisman institute. Hanna's group showed that naive hPSCs could be maintained using MEKi, GSK3i, JNKi, P38i, PKCi, and ROCKi, in a medium called NHSM (Gafni et al., 2013). This first publication has been heavily scrutinized over the observation of robust integration into mouse chimeras, as well as a gene expression profile that mismatches various lineage-specific differentiation markers expressed during supposed pluripotency, thus this work still remains debated (Theunissen et al., 2016). Despite the negative hype of these cells and their potential misclassification as 'naïve', they have shown promise in their ability to differentiate toward PGCLCs (Irie et al., 2015). In fact cells converted from primed ESCs to a reformulated NHSM, along with briefly culturing with half/half primate IVF medium, were able to achieve chimeric integration and germ line contribution in cynologous monkeys (Y. Chen et al., 2015). Another report has identified the origin of PGCs in cynologous monkeys to be the amnion (Sasaki et al., 2016). This indicates that indeed a subset of the factors used in NHSM medium does induce the ability for integration, despite the fact that at this stage, they have been more appropriately designated as a primed/differentiated mesoderm intermediate, due to their expression of factors such as brachyury, a marker of mesoderm (Takashima et al., 2014). The propensity to allow integration does indeed fit into the pluripotent definition, although in comparison to the transcriptional differences between naïve, primed, and endogenous epiblast in mouse, NHSM cells remain outliers. While these first derivation reports were likely premature in their classification, the field at large strived to reproduce the initial findings and continued

searching for representative human naïve pluripotency. Efforts in dissecting the transcriptional landscape of pre-implantation human embryos have allowed for profiling endogenous embryonic populations and identification of novel human pre-implantation epiblast specific markers such as KLF17 (Blakeley et al., 2015; Petropoulos et al., 2016)

Approximately a year later two landmark naïve pluripotency papers made their appearance with robust naïve culture systems. One of Beatrice Mintz's former students, Rudolph Jaenish and his group at MIT came up with a screening method to systematically test small molecule inhibitors on their ability to retain *in vitro* naïve pluripotency. In order to produce a generalized transcriptional profile for naïve vs. primed, they first looked at a set of most variable genes between primed and naïve in mouse PSCs. With a set of five inhibitors including MEKi, GSK3i, ROCKi, BRAFi, and SRCi, their 5i/L/A cocktail was able to recapitulate a stable naïve state in human pluripotent stem cells that is transcriptionally similar to naïve mouse cells in the 2i condition (Theunissen et al., 2014). Almost concurrently, the lab of Austin Smith & Jennifer Nichols at Cambridge University, produced a new method of maintaining human naïve pluripotency by the overexpression of KLF2 and NANOG to 'reset' cells in order to transition them from primed hPSCs into a stable naïve state maintained by a MEKi, GSK3i, ROCKi, and PKCi in their t2iGöY medium (Takashima et al., 2014). In 2016 the group reported derivation into this condition directly from isolated cells of the ICM (Guo et al., 2016). Further analysis of these two robust naïve systems has shown that they overlap with cleavage-stage embryos in sharing a unique transposable element expression profile, thus validating their relevance to pre-implantation. While the 5i/L/A and t2iGöY maintain this distinctive transposon signature, NHSM cells do not (Theunissen et al., 2016).

1.10 Human Chimeric Integration & Limitation

Chimeric integration experiments where conventional primed hPSCs are injected into postimplantation mouse embryos have led to another means of validating human pluripotent potential. While mouse and human maintain very different developmental timelines and completely different implantation schemas, conventional primed hPSCs have been shown to efficiently integrate into early and late mouse gastrula, thus validating the capacity for *in vivo* contribution to endoderm, ectoderm, and mesoderm lineages (Mascetti & Pedersen, 2016). Based off fate mapping and gestational relevance, conventional primed hPSCs are appropriately classified as a post-implantation cell type (Wu et al., 2015). This degree of developmental competency has not been observed in human naïve chimeric integration

experiments performed using preimplantation mouse embryos. Human naïve hESCs integrate into mouse morula and blastocysts at very low efficiency (Theunissen et al., 2016). It is possible that at this pre-implantation stage of development, the mouse/human chimeric capacity for integration is insufficient as a model for validating pluripotency, due to differences in developmental timing, epigenetic inconsistencies, dissimilarities in embryonic lineage segregation, or blatant species-specific genomic incompatibility. It is also completely plausible that the current state of naive cultured cells are incompetent for ICM integration at all, although this could not be thoroughly analyzed without performing human/human integration attempts into developing morula or blastocysts (De Los Angeles et al., 2015). Human naïve stem cell integration experiments are the remaining unvalidated proof in principal to provide functional confirmation that naïve hPSCs are definably pluripotent and developmentally parallel preimplantation.

As our mechanistic understanding of endogenous pluripotency burgeons, new information is constantly reshaping our perspective of undifferentiated states. Clearly, there is still much to discover before we can safely and efficiently wield the full potential of pluripotency toward better understanding development, treating infertility, and curing diseases at it's root.

2 AIMS

The aim of this thesis was to utilize single-cell whole-genome technologies to better compare embryonic stem cells in different *in vitro* and *in vivo* states.

The specific aims of the four projects were:

I. To evaluate allele specific transcription in pre- and postimplantation mouse stem cell states in order to better evaluate X chromosome inactivation dynamics.

II. To screen human naïve and primed pluripotent stem cells to elucidate surface markers that can be utilized for sorting and filtering during state conversion.

III. To study the timing of the first lineage specification and commitment in mouse embryos, and to better understand their underlying signalling.

IV. To create a set of criteria using whole-genome analysis and functional integration competency for the evaluation of totipotency.

3 MATERIALS AND METHODS

This section summarizes the materials and methods implemented in this thesis. Please see Paper I – IV for a more detailed description of each technique. Here we also describe several unpublished efforts, devised for the purpose of better modeling and evaluation of pluripotency, that we feel are worthy of mention in this section of the doctoral thesis.

3.1 STEM CELL CULTURE

In vitro stem cells are a useful tool for modeling development and have been fundamental to the experiments performed in this thesis. In many ways, stem cells allow us to capture *in vivo* physiology in reproducible and genetically static backgrounds. In this thesis we assay five different *in vitro* stem cell culture conditions in mouse, and two states of pluripotency in human. Our studies in mouse allow a functional evaluation of stem cells using embryos, while our study in human provides insight into human embryo biology using stem cells. We use these two methods to better elucidate developmental understanding of how *in vitro* states recapitulate *in vivo* cellular dynamics.

3.1.1 Mouse Embryonic Stem Cells

mESCs were maintained in multiple states using different culture methodologies. In Paper I we generate 2i/LIF (naïve ground state), LIF/Serum (metastable naïve), and EpiSCs (primed). Paper IV reanalyzed 2i and Primed cells from Paper I, and added additional 2i cells, and two examples of expanded/extended pluripotent stem cell conditions.

2i, naïve ground state – CHIR (GSK3i), PD0325901 (MEKi), LIF (Leukemia Inhibitory Factor). Maintained on a substrate of MEFs. Passaged using Passaged using .05% Trypsin-EDTA or TrypLE. Used in Paper I and IV.

LIF/Serum, metastable naïve state – LIF. Maintained on .1% Gelatin. Passaged using Passaged using .05% Trypsin-EDTA. Used in Paper I.

EpiSCs, primed state – bFGF, Activin. Maintained on substrate of Fibronectin. Passaged using Collagenase IV or Accutase. Used in Paper I and IV.

D-EPSCs, extended state – LIF, CHIR, DiM ((S)-(+)- Dimethindene maleate), and MiH (Minocycline hydrochloride). Maintained on MEFs. Passaged using .05% Trypsin-EDTA. Used in Paper IV.

L-EPSCs, expanded state- LIF, CHIR, PD0325901, JNK Inhibitor VIII, SB203580, A-419259 and XAV939. Maintained on SNL-feeders. Passaged using Accutase. Used in Paper IV.

3.1.2 Human Pluripotent Stem Cells

In Paper III, hPSCs were maintained in two states: naïve and primed. Stem cell lines were maintained and transitioned between these culture conditions. Human stem cell lines are validated using teratoma assay and by their *in vitro* capacity to differentiate into all three germ lineages. Human naïve and primed culture conditions are maintained at 37C in 5% O2 (hypoxia), 5% CO2.

Primed hPSCs- supplemented with bFGF and TGFB. Maintained on a substrate of 10ug/ml recombinant laminin 521, in a medium of NutriStem XF. Passaged using TrypLE or Accutase (Rodin et al., 2014).

Naïve t2iGöY- CHIR, PD0325901,LIF, PKCi, ROCKi, toggled by using overexpression of KLF2 and NANOG. Maintained on a substrate of MEFs or Matrigel (Geltrex). Passaged using Accutase (Takashima et al., 2014).

Naive 5iLFA- CHIR, PD0325901, LIF, BRAFi, JNKi, bFGF, ActivinA. Dependent on substrate of E12.5 MEFs. Passaged using Accutase (Theunissen et al., 2014)

3.2 EMBRYO CULTURE

Availability of numerous mouse embryos is not a limitation in research, while access to human embryos remains extremely limited and relies on donations from informed couples. Despite how well stem cells recapitulate pre- and postimplantation pluripotency, their efficacy will always need to be compared to natural embryogenesis.

3.2.1 Mouse Embryo Culture

Mouse embryos were utilized in Paper I, III, and IV. Embryos were flushed from fertilized super-ovulated females and allowed to develop in KSOM medium (Millipore). In Paper I mouse embryos of C57BL/6J x CAST/EiJ background are used for the derivation of mouse

embryonic stem cell lines specifically for evaluating maternal and paternal allele specific transcription. Paper III transcriptionally dissects pre-lineage mouse embryogenesis and uses re-aggregation of developing embryos as an assay to evaluate lineage specification and commitment. Paper IV expands the transcriptional analysis to include preimplantation to postimplantation and gastrulation (Zygote – E7.5), employing the embryo aggregation assay for evaluation of any putative totipotent cell types.

3.2.2 Mouse Embryo Aggregation Assay

Aggregation assay is the method of dissociating blastomeres and reaggregating them to test developmental plasticity and commitment. The mouse embryo aggregation assay is first utilized in Paper III as a means to determine the specific timing of ICM and TE divergence In Paper IV the aggregation assay is used as a criteria to define totipotency. If the cell type being aggregated is capable of giving rise to both embryonic and extraembryonic lineages, it is considered totipotent. This is the strictest level of stringency for defining totipotency.

3.2.3 Human Embryo Culture

Human embryo samples used in Paper III were donated by patients of Karolinska University Hospital in Huddinge and Carl von Linné Clinic in Uppsala. All experiments were approved by Stockholm's Regional Ethics Board (2012/1765-31/1), and with full disclosure and informed consent from the donating couples.

3.2.4 Thawing Human Embryos

Embryos were thawed using ThawKit Cleave (Vitrolife). G-1 Plus (Vitrolife) was used from thaw to E3, then G-2 Plus (Vitrolife) was used from E3 to E6-7. G-1 Plus and G-2 Plus were more recently reformulated into a combined version called GTL (Vitrolife), which also efficiently sustained development. Medium was covered by Ovoil (Vitrolife) and maintained at a strict 5% O2, 5% CO2, at 37C.

3.2.5 Human Embryo Immunosurgery

After zona pellucidae has been removed using Tyrode's (SigmaAldrich), immunosurgery can be performed by plating a drop (30µl) anti-human antibody (Sigma, H8765) diluted 1:3 on a pre-warmed dish. The embryos are placed into the droplet for at least one hour in incubator at 37C. After incubation embryos can be washed three times using KSOM medium (Millipore), then put in a drop (30µl) complement guinea pig serum (Sigma, S1639) for 15 minutes. Embryos are then placed back into KSOM medium, triturated, and returned to the

incubator for 20 minutes to 45 minutes (triturating every 10 minutes), until the ICM can be seperated.

3.2.6 Human Postimplantation Embryo Model

In Paper II we opted to implement the postimplantation human embryo model (Shahbazi et al., 2016) to test our primed specific cell surface markers. While the results were somewhat conflicting due to non-specific binding of antibodies in the 3D outgrowths, ultimately we scrapped this prospective angle of Paper II because it would take too many embryos to troubleshoot properly. Consistent with the original reports of the method: of the embryos thawed about 50% hatch, about 50% of those will go onto attach, and of those about 50% will contain OCT4 positive epiblast cells. Postimplantation outgrowths are limited by the international 14 day rule and must be fixed by E14.

The postimplantation model plates a hatching blastocyst onto Ibidi cover-slip chamber slide. The embryo is kept in IVC-1(20%FBS) until E7-E8, when it is changed to IVC-2 (30% KOSR). Both culture mediums contain 200ng/ml progesterone, 8nM estradiol, and 25uM N-acetyl-L-Cysteine. Outgrowths are grown in 21% O2, 5% CO2 at 37C.

3.3 MOLECULAR & COMPUTATIONAL METHODS

3.3.1 Immunofluorescent Microscopy

Immunofluorescent staining is a means to visualize protein specific expression in individual cells. By producing an antibody in another species, we can generate antigens that are exclusive to any specific protein of interest. Before using our antibody we first administer a fixation solution which quickly cross-links proteins, immediately killing the cells. Permeabilization buffers are mild detergents used to allow antibodies access to cytoplasmic and nuclear compartments. The cells are also exposed to a blocking buffer which is usually protein rich and is meant to displace non-specific binding of the antibody. The primary antibody specific to our protein of interest is now administered to the cells and during incubation, antigen specific binding will take place. A secondary antibody contains a primary antibody recognition unit, as well as a conjugated fluorophore that allows fluorescence detection/imaging.

Cells are generally washed with PBS -/- once, then 4% formaldehyde is added for 15 minutes of fixation. The fixative is removed and cells are permeabilized with 0.3% Triton X-100 in PBS for 10 minutes. Cells are blocked for one hour in 4% donkey or fetal bovine serum, 1% BSA, and 0.1% Tween-20 in PBS. Antibodies are diluted in blocking buffer. Primary antibodies are left overnight at 4C. Cells are washed with PBS containing .1% BSA, .1% Tween-20. Secondary antibodies are left for 1 hour room temperature or overnight at 4 degrees. DNA is stained with 1ug/ml Hoechst for 15 minutes in PBS.

3.3.2 FACs & Cell Sorting

FACs (fluorescence activated cell sorting) was used in Paper I, II, and IV for various seeding/sorting applications. FACs offers a method of automated cell sorting using various criteria such as viability, cell cycle, or surface marker composition. Importantly, it provides a quantitative method for evaluating cellular heterogeneity. In Paper I and IV we use FACs for single-cell sorting of cells into plates for sequencing. In Paper II we perform extensive cell sorting for screened state specific surface antibodies and define co-expressing signatures. One limitation of this technology is the reliance on high numbers of cells as input, although improved hardware/software and more efficient labeling strategies are constantly improving.

3.3.3 RNA FISH

RNA is single stranded molecular sequence that is possible to visualize using specific probes that bind like oligonucleotides to RNA. These RNA binding units can be fluorescently labeled, and their signal amplified for visual quantification.

Samples were fixed using 4% formaldehyde for 15 mins at RT, then washed PBS. Large fixed samples in suspension such as embryos are transferred to silanized glass coverslips and dried for 2 mins prior to continuing. Coverslips holding samples are placed in pre-chilled (-20°C) methanol for 10 mins at -20°C to permeabilize cells, and left out to air dry for 30 minutes at RT. Once dried the samples are heat shocked with TE buffer, pH 8.0 (Promega) at 70°C for 10 minutes. Samples are washed with 2X SSC (Thermo). Next samples are hybridized for 6h at 38.5°C in a humidity chamber with RNA FISH probes in hybridization buffer using of RNase free water, 2X SSC, 10% dextran sulfate (Sigma), 10% formamide (Thermo), 2 mg/ml E. coli tRNA (Sigma), 2 mM ribonucleoside vanadyl complex (New England Biolabs), and 2 mg/ml BSA (Jackson ImmunoResearch). After 6h hybridization, samples are washed with 20% formamide in 2X SSC four times, each 15 minutes at 38.5°C. In the last 15 minute wash, use Hoescht 33342 (1µg/ml; ThermoFisher Scientific) in the wash buffer for nuclear stain. Samples are further washed with 2X SSC and mounted with Prolong antifade (Thermo), left to dry for 24 hours in darkness at room temperature before imaging.

3.3.4 RNA-Seq Transcriptomics

Whole genome RNA analysis has revolutionized our ability to interpret gene expression maintained in different cell types. Paper II implements bulk RNA-Seq of sorted subgroups, while Paper I, III, and IV use single-cell SMART-Seq2 data. The single-cell analysis reveals similarities and dissimilarities between stem cells and endogenous cell populations.

The raw binary format output of a sequencing run is a BCL file. These binary files are re-formatted into text files called FASTQ which contain the binary raw data as well as their associated quality scores. After BCL to FASTQ conversion we align the reads to STAR reference genome of choice. After trimming, along with barcode and UMI demultiplexing, BAM files containing this indexed information are annotated into cell/gene expression counts. Expression matrixes were generated from each experiment and used for downstream analysis in each respective project. Paper I used the allelic resolution provided by single-cell transcriptomics to calculate a ratio of active to inactive gene expression along individual chromosomes. Paper III used single-cell transcriptomics to corroborate CDX2 fluorescence

intensity tied to Hippo dynamics. Paper IV combines datasets to track mouse embryogenesis from zygote to E7.5 gastrulation stage, in order to compare stem cell and blastoid data with endogenous cell types. Using regulon analysis tool SCENIC we also tracked expression of transcription factors and their target genes to better validate our observed trajectories.

4 RESULTS AND DISCUSSION

4.1 SINGLE-CELL XCI AND PLURIPOTENCY TRANSITION (PAPER I)

In this project we found that mESCs express transcriptional profiles that parallel developmental progression. Our analysis identified that ES2i cells were transcriptionally closest to E4.5 epiblast cells. We identified gene expression profiles for each developmental stage and recognized gender specific differential expression patterns, including the expression of the Y chromosome gene Uba1y which we found to be exclusively expressed in male ES2i cells. Utilizing our moving window methodology to calculate and classify expression output across the entire X chromosome, we identified heterogeneous XCI dynamics in all female stem cell states. Allele specificity allowed us to calculate the ratio of X chromosomal output, that was classified as: “uninitiated”, “ongoing”, and “finished”. This study strongly disproves the dogma that 2i cells have completely active X chromosomes, and in fact we observed a large fraction that expressed “ongoing” XCI dynamics. X chromosome inactivation is asynchronous at single-cell resolution and not highly correlated with loss of pluripotency. Interestingly simultaneous expression from both alleles correlates with a higher transcriptional output in general.

Discussion

Allele specific resolution adds a level of unprecedented analysis of XCI dynamics between different states of mouse pluripotency. This study allowed us to better recognize the heterogenous nature of X inactivation found in all female pluripotent stem cell states assayed. A deeper investigation into allele preference will likely lead to a more comprehensive understanding of how dominant and recessive genes are regulated. Since we correlate increased expression with biallelic nature, it is quite provocative to speculate on the underlying molecular kinetics driving allelic preference in the first place. Systematic editing of individual alleles, as well as research involving haploid genomes, will help elucidate these still mysterious cellular behaviors.

4.2 SURFACE MARKERS FOR HUMAN STEM CELL STATES (PAPER II)

In this study we profiled surface markers in naive and primed hPSCs by using 377 cell surface proteins with 487 antibodies. We identified CD46, CD151, PDPN, and MCAM as markers conserved in both naïve and primed states of pluripotency. Primed specific markers identified and confirmed with microscopy include: CD90, CD57, CD24, and HLA-A, B, C. The naïve specific markers identified and confirmed to also be expressed in preimplantation human embryos include: CD75, CD7, CD77, CD130. We found that by using these state specific surface markers as a panel in combination with each other, we could efficiently sort naïve from primed cells using FACs. A primed culture spiked with 10% naïve cells was able to be discriminated using gating of CD75+/CD130+, CD57-/CD24-. We demonstrate the utility of this panel for tracking state conversion from naïve to primed, and primed to naïve, by implementing a classification based on how many of the four naïve markers are being expressed. Through transcriptomic analysis of transitioning naïve sub-populations we deduced that DPPA3, TBX3, FGF18, and FOXC1 were genes indicative of early naïve conversion, while MEG3, XIST, and ZNF729 are associated with late stage naïve conversion. Complementing these findings, we were able to identify transposable elements LTR5B, LTR7Y, HERV9NC-int, associated with early naïve conversion, and BSR/Beta, MER47C, MER57E3, associated with late naïve stage cells. Analysis of XCI dynamics during different points of state conversion showed a reactivation of X chromosome genes primarily during late stage naïve conversion.

Discussion

The capacity to sort state specific cells is a useful tool for basic research and potential future translational applications. This study identified discriminating surface markers that can separate two *in vitro* pluripotent states. The classification of transitioning subpopulations during primed to naïve conversion has allowed us to create a resource for future isolation of these specific cell types. Curiously, when assaying naïve specific surface markers in preimplantation human blastocysts we did not observe an exclusively epiblast signal, indicating that our surface protein signature may track all preimplantation lineages. Future investigation should be initiated to separate human preimplantation lineages based on surface protein composition. This will likely go unresolved until a vast number of donated embryos can be screened using a similar methodology to our approach, or synthetic embryos can be utilized to take their place in basic research.

4.3 THE FIRST LINEAGE SPECIFICATION AND COMITMENT (PAPER III)

In this project we utilized a knock-in Cdx2-eGFP reporter mouse line as a readout for profiling expression of developing TE. We quantified this expression by fluorescence to directly evaluate single-cell Cdx2 expression in morula aggregation experiments, and prior to picking for single-cell transcriptomics. Our analysis revealed that Cdx2 expression emerges at the 16-cell morula stage, and exhibits a gradual restriction to the outer cells which compose TE. This coincides with the onset of active Hippo signaling in response to Cdx2. We found that beyond the morula stage there was an increasing positive correlation between nuclear and cytoplasmic Yap ratios, further corroborating this trend. After tracking single cell fate using morula aggregation, we validated these findings by analyzing lineage specification based on single-cell transcriptomics. The 8-cell embryo is totipotent and expresses no Cdx2. At the 16-cell stage, many ICM and TE destined cells are expressing Cdx2. Based on plasticity due to position, commitment occurs between the late 32-cell to 64 cell stage embryos. Transcriptionally this is specified by the early 32-cell stage. Supporting this, we observe that the embryo cells are only responsive to ROCK inhibitor up until the 32-cell stage morula. Interestingly, the plasticity of ICM commitment remains more promiscuous than TE. Live imaging of lineage specification allowed precise dissection of inside-outside cellular dynamics and integration events.

Discussion

Understanding how and when lineage specification and commitment occur during mammalian embryogenesis is crucial for our classification of specific cell states and lineages. The capacity to generate TE and ICM is a defining trait of totipotency. This study tracks the molecular dynamics underpinning plasticity due to polarity versus placement. Extraembryonic differentiation is initiated and already stabilized by the late 32-cell stage. Understanding which features are responsible for maintaining the TE trajectory is of particular interest for attempts at using embryonic cell types to generate extraembryonic lineages. It still remains uncertain if embryonic lineages can actively give rise to genuine extraembryonic differentiation *in vitro*. This study provides evidence that lineage specific gene expression determinants can have state dependent sensitivity. Just because plasticity of the genome may allow higher transcriptional output from certain early pre-blastocyst expressed genes, these differences do not automatically imply that a totipotent state was activated.

4.4 AN EVALUATION OF TOTIPOTENTCY (PAPER IV)

Recent publications have made claim that stem cells can be maintained in extended/expanded pluripotent states that harbor totipotent qualities (Yang et al., 2018, Yang et al., 2018). One EPSC condition (D-EPSCs) was recently used to generate blastoids from exclusively that cell state (Li et al., 2019). This differs from other previously described blastoids generated by aggregating embryonic stem cells with extraembryonic competent trophoblast stem cells (TSCs) (Rivron et al., 2018). Despite blastoids being morphologically similar to embryos, they have not yet been shown to properly implant *in vivo*. This analysis shows that while blastoid EPI and PE transcriptionally align with E4.5 blastocyst cells, blastoid TE is transcriptionally divergent from native TE, lacking expression of key TE markers such as Elf5 and Cdx2. Based on a set of robust assays that evaluate genome-wide transcriptional associations, and the previously described aggregation assay from Paper III, it is apparent that EPSCs are not maintaining a cell state aligned with totipotency or extraembryonic cell types. Generalized, both L-EPSCs and D-EPSCs cluster along the embryonic lineage between naïve and primed stem cell states. D-EPSCs cluster especially close to E5.5 epiblast cells, and share a similar gene expression profile. The single-cell transcriptomics results complemented the "gold standard" functional aggregation assay in disproving L-EPSCs and D-EPSCs putative capacity to give rise to extraembryonic TE *in vivo* at the appropriate developmental time. Instead these cells were found to integrate primarily into the epiblast compartment, even more robustly than 2i cells. While some cells express Our in-depth tracking of E4.5, E6.25, and E12.5 anatomy was key for properly evaluating lineage contribution.

Discussion

This benchmarking has strictly defined totipotency as a capacity for generating both embryonic and extraembryonic cell types within the context of stage matched embryo dynamics. From our findings it is clear that EPSCs are most similar to conventional embryonic pluripotent states and do not resemble anything remotely totipotent. In fact many of the genes described as being highly expressed in blastomeres, were in fact more indicative of an E5.5 formative state, including the upregulation of DNA methyltransferases Dnmt1, Dnmt3a, and Dnmt3b. This evidence that D-EPSCs maintain a peri-implantation embryonic identity brings into question the blastoids that were produced solely from D-EPSCs. Since the B-blastoid dataset contained a large transitioning Cdx2 positive population that co-expressed T brachyury, and directly clustered with embryonic mesoderm, it presents the possibility that B-blastoid TE is embryonic mesoderm derived. Mesoderm and TE are

transcriptionally quite similar, and morphologically capable of producing analogous-looking cyst like structures. This brings into question if the overexpression of O,S,K,M can actually reprogram to totipotency, or if Abad et al., 2013 essentially produced mesodermal cell types that perhaps only share charateristics with extraembryonic outgrowth. Our increased stringencies for evaluation presented here would be capable of testing this potential if live cells could be isolated for aggregation experiments.

5 CONCLUDING REMARKS

This thesis highlights the incredible progress made in the field of stem cell biology over the last few decades. Advancement in cellular interrogation technologies is allowing unprecedented dimensions of complexity to be resolved. In a single generation we a shift where a PhD no longer studies a single gene but studies all genes simultaneously. Our focus has now transitioned toward identifying cellular states that can be characterized by specific gene expression profiles based on numerous genes. As our mechanistic understanding of cellular life continuous to improve, the biological questions we can ask grow with it. Each of the studies included in this thesis are all based on much older research topics that have evolved with the improving technologies.

- In Paper I we evaluated allele specific transcription patterns in pre- and postimplantation stem cell states, and deduced X inactivation dynamics at single cell resolution across the entire X chromosome.

- In Paper II we effectively screened human naïve and primed stem cells and identified consistent state specific surface marker combinations. We use the markers to track subpopulations during conversion.

- In Paper III we were able to functionally and transcriptionally deduced the timing of specification and commitment during the first lineage segregation in mouse.

- In Paper IV we succesfully benchmarked the criteria for unambiguously evaluating totipotency.

While the studies presented in this thesis have focused on basic research, the implications of their relevance in regenerative medicine are intimately tied to the future of science, medicine, and humanity as a whole.

6 FUTURE PERSPECTIVES

The future holds high potential for stem cell biology. The advent of bona fide synthetic embryos in model species is undoubtedly on the horizon. With it will come a plethora of basic research findings that will allow us to understand human preimplantation with such molecular precision that germ editing technologies will become a realistic solution for monogenic heritable disease. The ability to scale the production of model embryos will allow vast screening platforms to test and evaluate various systems using an identical genetic background. This will allow cleaner analysis than evaluating even sibling embryos, when taking into account the individual to individual genetic variation. When evaluating off-target mutations without genetically identical samples for comparison, it is difficult to separate inappropriate editing events from natural variants. Stable, genetically homogenous, stem cell lines will be necessary for production of synthetic embryo screening platforms.

Large-scale single-cell multi-omics approaches will resolve the efficacy of site-specific gene editing tools. Once these technologies are refined to a resolution that adequately assesses safety and efficiency, they will be used to define best practice solutions in treating heritable genetic disorders. Initially stem cell lines carrying the exact mutation or variant sequence will be produced for use in generating the synthetic embryos, which then are utilized to benchmark the safest and most efficient editing strategy. To prove worth in natural embryos, the top editing candidates will be tested on donated samples harboring the specific genetic disease variant. To ensure precision editing, there will be numerous evaluations with rigorous multi-omics testing. Once these heritable solutions become routine in terms of basic research, their implementation into society as a medical intervention to cure debilitating diseases will inevitably follow.

Stem cell biology will not only aid with curing disease in future generations, but it’s role in regenerative medicine is expanding at such a rate that stem cell therapies are already being approved and tested in humans. As our understanding of cellular behavior and physiology burgeons, so will our ability to replenish adult stem cell niches, reinnervate damaged organs, and completely replace diseased tissues. Stem cells are a building block for cell replacement strategies, as long as they can be safely wielded.

7 ACKNOWLEDGEMENTS

A quick acknowledgement here to thank Sweden, my Supervisor, and my Sweetheart.
I'd like to thank the country of Sweden for allowing me to pursue my education and passion at one of the most prestigious academic establishments to ever exist. A huge appreciation for my main supervisor Fredrik Lanner who has believed in me and allowed a real melding of minds for how we perceive cellular biology. We still can disagree on points, but his door is always open for debate on conjecture. The Lanner lab, Damdimopoulou lab, and the KI crew, I'm gonna have to say my peace to everyone individually, too many friends for one liners and this thesis needs to be turned in with a quickness.. You guys have made this experience so fun and rich and full of life, this is definitely not a goodbye. The last and most important acknowledgement goes to my wife to-be Margherita Zamboni. I don't know how you do it, but your love drives me. I thank you for the endless hours of scientific debate, and listening to my crazy rants (like every day), and especially for going through the trouble of helping me improve my bioinformatics skills. I hope to someday aspire to your brilliance in that field. A shout out to my boy to-be Oliver Schell, can't wait to share the wonders of science with you.

8 REFERENCES

Abad, M., Mosteiro, L., Pantoja, C., Canamero, M., Rayon, T., Ors, I., . . . Serrano, M. (2013). Reprogramming *in vivo* produces teratomas and iPS cells with totipotency features. *Nature, 502*(7471), 340-345. doi: 10.1038/nature12586

Alexandrova, S., Kalkan, T., Humphreys, P., Riddell, A., Scognamiglio, R., Trumpp, A., & Nichols, J. (2016). Selmbryos. *Development, 143*(1), 24-34. doi: 10.1242/dev.124602

Bermejo-Alvarez, P., Ramos-Ibeas, P., & Gutierrez-Adan, A. (2012). Solving the "X" in embryos and stem cells. *Stem Cells Dev, 21*(8), 1215-1224. doi: 10.1089/scd.2011.0685

Blakeley, P., Fogarty, N. M., Del Valle, I., Wamaitha, S. E., Hu, T. X., Elder, K., . . . Niakan, K. K. (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq.*Development, 142*(20), 3613. doi: 10.1242/dev.131235

Chan, Y. S., Goke, J., Ng, J. H., Lu, X., Gonzales, K. A., Tan, C. P., . . . Ng, H. H. (2013). Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell, 13*(6), 663-675. doi: 10.1016/j.stem.2013.11.015

Chen, G., Schell, J. P., Benitez, J. A., Petropoulos, S., Yilmaz, M., Reinius, B., . . . Deng, Q. (2016). Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. *Genome Res, 26*(10), 1342-1354. doi: 10.1101/gr.201954.115

Chen, Y., Niu, Y., Li, Y., Ai, Z., Kang, Y., Shi, H., . . . Li, T. (2015). Generation of Cynomolgus Monkey Chimeric Fetuses using Embryonic Stem Cells. *Cell Stem Cell, 17*(1), 116-124. doi: 10.1016/j.stem.2015.06.004

Davidson, K. C., Mason, E. A., & Pera, M. F. (2015). The pluripotent state in mouse and human. *Development, 142*(18), 3090-3099. doi: 10.1242/dev.116061

De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H., . . . Daley, G. Q. (2015). Hallmarks of pluripotency. *Nature, 525*(7570), 469-478. doi: 10.1038/nature15515

Durruthy-Durruthy, J., Wossidlo, M., Pai, S., Takahashi, Y., Kang, G., Omberg, L., . . . Sebastiano, V. (2016). Spatiotemporal Reconstruction of the Human Blastocyst by Single-Cell Gene-Expression Analysis Informs Induction of Naive Pluripotency. *Dev Cell, 38*(1), 100-115. doi: 10.1016/j.devcel.2016.06.014

Eakin, G. S., & Hadjantonakis, A. K. (2006). Production of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Nat Protoc, 1*(3), 1145-1153. doi: 10.1038/nprot.2006.173

Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature, 292*(5819), 154-156.

Extavour, C. G., & Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development, 130*(24), 5869-5884. doi: 10.1242/dev.00804

Faridani Omid R, A. I., Haggemann-Jensen Michael, Schell John P, Lanner Fredrik, Sandberg Rickard. (2016). Single-cell sequencing of the small-RNA transcriptome. *Nature Biotechnology, Advance Online Publication*(Advance Online Publication).

Gafni, O., Weinberger, L., Mansour, A. A., Manor, Y. S., Chomsky, E., Ben-Yosef, D., . . . Hanna, J. H. (2013). Derivation of novel human ground state naive pluripotent stem cells. *Nature, 504*(7479), 282-286. doi: 10.1038/nature12745

Gu, W., Gaeta, X., Sahakyan, A., Chan, A. B., Hong, C. S., Kim, R., . . . Christofk, H. R. (2016). Glycolytic Metabolism Plays a Functional Role in Regulating Human Pluripotent Stem Cell State.*Cell Stem Cell, 19*(4), 476-490. doi: 10.1016/j.stem.2016.08.008

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., . . . Nichols, J. (2016). Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports, 6*(4), 437-446. doi: 10.1016/j.stemcr.2016.02.005

Hikabe, O., Hamazaki, N., Nagamatsu, G., Obata, Y., Hirao, Y., Hamada, N., . . . Hayashi, K. (2016). Reconstitution *in vitro* of the entire cycle of the mouse female germ line. *Nature*. doi: 10.1038/nature20104

Iles, S. A. (1977). Mouse teratomas and embryoid bodies: their induction and differentiation. *J Embryol Exp Morphol, 38*, 63-75.

Irie, N., Weinberger, L., Tang, W. W., Kobayashi, T., Viukov, S., Manor, Y. S., . . . Surani, M. A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. *Cell, 160*(1-2), 253-268. doi: 10.1016/j.cell.2014.12.013

Lanner, F., & Rossant, J. (2010). The role of FGF/Erk signaling in pluripotent cells. *Development, 137*(20), 3351-3360. doi: 10.1242/dev.050146

Li, R.,…Belmonte J.C.I. (2019). Generation of Blastocyst-like Structures from Mouse Embryonic and Adult Cell Cultures. *Cell*. **179** (3),687–702 e618,

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature, 190*, 372-373.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A, 78*(12), 7634-7638.

Mascetti, V. L., & Pedersen, R. A. (2016). Contributions of Mammalian Chimeras to Pluripotent Stem Cell Research. *Cell Stem Cell, 19*(2), 163-175. doi: 10.1016/j.stem.2016.07.018

McLaren, A. (1999). Signaling for germ cells. *Genes Dev, 13*(4), 373-376.

McLaren, A. (2000). Germ and somatic cell lineages in the developing gonad. *Mol Cell Endocrinol, 163*(1-2), 3-9.

Mintz, B., & Illmensee, K. (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. *Proc Natl Acad Sci U S A, 72*(9), 3585-3589.

Molyneaux, K. A., Stallock, J., Schaible, K., & Wylie, C. (2001). Time-lapse analysis of living mouse germ cell migration. *Dev Biol, 240*(2), 488-498. doi: 10.1006/dbio.2001.0436

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., & Roder, J. C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A, 90*(18), 8424-8428.

Nakamura, T., & Extavour, C. G. (2016). The transcriptional repressor Blimp-1 acts downstream of BMP signaling to generate primordial germ cells in the cricket Gryllus bimaculatus.*Development, 143*(2), 255-263. doi: 10.1242/dev.127563

Nichols, J., & Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell, 4*(6), 487-492. doi: 10.1016/j.stem.2009.05.015

Okamoto, I., Patrat, C., Thepot, D., Peynot, N., Fauque, P., Daniel, N., . . . Heard, E. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development.*Nature, 472*(7343), 370-374. doi: 10.1038/nature09872

Pera, M. F., & Trounson, A. O. (2004). Human embryonic stem cells: prospects for development. *Development, 131*(22), 5515-5525. doi: 10.1242/dev.01451

Peterson, S., & Loring, J. F. (2012). *Human Stem Cell Manual: A Laboratory Guide*: ACADEMIC PressINC.

Petropoulos, S., Edsgard, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., . . . Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell, 167*(1), 285. doi: 10.1016/j.cell.2016.08.009

Ptak, G. E., Tacconi, E., Czernik, M., Toschi, P., Modlinski, J. A., & Loi, P. (2012). Embryonic diapause is conserved across mammals. *PLoS One, 7*(3), e33027. doi: 10.1371/journal.pone.0033027

Qin, H., Hejna, M., Liu, Y., Percharde, M., Wossidlo, M., Blouin, L., . . . Ramalho-Santos, M. (2016). YAP Induces Human Naive Pluripotency. *Cell Rep, 14*(10), 2301-2312. doi: 10.1016/j.celrep.2016.02.036

Reyes-Lamothe R, Sherratt D, Leake M (2010). Stoichiometry and Architecture of Active DNA Replication Machinery in Escherichia Coli. *Science*. **328** (5977): 498–501. doi:10.1126/science.1185757

Rivron, N.C,…Geijsen N. (2018). Blastocyst-like structures generated solely from stem cells. *Nature*. **557** (7703), 106–111,

Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., . . . Saitou, M. (2016). The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. *Dev Cell*. doi: 10.1016/j.devcel.2016.09.007

Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., . . . Saitou, M. (2015). Robust *In Vitro* Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell, 17*(2), 178-194. doi: 10.1016/j.stem.2015.06.014

Scognamiglio, R., Cabezas-Wallscheid, N., Thier, M. C., Altamura, S., Reyes, A., Prendergast, A. M., . . . Trumpp, A. (2016). Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. *Cell, 164*(4), 668-680. doi: 10.1016/j.cell.2015.12.033

Stevens, L. C. (1958). Studies on transplantable testicular teratomas of strain 129 mice. *J Natl Cancer Inst, 20*(6), 1257-1275.

Stevens, L. C. (1959). Embryology of testicular teratomas in strain 129 mice. *J Natl Cancer Inst, 23*, 1249-1295.

Stevens, L. C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. *Dev Biol, 21*(3), 364-382.

Surani, A., & Smith, J. (2007). Obituary: Anne McLaren (1927-2007). *Nature, 448*(7155), 764-765. doi: 10.1038/448764a

Surani, M. A. (2016). In retrospect: Thirty-five years of endless cell potential. *Nature, 535*(7613), 502-503. doi: 10.1038/535502a

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell, 126*(4), 663-676. doi: 10.1016/j.cell.2006.07.024

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., . . . Smith, A. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell, 158*(6), 1254-1269. doi: 10.1016/j.cell.2014.08.029

Tang, W. W., Kobayashi, T., Irie, N., Dietmann, S., & Surani, M. A. (2016). Specification and epigenetic programming of the human germ line. *Nat Rev Genet, 17*(10), 585-600. doi: 10.1038/nrg.2016.88

Tesar, P. J. (2016). Snapshots of Pluripotency. *Stem Cell Reports, 6*(2), 163-167. doi: 10.1016/j.stemcr.2015.12.011

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., . . . McKay, R. D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature, 448*(7150), 196-199. doi: 10.1038/nature05972

Theunissen, T. W., Friedli, M., He, Y., Planet, E., O'Neil, R. C., Markoulaki, S., . . . Jaenisch, R. (2016). Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell, 19*(4), 502-515. doi: 10.1016/j.stem.2016.06.011

Theunissen, T. W., Powell, B. E., Wang, H., Mitalipova, M., Faddah, D. A., Reddy, J., . . . Jaenisch, R. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell, 15*(4), 471-487. doi: 10.1016/j.stem.2014.07.002

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science, 282*(5391), 1145-1147.

Valamehr, B., Robinson, M., Abujarour, R., Rezner, B., Vranceanu, F., Le, T., . . . Flynn, P. (2014). Platform for induction and maintenance of transgene-free hiPSCs resembling ground state pluripotent stem cells. *Stem Cell Reports, 2*(3), 366-381. doi: 10.1016/j.stemcr.2014.01.014

Vallier, L., Alexander, M., & Pedersen, R. A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci, 118*(Pt 19), 4495-4509. doi: 10.1242/jcs.02553

Vallot, C., Ouimette, J. F., Makhlouf, M., Feraud, O., Pontis, J., Come, J., . . . 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*(5), 533-546. doi: 10.1016/j.stem.2015.03.016

Wu, J., Okamura, D., Li, M., Suzuki, K., Luo, C., Ma, L., . . . Izpisua Belmonte, J. C. (2015). An alternative pluripotent state confers interspecies chimaeric competency. *Nature, 521*(7552), 316-321. doi: 10.1038/nature14413

Yamanaka, Y., Lanner, F., & Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development, 137*(5), 715-724. doi: 10.1242/dev.043471

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., . . . Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature, 453*(7194), 519-523. doi: 10.1038/nature06968

Yu, Miao; Ren, Bing (2017). "The Three-Dimensional Organization of Mammalian Genomes". *Annual Review of Cell and Developmental Biology*. **33**: 265–289. doi:10.1146/annurev-cellbio-100616-060531