Embryonic Diapause as a Model for Pluripotent Stem Cell Quiescence

Lilyana Margaretha

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

C. Anthony Blau, Chair

Carol Ware

Hannele Ruohola-Baker

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Embryonic Diapause as a Model for Pluripotent Stem Cell Quiescence

Lilyana Margaretha

Chairperson of the Supervisory Committee:
Professor C. Anthony Blau
Department of Hematology

Pluripotent stem cells inhabit, and can transition between, distinct metastable states. Embryonic stem cells (ESC), derived from the pre-implantation inner cell mass (ICM) of mouse embryos, retain a relatively more primitive (or "naïve") phenotypic and functional profile compared to "primed" pluripotent cells derived from the post-implantation epiblast. Based on the hypothesis that a physiological state of stem cell dormancy, known as diapause, might represent a novel pluripotent state, I characterized the transcriptional programs of pluripotent stem cells from: i) the ICM of pre-implantation embryos; ii) the epiblast of post-implantation embryos; and iii) the ICM of embryos induced to enter diapause. The results presented in this dissertation demonstrate significant differences in the molecular signatures of these states and define diapause as a distinct "quiescence-associated" pluripotent state. Using this information I constructed a "developmental yardstick" for pluripotent stem cells and found that the quiescence-associated pluripotent state can be modeled *in vitro*. My findings also point to *Runx1* as a likely regulator of the quiescence-associated state and provide a roadmap for dissecting the molecular basis of stem cell quiescence.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I wish to express my sincere gratitude to my two advisors, Tony Blau and Carol Ware, for all of their invaluable input, support and excellent mentorship throughout my graduate work. I also greatly appreciate the advice and guidance from the other members of my other supervisory committee: Hannele Ruohola-Baker, Karol Bomsztyk, and Denny Liggitt.

I am grateful for collaborating with Brig Mecham, Larry Ruzzo, Daniel Jones, Jay Shendure, Ruolan Qiu, Charlie Lee, John Scott and Jennifer Whiting. I would also like to acknowledge all of the past and present members of the Blau laboratory and the UW ISCRM ESC core staff, with a special thanks to Chris Cavanaugh for tremendous help in the mouse work. I would like to thank Ron Seifert from the UW ISCRM Garvey Imaging Lab for assistance in confocal imaging. I greatly appreciate the administrative help from the UW Molecular and Cellular Biology program and the UW Institue for Stem Cell and Regenerative Medicine.

This thesis would never have been completed without the immense support and encouragement from my family and friends.

Dedication

This doctoral thesis is dedicated to my parents, Alvin and Anita, for their love, support, encouragement throughout my life

To my husband Erwin for his love and support throughout my graduate career

To my daughter Vanessa for brightening my days.

Chapter 1 : Introduction

Aims

The goal of my work has been to study stem cell quiescence in pluripotent stem cells. My dissertation addresses four main questions: (1) How distinct are the transcriptional profiles of pluripotent stem cells obtained directly from embryos at various stages of development? (2) What are the defining transcriptional features of embryonic diapause, a quiescent-associated state of development? (3) What are the mechanisms underlying diapause? (4) Do ESC treated with HDAC inhibitors have any developmental equivalence to diapause?

Different pluripotent stem cell states

Stem cells possess two remarkable properties: an enormous capacity for self-renewal and the ability to give rise to multiple types of daughter cells - generating tissues, organs and whole organisms. Attempts to learn how stem cells function in a physiological context are confounded by the difficulty of investigating their behavior under basal conditions *in vivo*. Stem cells are rare, inaccessible to real-time monitoring, and heterogeneous in their responses to biological signals. These difficulties have motivated efforts to develop physiologically relevant *in vitro* models for investigating stem cell function. Pluripotent stem cell cultures provide a means for examining stem cell biology; however, the *in vivo* relevance of *in vitro* observations is often uncertain (Pera and Tam, 2010, Smith, 2001).

Mounting studies show that the notion of a “pluripotent stem cell state” is misleading in its connotation of uniformity. Two major pluripotent stem cell states have been defined in mice: the naïve state, characteristic of pluripotent cells from the inner cell mass (ICM) of late (day 4.5) pre-implantation embryos and reflected in mouse embryonic stem cells (mESC) insulated from differentiation signals (Ying et al., 2008), and the primed state, characteristic of pluripotent cells from post-implantation embryos and reflected in mouse epiblast stem cells (EpiSC) and human embryonic stem cells (hESC) (Hanna et al., 2010). The differences between these states are not subtle. Naïve ESC have a domed appearance, are tolerant of trypsin exposure and single cell passage, express both X chromosomes and *Nanog* alleles, require protection from differentiation signals to self-renew, and are capable of contributing to virtually all tissues

(including the germline) in chimeric mice (Bradley et al., 1984, Evans and Kaufman, 1981, Li et al., 2012, Miyanari and Torres-Padilla, 2012, Martin, 1981, Williams et al., 1988, Ying et al., 2003, Ying et al., 2008). In contrast, human ESC and mouse epiblast stem cells (EpiSC) have a flattened appearance, self-renew in response to Activin and fibroblast growth factor – 2 (FGF2), variable X chromosome inactivation, and the latter fail to form chimeric mice (Brons et al., 2007, Dvash et al., 2010, Tesar et al., 2007, Thomson et al., 1998).

Extensive efforts have been carried out to induce pluripotent stem cells to transition between states, e.g. from the naïve to the primed state, (Bao et al., 2009, Guo et al., 2009, Hanna et al., 2009, Zhou et al., 2010), from the pluripotent state to more differentiated cell types (e.g. somatic cells), and more recently, from differentiated somatic cells back to the pluripotent state (i.e. induced pluripotent stem cells or iPSC) (Takahashi and Yamanaka 2006). Promoting these transitions requires genetic manipulation and/or the addition of exogenous factors. One groundbreaking study was the use of histone deacetylase inhibitors (HDACi) to promote stem cell self-renewal across different species and across different pluripotent states (Ware et al., 2009). An intriguing feature of this study was that butyrate, a 4-carbon fatty acid HDACi, promoted self-renewal only at a relatively low and narrow range of concentrations (0.2 – 0.3 mM), further suggesting physiological relevance.

Global expression profiling of butyrate-induced genes reveals very little overlap between mouse and human ESC. This observation can be explained by the fact that mouse and human ESC appear to represent distinct developmental stages *in vitro* (Brons et al., 2007, Tesar et al., 2007). While mouse and human ESC are similarly derived from the pre-implantation blastocysts, they differ with respect to morphology, gene expression and signaling pathways. In contrast, pluripotent stem cells from the epiblast layer of post-implantation mouse embryos (epiblast stem cells or EpiSC) more closely resemble human ESC. Mouse ESC treated with butyrate shift their gene expression profiles toward a slightly later stage analogous to EpiSC, whereas human ESC treated with butyrate repress genes associated with mouse EpiSC and induce genes associated with mouse ESC. Thus butyrate (and other HDACi) induces a convergence in the gene expression profiles of developmentally earlier (mouse) and later (human) ESC toward a common intermediate state albeit not exactly the same state (Figure 1.1). A remaining question from the Ware study is whether there exists an *in vivo* corollary of the HDACi-treated ESC. Answers to this question not only will help in understanding the biology of

a particular developmental stage in nature, but will also confirm the usefulness of this system as an *in vitro* model to study an *in vivo* phenomenon.

Embryonic diapause: Overview

Results described above, combined with the observation that butyrate reduces 5-bromo-2'-deoxyuridine (BrdU) incorporation in both mouse and human ESC (Ware et al., 2009), led us to hypothesize that butyrate (and other HDACi)-treated ESC might mimic a naturally occurring phenomenon called "embryonic diapause" (Figure 1.2).

Embryonic diapause, or delayed implantation, is an evolutionarily conserved reproductive strategy adopted by nearly 100 different mammalian species that favors the survival of newborn (Renfree and Shaw, 2000). During unfavorable conditions, resulting either from environmental or maternal stimuli (e.g. lactation), embryo implantation is delayed. Instead, the pre-implantation embryo closely associates with the uterine wall without implanting and is maintained in a state of developmental dormancy until more favorable conditions arise. Diapause is accompanied by a significant reduction in the rate of cell division. This reproductive strategy thus coordinates the birth of offspring with relatively more favorable environmental conditions (e.g. available food, mild weather, etc). As reproduction exacts a large energy cost, it would be essential to ensure that offspring survives and is born into optimal conditions.

Two types of embryonic diapause have been identified in mammals: obligate and facultative. Obligate diapause occurs in every gestation of some species, such as armadillos, all species of pinniped, many mustelids, all ursids, one species of fruit bat and the roe deer (Daniel, 1970, Lindenfors et al., 2003, Lopes et al., 2004). Also known as seasonal delayed implantation, the phenomenon is induced by environmental factors (e.g. photoperiod and seasonal temperature variations) and occurs as a regular part of the reproductive cycle (Sandell, 1990). In roe deer, for example, embryonic diapause lasts for 5 months, extending the total period of pregnancy to 9 months until optimal conditions for delivery arrive in the spring. Facultative diapause, on the other hand, is usually associated with metabolic stress such as starvation and lactation. In some rodents, insectivores and marsupials, facultative embryonic diapause is induced in pregnant females that are lactating a previous litter (Daniel, 1970, Vogel, 1981).

Embryonic diapause is evolutionarily conserved across mammals

Rodents are the most commonly used mammalian models for reproductive physiology, and the occurrence of embryonic diapause was already reported in lactating rodents 120 years ago (Lataste, 1891). Various stimuli might induce the facultative diapause in rodents: pregnancy associated with heavy lactation, pre-puberty, elevated environmental temperature and even social stress such as overcrowding or presence of strange males (Bellringer et al., 1980, Marois, 1982). While this response to environmental or metabolic stress occurs naturally *in vivo*, there is as yet little understanding of the genes that control the process of embryonic diapause. Studies of this phenomenon using rodent models have benefited from findings that embryonic diapause can be experimentally induced by ovariectomy, which prevents the ovarian estradiol surge that is required for implantation (Hamatani et al., 2004, Hunter and Evans, 1999, Psychoyos, 1973, Renfree, 1981).

A recent study challenged the general belief that embryonic diapause has evolved independently within different taxonomic groups of mammals (Conaway, 1971, Lopes et al., 2004, Mead, 1993, Ptak et al., 2012, Sandell, 1990). By successfully inducing diapause in normally non-diapausing mammals, Ptak and colleagues proved that embryonic diapause is conserved across mammals. Blastocysts from sheep (a non-diapausing species) were transferred into mouse uteri in which diapause conditions were induced. Not only could the sheep embryos enter into diapause, as demonstrated by growth arrest, viability maintenance and embryonic diapause-specific pattern of gene expression, they could eventually develop into normal lambs when transferred to surrogate ewe recipients. These findings suggest that embryonic diapause may be more widespread than previously thought. Embryonic diapause could be a fundamental, evolutionarily conserved phenomenon that is “neglected” in many species because it is no longer necessary. Ptak elaborated that if a female is living under controlled conditions with food, nice temperature and no overcrowding, her embryos will not go through diapause because it is not necessary.

A detailed investigation of embryonic diapause has not been previously reported in primates. Studies in humans are obviously not feasible due to ethical constraints. Although it is not known whether diapause can occur in humans, anecdotal cases suggest that human embryonic development can be temporarily arrested prior to implantation. In the first case study, a woman

undergoing *in vitro* fertilization experienced a five-week delay from the time of oocyte aspiration to the appearance of human chorionic gonadotropin (hCG), which is released at the time of implantation (Grinsted and Avery, 1996). In the second study conducted on 189 naturally occurring pregnancies, implantation occurred between 6 and 12 days after ovulation (Wilcox et al., 1999). Variation in the length of time prior to implantation might be attributable to stress around the time of conception. Stress could trigger a change in the hormone levels, which will subsequently affect the receptivity of the uterus and induce diapause.

The potential clinical importance of investigating embryonic diapause: If human embryos are capable of entering diapause, it may explain inaccurate due dates that are usually determined by the date of the last menstrual period. By understanding this phenomenon, the erroneous classification of infants that are "small for gestational age" and unnecessary Caesarean sections may be avoided.

A recurring goal in refining *in vitro* fertilization (IVF) techniques is to improve pregnancy success rates following IVF by enhancing embryo survival and implantation after embryo freeze-thaw or after an extended embryo culture period prior to uterine transfer. Previously, *in vitro* fertilized embryos were transferred into the uterus at the 6-8 cell stage (i.e. 3 days after oocyte retrieval). With refinement of embryo culture systems, human embryos can be cultured up to the blastocyst stage prior to uterine transfer. Extended culture period increases implantation rates while minimizing multiple births. However, only 38-66% of fertilized embryos can develop into blastocysts regardless of the quality of the *in vitro* culture system (Gardner et al., 1998). Furthermore, ovarian stimulation with exogenous gonadotrophins is associated with decreased endometrial receptivity (McKiernan and Bavister, 1998, Paulson et al., 1997). It will be of a great interest to artificially induce *in vitro*-fertilized embryos to enter into a diapause-like state until conditions become more favorable for implantation.

Embryonic diapause as a model of stem cell quiescence

Stem cells strike a balance between the competing interests of their two defining properties: self-renewal and multilineage differentiation. Terminally differentiated cells of most tissues in multicellular organisms are usually short-lived and require perpetual replenishment from tissue specific stem cell pools, both for daily homeostasis and for tissue repair. A growing number of

studies in the last decade report that there exists a subpopulation of stem cells that cycle infrequently even in tissues with high turnover rates. These stem cells are categorized as quiescent or dormant stem cells and may be reserved for emergency situations such as injury or stress (Fuchs, 2009, Wilson et al., 2008).

In adult tissues, quiescent stem cells are of inarguable importance. In the hematopoietic system, more than 70% of long term-hematopoietic stem cells (LT-HSC), accounting for ~1-2% of the entire stem cell pool, reside in the G0 phase of the cell cycle (Cheshier et al., 1999, Foudi et al., 2009, Passegue et al., 2005, Wilson et al., 2008). The dormant hematopoietic stem cells (HSC) are efficiently activated to self-renew in response to bone marrow injury or granulocyte colony-stimulating factor (G-SCF) stimulation. After homeostasis is re-established, the activated HSC can return to dormancy, suggesting the capacity of a reversible switch from dormancy to self-renewal under conditions of hematopoietic stress (Wilson et al., 2008). Another type of quiescent stem cell is muscle-specific, the satellite cell. During postnatal development, satellite cells divide to provide new myonuclei to growing muscle fibers (Moss and Leblond, 1971), and subsequently transition into an undifferentiated, quiescent state in adult skeletal muscle (Schultz et al., 1978). In the event of muscle damage, satellite cells exit from quiescence and proliferate as myoblasts. Myoblasts then differentiate and fuse with each other or to the injured myofibers to regenerate mature myofibers (Charge and Rudnicki, 2004).

The quiescent state provides stem cells with several advantages (Fuchs, 2009, Wilson et al., 2008). First, by reducing excessive proliferation, oxidative stress and telomere shortening are minimized. These aspects of quiescence are important to preserve the self-renewal capacity of stem cells and to prevent exhaustion. Second, lowering the rate of DNA replication minimizes mutations due to spontaneous DNA replication errors. Meanwhile, a slower division rate also provides stem cells with more time to repair DNA damage. Overall, quiescence or dormancy protects the genomic integrity of stem cells and reduces the risk of cancer (Wang et al., 2010).

A large variety of tumors have a small population of cells in the quiescent state, called cancer stem cells (CSC), that are responsible for initiating and maintaining the tumor bulk. Quiescence may provide a means for cancer stem cells to evade many types of cytotoxic treatment (Li and Bhatia, 2011, Sang et al., 2010). Understanding the mechanisms that support stem cell quiescence may shed light on how cancer stem cells survive chemotherapy.

Despite the inarguable importance of quiescence in many aspects of cell biology, there is no consensus as to what the term means. In this study I define quiescence as a reversible state of slowed proliferation in a stem cell that retains the capacity for pluripotent differentiation. Attempts at a more precise definition, for example reversible entry into the G0 phase of the cell cycle, fail to encompass many examples of quiescence (Coller, 2011), as highlighted by the recent discovery in budding and fission yeast that quiescence can occur in all stages of the cell cycle (Laporte et al., 2011). Indeed, Coller and colleagues have argued for a quiescence-associated transcriptional profile, uncoupled from distinct phases of the cell cycle (Coller et al., 2006).

Previous studies reported that the ICM of diapause mouse blastocysts continues to divide, but the rate of division is greatly reduced (Sherman and Barlow, 1972). Embryos from this quiescent state are still able to differentiate into all three embryonic germ layers and generate chimeric mice, showing they retain their pluripotency (Brook and Gardner, 1997, Evans and Kaufman, 1981). I characterized the transcriptional profile of pluripotent stem cells from diapause mouse embryos by RNA-Sequencing (RNA-Seq) as a means for interrogating this state of stem cell quiescence. I also extracted the pluripotent stem cells from pre-implantation and post-implantation mouse embryos from the same genetic background as experimental controls. The study of stem cell quiescence using pluripotent stem cells from embryos is relatively new and will enrich the existing knowledge about stem cell quiescence that came from studies of dormant multipotent stem cells in adult tissues.

Figures

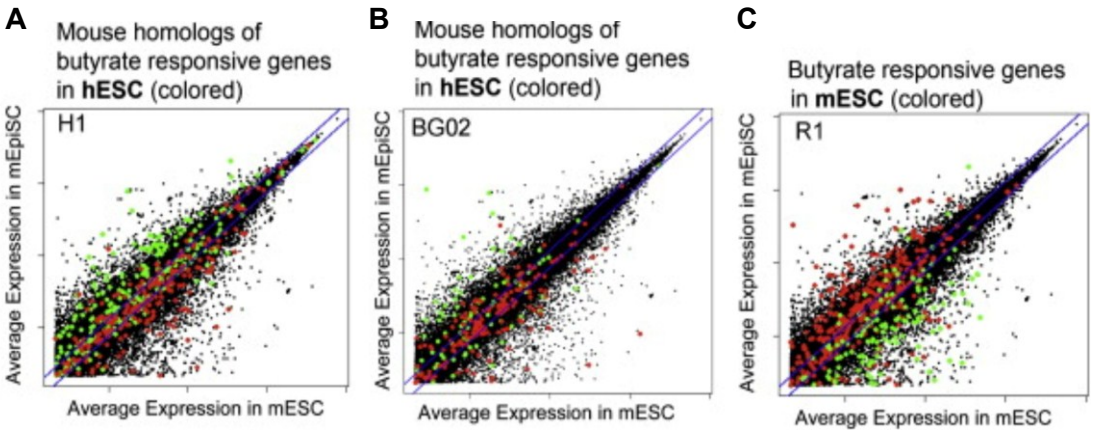


Figure 1.1 (adapted from Ware et al., 2009, Figure 3). Butyrate induces transcriptional response in human and mouse ESC towards a developmental stage intermediate between mESC and hESC/mEpiSC. Scatter plots (*black*) represent the average expression levels of genes in mouse EpiSC (y-axis) versus mESC (x-axis) (from Tesar et al., 2007; identical for all three panels). Colored dots indicate genes that were significantly upregulated (*red*) or downregulated (*green*) in response to butyrate. (A) Butyrate-responsive homologous genes in H1 hESC line. (B) Butyrate-responsive homologous genes in BG02 cells. (C) Butyrate-responsive genes in R1 mESC. Note that butyrate pulls the gene expression profile of hESC toward mESC (x-axis) and away from mEpiSC (y-axis) while pushing mESC toward EpiSC; thus, the relative orientation of red and green dots between (A) and (B) versus (C) is reversed.

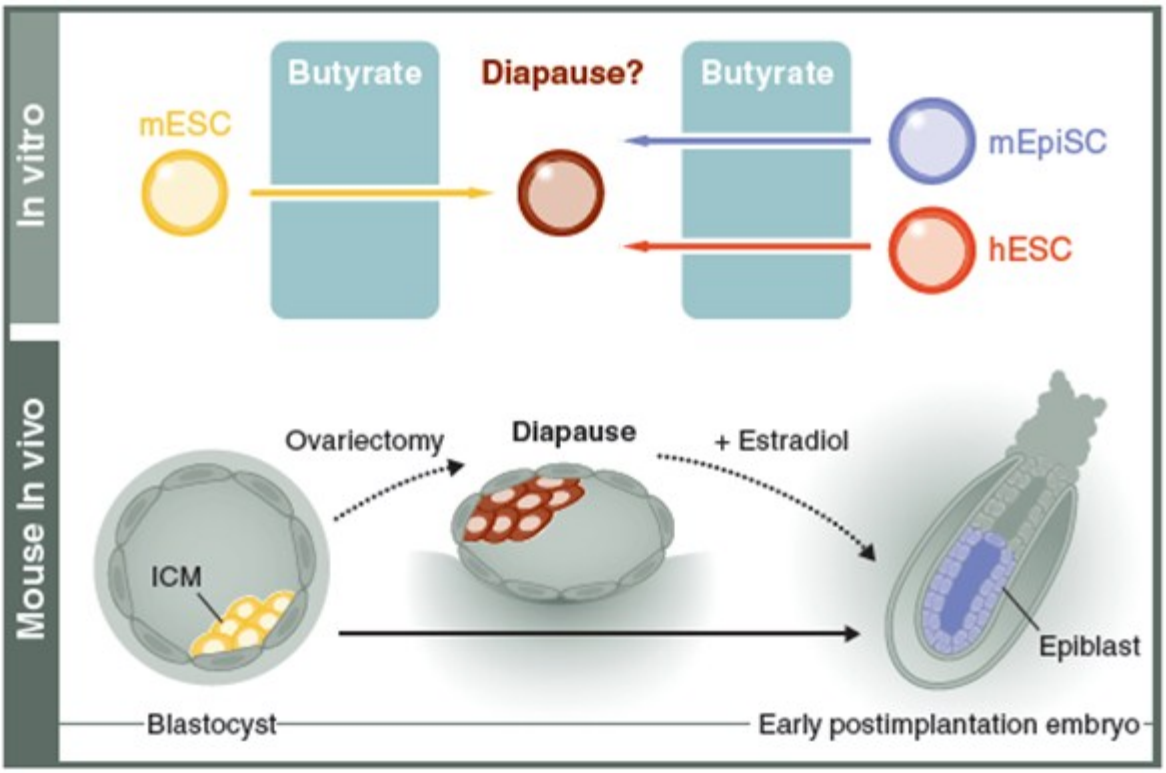


Figure 1.2. Schematic depiction of the hypothesis that butyrate (HDACi) exposure induces pluripotent stem cells to enter a state that mimics a physiological phenomenon known as diapause. Summary of *in vitro* findings from Ware et al., 2009 (top panel) shows that butyrate exposure induces mESC to move forward and hESC and mEpiSC to move backward to a developmental stage intermediate between mESC and hESC/mEpiSC. Lower panel illustrates the hypothesis that butyrate exposure mimics embryonic diapause, which is known to occur in mice. I used RNA-Sequencing (RNA-Seq) to characterize the transcriptional profiles of pluripotent stem cells isolated from pre-implantation (day 3.5), post-implantation (day 6.5) and diapause embryos.

Chapter 2 : A Developmental Yardstick for Pluripotent Stem Cells

Summary

A burgeoning menu of options exists for deriving and maintaining pluripotent stem cells in culture, however there is currently no way to place a given pluripotent stem cell line into a broader developmental context. I generated such a resource by characterizing the transcriptional programs of pluripotent stem cells from: i) the inner cell mass (ICM) of pre-implantation embryos; ii) the epiblast of post-implantation embryos; and iii) the ICM of embryos induced to enter a state of dormancy called diapause. My results demonstrate significant differences in the molecular signatures of these states and provide a "developmental yardstick" for pluripotent stem cells. With this yardstick I identified a novel pluripotent stem cell state that is most closely aligned with diapause. My results also provide a resource for dissecting mechanistic differences underlying these distinct pluripotent states.

Introduction

Many of the properties of mouse ESC and EpiSC reflect their tissues of origin, suggesting that mouse ESC, EpiSC and (by inference) human ESC might provide surrogates for interrogating these structures. Indeed, microarray profiling shows significant overlap between the transcriptional programs of mouse ESC and stem cells from the ICM, while mouse EpiSC substantially overlap with stem cells from the post-implantation epiblast (Epi) (Brons et al., 2007). However the tissues from which pluripotent stem cells originate do not strictly preordain their properties in culture, since adjustments in culture conditions can convert mouse ESC into EpiSC (Guo et al., 2009) and can revert mouse EpiSC back to ESC (Bao et al., 2009, Guo et al., 2009, Hanna et al., 2009, Zhou et al., 2010).

A previous study reported that sodium butyrate, a histone deacetylase inhibitor (HDACi), supports the self-renewal of both mouse and human ESC (Ware et al., 2009). Although butyrate affects gene expression globally, there was a minimal overlap in butyrate-regulated genes between mouse and human ESC. A comparison of butyrate-regulated genes with the published transcriptional profiles of mouse ESC and EpiSC revealed that human ESC exposed to butyrate induce genes associated with mouse ESC and repress genes associated with

EpiSC; i.e., butyrate pulls human ESC a step backward in development. Conversely, butyrate nudges mouse ESC forward, inducing a gene expression profile that more closely resembles EpiSC. Collectively, butyrate appeared to promote a convergence of mouse and human ESC toward an "intermediate" stem cell state. In addition to promoting stem cell self-renewal across different species and pluripotent states, two other features suggested physiological relevance – that butyrate's effect occurs only at a low and narrow range of concentrations (0.2 – 0.3 mM), and butyrate exposure reduces BrdU incorporation into DNA. Based on these observations I hypothesized that these responses to butyrate may be related to a physiological state of stem cell dormancy known as diapause. In diapause, the pre-implantation embryo suspends development for periods of days to weeks in response to environmental stresses such as starvation, lactation or cold (Lopes et al., 2004, Renfree and Shaw, 2000). Resolution of the stress allows implantation (and development) to proceed, thereby coordinating improved environmental conditions with the delivery of offspring.

To test whether butyrate elicits a diapause-like state in ESC, I used RNA sequencing (RNA-Seq) to characterize the transcriptional profiles of pluripotent stem cells isolated from pre-implantation (day 3.5), post-implantation (day 6.5) and diapause (day 8.5, which consists of day 2.5 + 6 days in diapause) embryos. The differences in transcriptional profiles between these developmental stages were plotted using principal component analysis, and the results were compared with previously published *in vivo* data and microarray profiles of ESC cultured in naïve conditions (using LIF plus inhibitors of GSK3β and ERK signaling), EpiSC cultured in primed conditions (FGF2 plus Activin) and ESC cultured in HDACi. My results define a novel, quiescence-associated pluripotent state, supporting the conclusion that butyrate and other select HDACi induce a diapause-like state in mouse ESC. My data provide a resource for defining distinct states of pluripotency and for connecting pluripotent stem cells in culture to their developmental corollaries *in vivo*.

Results

RNA-Seq of pluripotent stem cells isolated from embryos

I isolated the inner cell mass (ICM) from day 3.5 pre-implantation embryos, the epiblast (Epi) of day 6.5 post-implantation embryos, and the ICM of day 8.5 diapause embryos induced to delay

implantation via estrogen blockade (dICM) on day 2.5. For each collection, isolates from ~50 embryos (totaling ~1000 cells) were aggregated. For ICM and diapause ICM, each timepoint was examined in independent triplicate, while Epi was examined in duplicate. Reverse transcription, cDNA amplification, library construction and RNA-Seq were performed as described in Experimental Procedures and shown schematically in Figure 2.1. Raw reads were mapped to the mouse genome and a library of splice junction sequences. After filtering for rRNA, a total of 33,094,892 uniquely mapped reads were used for subsequent analysis (Table 2.1). Results can be downloaded at the link provided in the Experimental Procedures. I established the reliability of my assay by assessing: (1) RNA quality and cDNA amplification; (2) trophectoderm contamination; and (3) differentiation markers (Figure 2.2).

To contrast the transcriptional differences between the three developmental stages, a Principal Component Analysis (PCA) was performed. PCA is a transformation that reduces the dimensionality of the data while retaining the characteristics that contribute maximally to the overall variance. This approach was used to build basis vectors (eigengenes) that represent the molecular variation across the three developmental stages. The first two eigenarrays describe 65% of the total variation in my RNA-Seq data and clearly separate the three developmental stages (Figure 2.3A) – each open circle represents one of my 8 samples (3 ICM, 3 dICM and 2 Epi). The general pattern revealed by this representation of the data suggests that these stages are distinct molecular states. Hereafter, I refer to eigenarrays obtained from PCA of my RNA-Seq data as developmental basis vectors.

Estimating the contribution of primitive endoderm to the diapause signature

During uninterrupted mouse development implantation occurs after the late blastocyst (day 4.5) hatches from the zona pellucida. Since my ICM samples were collected from early (day 3.5) blastocysts, the transcriptional differences between my ICM and dICM samples reflect both changes attributable to diapause and those that accompany the transition from early to late blastocyst. A prominent morphological change that occurs during the interval between days 3.5 and 4.5 is formation of the primitive endoderm (PE). In the day 3.5 ICM, cells destined to form the PE intermingle with cells fated to form the epiblast in a "salt and pepper" distribution (Chazaud et al., 2006, Rossant et al., 2003); by day 4.5 PE cells have migrated to form a discrete epithelium lining the blastocoel cavity (Yamanaka et al., 2006). My list of dICM-

associated transcripts includes several encoded by genes expressed in the PE, including *Gata4* and *Dab2* (Gerbe et al., 2008). To gauge the extent to which PE-associated genes account for the transcriptional differences between my ICM and dICM samples, I examined genes preferentially expressed in PE. A previous report used microarrays to compare the transcriptional profiles of a cell line (eXtraembryonic ENdoderm [XEN]) derived from the PE layer of a late blastocyst with pre-implantation epiblast derived ESC (Kunath et al., 2005). I found that relatively few of the differences between ICM and dICM are attributable to XEN-associated genes and, after subtracting 480 genes reported to be differentially expressed in XEN cells, the differences between ICM and dICM in my PCA were retained (data not shown). These results suggest that pluripotent cells within the diapause embryo occupy a distinct transcriptional state.

Comparisons to Prior Knowledge

To evaluate my results in the context of prior knowledge, a previously published projection methodology was utilized to compare the three developmental stages measured in my RNA-Seq study with data from two previously described high-throughput molecular profiling studies (Kho et al., 2004). The first of these studies used mRNA microarrays to profile pluripotent stem cells from the pre-implantation, peri-implantation and diapause ICM and the post-implantation epiblast (Hunter et al., 2008; Table 2.2), while the second study analyzed the transcriptional profiles of whole (unfractionated) diapause blastocysts in dormancy and activated for implantation (Hamatani et al., 2004). After filtering the microarray and RNA-Seq data to identify genes highly varying in both experiments, new developmental basis vectors were calculated. The microarray data were then projected into the resulting space (Figure 2.3B). Each point corresponds to a single sample, and points with the same symbol and color depict experimental replicates.

Results allow for three conclusions. First, Figure 2.3B shows that my RNA-Seq samples still partition into three distinct regions in the context of the new developmental basis vectors. This suggests the filtering procedure did not disturb the major patterns that distinguish these groups. Second, the replicate samples from each of the 6 conditions explored in the Hunter study cluster tightly with one another in this space, indicating a high-degree of similarity in gene expression variation across both data sets. Third, the relationships between samples from my RNA-Seq

data and the Hunter study closely reflect the expected relationships based on their actual age. The early pre-implantation ICM samples from Hunter and my RNA-Seq data (days 3.7 and 3.5, respectively) occupy the left upper quadrant; the diapause samples from Hunter (days 5.7 and 7.5) and from my RNA-Seq study (day 8.5) localize to the left lower quadrant, while all of the post-implantation epiblast samples from Hunter (day 5.5 and day 6.5) and from my dataset (also day 6.5) appear on the far right side of the graph. These findings clearly distinguish these three developmental stages and support the conclusion that my RNA-Seq results accurately reflect global changes in gene expression across early stem cell development.

This analysis points to the existence of a third pluripotent state, associated with quiescence, that is distinct from the naïve and primed states. It is interesting to consider whether this quiescence-associated state represents a detour from development, or is more accurately regarded as a checkpoint through which developing embryos must pass. In addressing this question one of the timepoints assessed in the Hunter study (day 4.4) may be informative, since these samples are from the period immediately prior to implantation. If diapause represents a developmental detour, datapoints from the day 4.4 samples would be expected to extend diagonally from the left upper quadrant (ICM) to the right lower quadrant (Epi). Instead, Figure 2.3B shows that the transcriptional profiles of the day 4.4 samples extend from the ICM to the diapause quadrants, suggesting that a transient diapause-like state accompanies normal embryonic development.

I also analyzed my results in the context of a previous transcriptional comparison of embryos induced to enter diapause (by ovariectomy) or re-activated through the provision of estrogen (Hamatani et al., 2004). Unlike my study and the Hunter study, the Hamatani study analyzed whole (unfractionated) embryos. A PCA of genes that were differentially regulated both in my study and the Hamatani study confirmed that the transcriptional profiles of diapause embryos align to the left lower (diapause) quadrant, while activated embryos align to a distinct location within the right upper quadrant (Figure 2.3C).

Next I examined how ESC markers varied across my RNA-Seq samples. As described in the Methods, a list of 1076 genes that were expressed at significantly higher levels in undifferentiated ESC than in other cell types was assembled (see Table S2.1). The average expression levels for these ESC markers and 5761 non-ESC-associated genes were calculated

across all 8 RNA-Seq samples, and their frequency distributions were plotted. The joint distribution of the average expression of ESC markers is significantly higher than for all other genes, indicating that my RNA-Seq data captures molecular variation related to the stem cell paradigm (Figure 2.3D).

Comparisons to Cultured Pluripotent Stem Cell Lines

Previous studies have shown that ESC maintain a gene expression profile that is similar to the ICM, whereas EpiSC have a transcriptional profile that more closely aligns with the epiblast (Brons et al., 2007). To test whether these relationships are evident in my dataset, I collaborated with the lab of Dr. Carol Ware who performed Agilent RNA microarrays on quadruplicate samples of mouse ESC cultured under naïve conditions (2iL) and mouse EpiSC cultured in primed conditions (Activin plus FGF2). They also performed microarrays on mouse ESC cultured in HDACi. As mentioned above, ESC are exquisitely sensitive to even subtle changes in butyrate concentration, and to enhance culture stability butyrate was combined with a second HDACi (vorinostat - SAHA) (2-HDACi- see Methods). To compare results from these cultured cells with my RNA-Seq data, the intersection of genes that varied among conditions within each of the two datasets was used. PCA was performed on the filtered RNA-Seq data set, and the microarray data from the cultured cells were then projected onto the updated eigengenes (Figure 2.4A). The RNA-Seq samples again clustered tightly within conditions, suggesting that the filtering procedure did not disturb the major patterns described above. I also observe tight clustering of the replicate samples from mouse ESC or EpiSC, suggesting that the variation in gene expression across the culture conditions is similar to the variation across early stem cell development.

Mouse ESC cultured in 2iL align in the same left upper quadrant as ICM, while mouse EpiSC align to the same right lower quadrant as the Epiblast. Similar results were obtained when comparing my RNA-Seq data with previously published microarrays of mouse ESC and EpiSC (Tesar et al. 2007) (Figure 2.4B). These differences were also readily discernible at the level of individual genes (Figure 2.4C). Eight of the 10 genes that are most differentially expressed in ICM relative to epiblast are also expressed at significantly higher levels in ESC cultured in 2iL compared to EpiSC, while 9 out of the 10 genes that are most differentially expressed in epiblast relative to ICM are also expressed at significantly higher levels in mouse EpiSC

compared to ESC (p-value < 0.01). These findings confirm that differences between the transcriptional profiles of ESC and EpiSC cultured in their respective basal conditions correspond to differences that exist between ICM and epiblast *in vivo*.

I next tested whether the stem cell state that is induced by 2-HDACi is transcriptionally similar to diapause. Indeed, ESC cultured in 2-HDACi localized to the same quadrant as the diapause ICM (Figure 2.4A), demonstrating that 2-HDACi exposure elicits a transcriptional program that aligns with diapause.

Forming a Developmental Yardstick

The PCA comparison of my RNA-Seq results with the Hunter dataset (Figure 2.3B) demonstrates how samples of similar ages distribute in a non-random fashion on the projected space. This two-dimensional plot suggests that development proceeds counterclockwise, starting from the left upper quadrant occupied by ICM and ending at the far right side of the graph where the epiblast samples are located. To find the developmental ages to which ESC cultured in various conditions most closely correspond, the two-dimensional projections of my RNA-Seq results were transformed into a linear yardstick as described in the Methods section. Figure 2.5 shows the estimated developmental ages of the samples described above relative to the pre-implantation ICM. The result is consistent with the projections depicted in Figures 2.3B and 2.4A, whereby the predicted age of a given sample correlates well with its true developmental age.

Discussion

Embryonic diapause as a model of stem cell quiescence

A large body of evidence supports the existence of a relationship between a stem cell's ability to self-renew and preservation of its capacity for quiescence (Gan et al., 2010, Gurumurthy et al., 2010, Nakada et al., 2010, Qian et al., 2007, Yilmaz et al., 2006, Yoshihara et al., 2007, Zhang et al., 2006). In the context of malignancy, quiescence may provide a means for cancer stem cells to evade many types of cytotoxic treatment (Li and Bhatia, 2011, Sang et al., 2010). Understanding the mechanisms that support stem cell quiescence will provide insight into

developing methods for modulating quiescence in cultured stem cells and can shed light on how cancer stem cells survive chemotherapy.

My RNA-Seq characterization of the pluripotent stem cells from pre-implantation, post-implantation and diapause embryos provides a unique resource for evaluating the molecular differences between these distinct pluripotent states and for investigating the molecular basis of diapause specifically and quiescence generally.

A developmental yardstick as a gold standard for evaluation of pluripotent stem cell lines

Previous studies have defined two major pluripotent stem cell states: a naïve state, characteristic of mouse ESC, and a primed state, characteristic of mouse EpiSC and human ESC (Hanna et al., 2010). In this study I have further shown that there exists a third stem cell state, which I term the **quiescence-associated** state, characteristic of pluripotent cells from the diapause ICM, and mimicked in pluripotent stem cells cultured in 2-HDACi. How these *in vitro* defined states relate to pluripotent states *in vivo* is unclear, partly because pluripotent stem cells can be induced to transition between metastable naïve and primed states by adjustments in culture conditions (Bao et al., 2009, Guo et al., 2009, Hanna et al., 2009, Zhou et al., 2010). Therefore, a pressing problem in the evaluation of pluripotent stem cell lines is the lack of a gold standard.

I used RNA-Seq of stem cells isolated from ICM, dICM and Epi to provide a gold standard against which cultured pluripotent stem cells can be compared. My resulting yardstick provides a framework for comparing pluripotent stem cell lines that differ in tissue source, method of derivation, or culture conditions. Comparing results from different platforms (RNA-Seq versus microarrays) can be challenging due to technical variation between platforms that frequently overwhelms biological signal. I established the yardstick's validity by demonstrating its agreement with previously published microarrays of pluripotent stem cells extracted from mouse embryos (Figure 2.3). Furthermore, the transcriptional programs of the pre-implantation ICM and post-implantation epiblast align to the mouse ESC and EpiSC stages, respectively. Therefore my study demonstrates some degree of coherence of variations in gene expression between the distinct developmental stages across data sets. As shown by my finding that

mouse ESC exposed to 2-HDACi elicits a diapause-like transcriptional response, my developmental yardstick may be used to append cultured pluripotent stem cell lines to their developmental corollaries *in vivo*.

Previous studies reported that the ICM of diapause mouse blastocysts continue to divide, but its rate is greatly reduced (Sherman and Barlow, 1972). Embryos from this quiescent state are still able to differentiate into all three embryonic germ layers and generate chimeric mice, showing they retain their pluripotency (Brook and Gardner, 1997, Evans and Kaufman, 1981). My findings corroborate this existing knowledge about diapause. Pluripotency genes such as *Oct4*, *Nanog* and *Sox2* are expressed at comparable levels in ICM and dICM. The Gene Ontology (GO) enrichment shows that the genes downregulated in dICM are enriched for cell cycle processes, suggesting that stem cells in diapause proliferate more slowly. In particular, p21 (*Cdkn1a*), which mainly inhibits the activity of cyclin/cdk2 complexes and negatively modulates cell cycle progression (Brugarolas et al., 1999), is upregulated in dICM by at least 2 fold (p-value < 0.05) compared to ICM and Epi; *Brca1*, known to be responsive to estrogen and downregulated in whole dormant blastocysts, is repressed in dICM by ~3-fold (p-value < 0.01) compared to ICM and Epi. *Brca1*-/- embryos were also previously shown to have increased expression of p21 (Hakem et al, 1996). Collectively, my RNA-Seq results of diapause ICM representing stem cells in a quiescence mode are consistent with previous study using whole diapause blastocysts (Hamatani et al., 2004). Furthermore, as shown by the PCA plot of my RNA-Seq data, I present the quiescent state as the third pluripotent stem cell state that is distinct from the other two states: naïve and primed.

Is diapause a pit stop or detour?

A fundamental question about diapause is whether it occurs naturally as part of the developmental continuum. I propose two models of the quiescent state in mammalian embryonic development: 1) diapause as a developmental detour, or 2) diapause as a developmental pit stop. In the “detour” model, diapause is an **alternative** state, adopted by the pre-implantation embryos only when they encounter challenging circumstances. Embryos can return to the normal developmental pathway and resume implantation once stress is resolved. In the “pit stop” model, diapause is an **obligatory** state whose length is variable depending on the circumstances during the peri-implantation window. The PCA results of RNA-Seq of

pluripotent stem cells together with the Hunter's microarray data (Figure 2.3B) may give an insight into which diapause model is adopted by nature. With the first model where diapause is viewed as a branch-off of the developmental continuum, the left lower quadrant occupied by dICM samples may be skipped completely in normal development. Alternatively, with the second model whereby diapause is an obligatory state, the dICM quadrant is a transition phase that all embryos have to pass before implantation. The day 4.4 peri-implantation ICM samples from the Hunter study, which cross the boundary separating the ICM and diapause quadrants, suggest the latter. Previous studies have shown that intrauterine oxygen tension and glycogen storage are declining as the embryos develop, reaching the lowest level prior to implantation (Fischer and Bavister, 1999, Thomson and Brinster, 1966). It is possible that all peri-implantation embryos have to undergo the process of diapause and that additional stress plays a role in lengthening this otherwise brief snapshot of the developmental stage. Thus the period immediately preceding implantation may represent a metabolically responsive developmental checkpoint.

Mouse ESC exposed to 2-HDACi as *in vitro* model of stem cell quiescence

A major challenge to the investigation of stem cell quiescence is the trace amount of materials available for analysis. The availability of a cell culture system that mimics diapause would therefore provide a useful resource for exploring the molecular mechanisms that support this novel state of stem cell dormancy.

A previous study reported that mouse and human ESC could be induced to converge toward a state intermediate between naïve and primed through the addition of butyrate and other histone deacetylase inhibitors (Ware et al., 2009). I hypothesized that this "intermediate" stem cell state represented an *in vitro* reflection of diapause. Interestingly, ESC cultured in 2-HDACi adopts a gene expression profile that most closely aligns with diapause ICM. In keeping with my prediction of an *in vitro* diapause-like effect, butyrate exposure reduces BrdU uptake in mouse and human ESC (Ware et al., 2009). The PCA analysis of my RNA-Seq result and microarray study of mESC treated with 2-HDACi suggests that these cells align with and may be used as a model for stem cells in quiescence. My findings also point to the potential use of cultured pluripotent stem cells as a tool for exploring potential mechanisms of quiescence *in vitro*.

Methods

Isolation of pluripotent stem cells

All embryos were recovered from C57BL/6 females (Charles River). Noon on the day of vaginal plug detection was considered as day 0.5. Day 3.5 and diapause blastocysts were flushed from the uterus of superovulated pregnant females. Embryonic diapause was induced by intraperitoneal injection of tamoxifen and subcutaneous administration of 0.5 mg of Depo provera on day 2.5 ("chemical ovariectomy", Hunter and Evans, 1999). Diapause blastocysts were harvested 6 days later.

For the isolation of ICM, the zona pellucida was first removed from day 3.5 embryos by incubation in acid tyrode's solution. Blastocysts were placed in a rabbit anti-mouse polyclonal antibody (Rockland Immunochemicals) for 20 min at 37^{o}C followed by guinea pig serum complement for 20-30 min at 37^{o}C. The lysed trophectoderm (TE) cells were removed and the isolated ICM was placed in lysis buffer. The derivation of epiblast (Epi) from day 6.5 post-implantation embryos has been described previously (Brons et al., 2007). Biological triplicates for ICM and dICM samples and duplicates for Epi samples comprise the lysate from approximately 50 embryos.

RNA-Seq Library Preparation, Sequencing, and Alignment

Total RNA was extracted using QIAGEN Micro RNeasy kit and amplified following protocols of WT-Ovation Pico kit (Nugen). The quality and quantity of RNA were determined on an Agilent 2100 Bioanalyzer (Agilent Technologies). cDNA amplification products were made into double-stranded cDNA using WT-Ovation Exon Module kit (Nugen). 3 µg of cDNA products were sheared by sonication (Bioruptor UCD-200) for 2 x 15 min at 4^{o}C, H speed (320 W). cDNA fragments were then blunt-ended with End-It DNA End-Repair kit (Epicentre Biotechnologies) and A-tail was added to the end of the fragments. Adapters (33-34 nucleotides) were ligated to each end. The fragments were subjected to size selection by gel electrophoresis to isolate 200-300 bp fragments, which were further enriched by 18 cycles of PCR amplification. Cluster generation of the cDNA library was performed on a cBot Cluster Station (Illumina, Inc.) and the

samples were sequenced on Genome Analyzer IIx (Illumina, Inc.). Reads were aligned using Bowtie (version 0.12.0, Langmead et al., 2009) to the mouse genome (version mm9), obtained from the UCSC genome browser (Fujita et al., 2010). To account for reads crossing splice junctions, the remaining reads were aligned to a set of sequences taken from upstream of the donor site and downstream of the acceptor site of every splice junction annotated in version 58 of the Ensembl gene annotations (Hubbard et al., 2009). Only reads that were aligned uniquely to the genome were used in the subsequent analysis. Reads that were mapped to rRNA sequence (5S, 5.8S, 18S, 28S, mitochondrial 12S and 16S rRNA, obtained from NCBI website) were removed. Additional precaution of filtering out any reads overlapping low complexity regions or simple repeats as defined by RepeatMasker (Smit et al., 1996-2010; http://www.repeatmasker.org) was made. After summarizing the expression of each gene by the number of reads aligned unambiguously within any of its exons, the DESeq package was used to perform differential expression analysis (Anders and Huber, 2010). Transcripts expressed in ICM, dICM or Epi with >1.4-fold change and a p-value < 0.05 (adjusted for multiple testing with the Benjamini-Hochberg procedure) were considered to be differentially expressed. The UCSC genome browser tracks of RNA-Seq reads are available at http://genome.ucsc.edu/cgi-bin/hgTracks?db=mm9&position=chr7:54075001-54115000&hgct_customText=http://www.cs.washington.edu/homes/ruzzo/papers/Margaretha/data/tracks.txt.

Mouse pluripotent stem cell culture

R1 mouse ESC were plated without feeders on Matrigel (BD Biosciences) diluted according to manufacturer's instructions. Basal Mouse ESC culture medium consisted of DMEM-containing GlutaMax, 20% FBS (ES Cell Qualified, Invitrogen), 1 mM sodium pyruvate, 0.1 mM nonessential amino acids, 50 U/mL penicillin, 50 mg/mL streptomycin (all from Invitrogen, Carlsbad, CA), and 0.1 mM β-mercaptoethanol (Sigma). For the 2iL condition, mESC media was supplemented with 1000 units/mL mouse LIF (ESGRO, Millipore) plus 3 μM GSK3 inhibitor (CHIR99021, Selleck Chemicals) and 0.4 μM MEK inhibitor (PD0325901, Stemgent). For the 2-HDACi condition, the media was supplemented with 0.1 mM sodium butyrate (Sigma) and 25 nM vorinostat/SAHA (Cayman Chemical). Mouse ESC were passaged by washing once with PBS without calcium and magnesium (Invitrogen), followed by exposure to trypsin-EDTA (Invitrogen) until the cells were digested to single cells.

Mouse EpiSC line no. 5 (a gift from Paul Tesar and Ron McKay) was cultured as described (Tesar et al., 2007). Briefly, mouse EpiSC medium consisted of DMEM-F12-containing Glutamax (Invitrogen), 20% knockout serum replacement (KSR), 5 ng/mL FGF2 (R&D Systems), 10 ng/mL Activin A (Humanzyme), 0.1 mM β-mercaptoethanol (Sigma), 1 mM sodium pyruvate, 0.1 mM nonessential amino acids, 50 U/ml penicillin, 50 mg/ml streptomycin. Mouse EpiSC were passaged in clusters using dispase rather than as a single-cell suspension.

Microarray analysis of mouse pluripotent stem cell culture

Agilent Whole Mouse Genome Oligo Microarrays containing 41,000+ spots were hybridized with total RNA from cells cultured in each of the following conditions in independent quadruplicate: A) R1 cells grown with 2iL supplementation; B) R1 cells grown with 2-HDACi supplementation; C) mEpiSC grown with Activin plus FGF supplementation. Data used for the microarray analysis were defined as red and green processed signals from the Agilent output files. Annotation for the array data was provided by the Bioconductor annotation packages for the hgug4112a and mgug4112a platforms.

Data normalization and filtering

Any intensity-dependent biases due to array and dye effects from the microarray data were removed using the *snm* function in the supervised normalization of microarrays process (Mecham et al., 2010). Genes were matched across platforms using the Gene Symbol. Annotation for the RNA-Seq data was defined by getGeneExp function in DEGseq Bioconductor package. The RNA-Seq data were filtered to remove low expressors, which were defined as any gene whose maximum count across the RNA-Seq experiment was less than 4. A total of 6837 genes in our RNA-Seq data passed through this filtering process. For the genes that had multiple probes in microarray data, the probe with the maximal difference in expression values between conditions was selected for further analysis. Detailed filtering criteria for various datasets used in this chapter are available in Supplemental Methods below.

Data Projections

Principle Component Analysis of RNA-Seq data was done by Singular Value Decomposition (SVD) approach to form basis vectors. To compare various datasets, the microarray data were pre-processed using the criteria as described in the Supplemental Methods section. After filtering the microarray and RNA-Seq data to identify genes highly varying in both experiments, new developmental basis vectors were calculated. Microarray data were then projected onto these basis vectors following the methods initially described by Kho et al. (2004).

Stem cell markers in RNA-Seq data

A list of genes from 20 studies involving differentiation of ESC was compiled. The information about the studies used here is available in Table S2.1. The average ratio of gene expression levels in undifferentiated versus differentiated ESC was determined for each study. Stem cell markers were defined as genes that were expressed at least 2-fold higher in undifferentiated ESC than in other cell types in at least 3 out of 20 studies examined. The average expression levels of 1076 stem cell markers and other 5761 non-ESC-associated genes were calculated across all 8 RNA-Seq samples. The frequency distributions of the average gene expression were plotted separately for stem cell markers and other non-ESC-associated genes.

Developmental yardstick

The RNA-Seq ICM samples were set as the beginning of my yardstick. The *x,y* coordinates of dICM and Epi samples were then used to calculate the angle formed by these two stages with respect to the ICM samples. The *x,y* values obtained from the basis projection were transformed into radians using atan2(y,x). The radians were then translated into degrees, with degree 0 set for ICM samples.

Placing the ICM, dICM and Epi samples on a single-dimensional scale, a linear yardstick was constructed to estimate the developmental age of samples of interest. A similar approach was employed to find the average angle formed between the samples from the microarray studies described above (Hunter study and cultured pluripotent stem cell array) and the ICM samples from my RNA-Seq data.

Quantitative real-time PCR

Total RNA was purified using the RNeasy Micro Kit (Qiagen) following the manufacturer-specified protocol. Reverse transcription of total RNA was performed using random hexamers with the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). qPCR was performed in triplicate using TaqMan Universal PCR Master Mix, No AmpErase UNG (Applied Biosystems) or SYBR Green PCR Master Mix (Applied Biosystems) in 10 µl reactions using an Applied Biosystems 7900HT Fast Real-Time PCR System set to the following conditions: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 sec and 60°C for 1 min. *β-actin* was used as endogenous control for normalization of target mRNA expression. SDS RQ Manager 1.2 Software (Applied Biosystems) was used for data analysis.

The following gene specific primer pairs were used in this study:

Genes	Forward sequence (5' to 3')	Reverse sequence (5' to 3')
Oct4	CGTTCTCTTTGGAAAGGTGTTC	GAACCATACTCGAACCACATCC
Sox2	GCGGAGTGGAAACTTTTGTCC	CGGGAAGCGTGTACTTATCCTT
Nanog	CTCAAGTCCTGAGGCTGACA	TGAAACCTGTCCTTGAGTGC
Bnc1	ATGGCTGAGGCTATCGGCT	AGGCTGCTAATATCGAACACCA
Eomes	CCGCCCACTACAATGTTTTCG	GAGAAGGTGAAGGTCTGAGTCTTGG
Hprt1	GCTGGTGAAAAGGACCTCT	CACAGGACTAGAACACCTGC
18s rRNA	AGGGGAGAGCGGGTAAGAGA	GGACAGGACTAGGCGGAACA
β-actin	GGCTGTATTCCCCTCCATCG	CCAGTTGGTAACAATGCCATGT

TaqMan Gene Expression Assays (Applied Biosystems) used in this study include the following: *Dppa3* Mm01184198_g1, *Zfp42/Rex1* Mm01194090_g1, *Gad1* Mm00725661_s1, *Dppa5a* Mm01171664_g1, and *Pgk1* Mm00435617_m1.

Supplemental Method

Detailed Analysis Description of Microarray Data Preprocessing

Each dataset had to be normalized and then filtered in order to make comparisons with the RNA-Seq study. The normalization strategies were based on the supervised normalization of microarrays (SNM) framework. The filtering strategies were designed to remove data within a

study that did not exhibit strong variation across the samples. The specific steps taken for each dataset were described below.

A. Hunter (GSE8881)

This dataset used a spotted cDNA microarray technology. The printing and immobilization of the slides were carried out by the Microarray Facility, School of Biosciences, Cardiff University. In general the microarrays used were of lower quality than the newer generation of microarrays provided by Affymetrix or Agilent. This was evident in the data itself as there were larger variances across probes in the reference channel than the experimental channel. This result suggests that there is more variation in intensity arising from the different aliquots of their uniform reference than in the samples from the different developmental stages. Low average correlation between probes aligning to the same gene (mean = 0.28) was also observed, further supporting the conclusion that this data is noisy and should be analyzed accordingly. To account for these issues, the data from the reference channel were ignored since they were not uniform. Data from the experimental probes were filtered based on differential expression. The model used to normalize the data simply removed intensity-dependent effects arising from the arrays. Probes with maximal difference in expression between any two conditions smaller than a two-fold change were removed. For genes measured by more than one probe, the probe with the largest variance was selected. The filtering strategy resulted in data for 1,759 genes that were in common with the RNA-Seq data. The annotation information for this study was obtained from GEO (GPL5771).

B. Hamatani (GSE1634)

This dataset used NIA Mouse 22K Microarray v1.0 (Development 60-mer Oligo). For this dataset, any control probes were first removed. The data were then normalized with a model that accounted for both intensity-dependent array and dye effects. Any probes, for which the corresponding gene symbol was either missing or set to NA, were removed. Probes with maximal difference in expression between the two conditions (i.e. dormant and activated) smaller than 1 and with the largest expression value smaller than 6 were removed. The latter decision was based on the observation that the differences between conditions were slightly biased at low intensity ranges. Note that this criteria were similar to those used in the original

paper (see their figure 2) and resulted in more genes than were identified as differentially expressed by the original paper. As there was not any major confounding problem in their reference samples as seen in GSE8881, the data used from this study for the projections consisted of the ratios between the experimental and reference condition. The filtering strategy resulted in data for 1,979 genes that were in common with the RNA-Seq data. The annotation information for this platform was obtained from NCBI GEO (GPL870).

C. Mouse ESC (2iL vs. 2-HDACi) and mouse EpiSC

This dataset used Agilent 4x44K Mouse (G4122F) whole genome microarrays. For this dataset, the control probes were first removed. The data were then normalized with a model that included the biological effects and terms for intensity-dependent array and dye effects. This dataset consisted of a fourth condition that was not included in this analysis but was included during the data normalization step to assure balance between dyes and conditions. I expect this to have little to no effect on the normalized values I obtained and it will certainly not bias the projection results. Probes with maximal difference in expression between any two samples greater than a two-fold change and with largest expression value smaller than 6 were removed. For genes measured by more than one probe, the probe with the largest variance was selected. The filtering strategy resulted in data for 4,323 genes that were in common with the RNA-Seq data. The annotation information for this dataset was available from Agilent.

D. Tesar (GSE7866)

This dataset used Agilent 4x44K Mouse (G4122F) whole genome microarrays. No reference channel was utilized in their experimental design. The data were normalized with a model that removes intensity-dependent array effects. Probes with maximal difference in expression between any two samples smaller than a two-fold change and with the largest expression value smaller than 5 were removed. For genes measured by more than one probe, the probe with the largest variance was selected. The filtering strategy resulted in data for 2,912 genes that were in common with the RNA-Seq data. The annotation information for this dataset was available from Agilent.

Projection Methodology

Let **R** denote the matrix of RNA-Seq data and **P** be the data matrix corresponding to the samples to be projected. Assume that the rows of **R** and **P** correspond to genes while the columns correspond to samples. The rows are ordered across **R** and **P** so that row *j* corresponds to the same gene in both matrices. The following is the algorithm used to project **P** onto **R**.

For both **R** and **P**, first center and scale their rows, then center and scale the columns. Let **Rs** and **Ps** correspond to the scaled matrices.

Take a SVD of $\mathbf{Rs} = \mathbf{U} * \mathbf{D} * \mathbf{V}^T$, where * denotes matrix multiplication.

Let $\mathbf{G} = (\mathbf{U}^T * \mathbf{Ps})^T / n$

The matrix **V** corresponds to the developmental basis vectors and **G** corresponds to the projected positions of the samples in **P**.

Relative Nature of Sample Projections

A key step in the projection methodology is centering data across probes and then samples. This step is necessary due to differences in probe dynamics between different microarray technologies and as a result of the massive differences between microarray and sequencing technologies. These transformation steps are necessary to make the variation across each probe normalized to the overall mean across samples. Therefore, the projection methodology compares relative expression changes across experiments. As a result of this step a single sample can land in different regions within the developmental basis vectors depending on which samples are included when mean centering probes. In other words, the position assumed by any projected sample is relative to the other samples included in the projection.

Figures

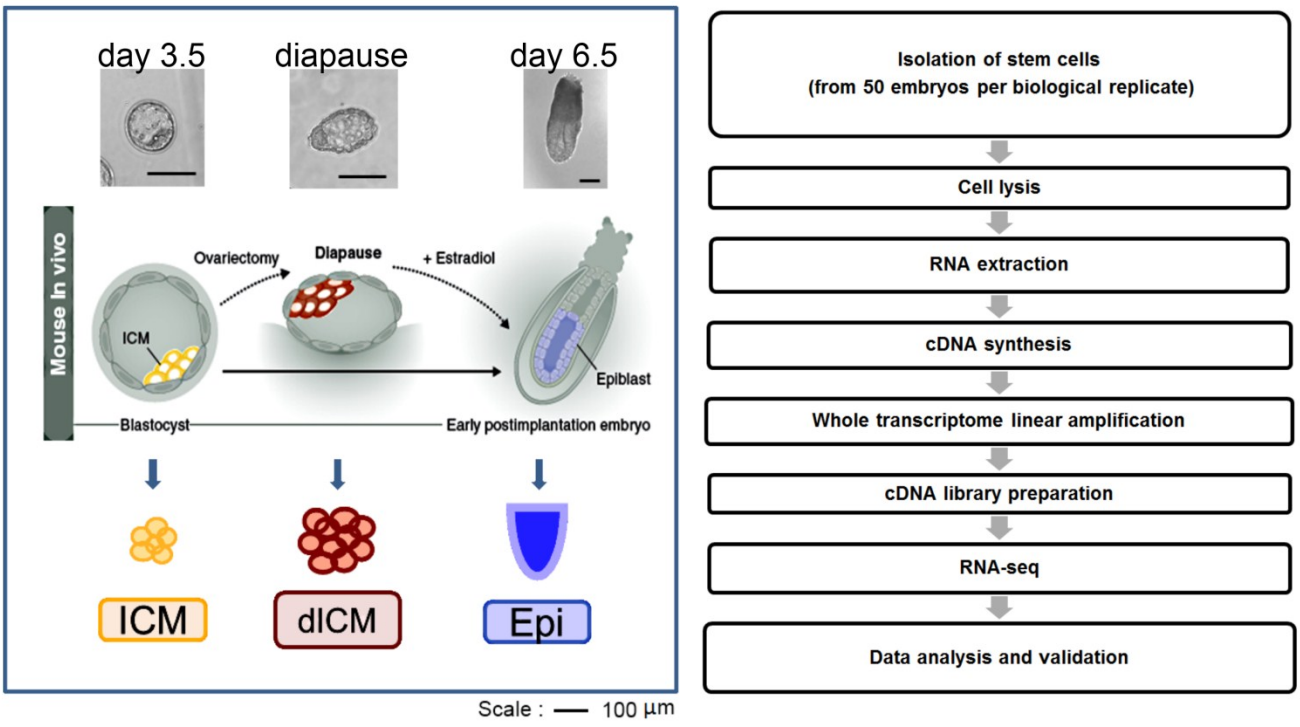
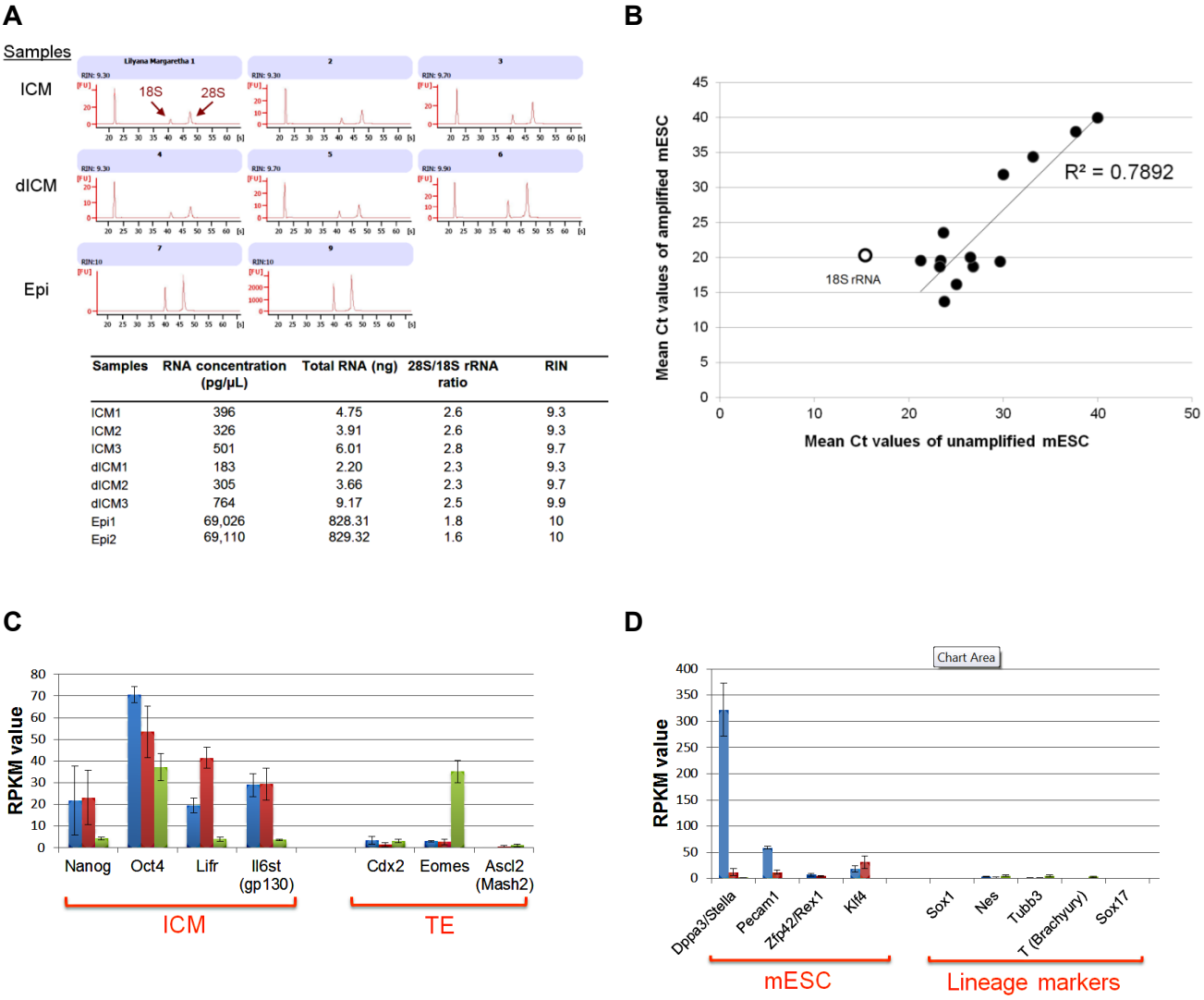


Figure 2.1. **Schematic diagram showing embryo morphology and experimental procedures.** Pluripotent stem cells from 3 different developmental stages were isolated and aggregated to make biological replicates from 50 embryos each. For day 3.5 ICM and diapause ICM, each timepoint was examined in independent triplicates; the day 6.5 Epi was examined in duplicates. Whole transcriptome linear amplification with minimal ribosomal RNA amplification was performed, followed by cDNA library preparation and RNA-sequencing using Illumina platform.



Samples	RNA concentration (pg/µL)	Total RNA (ng)	28S/18S rRNA ratio	RIN
ICM1	396	4.75	2.6	9.3
ICM2	326	3.91	2.6	9.3
ICM3	501	6.01	2.8	9.7
dICM1	183	2.20	2.3	9.3
dICM2	305	3.66	2.3	9.7
dICM3	764	9.17	2.5	9.9
Epi1	69,026	828.31	1.8	10
Epi2	69,110	829.32	1.6	10

Figure 2.2 Samples used for RNA-Seq from pluripotent stem cells isolated from mouse embryos were obtained from high quality RNA and contained minimal transcripts from ribosomal RNA as well as extraembryonic tissues. (A) RNA Quality Control was done on a pico chip with Agilent Bioanalyzer 2100. Two distinct ribosomal RNA (rRNA) peaks, corresponding to 18S and 28S rRNAs, were observed in all samples. No smaller peaks denoting degraded samples were found between the two rRNA peaks. All samples had RNA Integrity Number (RIN) greater than 9. The table below shows the concentration and the amount of total RNA, the ratio of 28S rRNA to 18S rRNA and RIN of each sample. (B) Whole transcriptome linear amplification minimized ribosomal RNA amplification. Quantitative real-time

PCR was done on cDNA samples from mESC that have been prepared either with or without amplification using WT-Ovation Pico system (NuGEN). Shown in the graph are the mean Ct values the following genes: *Oct4*, *Sox2*, *Nanog*, *Dppa3*, *Zfp42*, *Bnc1*, *Gad1*, *Inhba*, *Dppa5*, *Eomes*, *Hprt1*, *Pgk1*, *b-actin* (shown as filled circles) and *18s* rRNA (open circle). Mean Ct value of *18S* rRNA from amplified mESC showed a higher value than expected, suggesting that rRNA was not amplified as efficiently as the others. (C) Pluripotent stem cells from three developmental stages (ICM, *blue*; dICM, *red*; Epi, *green*) expressed pluripotency markers *Oct4* and *Nanog*. Other ICM markers normally expressed prior to implantation were expressed in both pre-implantation and diapause ICM as expected. Markers of trophectoderm (TE) were expressed at a very low level, indicating minimal extraembryonic TE contamination. Reads per kilobase of exon model per million mapped reads (RPKM) values were determined for the genes of interest as a way to quantify gene expression levels, normalizing the differences in total number of reads in each sample and gene lengths. (D) Some markers that were expressed higher in mESC than mEpiSC had higher RPKM values in pre-implantation ICM samples than those of Epi. Markers of lineage specification were expressed at a very low level, suggesting minimal contamination of differentiated cells in all samples.

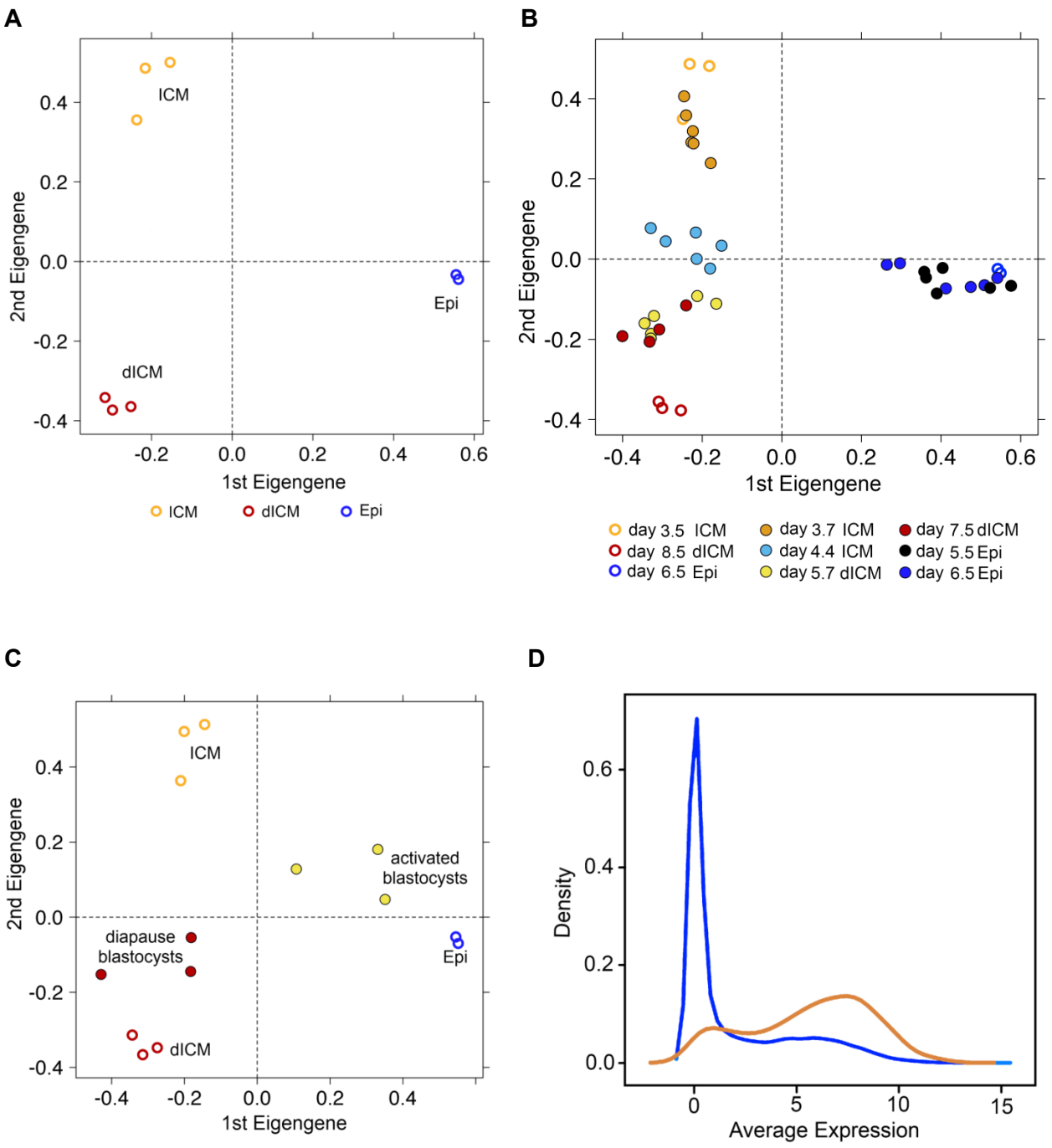


Figure 2.3 Comparison of RNA-Seq results with previously published studies shows reproducibility and enrichment of stem cell markers. (A) Principal Component Analysis (PCA) plot depicting the first two eigengenes of RNA-Seq data demonstrates reproducibility among biological replicates (ICM: *orange*, dICM: *red*, Epi: *blue*). The samples from distinct developmental stages reside in different quadrants. (B) Microarray data from Hunter et al., 2008

(filled circles) were projected onto RNA-Seq (open circles) PCA plot. The various stages of development occupy distinct regions in this space. The early pre-implantation ICM samples from Hunter (day 3.7, *orange*) and my RNA-Seq data (day 3.5, *orange*) occupy the left upper quadrant; the diapause samples from Hunter (day 5.7, *yellow,* and day 7.5, *red*) and from my RNA-Seq study (day 8.5, *red*) localize to the left lower quadrant, while all of the post-implantation epiblast samples from Hunter (day 5.5, *black,* and day 6.5, *blue*) and from my dataset (also day 6.5, *blue*) appear on the far right side of the graph. Note that day 4.4 samples from Hunter study (*light* blue) extend from the ICM to the diapause quadrants. Supporting information for the samples used in Hunter study is presented in Table 2.2. (C) Microarray data from Hamatani study (denoted by filled circles) confirmed that the transcriptional profiles of diapause embryos (*red*) align to the left lower (diapause) quadrant, while activated embryos (*yellow*) align to a distinct location within the right upper quadrant. (D) The frequency distribution of the average expression of stem cell markers (*orange)* in my RNA-Seq samples is higher than that of all other genes (*blue*).

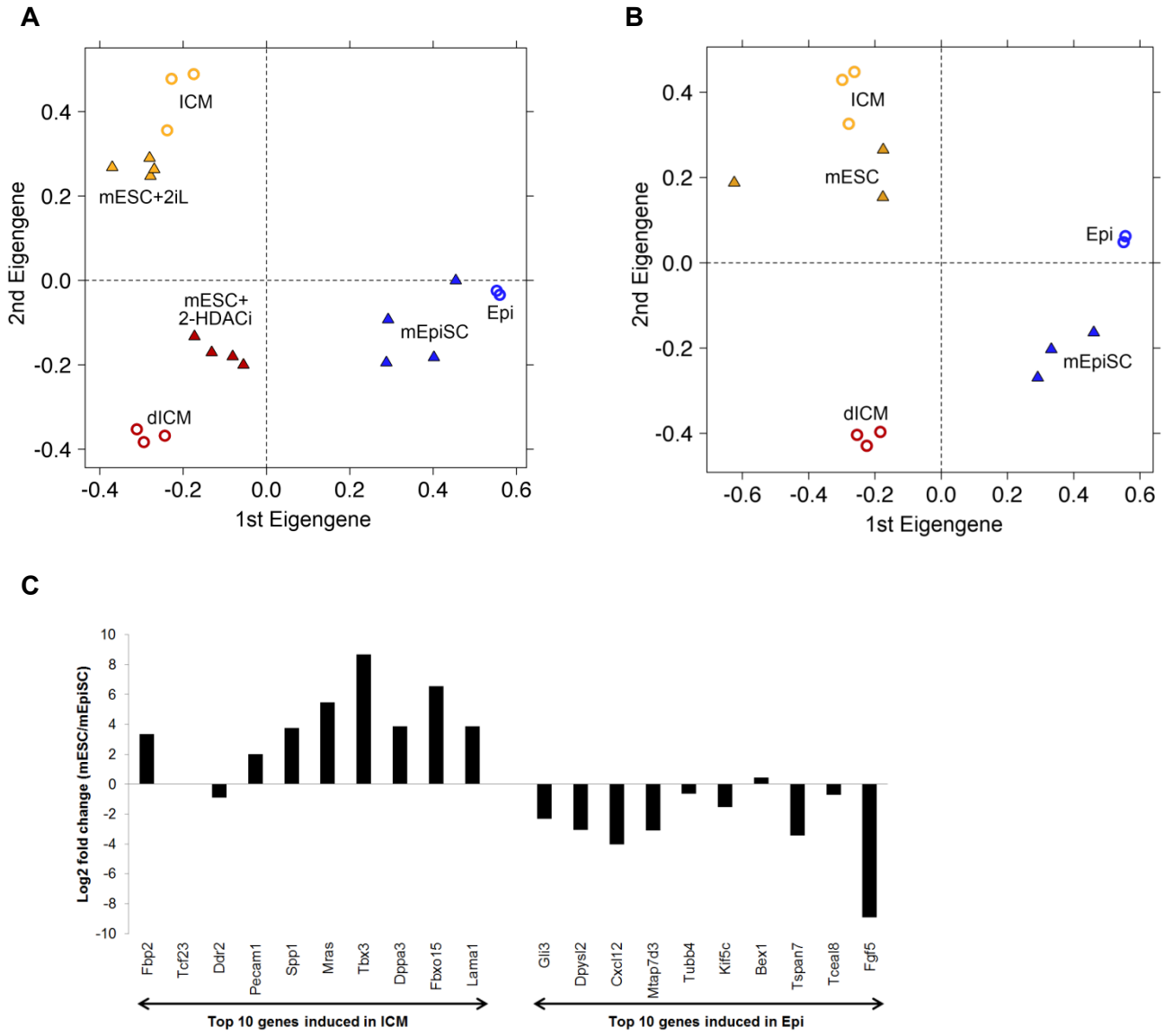


Figure 2.4 Adjusting culture conditions can induce mouse ESC and EpiSC to adopt diapause-like gene expression profiles. (A) Gene expression profiles of quadruplicate mouse ESC or EpiSC samples (denoted by triangles) projected onto the PCA plot of my RNA-Seq data (denoted by open circles). Note that mESC cultured in the presence of LIF plus 2 inhibitors (mESC+2iL, *orange*; Ying et al., 2008) align to the same left upper quadrant as ICM (*orange*), while mouse EpiSC cultured in the presence of FGF and Activin (*blue*) align to the same right lower quadrant as Epi (*blue*). In contrast, mESC cultured in 2-HDACi (*red*) align to the left lower quadrant corresponding to dICM (*red*). (B) A previously published microarray data comparing

mouse ESC and EpiSC (denoted by triangles; Tesar et al., 2007) were projected onto the RNA-Seq PCA plot. Three biological replicates from each culture conditions were used in Tesar study. Mouse ESC samples (*orange*) reside in the region occupied by the RNA-Seq ICM samples, while the mouse EpiSC samples (*blue)* are found closest to the Epi samples. (C) Genes that are highly induced in ICM and Epi samples from RNA-Seq data are showing a coordinate expression in mouse ESC and EpiSC, respectively. Shown in the graph is the fold change of average gene expressions in mouse ESC relative to EpiSC from the microarray data. Eight out of ten genes most highly expressed in ICM relative to Epi are significantly upregulated in mouse ESC compared to EpiSC, while 9 out of 10 genes most highly expressed in Epi relative to ICM are significantly downregulated (p-value < 0.05).

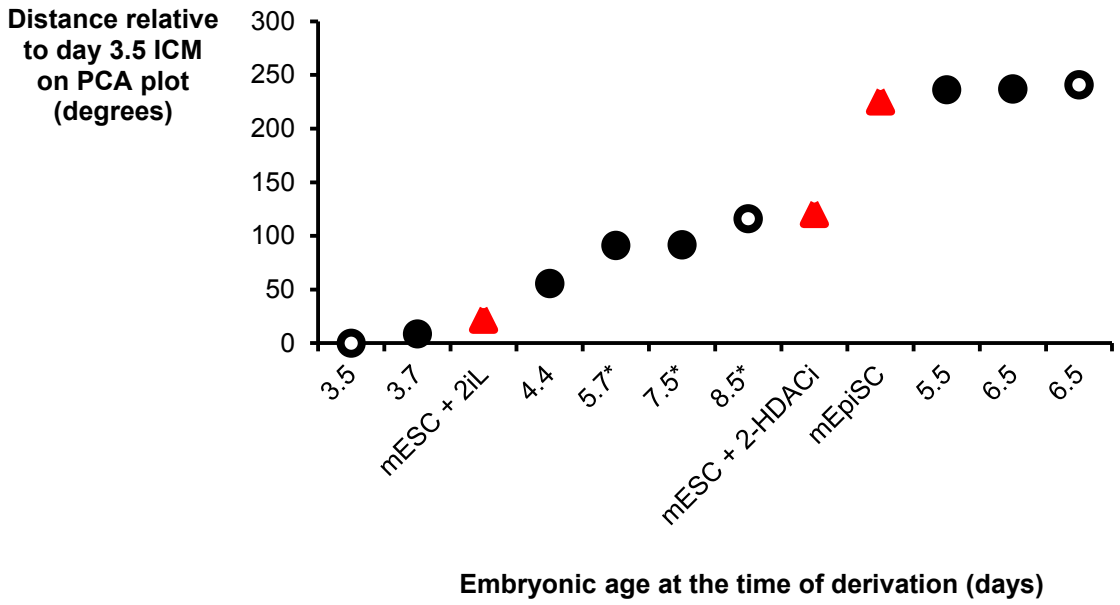


Figure 2.5 Various datasets projected onto a linear developmental yardstick. The samples from RNA-Seq (*open circles*), Hunter study (*filled circles*), and microarray of cultured pluripotent stem cells (*red triangles*) are combined into a single one-dimensional space. I made an assumption that development moves counterclockwise around the projected space of PCA plot, starting from the upper left and ending at the far right side of the plot. The angles between the samples relative to RNA-Seq day 3.5 ICM sample were obtained from the corresponding PCA plots. Samples were then rank-ordered on the x-axis based on their distance (in degrees) relative to day 3.5 ICM sample. Note how the relative distance obtained from the PCA plot (shown on the y-axis) positively correlates with the embryonic age at the time of sample derivation (shown on the x-axis). The graph presents the estimated developmental ages for the cultured pluripotent stem cells. * denotes pluripotent stem cells from diapause embryos.

Tables

Samples	Total no. of reads	No. of UNIQUE & SPLICE non-ribosomal RNA reads
ICM1	8,191,499	2,844,162
ICM2	22,947,915	3,023,359
ICM3	20,414,780	7,979,055
dICM1	11,968,001	3,743,553
dICM2	25,104,463	3,572,446
dICM3	31,222,479	7,581,714
Epi1	4,635,896	2,173,025
Epi2	8,036,358	2,177,578
TOTAL	**132,521,391**	**33,094,892**

Table 2.1. Summary of RNA-Seq reads mapped to mouse genome. Total number of RNA-Seq reads for each sample are shown. Reads were categorized into 4 groups: uniquely mapped to the mouse genome ("Unique"), mapped across splice junctions ("Splice"), mapped to more than one location in the genome ("Multi"), and not mapped to mm9 genome ("No Match"). Criteria used to obtain reads for the downstream analysis were as follows: (1) mapped either uniquely to the mm9 genome ("Unique") or to the splice junction library ("Splice") and (2) did not contain any rRNA sequence.

Samples	Approximate Embryonic Age	Note
ICM88	day 3.7	pre-implantation ICM
ICM105	day 4.4	peri-implantation ICM
DICM136	day 5.7	diapause ICM
DICM180	day 7.5	diapause ICM
EE5.5	day 5.5	post-implantation Epiblast
EE6.5	day 6.5	post-implantation Epiblast

Table 2.2 Description of samples used in Hunter study. Pre-implantation ICM samples came from two different timepoints (88 and 105 hours post-coitum [hpc]), which corresponds to days 3.7 and 4.4). The latter was harvested just prior to the time of implantation, and hence denoted as peri-implantation. Diapause ICM samples were harvested from 136 hpc (day 5.7) and 180 hpc (day 7.5). Post-implantation epiblast samples were from days 5.5 and 6.5. Details of methods can be found in Hunter *et al.* (2008). Briefly, each stage had 3 biological replicates, except DICM180 for which there were only 2 biological replicates. Each biological replicate had a technical duplicate on the array.

Supplementary Tables

Table S2.1 Consensus list of stem cell markers from 20 different studies comparing undifferentiated versus differentiated ESC.

Chapter 3 : An Evolutionarily Conserved Program for Stem Cell Dormancy

Summary

The ability to reversibly transition to a state of quiescence allows tissue specific stem cells to persist lifelong and cancer stem cells to resist therapy. However, the mechanisms that underlie stem cell quiescence are poorly understood. Here I outline the molecular framework of stem cell quiescence in a mouse model of delayed blastocyst implantation, also known as diapause. I performed RNA-sequencing (RNA-Seq) to characterize the transcriptional profiles of pluripotent stem cells extracted from embryos at three distinct developmental stages: pre-implantation, post-implantation, and during diapause. My analysis of diapause-associated gene expression points to a program for stem cell quiescence that includes the induction of *Runx1*, isoform switch involving the tumor suppressor *Lkb1*, and the induction of *Prkaa2*, which encodes the LKB1 target, AMPKα2. Portions of these same pathways support dormancy in lower organisms and quiescence in other stem cell types.

Introduction

Stem cells strike a balance between the competing interests of their two defining properties: self-renewal and multilineage differentiation. A large body of evidence supports the existence of a relationship between a stem cell's ability to self-renew and preservation of its capacity for quiescence. For example, in response to deleting single genes such as *Pten* (Yilmaz et al., 2006, Zhang et al., 2006), *Mpl* (Qian et al., 2007, Yoshihara et al., 2007) and *Lkb1* (Gan et al., 2010, Gurumurthy et al., 2010, Nakada et al., 2010), hematopoietic stem cells exit quiescence, proliferate transiently, and exhaust. In the context of malignancy, quiescence enables cancer stem cells to evade many types of cytotoxic treatment. I studied stem cell quiescence in the setting of diapause, a naturally occurring state of physiological stem cell dormancy in which implantation and embryonic development are temporarily suspended in response to stressful environmental conditions. Pluripotent stem cells from mouse blastocysts induced to enter diapause (i.e. diapause ICM [dICM]) were harvested, and their transcriptomes were characterized by RNA-Seq and compared to the pluripotent stem cells from normal pre- and post-implantation embryos (ICM and Epi, respectively). This study, described in chapter 2,

generated a list of candidate genes that are potentially important in the maintenance of stem cell quiescence. In this chapter, I will focus on three prominent and intriguing features of the transcriptional profile of diapause ICM: the diapause associated induction of *Runx1,* a diapause associated isoform switch involving the tumor suppressor, *Lkb1*, and the diapause associated induction of *Prkaa2*, which encodes the LKB1 target, AMPKα2.

The first gene that is expressed at a significantly higher level in dICM compared to ICM and Epi (6-fold and 431-fold, respectively) is Runt-related transcription factor 1 (*Runx1*). The *Runx* genes belong to a small family of evolutionarily conserved transcription factors that control critical cell fate decisions in a number of different cell lineages (Wang et al., 2010). Runx family members function in the context of heterodimers, Core Binding Factors (CBFs), that consist of an α subunit and a β subunit. In mammals, one of three different Runx family members (*Runx1*, *Runx2*, and *Runx3*) encodes the DNA-binding α subunit, while a single gene (*Cbfb*) encodes the non-DNA binding β subunit. *Runx1* is one of the most frequently mutated genes in human leukemia, occurring in approximately 30% of patients with acute myeloid leukemia (AML), and prompting its alternative designation: *AML1* (Look, 1997, Speck and Gilliland, 2002). *Runx1* is a key regulator of definitive hematopoiesis during development. *Runx1-/-* mice have a complete block in definitive hematopoiesis leading to death by embryonic day 12.5, while primitive hematopoiesis is only modestly affected (Okuda et al., 1996). *In vivo* studies involving transplantation of bone marrow cells overexpressing different isoforms of *Runx1* into lethally irradiated wild-type mice showed that *Runx1*b and *Runx1*c isoforms act as positive regulators of hematopoietic stem cell quiescence (Jacob et al., 2010, Tsuzuki et al., 2007). Loss of *Runx1* in adult mice results in an initial transient expansion of HSC, followed by decrease in HSC numbers in aged conditional *Runx1* knockout mice. This stem cell exhaustion phenomenon suggests a role of Runx1 in regulating the adult HSC in a state of quiescence (Growney et al., 2005, Jacob et al., 2009). Furthermore, as *Runx1* is a transcriptional repressor of *Bmi1*, an intestinal stem cell (ISC) marker, *Runx1* may also act as a positive regulator of quiescence in ISC (Motoda et al., 2007, Wang et al., 2010).

Other intriguing features of diapause ICM found in my RNA-Seq results are an isoform switch involving the tumor suppressor, liver kinase B1 (*Lkb1*), and the induction of *Prkaa2* which encodes the LKB1 target, 5'-AMP-activated protein kinase catalytic subunit α2 (AMPKα2). AMPK and LKB1 are required for mouse embryo fibroblasts (MEFs) to adapt to genotoxic stress

and oxidative stress levels (Bungard et al., 2010), and are known to regulate hematopoietic stem cell quiescence (Gurumurthy et al., 2010, Gan et al., 2010, Nakada et al., 2010). This makes them excellent candidates to regulate pluripotent stem cell quiescence during diapause.

Further detailed mechanistic studies of stem cell quiescence require *in vitro* models that are readily accessible and apt to genetic as well as biochemical modifications. Using a relatively low and narrow concentration of a 4-carbon fatty acid, butyrate, Ware and colleagues proposed that histone deacetylase inhibitors (HDACi) can induce mouse and human ESC to adopt an intermediate stem cell state (Ware et al., 2009). In the previous chapter, I have shown that the transcriptional profile of mouse ESC exposed to butyrate and another HDACi called SAHA (2-HDACi) aligns with the transcriptome of the pluripotent stem cells from diapause embryos (i.e. diapause ICM). These findings demonstrate that 2-HDACi supplementation to mouse ESC culture elicits a transcriptional program that most closely aligns with diapause, and hence can be used as an *in vitro* system to study the biology of stem cell quiescence in diapause. Using mouse ESC treated with 2-HDACi, I investigated whether the genes that are differentially expressed or regulated *in vivo* are functionally important in the *in vitro* model of diapause. Here I showed that the ablation of one allele of *Runx1* is sufficient to accelerate spontaneous differentiation in mouse ESC cultured in 2-HDACi.

I also examined the protein expression of RUNX1, LKB1 and phosphorylated form of AMPKα2 in day 3.5, diapause and peri-implantation mouse embryos. Interestingly, I found that induction of RUNX1 in the ICM of diapause embryos, particularly in the PE layer, is correlated with OCT4 expression pattern; however, there is a lack of detectable change in the phosphorylated AMPKα2 protein level. I also confirmed the absence of short isoform of *Lkb1* in diapause ICM. With antibodies that recognize both isoforms of LKB1, I showed that the overall LKB1 protein level is unchanged as expected from the RNA-Seq results.

Results

In conjunction with the genome-wide analysis on transcriptional profiles of ICM, dICM and Epi described in Chapter 2, I generated lists of mRNAs that were differentially expressed in each stage relative to the two other stages (see Methods). This analysis generated 728, 922 and 3,568 genes that were differentially expressed in ICM, dICM and Epi, respectively (Tables S3.1-

S3.6). Categorization of differentially expressed genes using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway annotation identified a number of interesting features (Tables S3.7-S3.12). *Tet1*, a previously described pluripotency biomarker picked from 25 microarray studies using a machine learning algorithm, was significantly upregulated in ICM (Scheubert et al., 2011); p21 was upregulated and cell cycle-associated genes were downregulated in dICM, while genes associated with RNA processing were predominant in Epi. These findings are consistent with prior knowledge (Hakem et al., 1996, Hamatani et al., 2004, Ji et al., 2009, Pritsker et al., 2004, Salomonis et al., 2010) further supporting the validity of my study.

To characterize the transcriptional profile of dICM, I focused on the lists of genes that were differentially expressed in diapause ICM relative to both ICM and Epi (Table S3.3 and S3.4). These lists consist of 719 and 203 genes were expressed at significantly higher and lower levels, respectively (adjusted p-value <0.05). Gene set enrichment analysis (GSEA) indicated a diapause-associated induction of genes regulated by p21, TGFβ signaling, hypoxia, and starvation, and a repression of genes regulated by c-Myc, YY1, and Elongin-A as well as genes associated with cell-cycle (Table 3.1). Upregulation of p21 and differential regulation of cell cycle genes in whole diapause blastocysts have been reported previously (Hamatani et al., 2004). The role of TGFβ in maintaining dormancy of prostatic stem cells has been established (Hatsfeld et al., 1991, Salm et al., 2005, Yamazaki et al., 2009). Hypoxia and starvation are known to suppress cell cycle, induce quiescence and suppress the activity of c-Myc (Angelo and van Gilst, 2009, Goda et al., 2003, Hermitte et al., 2006, Koshiji et al., 2004). Inactivation of c-Myc, which in turn derepresses p21, is associated with dormancy in cancer (Iriuchishima et al., 2011, Shachaf et al., 2004, Yu et al., 2005). Ablation of YY1 and Elongin-A causes mitotic defect, which leads to cell cycle arrest and slow growth phenotype, respectively (Affar et al., 2006, Yamazaki et al., 2003). In short, all of these genes and pathways have been previously implicated in dormancy or cell cycle arrest.

I also used a published database (GSEA c3) to identify transcription factors that, along with their predicted targets, were upregulated in dICM (Table 3.2). This list included *Runx1*, which was selected for further study based on extensive, albeit conflicting evidence regarding its role in regulating stem cell quiescence (Challen and Goodell, 2010, Tsuzuki et al., 2007, Wang et al., 2010).

Runx1 is induced in diapause

My RNA-Seq results showed a highly reproducible, diapause-associated induction of *Runx1*. *Runx1* transcript levels in the dICM are 6-fold higher relative to ICM, and 431-fold higher relative to epiblast (Figure 3.1A). The absence of reads aligned to the first two *Runx1* exons is consistent with induction of the embryonic isoform, Runx1b, as confirmed by quantitative real-time PCR (Figure 3.1B). Confocal microscopy also shows that RUNX1 is induced in the ICM of diapause embryos, and exhibits a nuclear pattern of distribution. RUNX1 expression was most prominent within the PE (Figure 3.1C, compare top and middle panels). As PE cells are fated to generate extra-embryonic yolk sac (Rossant et al., 2003), I hypothesized that RUNX1 may act to block PE differentiation. A recent report noted that PE cells lose developmental plasticity in association with the loss of Oct4 (Grabarek et al., 2012). Strikingly, found that the pattern and intensity of RUNX1 staining in diapause ICM correlates almost perfectly with OCT4 (Figure 3.1C, middle panel). To examine RUNX1 expression during uninterrupted mouse development, I examined late (day 4.4) blastocysts obtained immediately prior to implantation. While RUNX1 staining increases between days 3.5 and 4.4, and is also more intense within the PE layer, unlike diapause embryos I observed no correlation between the staining patterns of RUNX1 and OCT4. (Figure 3.1C, bottom panel). Collectively, these results suggest that RUNX1 and OCT4 may cooperate to support diapause.

Evaluating Runx1 function *in vitro* using 2-HDACi model

Since the minute size of diapause embryos complicates detailed investigations, I examined the effects of *Runx1* in the context of ESC. A previous study examined the effects of conditionally over-expressed transcription factors, including *Runx1*, on the gene expression profiles of mouse ESC (de Cegli et al., 2010, GEO: GSE19836). 147 genes whose expression is higher in dICM relative to day 3.5 ICM and day 6.5 Epi were defined as diapause signature genes. Using the microarray data from de Cegli study, the ratio of gene expression levels of these diapause signature genes in *Runx1* over-expressing mouse ESC was compared to the control samples. The result shows that *Runx1* over-expression in mouse ESC accounts for higher expression levels of diapause signature genes compared to other non-diapause-associated genes ($p < 10^{-10}$, Figure 3.2A). PCA also shows that *Runx1* over-expression induces a transcriptional profile that corresponds to diapause (data not shown).

In mouse ESC, 2-HDACi induces a 2.4-fold increase in the same *Runx1* isoform that is induced in diapause, *Runx1b* (Figure 3.2B). To further examine the 2-HDACi model and *Runx1*'s potential role in diapause, I tested whether *Runx1* modifies the response of mouse ESC to the 2-HDACi combination. The responses to 2-HDACi were compared in *Runx1*+/+, +/- and -/- mouse ESC lines (kindly provided by Georges Lacaud, University of Manchester - Lacaud et al., 2002). Clear morphological differences between *Runx1*+/+, +/- and -/- mouse ESC lines were apparent within 4 days of 2 HDACi exposure; *Runx1* +/+ ESC retained a partially mounded appearance, whereas the *Runx1*+/- and *Runx1*-/- lines adopted a flattened morphology (Figure 3.2C). The latter also displayed a tendency toward spontaneous differentiation, indicating a requirement for *Runx1* to maintain pluripotency and quiescence. Collectively, these results lend strong support to the validity of the 2-HDACi model, and to *Runx1*'s candidacy as a diapause regulator.

To investigate the extent of pluripotency loss in 2-HDACi condition, I analyzed the expression levels of pluripotency markers and mouse ESC-associated genes by qPCR (Figure 3.2D). Pluripotency markers, *Oct4* and *Sox2*, had a significant decrease after 2-HDACi treatment, with a greater repression in both Runx1+/- and -/- mouse ESC compared to wild-type ESC. A similar reduction was seen for *Zfp42/Rex1*, which is associated with mouse ESC and pre-implantation ICM. *Nanog*, which is another pluripotency marker, showed a >2-fold reduction in the 2-HDACi condition and mouse EpiSC. A recent study showed the importance of allelic regulation of *Nanog* in controlling ground-state pluripotency (Miyanari and Torres-Padilla, 2012). The authors also reported that the *Nanog*-heterozygous blastocysts had fewer ICM derivatives and delayed PE formation, indicating a role for the biallelic expression of *Nanog* in the timely maturation of the ICM into a fully reprogrammed pluripotent epiblast. My result suggests mouse ESC cultured under the naïve condition had a biallelic expression of *Nanog* but switched to a monoallelic expression in the 2-HDACi condition as well as in mouse EpiSC; the level of *Nanog* expression level was similar in all 3 ESC lines in the 2-HDACi condition. Monoallelic expression of *Nanog* in mouse ESC exposed to 2-HDACi confirmed that the 2-HDACi treatment shifted mouse ESC away from the naïve state. Collectively, the results of pluripotency marker analysis using *Runx1*+/+, +/- and -/- mouse ESC lines are consistent with a requirement for *Runx1* to maintain pluripotency and stem cell quiescence.

Exposure to HDACi is known to cause a flattened morphology of mouse ESC colonies (Ware et al., 2009). To differentiate whether the flat colonies were still in the 2-HDACi-associated state or have switched to the primed state, I checked the expression of the genes known to be induced in mouse ESC treated with 2-HDACi (*Gsn* and *Cryab*) and two other genes that are induced in mouse EpiSC as well as post-implantation epiblast (*Cer1* and *Ldb2*) (Figure 3.2D). The 2-HDACi markers were induced significantly in mouse ESC exposed to 2-HDACi compared to mouse EpiSC regardless of the *Runx1* level. All 3 ESC lines also had a higher expression of *Gsn* in the 2-HDACi condition relative to the naïve (2iL) condition. Although *Runx1+/-* and *-/-* ESC had a significant induction of *Cryab* expression as expected in 2-HDACi condition, *Cryab* expression level in wild-type mouse ESC was unchanged by day 4 of 2-HDACi exposure. My findings suggest that *Cryab* might be a late-response gene in wild-type ESC, while *Gsn* is an early marker of 2-HDACi-associated state. Both EpiSC markers were expressed at a very low levels in mouse ESC treated with 2-HDACi, but with higher expression in *Runx1-/-* mouse ESC compared to the other two ESC lines. The expression level of the EpiSC markers was not substantial in 2-HDACi condition compared to mouse EpiSC, suggesting that there is a drift toward the primed state in the *Runx1-/-* cultures. I further checked the expression of some genes associated with the germ line, including *Piwil2* and *Dppa3*. These germ line-associated genes are expressed by mouse ESC in naïve condition but undetected in mouse EpiSC. I detected the expression of these markers in all 3 ESC lines exposed to 2-HDACi, with lower level of expression in *Runx1+/- and -/-* ESC. Collectively, my results suggest that loss of *Runx1* allele is required to keep mouse ESC in the quiescence-associated state longer.

Butyrate exposure decreases BrdU incorporation in mouse ESC, suggesting a reduction in the rate of cell division (Ware et al., 2009). I found that Runx1 is not required for this effect since 2 HDACi slows cell growth to a similar extent in Runx1 +/+, +/- and -/- mouse ESC (Figure 3.2E). Similarly the cell cycle profile was unaffected by Runx1 genotype (Figure 3.2F).

Activation of the metabolic sensing pathway

AMP-activated protein kinase (AMPK) sits at a unique position as an energy sensor that can interface with diverse signaling molecules ranging from LKB1 to mammalian target of rapamycin (mTOR), affecting processes from ribosomal biogenesis to actin regulation (Williams and Brenman, 2007). Transcript levels of AMPKα2 (*Prkaa2*), the catalytic subunit of AMPK, were 12

fold higher in diapause ICM than in the pre-implantation ICM or epiblast (Figure 3.3A). AMPKα2 Thr172, required for AMPK activation, is phosphorylated by two major upstream kinases: LKB1 and Ca^{2+}/calmodulin-dependent protein kinase kinase 2 (CAMKK2) (Hawley et al., 1996, Hawley et al., 2003, Henin et al., 1996, Shaw et al., 2004, Stein et al., 2000, Sullivan et al., 1994, Woods et al., 2003). Confocal microscopy with an antibody against the phosphorylated form of AMPKα2 Thr172, however, did not show qualitative difference between day 3.5 and diapause embryos (Figure 3.3B)

While the overall *Lkb1* transcript levels did not change between pre-implantation and diapause ICM (Figure 3.4A), its expression profile suggested an isoform switch, with predominance of a previously described short *Lkb1* isoform in the pre-implantation ICM (Denison et al., 2009, Towler et al., 2008) (Figure 3.4B). *Lkb1/Stk11* is a tumor suppressor that is mutated in Peutz-Jeghers syndrome and, in addition to inhibiting cell proliferation, has profound effects on cell polarity and on the detection and response to low energy (Alessi et al., 2006, Williams and Brenman, 2007). The short isoform of LKB1 ($LKB1_{Short}$) arises from alternative splicing, with replacement of the carboxy-terminal 63 residues by a unique 39-residue sequence lacking known phosphorylation (Ser^{431}) and farnesylation (Cys^{433}) sites. $LKB1_{Short}$ levels are particularly high in the testis, and male mice lacking $LKB1_{Short}$ (while retaining full length LKB1) are viable but sterile (Towler et al., 2008). Confocal microscopy with an antibody against both LKB1 isoforms showed a predominantly nuclear staining pattern that did not differ significantly between pre-implantation and diapause embryos (Figure 3.4C). A second kinase that directs AMPK Thr172 phosphorylation is CaMKK2. According to my RNA-Seq data, this kinase is not transcribed in the pluripotent stem cells from all 3 stages analyzed (ICM, dICM and Epi). In aggregate, these findings suggest that LKB1 is the predominant upstream activator of AMPK in diapause ICM, and that its activity may be regulated via an isoform switch.

In contrast to the relative abundance of $LKB1_{Short}$ in the pre-implantation ICM and post-implantation epiblast, $LKB1_{Short}$ was negligible in cultured mouse ESC or EpiSC (Figure 3.5A). To examine the function of this LKB1 short isoform, it was overexpressed in mouse ESC and its effects were observed on gene expression and cell signaling. Chaozhong Song from the laboratory of Dr. Blau generated the TALEN construct to modify the endogenous *Lkb1* locus in mESC (Figure 3.5B). The modified *Lkb1* locus expresses only the short form of LKB1. After two rounds of transfection with the TALEN pairs and donor plasmid and selection for *Neo* resistant

clones, mESC (R1 line) expressed only the short isoform of *Lkb1*. Mouse ESC expressing either both isoforms of LKB1 or only $LKB1_{Short}$ are viable. They had more compact and moundy colonies than the wild-type mouse ESC (Figure 3.5C and data not shown). In general, they look healthy and proliferate as well as the parental cell lines.

Discussion

Especially striking is widely conserved nature of the molecular theme of stem cell quiescence. Here I investigated three genes whose roles in stem cell quiescence have been previously established.

Context-dependent roles of *Runx1* in stem cell quiescence

As discussed in chapter 2, an outstanding question about diapause is whether it represents a developmental detour or pit stop in normal embryonic development. The results of Principal Component Analysis (PCA) on my RNA-Seq of pluripotent stem cells together with the Hunter's microarray data may give an insight into which diapause model is adopted by nature. With the detour model where diapause is viewed as a branch-off of the developmental continuum, the left lower quadrant occupied by dICM samples may be skipped completely in normal development. Alternatively, with the pit stop model whereby diapause is an obligatory state, the dICM quadrant is a transition phase that all embryos have to pass before implantation. The day 4.4 peri-implantation ICM samples from the Hunter study, which cross the boundary separating the ICM and diapause quadrants, suggest the latter. The induction of *Runx1* protein in day 4.4 peri-implantation non-diapause blastocysts further implies that diapause-like state is a part of normal embryonic development.

The PE layer contains the earliest committed cells of the ICM, and is fated to generate the extra-embryonic yolk sac (Rossant et al., 2003). I hypothesize that high RUNX1 levels may be required to prevent PE cells from continuing along the pathway of differentiation to which they are committed. In keeping with this prediction, I find that cells with the highest RUNX1 levels in diapause ICM also express the highest levels of the pluripotency factor, OCT4. It has recently been reported that cells in the PE layer lose developmental plasticity in association with the loss of OCT4 (Grabarek et al., 2012). Another potential explanation for higher RUNX1 level in the PE

layer is that PE may serve as a niche to maintain the pluripotency of the epiblast population. Although RUNX1's function has been shown to be cell-autonomous, another member of Runx family, RUNX2, is required to generate the HSC-niche components, and hence is important for HSC quiescence (Calvi et al., 2003, Deguchi et al., 1999, Komori et al., 1997, Otto et al., 1997, Sato et al., 1998, Visnjic et al., 2004, Zhang et al., 2003).

Analysis of my RNA-Seq results suggested *Runx1* as a candidate regulator of diapause. I subsequently used the *in vitro* system to test my prediction that *Runx1* induction is required for diapause. In the absence of *Runx1*, ESC exposed to 2-HDACi differentiate, suggesting that *Runx1* is required to keep the cells in the quiescent, pluripotent and undifferentiated state. It is interesting to note that loss of even one functional allele of *Runx1* contributed to a faster change of morphology when the cells were exposed to 2-HDACi. Flattening of the colonies and spontaneous differentiation could be seen as early as 3 days of 2-HDACi treatment. The importance of gene dosage of *Runx1* and its physiological relevance have been reported in earlier studies. Haploinsufficiency of RUNX1 causes familial platelet disorder and predisposes to developing acute myeloid leukemia (Song et al., 1999). On the other hand, increased dosage of RUNX1 plays a role in the development of leukemia in myeloid and lymphoid lineages (Niini et al., 2000, Wotton et al., 2002, Yanagida et al., 2005). My results with the *in vitro* model of diapause suggest that the physiological function of *Runx1* in pluripotent ESC is sensitive to the dosage of the gene. My findings also point to the potential use of cultured pluripotent stem cells as a surrogate to study the mechanism of diapause *in vitro*.

The roles of *Runx* family genes in stem cell quiescence in various organisms and tissues have been conflicting. *Runx* family genes can have context-dependent functions in regulating stem cell behavior. Adding more to the complexity are the different alternatively spliced isoforms of the same gene and different proteins of the Runx family. For example, the transcription of *Runx1* gene is under the control of 2 alternative promoters, generating two different isoforms: *Runx1*c isoform expressed from a distal promoter and *Runx1*b isoform expressed from a proximal promoter. Another isoform of *Runx1*, called *Runx1*a, is also expressed from the proximal promoter but lacking the C-terminal domain shared by *Runx1*b and *Runx1*c. These different isoforms of *Runx1* were shown to have distinct roles in stem cell regulation in the hematopoietic system (Tsuzuki et al., 2007). In mammalian liver tissue, different family members of *Runx* are expressed at quiescent and activated states of liver stem cells (also

known as hepatic stellate cells). Under normal conditions, hepatic stellate cells remain in the quiescent state and express *Runx3*. Upon liver injury, hepatic stellate cells are activated to re-enter the cell cycle, associated with upregulation of *Runx1*a isoform and *Runx2* (Bertrand-Phillippe et al., 2004). The dual roles of RUNX can also be explained by the nature of RUNX as transcription factor. The functions of RUNX as either transcriptional activator or repressor may depend on other factors that are present together with RUNX in the transcription complex (Wang et al., 2010).

Runx1 and Notch signaling

The Notch signaling pathway may contribute to the role of *Runx1* in regulating stem cell quiescence. Interaction between *Runx1* and the Notch signaling pathway during developmental specification of hematopoietic stem cells (HSC) has been established (Burns et al., 2005). Forced expression of *Runx1* induces quiescence in hematopoietic stem cells (HSC) in a manner that may involve the Notch pathway (Challen and Goodell, 2010). Notch pathway activation is also essential for generation and maintenance of undifferentiated quiescent satellite cells in muscles (Fukada et al., 2011). The generation of undifferentiated quiescent satellite cells is impaired during postnatal development in mice where two of the Notch target genes (*Hesr1* and *Hesr3*) are knocked out.

I observed a variety of Notch associated genes being upregulated in diapause ICM. Genes upregulated at the mRNA level include *Notch2*, the Notch ligands *Jagged2* and *Dll-1* (Delta-like1), *Psen2* (which encodes the gamma secretase catalytic subunit that liberates the Notch intracellular domain), *Hes1* (a primary target of the Notch pathway), and *Maml3*, an essential positive regulator of Notch signaling (Oyama et al., 2011). Based on the microarray data, the Notch signaling might also be activated in mESC+2-HDACi, indicated by upregulation of its receptors *Notch2* and *Notch3*, ligands *Jag2, Dll3* and *Hey1* and the downstream transcriptional target *Hes5.* Mouse ESC treated with HDACi would then be a useful *in vitro* model to examine the role of Notch signaling in embryonic diapause and how it is associated with *Runx1*. To test whether inhibiting Notch signaling interrupts *Runx1* induction, one can use gamma secretase inhibitors or shRNAs directed against specific targets in the Notch pathway.

Metabolic sensing pathway in embryonic diapause

AMPK activation is usually marked by phosphorylation of 172 of the catalytic subunit. Although I see 12-fold increase of *Prkaa2* transcript in the RNA-Seq data, my immunofluorescence results did not show induction of phosphorylation in diapause embryos. It is possible that the post-translational modification is extremely fine-tuned such that the change in Thr172 phosphorylation is below the detection level of confocal microscopy, which is a qualitative approach.

Despite the lack of detectable change in phosphorylation level of AMPKα2 in my immunofluorescence analysis, other observations made from my RNA-Seq results suggest an activation of AMPK pathway activation in diapause ICM. Autophagy is the primary intracellular mechanism for degrading and recycling long-lived proteins and organelles (Levine et al., 2004). This cellular catabolic process is induced by either suboptimal extracellular environments (nutrient starvation, hypoxia, overcrowding) or intracellular stress (accumulation of damaged cytoplasmic components). For example, formation of dauer larvae in C. elegans in response to crowding or a shortage of food has been shown to involve autophagic activation in certain cells (Melendez et al., 2003). In mammals, autophagy also occurs in normal physiological processes (Mizushima and Levine, 2010), and deregulation of autophagy is associated with certain neurodegenerative disorders, cancers, and muscular diseases (Levine and Kroemer, 2008). A recent report showed that autophagy is activated in diapause mouse embryos (Lee et al., 2011). Consistent with AMPK activation, my RNA-Seq results showed that diapause ICM is associated with upregulation of genes involved in autophagy (*Gpnmb*, *Hummr*, *Gabarapl1*) as well as mitochondrial biogenesis (*Ppargc1a*, *Ppargc1b*) and glucose uptake (*Glut3*). These observations suggest an activation of AMPK signaling pathway despite the lack of detectable increase in phosphorylation of AMPKα2. Also induced in diapause ICM were a number of genes related to Alzheimer's disease (*App*, *Bace2*, *Mme*, *Apoe*), suggesting a previously uncharacterized biological role for this pathway.

One of the most striking findings from my RNA-Seq data is a >3 log diapause-associated reduction in mRNA levels of a previously described short isoform of LKB1 ($Lkb1_{Short}$), while corresponding levels of the long isoform ($Lkb1_{Long}$) change only negligibly (Figures 3.4A and 3.4B). As described above, $Lkb1_{Short}$ is widely expressed, with particularly high levels in the

testes, and the only known phenotype of *Lkb1*$_{Short}$ null mice is male infertility (Towler et al., 2008). In this context one may presume that LKB1$_{Short}$ performs a function essential to normal testicular development. But how might the striking association between diapause and the specific absence of *Lkb1*$_{Short}$ mRNA be explained? I hypothesize that the abrupt disappearance of *Lkb1*$_{Short}$ mRNA may start a clock that measures time based on declining levels of LKB1$_{Short}$ protein. A slow decline in LKB1$_{Short}$ protein levels may provide a mechanism that allows for the temporal regulation of metabolism. In this regard it is noteworthy that LKB1 and AMPK ration lipid reserves in C. elegans dauer larva, allowing for survival under harsh environmental conditions (Narbonne and Roy 2009). To test this prediction I would need a method for specifically measuring the abundance of LKB1$_{Short}$ in the dICM over a time course in diapause. While antibodies specific for the carboxy-terminus of LKB1$_{Short}$ have been described (Denison et al., 2009), none has proved suitable for immunofluorescence (Angela Woods, Hammersmith Hospital, London, personal communication). Further characterization of the mouse ESC expressing only the short isoform of LKB1 is still under investigation in the laboratory of Dr. Blau.

Methods

Embryo harvest

C57BL/6 embryos were collected by flushing the uterus of superovulated pregnant females on day 3.5 (pre-implantation), day 4.4 (peri-implantation), and 6 days after chemical ovariectomy on day 2.5(diapause). Detailed methods can be found in Methods section of Chapter 2. Noon on the day of vaginal plug detection was considered as day 0.5.

Whole-mount Immunofluorescence

The zona pellucida was removed from day 3.5 embryos by incubation in acid tyrode's solution. Embryos were fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) for 15-20 minutes at room temperature, rinsed in PBS+0.1% Triton X-100, permeabilized in 0.25% Triton X-100 for 15-20 minutes, rinsed in PBS+0.1% Triton X-100 and blocked in blocking buffer for at least 1 hour at room temperature or overnight at 4^{o}C. Blocking buffer was made of PBS supplemented with 10% FBS and 0.1% Triton X-100. Primary antibodies were: RUNX1

(OriGene, TA307515), OCT4 (Santa Cruz, sc-5279), p-AMPKα Thr172 (Cell Signaling, 2535), AMPKα2 (Santa Cruz, sc-19131), and LKB1 (Santa Cruz, sc-5638). Primary antibodies were diluted 1:100 in blocking buffer, and embryos were incubated with the appropriate antibodies at 4^{o}C overnight. They were rinsed three times in blocking buffer for 10 minutes each, and incubated with secondary antibodies for 1 hour at room temperature. Alexa Fluor secondary antibodies (Invitrogen) were used at 1:500 dilution in blocking buffer. Embryos were then incubated in Hoechst 33342 (Invitrogen, 1 μg/μL) for 10 minutes at room temperature, rinsed three times in blocking buffer for 10 minutes each, and mounted on a 35-mm glass-bottom dish (MatTek Corporation, P35G-1.5-14-C) in a PBS droplet overlayed with mineral oil (Sigma). Images were taken with a Nikon A1R confocal microscope (60x water immersion objective) or with a Zeiss LSM 510 Meta confocal microscope (40x water immersion objective). ImageJ 1.44 and NIS-Elements Viewer 3.20 software was used to visualize the data. All embryo images are individual laser confocal sections.

Mouse ESC and EpiSC culture

Runx1 +/+, +/- and -/- mouse ESC (J1 lines, gifts from Dr. Georges Lacaud) were plated without feeders on Matrigel (BD Biosciences) diluted according to manufacturer's instructions. Basal Mouse ESC culture medium consisted of DMEM-containing GlutaMax, 20% FBS (ES Cell Qualified, Invitrogen), 1 mM sodium pyruvate, 0.1 mM nonessential amino acids, 50 U/ml penicillin, 50 mg/ml streptomycin (all from Invitrogen, Carlsbad, CA), and 0.1 mM β-mercaptoethanol (Sigma). For the 2iL condition, mESC media was supplemented with 1000 units/mL mouse LIF (ESGRO, Millipore) plus 3 μM GSK3 inhibitor (CHIR99021, Selleck Chemicals) and 0.4 μM MEK inhibitor (PD0325901, Stemgent). For the 2-HDACi condition, the media was supplemented with 0.1 mM sodium butyrate (Sigma) and 25 nM vorinostat/SAHA (Cayman Chemical). mESC were passaged by washing once with PBS without calcium and magnesium (Invitrogen), followed by exposure to trypsin-EDTA (Invitrogen) until the cells were digested to single cells.

The phenol red used as pH indicator in the media is thought to have a weak estrogenic activity (Berthois et al., 1986; Welshons et al., 1988). Culture conditions in which phenol red and glucose are removed from the media allows mouse ESC to have a more homogeneous appearance (Ware et al., unpublished results). Mouse ESC were cultured in conditions

described above with the removal of phenol red and glucose from the medium. The deletion of exon 4 of *Runx1*, which encoded the last exon of DNA-binding domain, in the *Runx1*+/- and *Runx1* -/- mouse ESC had been confirmed previously (data not shown).

Mouse EpiSC line no. 5 (a gift from Paul Tesar and Ron McKay) was cultured as described (Tesar et al., 2007). Briefly, mouse EpiSC medium consisted of DMEM-F12-containing Glutamax (Invitrogen), 20% knockout serum replacement (KSR), 5 ng/mL FGF2 (R&D Systems), 10 ng/mL Activin A (Humanzyme), 0.1 mM β-mercaptoethanol (Sigma), 1 mM sodium pyruvate, 0.1 mM nonessential amino acids, 50 U/ml penicillin, 50 mg/ml streptomycin. Mouse EpiSC were passaged in clusters using dispase rather than as a single-cell suspension.

Cells were counted using a Nucleocounter NC-100, according to the manufacturer's protocol.

Quantitative real-time PCR

Total RNA was purified using the RNeasy Micro Kit (Qiagen) following the manufacturer-specified protocol. Reverse transcription of total RNA was performed using random hexamers with the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). qPCR was performed in triplicate using TaqMan Universal PCR Master Mix, No AmpErase UNG (Applied Biosystems) or SYBR Green PCR Master Mix (Applied Biosystems) in 10 µl reactions using an Applied Biosystems 7900HT Fast Real-Time PCR System. The qPCR conditions were set as follows: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 sec and 60°C for 1 min. *β-actin* was used as endogenous control for normalization of target mRNA expression. SDS RQ Manager 1.2 Software (Applied Biosystems) was used for data analysis.

The following gene specific primer pairs were used in this study:

Genes	Forward sequence (5' to 3')	Reverse sequence (5' to 3')
Runx1b	CCTCCGGTAGTAATAAAGGCTTCTG	CCGATTGAGTAAGGACCCTGAA
Runx1c	GTGTGCTGGAATTCGGCTTAG	AAGCCATCGTTTCCTTTCGA
Cer1	ACGAAGTACACTGGGAGACC	CGAATGGAACTGCATTTGC
Oct4	CGTTCTCTTTGGAAAGGTGTTC	GAACCATACTCGAACCACATCC
Sox2	GCGGAGTGGAAACTTTTGTCC	CGGGAAGCGTGTACTTATCCTT
Nanog	CTCAAGTCCTGAGGCTGACA	TGAAACCTGTCCTTGAGTGC
Piwil2	TTGGCCTCAAGCTCCTAGAC	GAACATGGACACCAAACCTACA

β-actin	GGCTGTATTCCCCTCCATCG	CCAGTTGGTAACAATGCCATGT

TaqMan Gene Expression Assays (Applied Biosystems) used in this study include the following: *Runx1* Mm01213404_m1 (probe spanning exons 4 and 5; exon 4 is deleted in Runx1 knockout mouse ESC) and Mm01213405_m1 (probe spanning exons 7 and 8), *Dppa3* Mm01184198_g1, *Zfp42/Rex1* Mm01194090_g1, *Gsn* Mm00456679_m1, *Cryab* Mm00515567_m1, *Ldb2* Mm00521915_m1, and *β-actin* Mm00607939_s1.

Cell cycle analysis

Cells were harvested, washed with PBS, and fixed in 70% ethanol overnight at 4^{o}C. Fixed cells were treated with RNase (20μg/mL) and stained with propidium iodide (PI) (20μg/mL) at room temperature for 30 minutes. DNA content was measured by the intensity of the fluorescence produced by PI using the BD FACSCanto™ II system (BD Biosciences). Data were analyzed with the FlowJo (version 9.4.11) software. The distribution of cells in G1, S and G2 phase was fitted with single Dean-Jett-Fox model.

Gene Set Enrichment Analysis (GSEA)

GSEA was performed as described in Subramanian et al., 2005. The c2 (curated) and c3 (motif) gene sets were from the MSigDB database v2.5 (released in April 2008). The parameters used were as follows: permutation=1000; permutation type=gene set; gene set filters: Min=15 and Max=500; FDR cut-off=0.25. SIZE = Number of genes in the gene set after filtering out those genes not found in the expression dataset. ES = Enrichment score for the gene set, reflecting the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes. NES = Normalized Enrichment Score, which is the enrichment score for the gene set after it has been normalized across analyzed gene sets. FDR = False Discovery Rate, which is the estimated probability that the normalized enrichment score represents a false positive finding; FDR values were adjusted for gene set size and multiple hypothesis testing. Gene sets were ordered by the NES before determining the rank number.

Figures

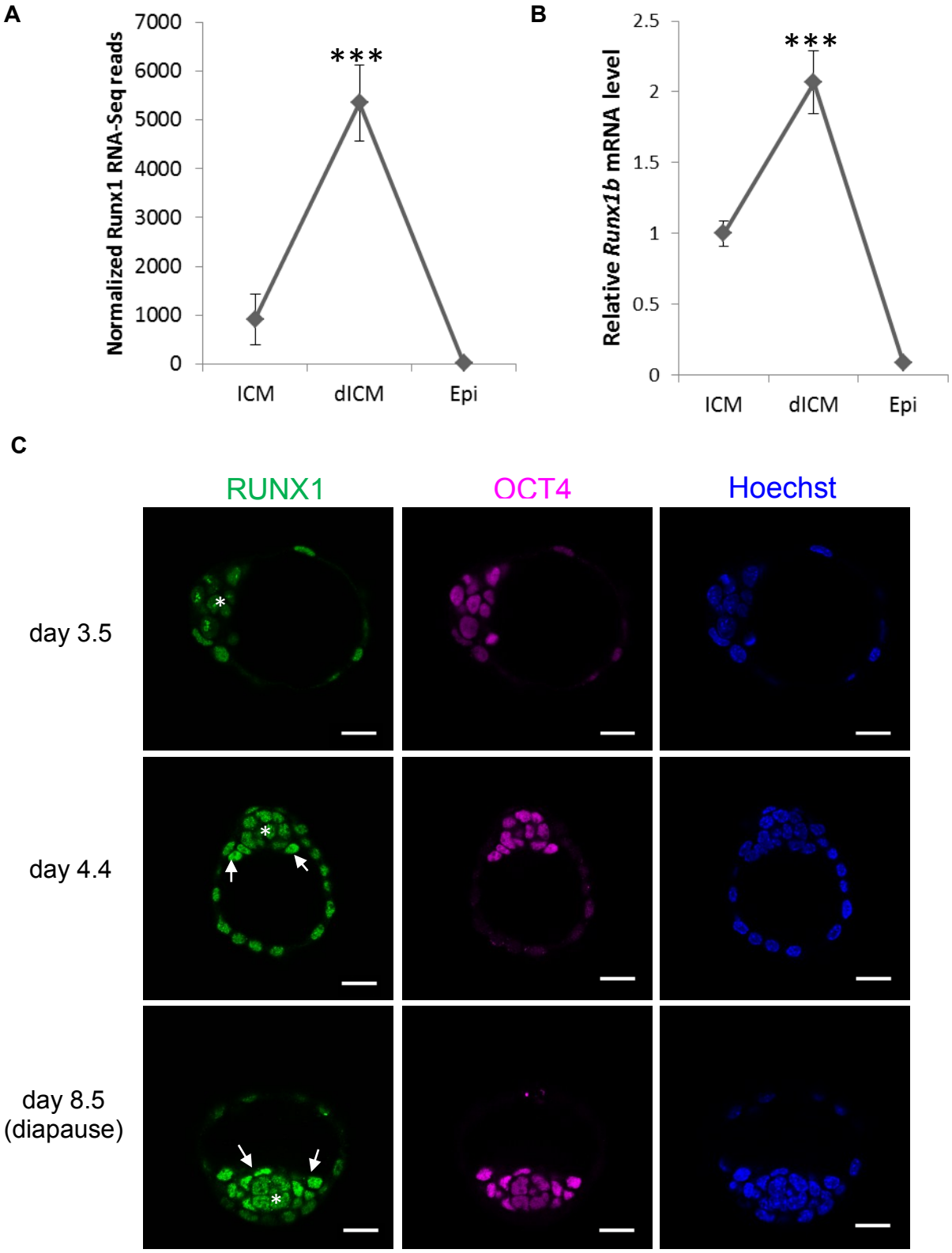


Figure 3.1

Figure 3.1 RUNX1 is induced in the ICM, notably in the Primitive Endoderm (PE) layer, of diapause embryos. (A) Average expression level of *Runx1* in ICM, dICM and Epi samples shown as number of RNA-Seq reads. (B) qRT-PCR confirmed the expression of *Runx1*b isoform in the pluripotent stem cells obtained directly from embryos. *Runx1b* transcript is induced in diapause ICM relative to day 3.5 ICM and day 6.5 Epi. Gene expression level was normalized with β-*actin*. (C) Immunostaining of day 3.5 and diapause embryos with RUNX1 (*green*) and OCT4 (*magenta*) antibodies. RUNX1 protein level is upregulated in diapause ICM, especially in the primitive endoderm (PE) layer (denoted by arrows). The epiblast subpopulation was marked by asterisks (*). OCT4 expression was upregulated in the PE cells of diapause embryos, corresponding to RUNX1 staining intensity and pattern. No correlation between RUNX1 and OCT4 staining was observed in day 4.4 peri-implantation embryos as OCT4 was expressed uniformly throughout the ICM. Hoechst staining (*blue*) shows nuclear staining. Scale: 20 μM. *** denotes p-value < 0.001 (Student's t-test).

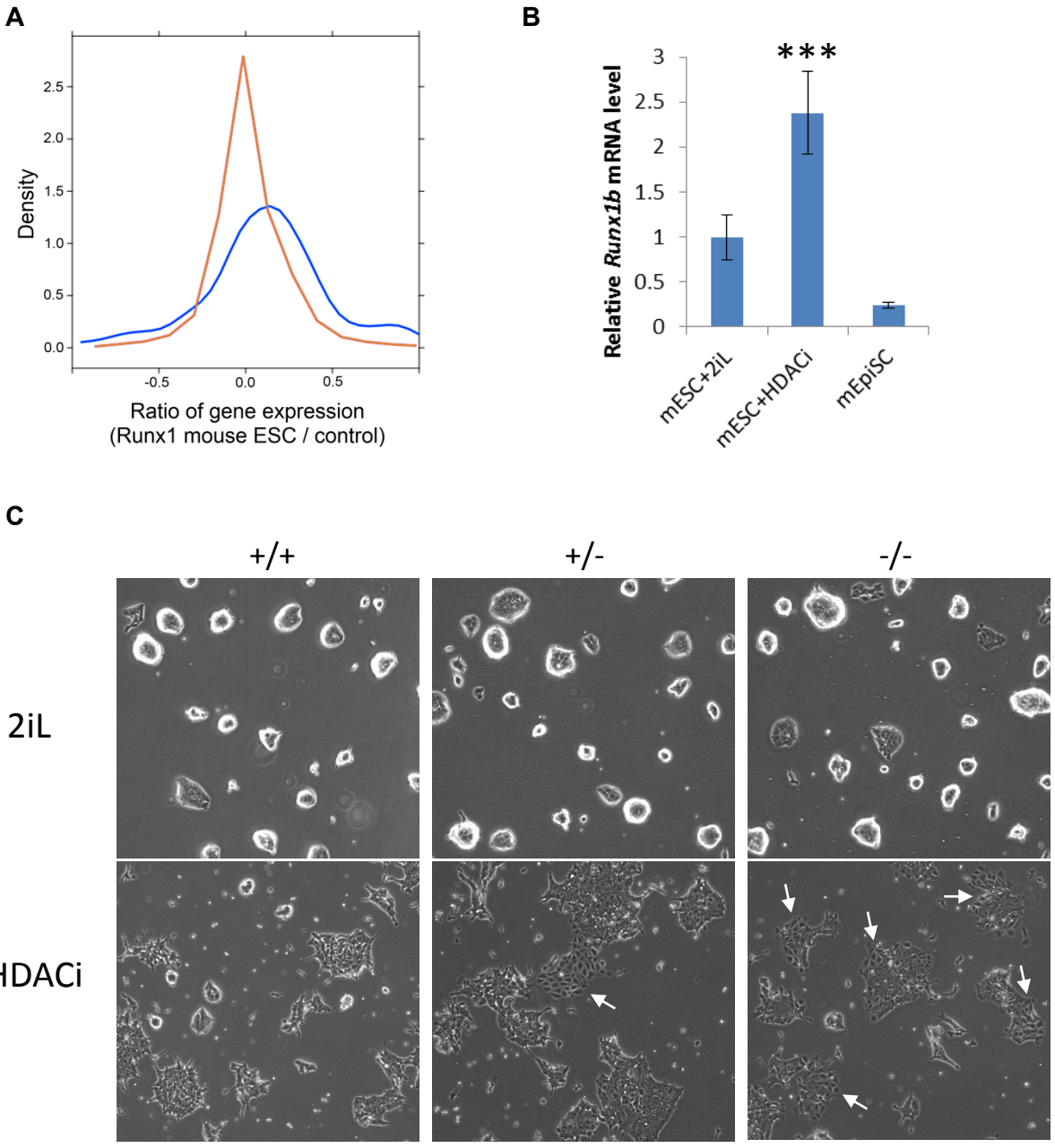

A
Density
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
0.0
0.5
Ratio of gene expression
(Runx1 mouse ESC / control)
B
Relative Runx1b mRNA level
***
3
2.5
2
1.5
1
0.5
0
mESC+2iL
mESC+HDACi
mEpiSC
C
+/+
+/-
-/-
2iL
2-HDACi

Figure 3.2

D

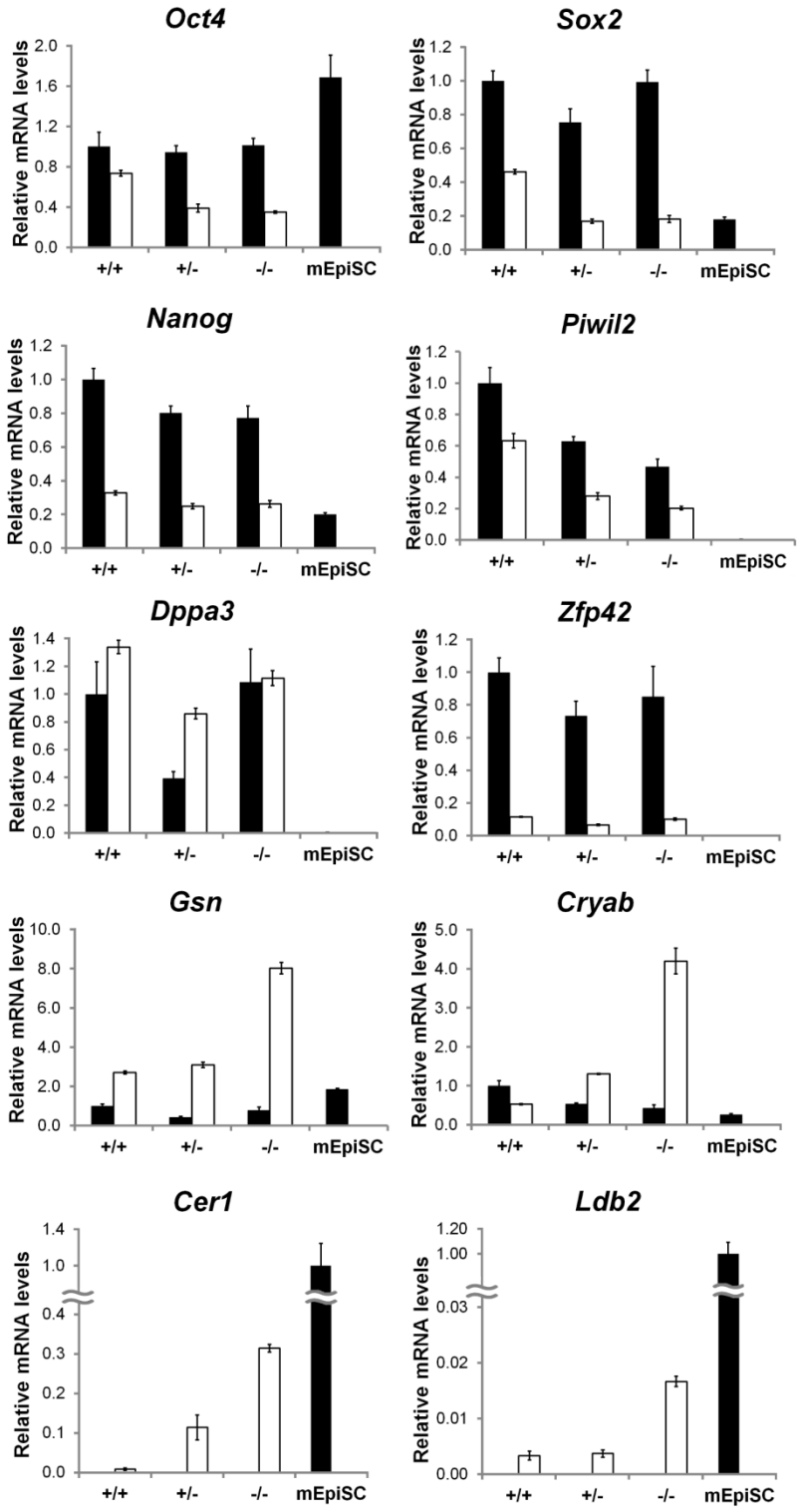
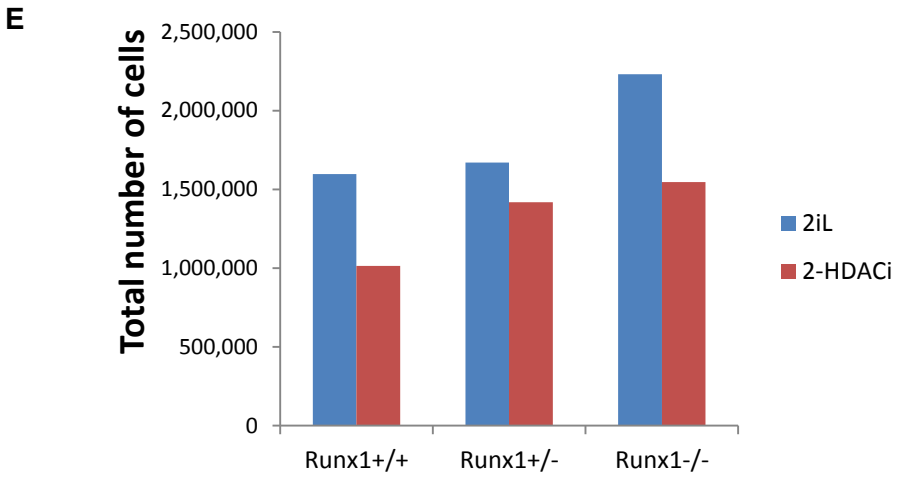

Oct4
Sox2
Nanog
Piwil2
Dppa3
Zfp42
Gsn
Cryab
Cer1
Ldb2
Relative mRNA levels
+/+
+/-
-/-
mEpiSC

Figure 3.2


E
Total number of cells
2,500,000
2,000,000
1,500,000
1,000,000
500,000
0
Runx1+/+
Runx1+/-
Runx1-/-
2iL
2-HDACi

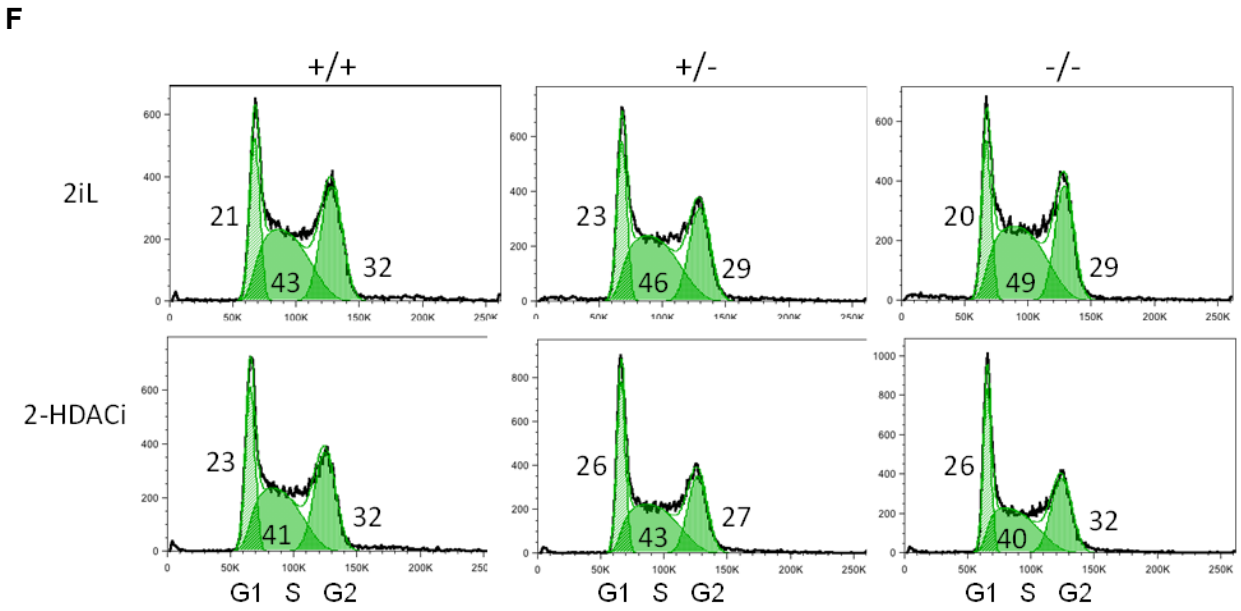

F
+/+
+/-
-/-
2iL
2-HDACi
21
43
32
23
46
29
20
49
29
23
41
32
26
43
27
26
40
32
G1 S G2
G1 S G2
G1 S G2

Figure 3.2

Figure 3.2 Mouse ESC exposed to 2-HDACi as a model to study pluripotent stem cell quiescence *in vitro*. (A) Density plots showing ratio of gene expression levels for diapause signature genes (*blue)* and all other genes (*orange*) in *Runx1*-over-expressing (*Runx1*-OE) mouse ESC compared to controls. Positive values in x-axis indicate increased expression levels in *Runx1*-OE mouse ESC. *Runx1*-OE mouse ESC upregulated a significant portion of the diapause signature (*blue*) relative to other genes (*orange*) (p-value < 10^{-10}). (B) qPCR showing *Runx1* is induced in mESC exposed to 2-HDACi. *** denotes p-value < 0.001 (Student's t-test). (C) Morphology of *Runx1* +/+, +/- and -/- mESC cultured either in 2iL or 2-HDACi conditions without phenol red and glucose. Note that *Runx1* +/+ still had several moundy colonies after 3-4 days post-passage in 2-HDACi, while all of the colonies in *Runx1*+/- and -/- mESC were flattening. The latter also displayed a tendency toward spontaneous differentiation (*white arrows*), suggesting that gene dosage of *Runx1* is important in protecting self-renewal property of mouse ESC in response to HDACi exposure. Experiments have been repeated three times with the same results. (D) qPCR with pluripotency, 2-HDACi and mEpiSC markers to check the state of ESC+2-HDACi after 4 days post-passage. Black bars represent mouse ESC in the 2iL condition and mouse EpiSC, while white bars are 2-HDACi culture conditions. Markers for pluripotency were Oct4, Sox2, and *Nanog.* Markers for mouse ESC in the naïve condition were *Zfp42/Rex1*. *Dppa3* and *Piwil2* are associated with germ line, and also are expressed in mouse ESC. Markers for ESC+2-HDACi were *Gsn* and *Cryab*. Markers for EpiSC were *Cer1* and *Ldb2*. (E) Cell proliferation assay of *Runx1* +/+, +/- and -/- mESC in naïve (2iL) and 2-HDACi condition. In all three cell lines, 2-HDACi exposure caused a decrease in cell proliferation rate. (F) Cell cycle profile of *Runx1* +/+, +/- and -/- mESC in naïve and 2-HDACi condition. As shown by Ware et al., 2009, butyrate slowed the rate of cell division of mESC without perturbing the cell cycle.

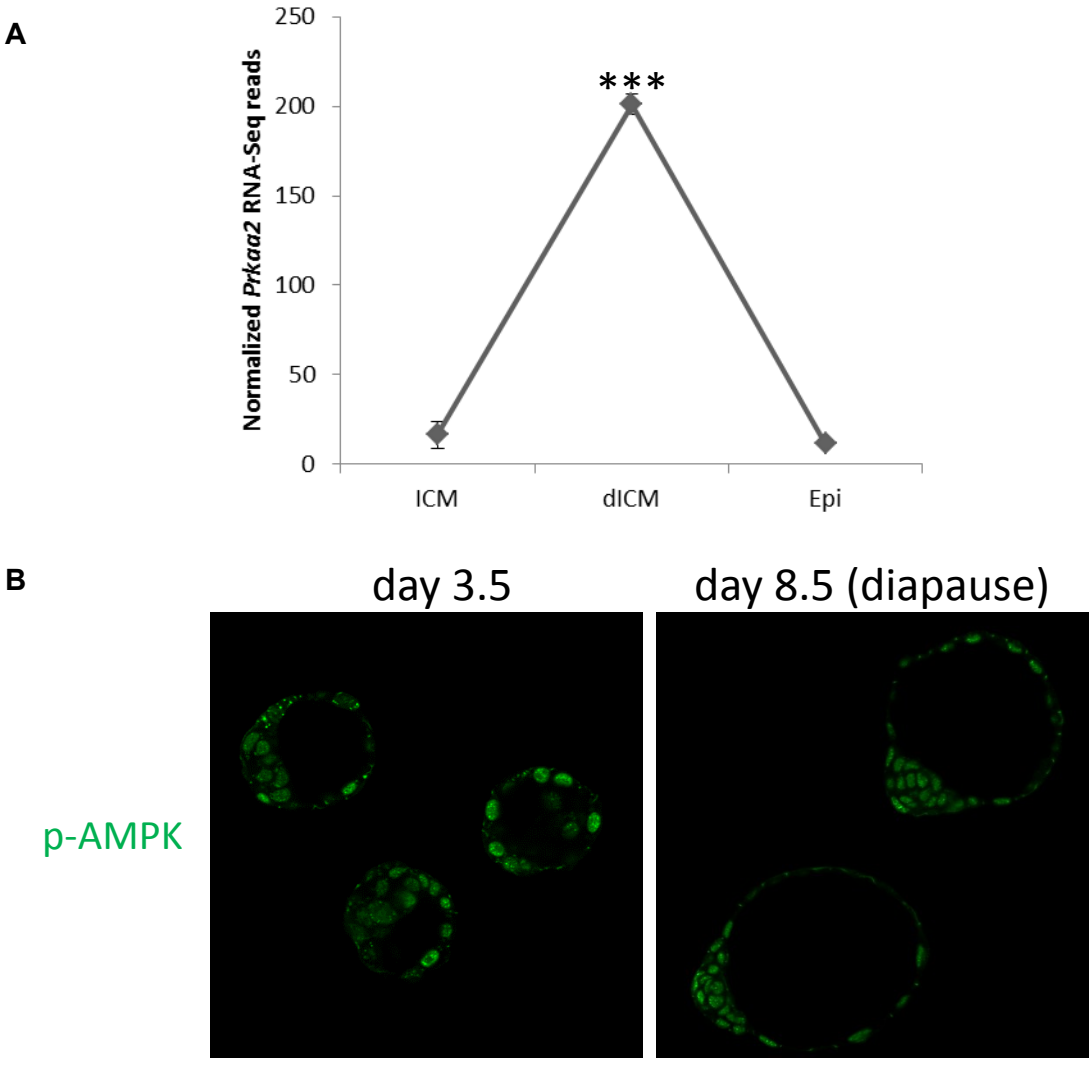


Figure 3.3 Regulation of AMPKα2 (*Prkaa2*) in the ICM of diapause embryos. (A) RNA-Seq reads of *Prkaa2* gene showing ~12-fold and ~18-fold induction in dICM relative to ICM and Epi, respectively. *** denotes p-value < 0.001 (Student's t-test). (B) Immunostaining of day 3.5 and diapause embryos with antibody against AMPKa2 p-Thr172. No significant induction of phosphorylated AMPKa2 protein level was observed in diapause embryos.

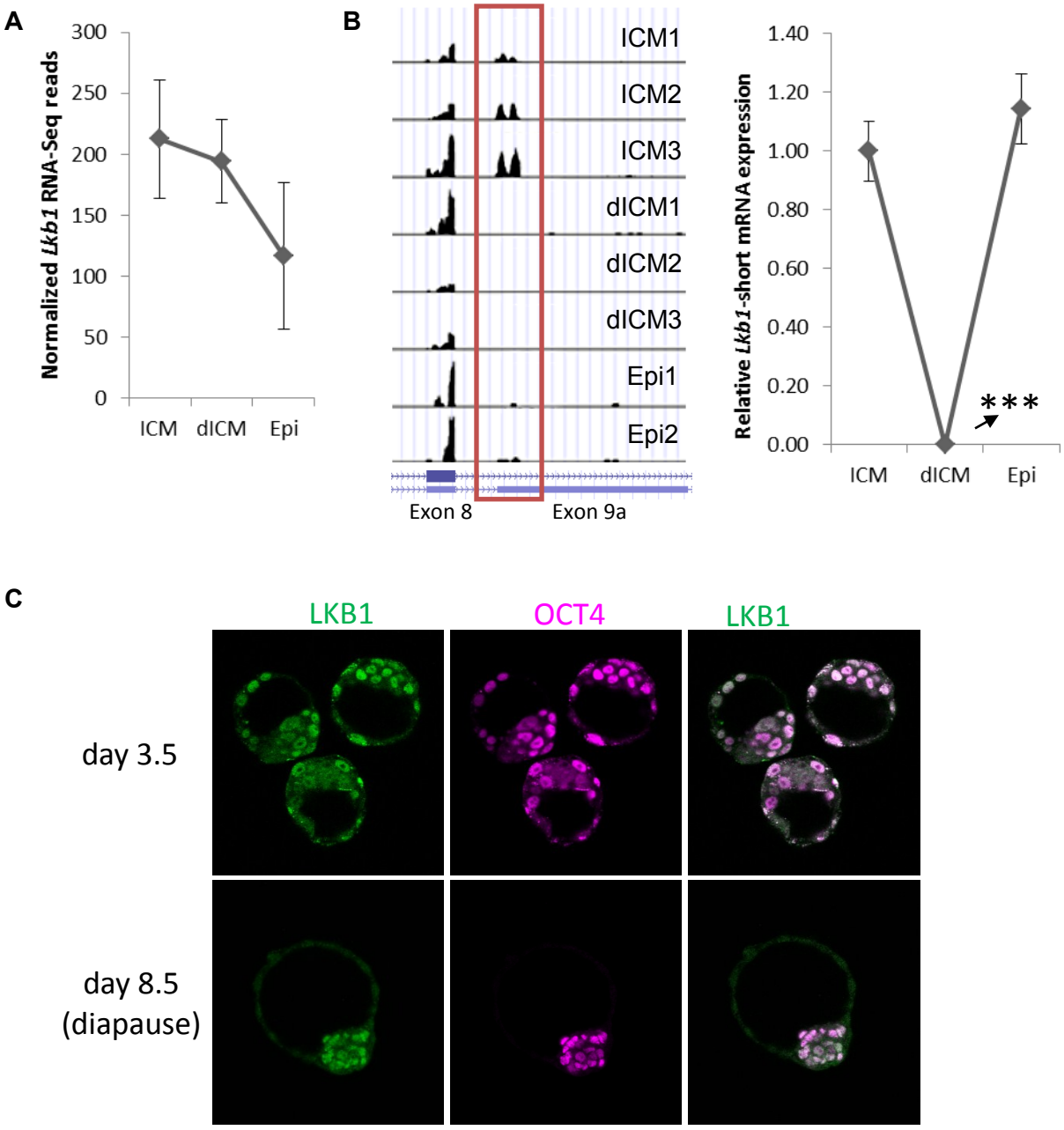


Figure 3.4 Regulation of the short isoform of LKB1 (*Lkb1*$_{Short}$) in the ICM of diapause embryos, suggesting an isoform switch of *Lkb1*. (A) The overall *Lkb1* transcript levels detected by RNA-Seq did not change between pre-implantation and diapause ICM. (B) UCSC genome browser snapshot of exons 8 and 9a of *Lkb1* showing an absence of transcript from exon 9a, which is exclusively expressed in the short isoform of *Lkb1*, in dICM samples (*left panel*). qPCR confirmation from *in vivo* samples revealed a striking, >1000 fold diapause-associated reduction in *Lkb1*$_{short}$ (*right panel*). *** denotes p-value < 0.001 (Student's t-test). (C)

Immunostaining of day 3.5 and diapause embryos with antibodies against both LKB1 isoforms and OCT4 showed a predominantly nuclear staining pattern that did not differ significantly between pre-implantation and diapause embryos.

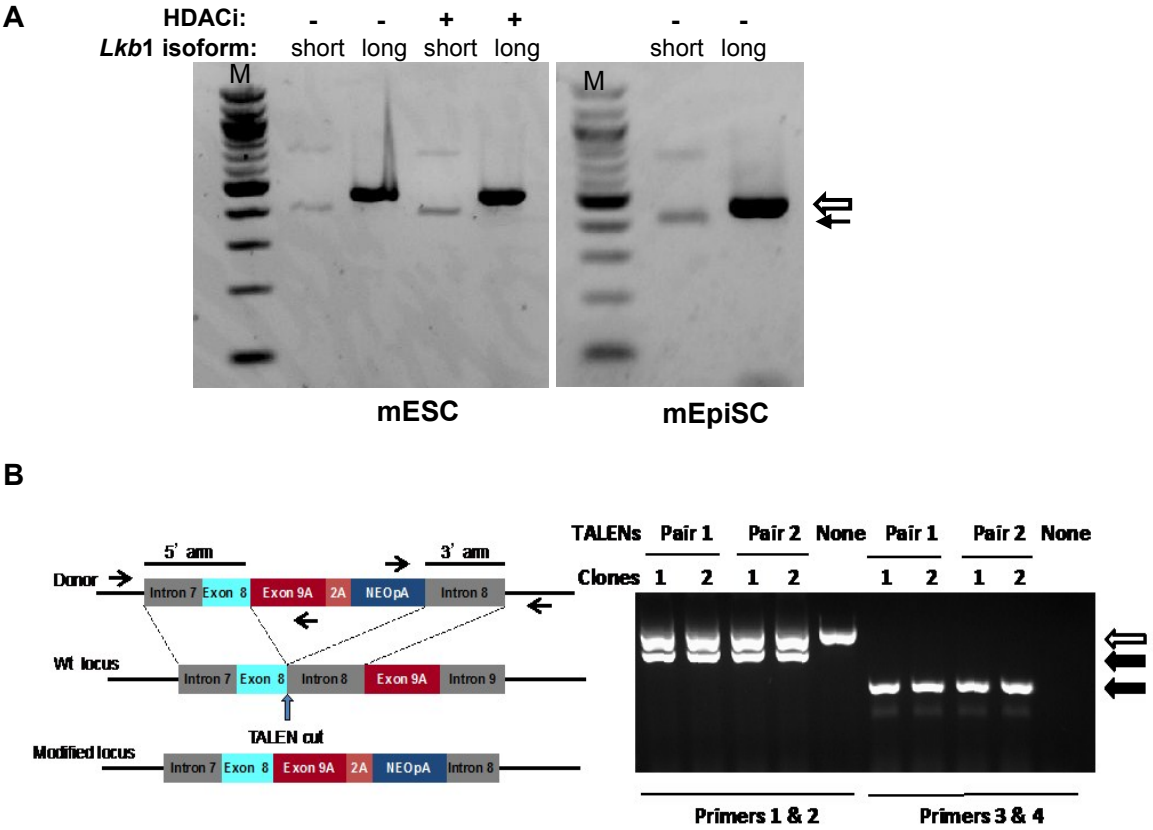


Figure 3.5 Overexpression of LKB1 short isoform in mouse ESC. (A) Semi-quantitative PCR showed that mESC in culture, either with or without exposure to HDACi (butyrate), and mouse EpiSC had a predominant expression of the long isoform of LKB1. Amplicons from long and short isoforms are indicated by white and black arrows, respectively (B) Generation of

mESC expressing $LKB1_{short}$ isoform only. Left Panel: TALEN-mediated gene editing of the endogenous *Lkb1* locus. Schematic diagram showed the donor construct (*top*), the wild-type *Lkb1* locus (*middle*) and the modified *Lkb1* locus following homologous recombination (*bottom*). Exon 8 is the last common exon of the long and short forms of LKB1. Exon 9A and 9B are short form- and long form-specific exons, respectively. The modified *Lkb1* locus will express only the short form of LKB1 since the short form-specific exon 9A is fused directly in-frame with exon 8 followed by 2A peptide and *Neo* followed by a poly A site. Arrows depict genotyping primers. Right Panel: Modification of the endogenous *Lkb1* locus in mESC. mESC were transfected with the TALEN pairs and donor plasmid. *Neo* resistant clones were analyzed for homologous recombination using PCR. Amplicons from WT and modified loci are indicated by white and black arrows respectively. Primers 1 & 2 amplify both the wild-type and modified loci. Primers 3 & 4 amplify only the modified locus. (C) Mouse ESC expressing $LKB1_{Short}$ isoform are viable, with more compact and moundy colonies than the wild-type mouse ESC.

Tables

	GSEA c3 (motif) results						RNA-Seq	
Genes	**RANK # (based on NES)**	**SIZE**	**ES**	**NES**	**p-value**	**FDR**	**dICM / ICM ratio**	**dICM / epi ratio**
Foxf2	5	98	-0.46	-1.80	0.000	0.070	10.9	14.0
Srf	7	108	-0.44	-1.78	0.000	0.062	2.6	4.3
Runx1	39	95	-0.42	-1.64	0.003	0.056	5.8	430.7
Tcf3	52	88	-0.41	-1.60	0.002	0.070	3.0	1.9
Atf6	89	60	-0.42	-1.52	0.012	0.095	1.6	4.2
Nfkb1	102	104	-0.38	-1.50	0.007	0.096	9.7	4.0
Egr2	137	79	-0.37	-1.43	0.029	0.132	142.4	1245.5
Atf3	145	232	-0.32	-1.43	0.006	0.132	32.8	96.0

Table 3.1 Summary of GSEA with the c3 motif database revealed enrichment of *Runx1* target genes in dICM compared to ICM. Gene sets were ordered by the Normalized Enrichment Score (NES) before determining the rank number. Only genes whose transcript levels were induced significantly in dICM relative to ICM and Epi, were included here (shown in the last two columns). The parameters used in GSEA and description of the result columns were noted in Methods.

Gene Sets INDUCED in diapause ICM

Gene Sets	Rank # (based on NES)	SIZE	ES	NES	p-value	FDR	Expression Dataset used
p21							
P21_P53_MIDDLE_DN	19	20	0.71	2.16	0.000	0.000	genes downregulated in dICM compared to ICM
P21_P53_ANY_DN	22	37	0.61	2.11	0.000	0.001	genes downregulated in dICM compared to ICM
P21_ANY_DN	50	28	0.54	1.82	0.003	0.017	genes downregulated in dICM compared to ICM
P21_P53_MIDDLE_DN	2	22	0.68	2.33	0.000	0.003	genes downregulated in dICM compared to Epi
P21_P53_ANY_DN	18	39	0.48	1.89	0.003	0.045	genes downregulated in dICM compared to Epi
TGFbeta							
TGF_BETA_SIGNALING_PATHWAY	56	31	-0.53	-1.66	0.008	0.103	genes upregulated in dICM compared to ICM
hypoxia							
MANALO_HYPOXIA_UP	12	54	-0.55	-1.95	0.000	0.014	genes upregulated in dICM compared to ICM
MENSE_HYPOXIA_TRANSPORTER_GENI	80	32	-0.50	-1.58	0.013	0.144	genes upregulated in dICM compared to ICM
HYPOXIA_REVIEW	179	42	-0.41	-1.42	0.062	0.229	genes upregulated in dICM compared to ICM
MENSE_HYPOXIA_UP	203	62	-0.39	-1.39	0.050	0.244	genes upregulated in dICM compared to ICM
MANALO_HYPOXIA_DN	14	70	0.55	2.23	0.000	0.000	genes downregulated in dICM compared to ICM
amino acid starvation							
PENG_LEUCINE_DEPRIVATION_UP	31	71	-0.41	-1.69	0.001	0.135	genes upregulated in dICM compared to Epi
PENG_GLUTAMINE_DEPRIVATION_DN	42	211	0.39	1.88	0.000	0.009	genes downregulated in dICM compared to ICM
PENG_LEUCINE__DEPRIVATION_DN	43	119	0.42	1.88	0.000	0.010	genes downregulated in dICM compared to ICM
PENG_RAPAMYCIN_DEPRIVATION_DN	52	158	0.39	1.79	0.000	0.021	genes downregulated in dICM compared to ICM

Gene Sets REPRESSED in diapause ICM

Gene Sets	Rank # (based on NES)	SIZE	ES	NES	p-value	FDR	Expression Dataset used
c-MYC							
YU_CMYC_DN	16	23	-0.63	-1.88	0.000	0.026	genes upregulated in dICM compared to ICM
LEE_MYC_UP*	64	34	-0.50	-1.64	0.010	0.110	genes upregulated in dICM compared to ICM
YU_CMYC_DN	8	23	-0.59	-1.87	0.002	0.101	genes upregulated in dICM compared to Epi
YU_CMYC_UP	7	28	0.71	2.40	0.000	0.000	genes downregulated in dICM compared to ICM
SCHUMACHER_MYC_UP	20	46	0.58	2.15	0.000	0.000	genes downregulated in dICM compared to ICM
COLLER_MYC_UP	92	15	0.54	1.50	0.062	0.125	genes downregulated in dICM compared to ICM
YU_CMYC_UP	29	29	0.47	1.70	0.008	0.140	genes downregulated in dICM compared to Epi
YY1							
GAY_YY1_UP	6	111	-0.52	-2.06	0.000	0.004	genes upregulated in dICM compared to ICM
GAY_YY1_DN	29	148	0.43	1.99	0.000	0.004	genes downregulated in dICM compared to ICM
GAY_YY1_DN	59	161	0.29	1.53	0.004	0.212	genes downregulated in dICM compared to Epi
ELONGIN A							
ELONGINA_KO_UP	1	135	-0.56	-2.31	0.000	0.000	genes upregulated in dICM compared to ICM
ELONGINA_KO_UP	1	139	-0.50	-2.30	0.000	0.001	genes upregulated in dICM compared to Epi
ELONGINA_KO_DN	123	140	0.30	1.35	0.011	0.235	genes downregulated in dICM compared to ICM
cell cycle							
SERUM_FIBROBLAST_CELLCYCLE	1	105	0.68	3.05	0.000	0.000	genes downregulated in dICM compared to ICM
CELL_CYCLE_KEGG	60	71	0.42	1.73	0.000	0.033	genes downregulated in dICM compared to ICM
HSA04110_CELL_CYCLE	63	93	0.40	1.70	0.000	0.040	genes downregulated in dICM compared to ICM
BRENTANI_CELL_CYCLE	82	65	0.40	1.58	0.006	0.080	genes downregulated in dICM compared to ICM
CELL_CYCLE	83	66	0.39	1.58	0.000	0.081	genes downregulated in dICM compared to ICM
SERUM_FIBROBLAST_CELLCYCLE	3	110	0.47	2.29	0.000	0.003	genes downregulated in dICM compared to Epi

Table 3.2 Summary of GSEA result using the c2 database for functional categories enriched or repressed in dICM. Gene sets were ordered by the Normalized Enrichment Score (NES) before determining the rank number. Categories of gene sets are shaded. The parameters used in GSEA were noted in Methods.

Supplementary Tables

Table S3.1. List of genes significantly induced in ICM compared to dICM and Epi.

Table S3.2. List of genes significantly repressed in ICM compared to dICM and Epi.

Table S3.3 List of genes significantly induced in dICM compared to ICM and Epi.

Table S3.4 List of genes significantly repressed in dICM compared to ICM and Epi.

Table S3.5 List of genes significantly induced in Epi compared to ICM and dICM.

Table S3.6 List of genes significantly repressed in Epi compared to ICM and dICM.

Table S3.7 Gene Ontology (GO) categories of significantly induced genes in ICM compared to dICM and Epi.

Table S3.8 GO categories of significantly repressed genes in ICM compared to dICM and Epi.

Table S3.9 GO categories of significantly induced genes in dICM compared to ICM and Epi.

Table S3.10 GO categories of significantly repressed genes in dICM compared to ICM and Epi.

Table S3.11 GO categories of significantly induced genes in Epi compared to ICM and dICM.

Table S3.12 GO categories of significantly repressed genes in Epi compared to ICM and dICM.

Chapter 4 : **Perspectives**

In this dissertation, I investigated the biology of embryonic diapause to study stem cell quiescence in pluripotent stem cells. The mechanisms that regulate diapause in mammals are poorly understood. I attempted to address several outstanding questions about the biology of embryonic diapause, including:

- Do the pluripotent stem cells from diapause embryos adopt a discrete transcriptional profile?

- Can I use an *in vitro* model to study embryonic diapause?

- What are the key regulators of embryonic diapause?

- Is embryonic diapause an obligatory state that is a part of the developmental continuum?

I used RNA-Seq to characterize the transcriptional profiles of pluripotent cells from 3 distinct developmental stages: day 3.5 pre-implantation embryos, day 6.5 post-implantation embryos, and day 8.5 diapause embryos. Prior to this dissertation, there were only two well-defined pluripotent stem cell states: naïve and primed. The RNA-Seq results established a third discrete pluripotent stem cell state characterized by pluripotent stem cells from diapause embryos. Therefore, I have defined a novel, quiescence-associated, pluripotent stem cell state, reflected by diapause ICM.

The transcriptional profiles of these *in vivo* embryonic pluripotent stem cells can be used to create a developmental yardstick that allows pluripotent stem cells in culture to be appended to their developmental corollaries *in vivo* (Figure 4.1). Based on the finding that butyrate and other HDACi induce mESC and hESC to converge toward an intermediate stem cell state (Ware et al., 2009), I tested the hypothesis that this “intermediate” state might reflect mechanisms similar to those that support diapause. Using the developmental yardstick, I have shown that the transcriptional profile of mouse ESC treated with HDACi (particularly a combination of butyrate and SAHA) aligns closely with that of pluripotent stem cells from diapause embryos. The results described in this dissertation suggest that the quiescence-associated state can be modeled *in vitro*. These findings open new avenues to investigate embryonic diapause using an *in vitro* model.

I used the embryonic diapause phenomenon as a paradigm for understanding stem cell quiescence (Figure 4.2). The RNA-Seq results identify *Runx1* as one of the key mediators of the diapause-associated transcriptional program. Other key signatures of pluripotent stem cells in diapause are the loss of a splice isoform of *Lkb1* and upregulation of the catalytic subunit of a well-known metabolic sensor AMPK. Using the *in vitro* model of diapause, I investigated the role of *Runx1* in embryonic diapause.

Using the publicly available resource (Hunter et al., 2008) and the developmental yardstick, I also attempted to answer a fundamental question whether embryonic diapause occurs as a detour or pit stop in normal development. Two findings described in my dissertation strongly suggest the latter. First, the peri-implantation embryos that have not undergone diapause from the Hunter study bridge the border between the ICM and the diapause quadrants. Second, induction of RUNX1, a potential key regulator in embryonic diapause, in day 4.4 peri-implantation blastocysts also supports the idea that diapause is an obligatory state that is a part of developmental continuum.

A previous study by Hamatani et al. (2004) attempted to investigate the genes that mediate embryonic diapause phenomenon and its activation with estrogen for implantation in the uterus. The genes differentially expressed between diapause blastocysts and activated blastocysts were categorized into six functional groups: genes involved in cell-cycle or cell-proliferation control; energy pathways and carbohydrate metabolism; signaling; nuclear transport; chromatin remodeling; and adhesion. Whole (unfractionated) blastocysts were used in the analysis. This study, therefore, could not distinguish which subset of cells contributed to the gene expression changes.

In order to investigate and compare ESC with their developmental equivalents *in vivo*, the pluripotent cell populations from which ESC were derived have to be isolated. Only then can the contribution from other cell populations in the gene expression profiles be eliminated. Using an established approach called immunosurgery, I isolated the ICM from day 3.5 pre-implantation and diapause blastocysts from the remaining extraembryonic trophectoderm (TE) layer. Brons et al. (2007) and Hunter et al. (2008) also followed a similar approach to characterize the transcriptional profiles of pluripotent stem cells from the pre-implantation and diapause embryos. This approach assumed that the ICM populations from these two embryonic stages

are still homogenous enough to present a transcriptional profile that is unique to the pluripotent stem cell subset. However, a few studies have shown that ICM of day 3.5 blastocysts is not homogenous, but composed of cells either fated to be epiblast or PE in a random "salt-and-pepper" pattern (Chazaud et al., 2006, Rossant et al., 2003). By day 4.5, the PE cells have differentiated and sorted to line the blastocoel cavity, whereas the epiblast remains as a mass of undifferentiated cells (Yamanaka et al., 2006). To my knowledge there is no technique to completely separate the PE layer from epiblast. Even the use of laser-drilling technology, which is similar to the laser-capture microdissection method used for isolating pockets of cancer cells from adjacent normal tissues, would not completely ablate the extra-embryonic tissue around the ICM cells (Cortes et al., 2006, Tanaka et al., 2006). Therefore, the RNA-Seq data generated for diapause ICM described here included cells that have differentiated into PE cells.

Despite the limitation in technique to isolate a "pure" population of pluripotent stem cell from diapause embryos, processing the acquired data *in silico* can serve as an acceptable alternative to obtain a list of diapause-associated genes that are unique to the stem cell population. Previous microarray studies using XEN cell line derived from the PE lineage and early pre-implantation embryos provide a resource for PE markers (Gerbe et al., 2008, Kunath et al., 2005). As expected, I observed several PE-associated genes being induced in dICM, for example: *Dab2* and *Lrpap1*. Interestingly, mouse ESC exposed to 2-HDACi also have elevated mRNA levels of some PE markers. Preliminary analysis identified less than 3% of the genes that are induced in dICM and mouse ESC+2-HDACi relative to their respective experimental controls are PE-associated genes. Further analysis filtering the PE markers from these two datasets and a repeat of projection studies described in Chapter 2 also confirmed that the observations described prior to *in silico* processing remained consistent. This analysis suggests that the gene expression profiles of dICM and mouse ESC exposed to 2-HDACi were contributed mainly by the pluripotent stem cell population and can be used to study stem cell quiescence.

As described in chapter 3, I noted that RUNX1 is induced prominently in the PE layer of both day 4.4 peri-implantation and diapause embryos. A number of questions related to this observation remain to be answered. First, is the induction of RUNX1 in PE a cell-autonomous or non-autonomous requirement in the embryos? The PE cells might require RUNX1 cell-autonomously to prevent them from differentiating further into a more committed lineage.

Alternatively, PE cells might regulate the pluripotency of the epiblast population through intercellular communication. It would be interesting to explore whether PE cells have any role in maintaining the stem cell quiescence of the epiblast population. Second, is RUNX1 required to maintain stem cell quiescence *in vivo*? Although the *in vitro* model of diapause suggests that RUNX1 is required to keep the cells undifferentiated longer in 2-HDACi condition, a true assessment of the role of RUNX1 in stem cell quiescence would be to see if embryos lacking RUNX1 can still enter diapause. Different possible outcomes might occur in the *Runx1* knockout embryos induced to go into diapause: i) The Runx1-/- genotype is not observed among diapause embryos; ii) The *Runx1*-/- genotype occurs at the expected frequencies in diapause embryos and *Runx1*-/- embryos are morphologically indistinguishable from +/- and +/+ genotypes; and iii) *Runx1-/-* embryos can enter diapause but are qualitatively abnormal. A variety of abnormalities may be envisioned, including an attempt by the embryo to progress through development during diapause, possibly leading to abnormal structures resembling teratomas or teratocarcinomas. Alternatively, progression through development in the absence of implantation may lead to significant cell death.

Considering the substantial advantages of using RNA-Seq compared to microarray for gene expression profiling, I realized that I have not utilized the full potential of my RNA-Seq data. Since the cDNA libraries used for my RNA-Seq study were generated using both poly-dT and random primers, this RNA-Seq data also contain information about non-coding RNA (ncRNA) from the 3 pluripotent stem cell states. ncRNA have been implicated in regulating many cellular processes, and therefore, characterization of the ncRNA in the pluripotent stem cells would give a novel insight into how ncRNA might play a role in stem cell quiescence (Mercer et al., 2009). The length of RNA included in this RNA-Seq study was limited to at least ~130 nucleotides due to size selection process during cDNA library preparation. Therefore, smaller ncRNA such as miRNA, piRNA and snoRNA are not included in the study. Sample preparation specifically tailored for small ncRNA is needed in order to study the small ncRNA. In addition to the extensive efforts of studying stem cell quiescence in embryonic diapause by characterizing the transcriptional profiles, it would also be interesting to study the regulation of its epigenetic status, such as histone methylation marks that are known to be fine-tuned in cultured pluripotent stem cells (Pan et al., 2007, Zhao et al., 2007).

There is still much to learn about stem cell quiescence and embryonic diapause. Further examination of embryos harvested at different time points in embryonic diapause, for example, would give us a clue as to the genes that are involved early or later in this suspended animation state. While my work presented here demonstrates some of the candidate key regulators in diapause, a number of other interesting observations remain to be investigated in the future. An increase of transcripts involved in cholesterol synthetic pathway (Figure 4.2), for instance, is an intriguing finding as I would expect the embryos to suppress any major anabolic pathway under energetic stress (Henin et al., 1995). There are two possible explanations for this seemingly controversial issue. Embryos in diapause state might be preparing for a major differentiation process following implantation once the environmental condition improves. Alternatively, the embryos with their declining, maternally-supplied cholesterol storage are forced to survive and keep dividing slowly; the latter process requires cholesterol for synthesis of cellular membranes. Further examination of this controversial process and other pathways regulated in diapause would enhance our understanding of stem cell quiescence in diapause.

The study of stem cell quiescence may have some merits in clinical applications. ESC are also known to employ overlapping signaling networks with cancer cells (Ben-Porath et al., 2008, Dreesen and Brivanlou, 2007, Wong et al., 2008), which survive well in hypoxic conditions (reviewed by Hanahan and Weinberg, 2000). Understanding the mechanism of stem cell quiescence in the embryonic diapause model system would give an insight into how cancer stem cells (CSC) adapt this strategy in order to survive in unfavorable conditions and how to improve cancer therapies that better target the CSC. A recurring goal in refining *in vitro* fertilization (IVF) techniques is to improve pregnancy rate following IVF by enhancing embryo survival and implantation rate after freezing-thawing or an extended culture *in vitro* prior to uterine transfer. The body of work here proposed an idea to solve the pressing problem in IVF by trying to artificially induce *in* vitro-fertilized embryos to enter into a diapause-like state until the maternal uterine environment is more receptive for implantation.

In conclusion, I have shown that the work described in this dissertation is replete with innovative ideas and approaches. Starting from the insight that butyrate and other HDACi induce mESC and hESC to converge toward an intermediate stem cell state (Ware et al., 2009), I demonstrated by using RNA-Seq technology on small aggregates of pluripotent stem cells dissected from early mouse embryos that this “intermediate” state reflects mechanisms similar

to those that support diapause. The use of my findings to construct a “yardstick” allows pluripotent stem cells in culture to be appended to their developmental corollaries *in vivo*. I used embryonic diapause as a paradigm for understanding stem cell quiescence and proposed the use of *in vitro* models to further our understanding of stem cell quiescence in embryonic diapause.

Figures

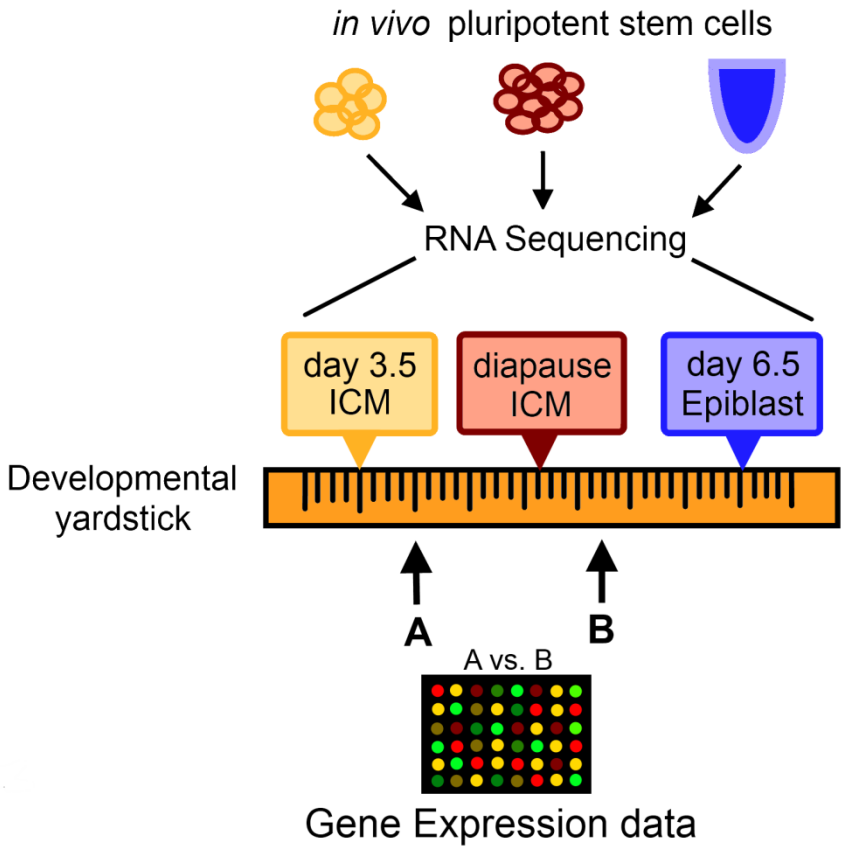


Figure 4.1 Graphical abstract of the use of developmental yardstick. RNA-Seq data of pluripotent stem cells isolated from pre-implantation (day 3.5), diapause and post-implantation (day 6.5) embryos were used to provide a gold standard against which cultured pluripotent stem cells can be compared. Placing the ICM, dICM and epiblast samples on a single-dimensional scale, I generated a linear developmental yardstick that can be used to estimate the developmental age of samples of interest. The resulting yardstick provides a framework for comparing pluripotent stem cell lines that differ in tissue source, method of derivation, or culture conditions.

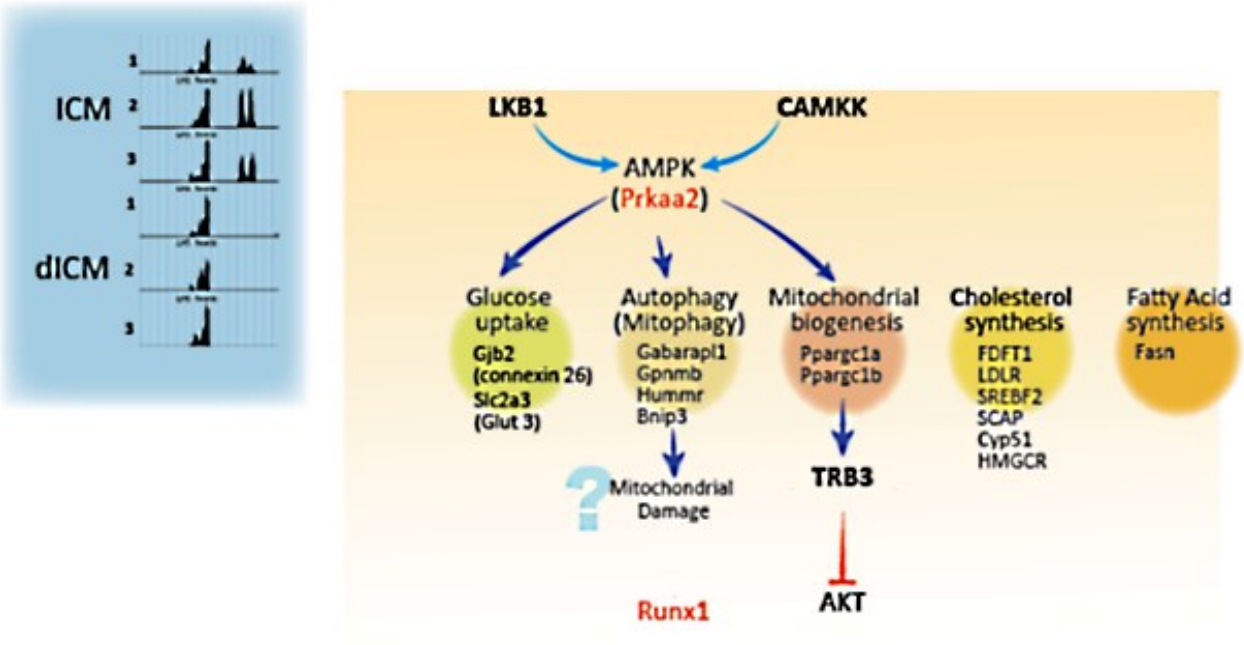


Figure 4.2 A proposed mechanism of stem cell quiescence using embryonic diapause as a model. LKB1 is the predominant upstream activator of AMPK in diapause ICM, and its activity may be regulated via an isoform switch (shown on the left panel). I proposed that AMPK is a key metabolic sensor in diapause. Consistent with AMPK activation, my RNA-Seq results showed that diapause ICM is associated with upregulation of genes involved in autophagy as well as mitochondrial biogenesis and glucose uptake. An increase of transcripts involved in the cholesterol and fatty acid synthetic pathway is an intriguing finding as I would expect the embryos to suppress any major anabolic pathway under energetic stress. Transcription factor RUNX1 might function as a key mediator of stem cell quiescence in diapause, regulating the expression of other genes supporting the function of diapause.

Bibliography

Affar, el B., Gay, F., Shi, Y., Liu, H., Huarte, M., Wu, S., Collins, T., Li, E., and Shi, Y. (2006). Essential dosage-dependent functions of the transcription factor yin yang 1 in late embryonic development and cell cycle progression. Mol. Cell Biol. *26*, 3565-3581.

Alessi, D.R., Sakamoto, K., and Bayascas, J.R. (2006). LKB1-dependent signaling pathways. Annu. Rev. Biochem. *75*, 137-163.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Angelo, G. and Van Gilst, M.R. (2009). Starvation protects germline stem cells and extends reproductive longevity in C. elegans. Science *326*, 954-958.

Appleford, P.J. and Woollard, A. (2009). RUNX genes find a niche in stem cell biology. J. Cell Bioc. *108*, 14-21.

Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., and Surani, M.A. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature *461*, 1292–1295.

Bellringer, J.F., Pratt, H.P.M., and Keverne, E.B. (1980). Involvement of the vomeronasal organ and prolactin in pheromonal induction of delayed implantation in mice. J. Reprod. Fertil. *59*, 223-228.

Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., Bell, G.W., Regev, A., and Weinberg, R.A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat. Genet. *40*, 499-507.

Berthois, Y., Katzenellenbogen, J.A., and Katzenellenbogen, B.S. (1986). Phenol red in tissue culture media is a weak estrogen: implications concerning the study of estrogen-responsive cells in culture. PNAS *83*, 2496-2500.

Bertrand-Phillippe, M., Ruddell, R.G., Arthur, M.J., Thomas, J., Mungalsingh, N., and Mann, D.A. (2004). Regulation of tissue inhibitor of metalloproteinase 1 gene transcription by RUNX1 and RUNX2. J. Biol. Chem. *279*, 24530-24539.

Bradley, A., Evans, M., Kaufman, M., and Robertons, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255-256.

Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., and Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.

Brook, F.A. and Gardner, R.L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. PNAS *94*, 5709-5712.

Brugarolas, J., Moberg, K., Boyd, S.D., Taya, Y., Jacks, T., and Lees, J.A. (1999). Inhibition of cyclin-dependent kinase 2 by p21 is necessary for retinoblastoma protein-mediated G1 arrest after γ-irradiation. PNAS *96*, 1002–1007.

Bungard, D., Fuerth, B.J., Zeng, P.Y., Faubert, B., Maas, N.L., Viollet, B., Carling, D., Thompson, C.B., Jones, R.G., and Berger, S.L. (2010). Signaling kinase AMPK activates stress-promoted transcription via histone H2B phosphorylation. Science *329*, 1201-1205.

Calvi, L.M., Adams, G.B., Weibrecht, K.W., Weber, J.M., Olson, D.P., Knight, M.C., Martin, R.P., Schipani, E., Divieti, P., Bringhurst, F.R., et al. (2003). Osteoblastic cells regulate the haematopoietic stem cell niche. Nature *425*, 841-846.

Challen, G.A. and Goodell, M.A. (2010). Runx1 isoforms show differential expression patterns during hematopoietic development but have similar functional effects in adult hematopoietic stem cells. Exp. Hematol. *38*, 403-416.

Charge S. B. and Rudnicki M. A. (2004). Cellular and molecular regulation of muscle regeneration. Physiol. Rev. *84*, 209-238.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev. Cell *10*, 615-624.

Cheshier, S.H., Morrison, S.J., Liao, X., and Weissman, I.L. (1999). *In vivo* proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. PNAS *96*, 3120-3125.

Coller, H.A., Sang, L., and Roberts, J.M. (2006). A new description of cellular quiescence. PLoS Biol. 4*,* e83.

Coller, H.A. (2011). Cell biology. The essence of quiescence. Science *334*, 1074-1075.

Conaway, C.H. (1971). Ecological adaptation and mammalian reproduction. Biol. Reprod. *4,* 239-247.

Cortes, J.L., Cobo, F., Catalina, P., Nieto, A., Cabrera, C., Montes, R., Barnie, A.H., and Concha, A. (2006). Evaluation of the laser technique method to isolate the inner cell mass of murine blastocysts. Biotechnol. Appl. Biochem. *46*, 205–209.

Daniel, J.C. Jr. (1970). Dormant embryos of mammals. BioScience *20*, 411-415.

De Cegli, R., Romito, A., Iacobacci, S., Mao, L., Lauria, M., Fedele, A.O., Klose, J., Borel, C., Descombes, P., Antonarakis, S.E., et al (2010). A mouse embryonic stem cell bank for inducible overexpression of human chromosome 21 genes. Genome Biology *11*, R64.

Deguchi, K., Yagi, H., Inada, M., Yoshizaki, K., Kishimoto, T., and Komori, T. (1999). Excessive extramedullary hematopoiesis in Cbfa1-deficient mice with a congenital lack of bone marrow. Biochem. Biophys. Res. Commun. *255*, 352-359.

Denison, F.C., Hiscock, N.J., Carling, D., and Woods, A. (2009). Characterization of an alternative splice variant of LKB1. J. Biol. Chem. *284*, 67–76.

Dreesen, O. and Brivanlou, A.H. (2007). Signaling pathways in cancer and embryonic stem cells. Stem Cell Rev. *3*, 7-17.

Dvash, T., Lavon, N., and Fan, G. (2010). Variations of X chromosome inactivation occur in early passages of female human embryonic stem cells. PLoS ONE *5*, e11330.

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

Fischer, B. and Bavister, B.D. (1999). Oxygen tension in the oviduct and uterus of rhesus monkeys, hamsters and rabbits. J. Reprod. Fertil. *99*, 673-679.

Foudi, A., Hochedlinger, K., Van Buren, D., Schindler, J.W., Jaenisch, R., Carey, V., and Hock, H. (2009). Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. Nat. Biotechnol. *27*, 84-90.

Fuchs, E. (2009). The tortoise and the hair: slow-cycling cells in the stem cell race. Cell *137*, 811-819.

Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., et al. (2010). The UCSC Genome Browser database: update 2011. Nucleic Acids Res. *39*, D876-882.

Fukada, S., Yamaguchi, M., Kokubo, H., Ogawa, R., Uezumi, A., Yoneda, T., Matev, M.M., Motohashi, N., Ito, T., Zolkiewska, A., et al. (2011). Hesr1 and Hesr3 are essential to

generate undifferentiated quiescent satellite cells and to maintain satellite cell numbers. Development *138,* 4609-4619.

Gan, B., Hu, J., Jiang, S., Liu, Y., Sahin, E., Zhuang, L., Fletcher-Sananikone, E., Colla, S., Wang, Y.A., Chin, L., et al. (2010). Lkb1 regulates quiescence and metabolic homeostasis of haematopoietic stem cells. Nature *20468*, 701-704.

Gardner, D.K., Vella, P., Lane, M., Wagley, L., Schlenker, T., and Schoolcraft, W.B. (1998). Culture and transfer of human blastocysts increases implantation rates and reduces the need for multiple embryo transfers. Fertil. Steril. *69*, 84-88.

Gerbe, F., Cox, B., Rossant, J., and Chazaud, C. (2008). Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. Dev. Biol. *313*, 594–602.

Goda, N., Ryan, H.E., Khadivi, B., McNulty, W., Rickert, R.C., and Johnson, R.S. (2003). Hypoxia-inducible factor 1alpha is essential for cell cycle arrest during hypoxia. Mol. Cell Biol. *23*, 359-369.

Grabarek, J.B., Zyzynska, K., Saiz, N., Piliszek, A., Frankenberg, S., Nichols, J., Hadjantonakis, A.K., and Plusa, B. (2012). Differential plasticity of epiblast and primitive endoderm precursors within the ICM of the early mouse embryo. Development *139*, 129-39.

Grinsted, J. and Avery, B. (1996). A sporadic case of delayed implantation after in-vitro fertilization in the human? Hum. Reprod. *11*, 651-654.

Growney, J.D., Shigematsu, H., Li, Z., Lee, B.H., Adelsperger, J., Rowan, R., Curley, D.P., Kutok, J.L., Akashi, K., Williams, I.R., et al. (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. Blood *106*, 494–504.

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 *136*, 1063–1069.

Gurumurthy, S., Xie, S.Z., Alagesan, B., Kim, J., Yusuf, R.Z., Saez, B., Tzatsos, A., Ozsolak, F., Milos, P., Ferrari, F., et al. (2010). The Lkb1 metabolic sensor maintains haematopoietic stem cell survival. Nature *468*, 659-663.

Hakem, R., de la Pompa, J.L., Sirard, C., Mo, R., Woo, M., Hakem, A., Wakeham, A., Potter, J., Reitmair, A., Billia, F., et al. (1996). The tumor suppressor Brca1 is required for embryonic cellular proliferation in the mouse. Cell *85*, 1009-1023.

Hamatani, T., Daikoku, T., Wang, H., Matsumoto, H., Carter, M.G., Ko, M.S.H., and Dey, S.K. (2004). Global gene expression analysis identifies molecular pathways distinguishing blastocyst dormancy and activation. PNAS *101*, 10326-10331.

Hanahan, D. and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

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.

Hanna, J.H., Saha, K., and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell *143*, 508-525.

Hatzfeld, J., Li, M.L., Brown, E.L., Sookdeo, H., Levesque, J.P., O'Toole, T., Gurney, C., Clark, S.C., and Hatzfeld, A. (1991). Release of early human hematopoietic progenitors from quiescence by antisense transforming growth factor beta 1 or Rb oligonucleotides. J. Exp. Med. *174*, 925-929.

Hawley, S.A., Davison, M., Woods, A., Davies, S.P., Beri, R.K., Carling, D., and Hardie, D.G. (1996). Characterization of the AMP-activated protein kinase kinase from rat liver and

identification of threonine 172 as the major site at which it phosphorylates AMP-activated protein kinase. J. Biol. Chem. *271*, 27879–27887.

Hawley, S.A., Boudeau, J., Reid, J.L., Mustard, K.J., Udd, L., Mäkelä, T.P., Alessi, D.R., and Hardie, D.G. (2003). Complexes between the LKB1 tumor suppressor, STRAD alpha/beta and MO25 alpha/beta are upstream kinases in the AMP-activated protein kinase cascade. J. Biol. *2*, 28.

Henin, N., Vincent, M.F., Gruber, H.E., and Van den Berghe, G. (1995). Inhibition of fatty acid and cholesterol synthesis by stimulation of AMP-activated protein kinase. FASEB J. *9*, 541–546.

Henin, N., Vincent, M.F., and Van den Berghe, G. (1996). Stimulation of rat liver AMP-activated protein kinase by AMP analogues. Biochim. Biophys. Acta. *1290*, 197–203.

Hermitte, F., Brunet de la Grange, P., Belloc, F., Praloran, V., and Ivanovic, Z. (2006). Very low O2 concentration (0.1%) favors G0 return of dividing CD34+ cells. Stem Cells *24*, 65-73.

Hubbard, T.J.P., Aken, B.L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., et al. (2009). Ensembl 2009. Nucleic Acids Res. *37*, D690-697.

Hunter, S.M. and Evans, M. (1999). Non-surgical method for the induction of delayed implantation and recovery of viable blastocysts in rats and mice by the use of tamoxifen and Depo-Provera. Mol. Reprod. Devel. *52*, 29–32.

Hunter, S.M., Mansergh, F.C., and Evans, M.J. (2008). Optimization of minuscule samples for use with cDNA microarrays. J Biochem. Biophys. Methods *70*, 1048-1058.

Iriuchishima, H., Takubo, K., Matsuoka, S., Onoyama, I., Nakayama, K.I., Nojima, Y., and Suda, T. (2011). Ex vivo maintenance of hematopoietic stem cells by quiescence induction through Fbxw7alpha; overexpression. Blood *117,* 2373-2377.

Jacob, B., Osato, M., Yamashita, N., Wang, C.Q., Taniuchi, I., Littman, D.R., Asou, N., and Ito, Y. (2010). Stem cell exhaustion due to Runx1 deficiency is prevented by Evi5 activation in leukemogenesis. Blood *115*, 1610-1620.

Ji, Z., Lee, J. Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. PNAS *106*, 7028-7033.

Lee, J.E., Oh, H.A., Song, H., Jun, J.H., Roh, C.R., Xie, H., Dey, S.K., and Lim, H.J. (2011). Autophagy regulates embryonic survival during delayed implantation. Endocrinology *152*, 2067-2075.

Kho, A.T., Zhao, Q., Cai, Z., Butte, A.J., Kim, J.Y., Pomeroy, S.L., Rowitch, D.H., and Kohane, I.S. (2004). Conserved mechanisms across development and tumorigenesis revealed by a mouse development perspective of human cancers. Genes Dev. *18*, 629-640.

Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y., Bronson, R.T., Gao, Y.H., Inada, M., et al. (1997). Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell *89*, 755-764.

Koshiji, M., Kageyama, Y., Pete, E.A., Horikawa, I., Barrett, J.C., and Huang, L.E. (2004). HIF-1alpha induces cell cycle arrest by functionally counteracting Myc. EMBO J. *23*, 1949-1956.

Kunath, T., Arnaud, D., Uy, G.D., Okamoto, I., Chureau, C., Yamanaka, Y., Heard, E., Gardner, R.L., Avner, P., and Rossant, J. (2005). Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. Development *132*, 1649-1661.

Lacaud, G., Gore, L., Kennedy, M., Kouskoff, V., Kingsley, P., Hogan, C., Carlsson, L., Speck, N., Palis, J., and Keller, G. (2002). Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro. Blood *100*, 458-466.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Laporte, D., Lebaudy, A., Sahin, A., Pinson, B., Ceschin, J., Daignan-Fornier, B., and Sagot, I. (2011). Metabolic status rather than cell cycle signals control quiescence entry and exit. JCB *192*, 949-957.

Lataste, M.F. (1891). Des variations de dureé de la gestation chez les mammifières et de circonstances qui déterminant ces variations: théorie de la gestation retardée. C. R. Seances Soc. Biol. Fil. *9*, 21-31.

Levine, B. and Klionsky, D.J. (2004). Development by self-digestion: molecular mechanisms and biological functions of autophagy. Dev. Cell *6*, 463-477.

Levine, B. and Kroemer, G. (2008). Autophagy in the pathogenesis of disease. Cell *132*, 27-42.

Li, L. and Bhatia, R. (2011). Stem cell quiescence. Clin. Cancer Res. *17*, 4936-4941.

Lindenfors, P., Dalen, L., Angerbjorn, A. (2003). The monophyletic origin of delayed implantation and its implications. Evolution *57*, 1952-1956.

Look, A.T. (1997). Oncogenic transcription factors in the human acute leukemias. Science *278*, 1059–1064.

Lopes, F.L., Desmarais, J.A., and Murphy B.D. (2004). Embryonic diapause and its regulation. Reproduction *128*, 669-678.

Marois, G. (1982). Inhibition of nidation in mice by modification of the environment and pheromones. Re-establishment by prolactin and thioproperazine. Ann. Endocrinol. (Paris) *43*, 41-52.

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

McKiernan, S.H.M. and Bavister, B.D. (1998). Gonadotrophin stimulation of donor female decreases post-implantation viability of cultured one-cell hamster embryos. Hum. Reprod. *13*, 72724-72729.

Mead, R.A. (1993). Embryonic diapause in vertebrates. J. Exp. Zool. *266*, 629-641.

Mecham, B.H., Nelson, P.S., and Storey, J.D. (2010). Supervised normalization of microarrays. Bioinformatics *26*, 1308-1315.

Melendez, A., Talloczy, Z., Seaman, M., Eskelinen, E.L., Hall, D.H., and Levine, B. (2003). Autophagy genes are essential for dauer development and life-span extension in C. elegans. Science *301*, 1387-1391.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nat. Rev. Genet. *10*, 155-159.

Miyanari, Y. and Torres-Padilla, M.E. (2012). Control of ground-state pluripotency by allelic regulation of Nanog. Nature *483,* 470-473.

Mizushima, N. and Levine, B. (2010). Autophagy in mammalian development and differentiation. Nat. Cell Biol. *12*, 823-830.

Moss F.P. and Leblond C. P. (1971). Satellite cells as the source of nuclei in muscles of growing rats. Anat. Rec. *170*, 421-435.

Motoda, L., Osato, M., Yamashita, N., Jacob, B., Chen, L.Q., Yanagida, M., Ida, H., Wee, H.J., Sun, A.X., Taniuchi, I., et al. (2007). Runx1 protects hematopoietic stem/progenitor cells from oncogenic insult. Stem Cells *25*, 2976-2986.

Nakada, D., Saunders, T.L., and Morrison, S.J. (2010). Lkb1 regulates cell cycle and energy metabolism in haematopoietic stem cells. Nature *468*, 653-658.

Nichols, J. and Smith, A. (2009). Naive and Primed Pluripotent States. Cell Stem Cell *4*, 487-492.

Niini, T., Kanerva, J., Vettenranta, K., Saarinen-Pihkala, U.M., and Knuutila, S. (2000). AML1 gene amplification: a novel finding in childhood acute lymphoblastic leukemia. Haematologica *85*, 362–366.

Okuda, T., van Deursen, J., Hiebert, S.W., Grosveld, G., and Downing, J.R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell *84*, 321-330.

Otto, F., Thornell, A.P., Crompton, T., Denzel, A., Gilmour, K.C., Rosewell, I.R., Stamp, G.W., Beddington, R.S., Mundlos, S., Olsen, B.R., et al. (1997). Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell *89*, 765-771.

Oyama, T., Harigaya, K., Sasaki, N., Okamura, Y., Kokubo, H., Saga, Y., Hozumi, K., Suganami, A., Tamura, Y., Nagase, T., et al. (2011). Mastermind-like 1 (MamL1) and mastermind-like 3 (MamL3) are essential for Notch signaling in vivo. Development *138*, 5235-5246.

Pan, G., Tian, S., Nie, J., Yang, C., Ruott, V., Wei, H., Jonsdottir, G.A., Stewart, R. and Thomson, J.A. (2007). Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. Cell Stem Cell *1*, 299-312.

Passegue, E., Wagers, A.J., Giuriato, S., Anderson, W.C., and Weissman, I.L. (2005). Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. J. Exp. Med *202*, 1599-1611.

Paulson, R.J., Sauer, M.V. and Lobo, R.A. (1997). Potential enhancement of endometrial receptivity in cycles using controlled ovarian hyperstimulation with antiprogestins: a hypothesis. Fertil. Steril. *67*, 321-325.

Pera, M.F. and Tam, P.P. (2010). Extrinsic regulation of pluripotent stem cells. Nature *465*, 713-720.

Pritsker, M., Donige, T.T., Kramer, L.C., Westcot, S.E. and Lemischka, I.R. (2004). Diversification of stem cell molecular repertoire by alternative splicing. PNAS *102*, 14290-14295.

Psychoyos, A. (1973). Endocrine control of egg-implantation. In Handbook of Physiology: Endocrinology Vol. 2 (American Society of Physiology), 187-215.

Ptak, G.E., Tacconi, E., Czernik, M., Toschi, P. Modlinski, J.A. and Loi, P. (2012). Embryonic diapause is conserved across mammals. PLoS ONE *7*, e33027.

Qian, H., Buza-Vidas, N., Hyland, C.D., Jensen, C.T., Antonchuk, J., Mansson, R., Thoren, L.A., Ekblom, M., Alexander, W.S., and Jacobsen, S.E. (2007). Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. Cell Stem Cell *1*, 671–684.

Renfree, M.B. (1981). Embryonic diapause in marsupials. J. Reprod. Fert. *29*, 67-68.

Renfree, M.B. and Shaw, G. (2000). Diapause. Annu. Rev. Physiol. *62*, 353-375.

Rossant, J., Chazaud, C., and Yamanaka, Y. (2003). Lineage allocation and asymmetries in the early mouse embryo. Philos. Trans. R. Soc. Lond, B. Biol. Sci. *358*, 1341-1348 (discussion 1349).

Salm, S.N., Burger, P.E., Coetzee, S., Goto, K., Moscatelli, D., and Wilson, E.L. (2005). TGF-{beta} maintains dormancy of prostatic stem cells in the proximal region of ducts. J. Cell Biol. *170*, 81-90.

Salomonis, N., Schlieve, C.R., Pereira, L., Wahlquist, C., Colas, A., Zambon, A.C., Vranizan, K., Spindler, M.J., Pico, A.R., Cline, M.S., et al. (2010). Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. PNAS *107*, 10514-10519.

Sandell, M. (1990). The evolution of seasonal delayed implantation. Q. Rev. Biol. *65*, 25-32.

Sang, L., Roberts, J.M. and Coller, H.A. (2010). Hijacking HES1: how tumors co-opt the anti-differentiation strategies of quiescent cells. Trends Mol. Med. *16*, 17-26.

Sato, M., Morii, E., Komori, T., Kawahata, H., Sugimoto, M., Terai, K., Shimizu, J., Yasui, T., Ogihara, H., Yasui, N., et al. (1998). Transcriptional regulation of osteopontin gene *in vivo* by PEBP2alphaA/CBFA1 and ETS1 in the skeletal tissue. Oncogene *17*, 1517-1525.

Scheubert, L., Schmidt, R., Repsilber, D., Lustrek, M. and Fuellen, G. (2011). Learning biomarkers of pluripotent stem cells in mouse. DNA Res. *18*, 233-251.

Schultz E., Gibson M. C., and Champion T. (1978). Satellite cells are mitotically quiescent in mature mouse muscle: an EM and radioautographic study. J. Exp. Zool. *206*, 451-456.

Shachaf, C.M., Kopelman, A.M., Arvanitis, C., Karlsson, A., Beer, S., Mandl, S., Bachmann, M.H., Borowsky, A.D., Ruebner, B., Cardiff, R.D., et al. (2004) MYC inactivation uncovers pluripotent differentiation and tumour dormancy in hepatocellular cancer. Nature *431*, 1112-1117.

Shaw, R,J,, Kosmatka, M., Bardeesy, N., Hurley, R.L., Witters, L.A., DePinho, R.A., and Cantley, L.C. (2004). The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress. PNAS *101*, 3329–3335.

Sherman, M.I. and Barlow, P.W. (1972). Deoxyribonucleic acid content in delayed mouse blastocyst. J. Reprod. Fertil. *29*, 123-126.

Smit, A.F.A., Hubley, R., and Green, P. (1996-2010). RepeatMasker Open 3.0. http://www.repeatmasker.org.

Smith, A.G. (2001). Embryo-derived stem cells: of mice and men. Annu. Rev. Cell Dev Biol. *17*, 435-462.

Song, W.J., Sullivan, M.G., Legare, R.D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I.C., Haworth, C., Hock, R., et al. (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. Nat. Genet. *23*, 166–175.

Speck, N.A. and Gilliland, D.G. (2002). Core-binding factors in haematopoiesis and leukaemia. Nat. Rev. Cancer *2*, 502–513.

Stein, S.C., Woods, A., Jones, N.A., Davison, M.D., and Carling, D. (2000). The regulation of AMP-activated protein kinase by phosphorylation. Biochem. J. *345*, 437–443.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS *102*, 15545-15550.

Sullivan, J.E., Brocklehurst, K.J., Marley, A.E., Carey, F., Carling, D., and Beri, R.K. (1994). Inhibition of lipolysis and lipogenesis in isolated rat adipocytes with AICAR, a cell-permeable activator of AMP-activated protein kinase. FEBS Letter *353*, 33–36.

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Tanaka, N., Takeuchi, T., Neri, Q.V., Sills, E.S., and Palermo, G.D. (2006). Laser-assisted blastocyst dissection and subsequent cultivation of embryonic stem cells in a serum/cell free culture system: applications and preliminary results in a murine model. J. Transl. Med. *4*, 20.

Tarin, J.J. and Cano, A. (1999). Do human concepti have the potential to enter into diapause? Hum. Reprod. *14*, 2434-2436.

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.

Thomson, J.L. and Brinster, R.L. (1966). Glycogen content of preimplantation mouse embryos. Anat. Rec.*155*, 97–102.

Towler, M.C., Fogarty, S., Hawley, S.A., Pan, D.A., Martin, D.M., Morrice, N.A., McCarthy, A., Galardo, M.N., Meroni, S.B., Cigorraga, S.B., et al. (2008). A novel short splice variant of the tumour suppressor LKB1 is required for spermiogenesis. Biochem. J. *416*, 1-14.

Tsuzuki, S., Hong, D., Gupta, R., Matsuo, K., Seto, M., and Enver, T. (2007). Isoform-specific potentiation of stem and progenitor cell engraftment of AML1/RUNX1. PLoS Med. *4*, e172.

Visnjic, D., Kalajzic, Z., Rowe, D.W., Katavic, V., Lorenzo, J., and Aguila, H.L. (2004). Hematopoiesis is severely altered in mice with an induced osteoblast deficiency. Blood *103*, 3258-3264.

Vogel, P. (1981). Occurrence and interpretation of delayed implantation in insectivores. J. Reprod. Fert. *29*, 51-60.

Wang, C.Q, Jacob, B., Nah, G.S. and Osato, M. (2010). Runx family genes, niche, and stem cell quiescence. Blood Cells Mol. Dis. *44*, 275-286.

Ware, C.B., Wang, L., Mecham, B.H., Shen, L., Nelson, A.M., Bar, M., Lamba, D.A., Dauphin, D.S., Buckingham, B., Askari, B., et al. (2009). Histone deacetylase inhibition elicits an evolutionarily conserved self-renewal program in embryonic stem cells. Cell Stem Cell *4*, 359-369.

Welshons, W.V., Wolf, M.F., Murphy, C.S., and Jordan, V.C. (1988). Estrogenic activity of phenol red. Mol. Cell Endocrinol. *57*, 169-178.

Wilcox, A.J. Baird, D.D. and Weiberg, C.R. (1999). Time of implantation of the conceptus and loss of pregnancy. N. Engl. J. Med. *340*, 1796-1799.

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*, 684-687.

Williams, T. and Brenman, J.E. (2008). LKB1 and AMPK in cell polarity and division. Trends Cell Biol. *18*, 193-198.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R.C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C.F., Eshkind, L., Bockamp, E., et al. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell *135*, 1118-1129.

Wong, D.J., Liu, H., Ridky, T.W., Cassarino, D., Segal, E., and Chang, H.Y. (2008). Module map of stem cell genes guides creation of epithelial cancer stem cells. Cell Stem Cell *2*, 333-344.

Woods, A., Johnstone, S.R., Dickerson, K., Leiper, F.C., Fryer, L.G., Neumann, D., Schlattner, U., Wallimann, T., Carlson, M., and Carling, D. (2003). LKB1 is the upstream kinase in the AMP-activated protein kinase cascade. Curr. Biol *13*, 2004–2008.

Wotton, S., Stewart, M., Blyth, K., Vaillant, F., Kilbey, A., Neil, J.C., and Cameron, E.R. (2002). Proviral insertion indicates a dominant oncogenic role for Runx1/AML-1 in T-cell lymphoma. Cancer Res. *62*, 7181–7185.

Yamanaka, Y., Ralston, A., Stephenson, R.O., and Rossant, J. (2006). Cell and molecular regulation of the mouse blastocyst. Dev. Dyn. *235*, 2301-2314.

Yamazaki, K., Aso, T., Ohnishi, Y., Ohno, M., Tamura, K., Shuin, T., Kitajima, S., and Nakabeppu, Y. (2003). Mammalian elongin A is not essential for cell viability but is required for proper cell cycle progression with limited alteration of gene expression. J. Biol. Chem. *278*, 13585-13589.

Yamazaki, S., Iwama, A., Takayanagi, S., Eto, K., Ema, H., and Nakauchi, H. (2009). TGF-beta as a candidate bone marrow niche signal to induce hematopoietic stem cell hibernation. Blood *113*, 1250-1256.

Yanagida, M., Osato, M., Yamashita, N., Liqun, H., Jacob, B., Wu, F., Cao, X., Nakamura, T., Yokomizo, T., Takahashi, S., et al. (2005). Increased dosage of Runx1/AML1 acts as a positive modulator of myeloid leukemogenesis in BXH2 mice. Oncogene *24*, 4477–4485.

Yilmaz, O.H., Valdez, R., Theisen, B.K., Guo, W., Ferguson, D.O., Wu, H., and Morrison, S.J. (2006). Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. Nature *441*, 475-842.

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.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Yoshihara, H., Arai, F., Hosokawa, K., Hagiwara, T., Takubo, K., Nakamura, Y., Gomei, Y., Iwasaki, H., Matsuoka, S., Miyamoto, K., et al. (2007). Thrombopoietin/MPL signaling regulates hematopoietic stem cell quiescence and interaction with the osteoblastic niche. Cell Stem Cell *1*, 685–697.

Yu. D., Dews, M., Park, A., Tobias, J.W., and Thomas-Tikhonenko, A. (2005). Inactivation of Myc in murine two-hit B lymphomas causes dormancy with elevated levels of interleukin 10 receptor and CD20: implications for adjuvant therapies. Cancer Res. *65*, 5454-5461.

Zhang, J., Niu, C., Ye, L., Huang, H., He, X., Tong, W.G., Ross, J., Haug, J., Johnson, T., Feng, J.Q., et al. (2003). Identification of the haematopoietic stem cell niche and control of the niche size. Nature *425*, 836-841.

Zhang, J., Grindley, J.C., Yin, T., Jayasinghe, S., He, X.C., Ross, J.T., Haug, J.S., Rupp, D., Porter-Westpfahl, K.S., Wiedemann, L.M., et al. (2006). PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. Nature *441*, 518-522

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.K., Shahab, A., Kuznetsov, V.A., et al. (2007). Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells. Cell Stem Cell *1*, 286-298.

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. J. Biol. Chem. *285*, 29676-29680.

Vita

Lilyana Margaretha was born in 1984 in Jakarta, Indonesia. She moved to the United States after finishing high school at Raffles Girls' Secondary School in Singapore. She went to Edmonds Community College and received her Associate of Science degree in 2003. In 2005, she earned a Bachelor of Science degree in Biochemistry from the University of Washington, Seattle. Upon completion of her undergraduate studies, she continued to work as a research technician in the laboratory of Dr. Cecilia Moens at Fred Hutchinson Cancer Research Center, Seattle for one year. She then joined the Molecular and Cellular Biology program at the University of Washington, Seattle and performed her doctorate work at UW Institute for Stem Cell and Regenerative Medicine (ISCRM) in the laboratory of Dr. C. Anthony Blau. In 2012, she earned a Doctor of Philosophy at the University of Washington in Molecular and Cellular Biology.