UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Molecular Determinants of Oocyte and Embryo Developmental Competence

Permalink

https://escholarship.org/uc/item/0kq611t3

Author

Dumdie, Jennifer Nicole

Publication Date

2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Molecular Determinants of Oocyte and Embryo Developmental Competence

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Jennifer Nicole Dumdie

Committee in Charge:

Professor Heidi Cook-Andersen, Chair
Professor Miles Wilkinson, Co-chair
Professor Louise Laurent
Professor Jens Lykke-Andersen
Professor Pamela Mellon
Professor Mana Parast

2019

The Dissertation of Jennifer Nicole Dumdie is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Co-chair

_____
Chair

University of California, San Diego

2019

DEDICATION

I dedicate this PhD to my dog, Cleopatra.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First of all, I want to thank all of my scientific mentors—both past and present—that have helped shape me into the scientist that I am today and that I will continue to grow to become. Thank you to Heidi Cook-Andersen and Miles Wilkinson for allowing me to join their laboratories and for helping me develop into a curious and thoughtful scientist with a greater understanding for what it means to think critically. I also want to acknowledge my committee members, some of which are also collaborators, Dr. Louise Laurent, Dr. Mana Parast, Dr. Jens Lykke-Andersen, and Dr. Pamela Mellon for their sustained enthusiasm for and love of science. It has truly been an honor and an inspiration to work alongside each and every one of them.

I would like to thank the Cook-Andersen and Wilkinson Lab members, past and present, who provided both technical knowledge, as well as emotional support, during these last four years. Enjoying coming in to the lab every day has been partly because of the amazing science, but mostly because of the amazing people I get to work with. It has also been a pleasure to work alongside and with my amazing colleagues in the Laurent and Parast laboratories; I appreciate all of their support, assistance with experiments, and the creation of a positive scientific atmosphere.

More specifically, I want to acknowledge Dr. Samantha Jones, a former Wilkinson Lab graduate student and my very good friend, who has been a source of scientific and personal support during my graduate school years. Sam provided countless hours of troubleshooting advice and proofreading during this time, as well as collaboratory mouse room hours that are very much appreciated. I also want to acknowledge Dr. Omar Farah of the Parast Lab. Omar has provided field-specific advice that has been instrumental in the development of these projects and in the development of myself as a critically-thinking scientist and well-rounded human being. Both Sam and Omar have been a valuable source of support during these years, and truly represent the kind of friendship that traverses graduate education. Also instrumental to

these projects coming to fruition have been Dr. Kyucheol Cho, Dr. Abhishek Sohni, Dr. Kun Tan, and Dr. Terra Plank-Brannan. Their comments, advice, and troubleshooting assistance have helped shape these research projects and the resulting publications.

I would like to thank Kristen Jepsen and her colleagues at the UCSD IGM Core for RNA sequencing and Eric Griffis at the UCSD Nikon Imaging Center for embryo imaging and downstream analysis.

I would like to thank my funding sources, the Ruth L. Kirschstein National Research Service Award (NRSA) Individual Predoctoral MD/PhD Fellowship (F30), NICHD, as well as the Achievement Rewards for College Scientists (ARCS) Scholarship (2016-2019) and the San Diego Matching Fellowship Graduate Scholarship, University of California, San Diego (2015).

Finally, I want to thank my family and friends for their amazing support during the acquisition of my PhD. Thank you to my brother, Derek, my mother, Debra, my father, Rob, and my grandmother, Laurena, for their love and support in the pursuit of my passion, and for their patience with my inquisitory, scientific nature. Thank you to Cleopatra, my best friend, for her tolerance of my long hours reading papers and writing this thesis, and for listening to every scientific talk I have ever given. Also, I want to thank Rafael, for helping me find my voice and my passion, and for directing me toward real value in life.

Chapter 2 contains material as a reprint of the following paper: Dumdie JN, Cho K, Ramaiah M, Skarbrevik D, Mora-Castilla S, Stumpo DJ, Lykke-Andersen J, Laurent LC, Blackshear PJ, Wilkinson MF, Cook-Andersen H. Chromatin Modification and Global Transcriptional Silencing in the Oocyte Mediated by the mRNA Decay Activator ZFP36L2. *Developmental Cell*. 2018 Feb 5; 44(3):392-402. The dissertation author was the primary investigator and author of this material.

Chapter 3 contains material that will be prepared for submission for publication. The dissertation author was the primary investigator and author of this material. Current co-authors

are as follows: Jennifer N Dumdie, Abhishek Sohni, Kristoffer Vitting-Seerup, Bo Porse, Heidi Cook-Andersen*, and Miles F. Wilkinson*. (* Authors contributed equally to this work)

Chapter 4 contains material that will be prepared for submission for publication. The dissertation author was the primary investigator and author of this material. Current co-authors are as follows: Jennifer N Dumdie, Cuong To, Srimeenakshi Srinivasan, Wei Zhang, Ana Lisa Yeo, Gabriel Garzo, Miles F. Wilkinson, Louise C. Laurent, and Heidi Cook-Andersen.

VITA

2013 Bachelor of Science, Biology and Mathematics, University of South Dakota, Vermillion, SD

2013-2015 University of California, San Diego School of Medicine

2019 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Ramaiah M, Tan K, Plank TM, Song HW, **Dumdie JN**, Jones S, Shum EY, Sheridan SD, Peterson KJ, Gromoll J, Haggerty SJ, Cook-Andersen H, Wilkinson MF. A microRNA cluster in the Fragile-X region expressed during spermatogenesis targets FMR1. *EMBO Rep.* 2018 Dec 20.

Dumdie JN, Cho K, Ramaiah M, Skarbrevik D, Mora-Castilla S, Stumpo DJ, Lykke-Andersen J, Laurent LC, Blackshear PJ, Wilkinson MF, Cook-Andersen H. Chromatin Modification and Global Transcriptional Silencing in the Oocyte Mediated by the mRNA Decay Activator ZFP36L2. *Developmental Cell.* 2018 Feb 5; 44(3):392-402.

Huang L, Shum EY, Jones SH, Lou CH, **Dumdie J**, Kim H, Roberts AJ, Jolly LA, Espinoza JL, Skarbrevik DM, Phan MH, Cook-Andersen H, Swerdlow NR, Gecz J, Wilkinson MF. A Upf3b-mutant mouse model with behavioral and neurogenesis defects. *Molecular Psychiatry.* 2017 Sep 26.

Lou CH, **Dumdie J**, Goetz AE, Shum EY, Brafman D, Liao X, Mora-Castilla S, Ramaiah M, Cook-Andersen H, Laurent L, Wilkinson MF. NMD Dictates Embryonic Stem Cell Fate. *Stem Cell Reports.* 2016; 6(6): 844-57.

Shum EY, Jones SH, Shao A, **Dumdie J**, Krause MD, Chan WK, Lou CH, Espinoza JL, Song HW, Phan MH, Ramaiah M, Huang L, McCarrey JR, De Rooij DG, Cook-Andersen H, Wilkinson MF. The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. *Cell*. 2016; 165(2): 382-95.

Mora-Castilla S, To C, Vaezeslami S, Morey R, Srimeenakshi S, **Dumdie JN**, Cook-Andersen H, Jenkins J, Laurent L. Miniaturization technologies for efficient single-cell library preparations and Next-Generation Sequencing. *J Lab Autom.* 2016 Aug; 21(4):557-67.

Chalpe AJ, Law CD, **Dumdie JN**, Hansen KA, Eyster KM. TNFα and IL1β Stimulate Differential Gene Expression in Endometrial Stromal Cells. *Advances in Biological Chemistry.* 2015; 5:126-141.

ABSTRACT OF THE DISSERTATION

Molecular Determinants of Oocyte and Embryo Developmental Competence

by

Jennifer Nicole Dumdie

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2019

Professor Heidi Cook-Andersen, Chair

Professor Miles F. Wilkinson, Co-Chair

The earliest stages of life, including the transition from the fully differentiated oocyte to the totipotent zygote, the first days of embryo cleavage and cell differentiation to form a blastocyst, and implantation of that blastocyst in the wall of the uterus, are somehow beautifully simple and strikingly complex at the same time. These stages, which represent the beginning of life for us and other vertebrates, are difficult to study, owing to a number of experimental and ethical considerations. However, advances in our understanding of nuclear reprogramming, transcriptional quiescence and activation, the maintenance of pluripotency, and what it takes for an embryo to initiate contact with the endometrium to generate a successful pregnancy, will

without question have far-reaching influence for many scientific and medical disciplines. Here, I attempt to unravel some of these concepts, combining molecular and developmental biology with genome-wide analysis only recently made possible through advances in techniques with low input. These approaches, in combination, allow us to ask questions about early developmental systems that would not have been possible only years prior.

In the oocyte, global transcriptional silencing is a highly conserved mechanism that is essential for the transition from the differentiated oocyte to the totipotent zygote. Here, I report that global transcriptional silencing in the mouse oocyte depends on an mRNA decay activator. By downregulating master regulators of transcription during oocyte growth, particularly a group of mRNAs encoding demethylases for H3K4 and H3K9, ZFP36L2 enables increased histone methylation that is associated with transcriptional silencing. These results uncover a mRNA decay mechanism that acts a developmental switch during growth of the mammalian oocyte, resulting in wide-spread shifts in chromatin modification, and mediating silencing of transcription in the oocyte.

The pluripotent population of cells in the blastocyst, the inner cell mass, is established in the mouse embryo approximately three days after fertilization. These cells will undergo gastrulation to form the entire organism and can be maintained in culture as embryonic stem cells. I report here that UPF2, a mRNA decay activator, is needed specifically within the pluripotent inner cell mass of the mouse embryo for maintenance of pluripotency in the embryo *in vivo,* as well as for embryonic stem cells *in vitro.* That mRNA decay may underly the establishment or maintenance of this intriguing and complex population of cells, is an exciting possibility.

Reproductive success depends on embryo implantation in the uterus, and in fertile and infertile couples alike, failure of the embryo to implant in the wall of the uterus accounts for up to 75% of all lost pregnancies. Here, I provide a qualitative assessment of gene expression and cellular communication networks within the major compartments of the human blastocyst that

are most closely associated with successful implantation and ongoing pregnancy. Most strikingly, establishment and/or maintenance of the extraembryonic primitive endoderm lineage—following the second major embryonic differentiation event—most strongly differentiates embryos of high and low implantation potential. Unbiased machine learning identified genes within each embryo compartment most closely associated with implantation. Taken together, this data supports a model in which successful implantation and ongoing pregnancy predominantly depends upon the inner cell mass and highlights a potentially novel role for the extraembryonic primitive endoderm in early pregnancy success.

Chapter 1

Introduction

The oocyte: a tale of transcriptional silence

Oocytes are unique -- they begin their life terminally differentiated, but upon fertilization, lose their identity and become totipotent, initiating the early stages of life. Oocytes grow within organized follicles, which include granulosa and theca cells that support the oocyte during its growth and preparation for ovulation. Folliculogenesis begins with a small (approximately 20 um) primordial follicle, containing an oocyte that has formed from a primordial germ cell. During folliculogenesis, the primordial follicle progressively expands to form primary, secondary, and finally, a large fluid-filled periovulatory antral follicle. In concordance, the oocyte progresses through a phase of cell growth, all the while arrested in meiosis I (Schultz et al., 2018). During this time, the early oocyte is classified as an immature germinal vesicle (GV) oocyte. At ovulation, the GV oocyte is released from the antral follicle and meiosis I resumes. Upon resumption of meiosis I, the germinal vesicle (nuclear envelope) breaks down, and the oocyte is now referred to as an MI oocyte. After the completion of meiosis I, the first polar body is extruded, and meiosis II begins (now an MII oocyte). The oocyte then pauses again, this time during metaphase of meiosis II, until fertilization occurs (Schultz et al., 2018). mRNA sequencing from oocytes at each of these stages revealed some important transitions during oocyte growth. During the secondary-to-early antral follicle transition, oocytes acquire meiotic competence, or the ability to progress through meiosis (Sorensen and Wassarman, 1976; Wickramasinghe et al., 1991). As the antral follicle grows, the oocyte progressively acquires developmental competence (Eppig and Schroeder, 1989; Sorensen and Wassarman, 1976), and during the last days of oocyte growth, oocytes acquire a competence associated with their ability to progress through the blastocyst stage of embryo development (Eppig and Schroeder, 1989; Sorensen and Wassarman, 1976) (Figure 1.1).

Interestingly, the acquisition of meiotic competence during antrum formation is accompanied by a rapid reduction in transcription, resulting in transcriptional silence in the

periovulatory GV oocyte (Schultz et al., 2018). Transcription is silent throughout the remainder of oocyte maturation, and only fully resumes in the late 2-cell embryo in the mouse (between the 4-8-cell stage in the human embryo), after fertilization (Figure 1.1).

Developmental competence of the oocyte is dependent on transcriptional silencing, as evidenced by a lack of maturation (meiotic progression) of the majority of transcriptionally active oocytes (Zuccotti et al., 1998). Transcriptional silencing is also a conserved feature of oocytes in all species examined to date; however it remains a poorly understood aspect of oocyte development (Andreu-Vieyra et al., 2010; Bouniol-Baly et al., 1999; Li et al., 2010; Liu and Aoki, 2002; Zuccotti et al., 2002). Effectively, in the absence of transcription, this implies that oocyte maturation, fertilization, nuclear reprogramming, and the first embryo cleavage event, all happen in the absence of *de novo* transcription. This makes post-transcriptional pathways an attractive target for analyses examining regulation of maternal mRNA pools in the oocyte across these stages of development, which will be introduced later.

Global transcriptional silencing and developmental competence are also accompanied by extensive chromatin structural changes and modifications (Schultz et al., 2018). At the time of transcriptional silencing, the chromatin pattern in the oocyte nucleus shifts from a diffuse configuration ("non-surrounded nucleolus" [NSN]) to a condensed structure that forms a distinct ring around the nucleolus ("surrounded nucleolus" [SN]). While this condensation in normal oocytes is commonly used as a visual indication of transcriptional silencing, these two events appear to be driven by distinct but parallel pathways as demonstrated in knock-out oocyte models (Andreu-Vieyra et al., 2010; Burns et al., 2003).

Mammalian preimplantation embryo development

Pre-implantation development encompasses ~4 days in the mouse and ~7 days in humans, and consists of successive rounds of cell division and differentiation, culminating in blastocyst implantation in the uterine cavity (Bedzhov et al., 2014; Eckert et al., 2015;

Lorthongpanich and Issaragrisil, 2015) (Figure 1.2). The very first differentiation events in the mammalian embryo segregate the extra-embryonic lineages from the pluripotent cells that will form the entire organism. The embryo first differentiates into the trophectoderm (TE), which will become the placenta, and the inner cell mass (ICM). Multiple regulatory events have been demonstrated as having concrete roles in dictating this first lineage specification. It is widely accepted that positioning of a cell within the embryo affects cell fate, and, more recently, the Hippo pathway has been shown to translate positional information into lineage specification cues (Lorthongpanich and Issaragrisil, 2015). Soon after, the ICM undergoes an additional round of differentiation, forming the pluripotent epiblast (EPI), which will form all 3 layers of the developing fetus, and the extra-embryonic primitive endoderm (PrE), which contributes mostly to the yolk sac (Figure 1.3). Differentiation of the ICM into EPI and PrE is a multistep process, beginning with lineage specification and subsequent maturation, and finally the sorting of cells into distinct compartments; i.e., the PrE epithelium surrounds a cluster of EPI cells. Much of what we know about these differentiation events in the early embryo come from studies in the mouse, as embryos in this model system can be genetically manipulated and studied. Mouse embryo studies will be discussed here. In the next section, I mention important differences that have been discovered in recent years between mouse and human development at this early preimplantation stage.

Following the transcriptionally quiescent state of the oocyte and zygote, transcription fully resumes in the early embryo (late 2-cell stage in mice and the 4-8 cell stage in humans). The next major embryonic event is compaction to form the morula. This occurs around the 16-cell stage and is characterized by the transformation of the embryo from a loose cluster of cells into a tightly packed mass in which individual cells cannot be distinguished. After compaction, the embryo begins to differentiate. Cells on the inside of the morula receive positional signals, sensing more tension from surrounding cells. This tension is relayed to these cells through

signaling pathways. For example, AMOT is phosphorylated and forms an active complex with NF2 and LATS kinases (Chazaud and Yamanaka, 2016). This complex phosphorylates YAP, which stays sequestered in the cytoplasm, preventing it from translocating to the nucleus to activate TE-specific genes (Chazaud and Yamanaka, 2016). In the outside cells, if AMOT is not phosphorylated and YAP is therefore not phosphorylated, YAP can translocate to the nucleus, leading to expression of the TE-specific factors TEAD4 and CDX2. Mature TE cells then act to pump fluid into the embryo, forming the blastocoele cavity. At this time, the embryo is called a blastocyst.

The ICM then undergoes an additional round of differentiation to form the PrE, as well as to maintain the pluripotent population of cells, the EPI. Bimodal expression of *Fgf4* at E3.25 is the earliest sign of heterogeneity within ICM cells (Guo et al., 2010; Kurimoto et al., 2006; Ohnishi et al., 2013), and several studies have demonstrated that FGF signaling plays an essential role in the establishment of the PrE lineage in the mouse. At the protein level, the earliest markers of the EPI and PrE lineages are NANOG and GATA6, respectively (Chazaud et al., 2006; Guo et al., 2010; Kurimoto et al., 2006; Plusa et al., 2008). By E3.75, these factors are exclusively expressed in early EPI and PrE cells, generating a ‘salt and pepper’ expression pattern. It has been hypothesized that the restriction of these factors to specific cells might rely on selective mRNA decay, as the specification event can be visualized by a combination of increased expression of one factor and a subsequent fading of the other (Bessonnard et al., 2014; Guo et al., 2010). However, this remains to be directly shown. It is also unclear what specific event triggers the first differences observed in ICM cells. What is clear is that an interplay between FGF signaling, NANOG, and GATA6 is necessary for correct EPI/PrE lineage establishment.

After PrE specification, maturation and maintenance of the PrE ensues. Mature PrE markers, including SOX17 and GATA4 are sequentially expressed, and work in concert with

FGF signaling for further maturation of this extra-embryonic lineage (Bessonnard et al., 2014; Frankenberg et al., 2011; Schrode et al., 2014).

At the time of implantation, the blastocyst, which consists of a ball of TE cells surrounding a fluid-filled cavity, and an EPI encased by adjacent polar TE cells and a PE epithelium, makes contact with the wall of the uterus. A competent blastocyst is just one essential piece to the puzzle, however. The endometrium must also be receptive, primed with the right hormone signals, for the blastocyst to implant. After contact is made, the TE secretes enzymes that degrade the extracellular matrix of the endometrium, the embryo burrows within these cells of the uterus, and implantation has commenced.

Exploring human-specific differences in preimplantation development

Until very recently, studies examining human embryo development were extremely scarce. This is due to a number of barriers. For example, there are the ethical considerations that must be considered when studying human development. Both ethics committee consent, as well as patient informed consent, must be documented before embryos can be used for research (Kalista et al., 2011). In addition, material is severely limited. Human embryos are generally either obtained as supernumerary embryos of poor or unknown quality, or those that are in excess of the couple's reproductive need. Further, because these embryos are donated and not transferred, along with cases of unknown quality, it is difficult to design relevant studies examining normal human development.

Despite these limitations, research in recent years has revealed some important differences between humans and other model systems, even at these early stages of development. The first and most obvious difference is the delayed activation of transcription in humans, which happens after an additional 1-2 rounds of cell division in humans at the 4-8 cell stage, as compared to the late 2-cell stage in mice (Figure 1.2). In addition, expression of transcription factors known to direct lineage specification of the TE versus the ICM is protracted

in human blastocysts as compared to mouse. In the mouse, CDX2 and OCT4 are restricted to the TE and ICM, respectively, at the time of segregation. In human blastocysts, however, these factors remain coexpressed in the TE (Berg et al., 2011; Chen et al., 2009; Niakan et al., 2012). Further, OCT4, while expressed throughout the embryo in earlier stages, as in mouse, only becomes restricted to the ICM around E6.0 in human blastocysts (Cauffman et al., 2009; Chen et al., 2009). Further, CDX2 is not expressed until E5.0 in the human embryo, after cavitation has already occurred (Chen et al., 2009). This suggests that other factors initiate this lineage specification event in humans.

The next differentiation event—the specification of PrE and EPI within the ICM—is initiated by FGF signaling in mouse blastocysts. However, this does not seem to be the case in humans because inhibiting FGF/Erk signaling in human embryos through inhibition of the downstream Erk pathway, or though FGFR inhibition, has been shown to have no effect on either EPI or PrE formation (Kuijk et al., 2012; Roode et al., 2012), indicating that this lineage specification is also differentially regulated in human.

Some studies have attempted to define human-specific signals that might regulate these early stages in the human embryo. For example, components of the TGF-β pathway have been demonstrated to play a role in maintaining pluripotency both in the human EPI (Blakeley et al., 2015) and hESCs (Bertero et al., 2015; Besser, 2004; James et al., 2005). For example, Blakeley et al. showed that blockage of Activin signaling in developing human blastocysts led to ablation of NANOG expression, as well as another marker of the PE lineage, SOX17. The same inhibitor treatment failed to change protein expression of these factors in the mouse blastocyst (Blakeley et al., 2015). A study in a non-human primate model system (marmoset) demonstrated that specification of the EPI and PrE requires cooperation between Wnt and FGF signaling (Boroviak et al., 2015), making the Wnt pathway an exciting potential target for future

studies aimed at teasing out the specifics of PrE and EPI specification in humans and other primates.

Gene expression differences between preimplantation human and mouse blastocysts have also been documented (Blakeley et al., 2015). In particular, TE-specific genes in mouse, including *Elf5, Eomes, Tcfap2c,* and *Id2,* display altered expression in human TE. *ELF5* and *EOMES* are not expressed in human TE at all, and *ID2* is more highly expressed in human PrE than TE. *TFAP2C* is expressed not only in human TE, but also in EPI, indicating altered roles for this mouse TE-specific factor in humans. In addition, other genes (*CLDN10, TRIML1, PLAC8*) that are not expressed in mouse TE, are specific for TE in human blastocysts. Similarly, genes identified as EPI-specific in the mouse embryo, such as *Esrrb,* and *Klf2,* display altered expression in human samples. While *KLF4* is not expressed in human blastocysts at all, *ESRRB* is more highly expressed in extra-embryonic cell types than the EPI. Further, human-specific EPI markers that are not expressed in the mouse have been documented, including *LEFTY1, NODAL, ACVRL1,* and *KLF17.* Interestingly, many of the PrE-specific markers that are known in mouse, are also expressed at the RNA level in humans; however, some human-specific PrE markers, such as *FOXA2,* are not expressed in mouse blastocyst samples (Blakeley et al., 2015).

The process of implantation also appears to be different for human blastocysts as compared to mouse. The mouse embryo makes contact with the endometrium on the mural TE (mTE) side, opposite the ICM, and mTE cells immediately differentiate into trophoblast giant cells to invade into the endometrium. In contrast, human (and bovine) blastocysts have been shown *in vitro* to adhere to an endometrial sample on the polar TE (pTE) side, adjacent to the ICM.

Together, these data suggest that while the many similarities between mouse and human pre- and peri-implantation development continue to make mice a useful model system

for understanding some aspects of human development, mice also have limitations in this regard. Thus, there is also a necessity for studying preimplantation development in humans. Advancement of our understanding of the biology underlying these differences will be key to future advances in stem cell biology, as well as treating infertility.

The opposite side of the coin: RNA decay as a regulator of gene expression

Research studies on the gene regulatory programs that guide development and maturation of the oocyte, as well as the preimplantation embryo, have largely focused on transcriptional regulation. While these studies have laid the foundation for our understanding of development, they have failed to account for the potentially important role of post-transcriptional regulation in these developmental events. One key post-transcriptional step that is subject to regulation is mRNA decay. This is important, as the steady-state level of an mRNA is dependent not only on transcriptional synthesis rates, but also on the rate at which it is degraded (Figure 1.4). Indeed, increasing evidence suggests that coordination between these two regulatory mechanisms is critical for complex biological processes during development to occur (Alonso, 2011; Dumdie et al., 2018; Li et al., 2015; Lou et al., 2016; Medghalchi et al., 2001; Shum et al., 2016; Thoren et al., 2010; Weischenfeldt et al., 2008) (Figure 1.4).

Eukaryotic mRNA decay rates differ up to 100-fold between transcripts (Thomsen et al., 2010; Wang et al., 2002), and these rates can be modified in response to internal or external signals (Tucker et al., 2002). If degradation rates are too high, an mRNA might reach steady-state levels that are too low for viability, and if degradation rates are too low, expression level of the gene could be too high or even indifferent to changes in transcriptional regulation. Both of these scenarios would likely have a detrimental effect on development.

There are a plethora of RNA decay regulators which have already been shown to play vital roles in varying developmental systems. Some of these that will not be discussed here include RNA regulation by small RNAs, including miRNAs and piRNAs, as well as long non-

coding RNAs. One of the main pathways by which mRNA decay is regulated is by selective modulation of polyA-tail length, where tail length correlates with mRNA stability (Alonso, 2011; Houseley and Tollervey, 2009; Labno et al., 2016). In fact, computational studies have demonstrated that a high proportion of mRNAs in vertebrates have alternative polyA-tail lengths – at least 50% of human genes (Tian et al., 2005). Other 3' end modifications such as uridylation, have also been shown to play a role in mediating the rate by which RNAs are degraded (Labno et al., 2016).

Another class of RNA decay modulators is RNA-binding proteins (RBPs). These proteins act in RNA decay pathways, including AU-rich element (ARE)-mediated RNA decay and Nonsense-mediated RNA decay (NMD), both of which have demonstrated roles for regulating gene expression in development, as detailed below.

ARE-mediated RNA decay

Adenylate/uridylate-(AU) rich element (ARE) motifs are present in the 3' untranslated region (UTR) of a subset of cellular mRNAs. The TIS11 RBP family regulates the expression of mRNAs containing AREs by binding to these elements and initiating rapid decay. In the absence of one of the TIS11 family members, which includes ZFP36, ZFP36L1, and ZFP36L2, expression of mRNAs containing AREs dramatically increases due to their role as RNA decay-promoting factors. Some of the critical genes that contain AREs and are thus regulated by ARE-mediated decay include those important for initiating inflammatory responses and cancer (Sanduja et al., 2010). It is currently estimated that approximately 8% of the human transcriptome contains AREs (Bakheet et al., 2006; Brooks and Blackshear, 2013a), highlighting the far-reaching capabilities of this RNA decay pathway in regulating normal gene expression.

Regulation of TIS11 family members occurs at multiple levels, including their phosphorylation, their subcellular localization, and their association with other proteins. These RBPs have been shown to be extensively phosphorylated, particularly at serine residues (Cao

et al., 2006). In the unphosphorylated state, TTP (ZFP36) is able to bind to ARE-containing mRNAs and targets them for decay. Interestingly, in the unphosphorylated state, while TTP is active, it is also subject to proteosomal degradation, a fate that is unlikely while in the phosphorylated state (Benjamin et al., 2006; Carballo et al., 2001). After binding to ARE-containing mRNAs, TTP (as well as other family members) directly promotes deadenylation (via direct association with CCR4), and via recruitment of PARN (Lai et al., 2003; Lykke-Andersen and Wagner, 2005; Sanduja et al., 2010). Deadenylation by CCR4 and PARN then lead to rapid mRNA decay.

Each TIS11 family member appears to have non-redundant, tissue-specific functions. For example, *Ttp* and *Zfp36l1* knockout models exhibit hyperinflammatory (Taylor et al., 1996) and vascular (Stumpo et al., 2004) phenotypes, respectively. Complete loss of ZFP36L2 in mice results in severe defects in hematopoiesis and death within a few weeks of birth (Stumpo et al., 2009). Female *Zfp36l2*-mutant mice that lack the 5' end of the *Zfp36l2* gene and express a 29 amino acid N-terminal truncated version of ZFP36L2, are infertile, with genetic background-specific blocks in ovulation or embryo development past the 2-cell stage (Ball et al., 2014; Ramos et al., 2004), which hinted at a role for ZFP36L2 in the oocyte-to-embryo transition.

Nonsense-mediated RNA decay

NMD was originally discovered by virtue of its ability to degrade aberrant RNAs harboring premature termination codons and thereby protect cells from truncated dominant-negative proteins (Chang et al., 2007; Huang and Wilkinson, 2012; Popp and Maquat, 2013; Schweingruber et al., 2013). Subsequently, it was discovered that NMD degrades subsets of normal RNAs, with loss or disruption of NMD leading to dysregulation of 5-20% of the normal transcriptome in species spanning the phylogenetic scale (Chan et al., 2007; Gatfield et al., 2003; He et al., 2003; Lelivelt and Culbertson, 1999; Mendell et al., 2004). In higher eukaryotes,

it is not known what proportion of these dysregulated RNAs are direct NMD targets (also known as “NMD substrates”).

NMD is well studied at the biochemical level, with over 15 proteins involved. Their molecular choreography is relatively well defined (Chang et al., 2007; Kervestin and Jacobson, 2012; Popp and Maquat, 2013; Smith and Baker, 2015). The UPF1, UPF2, and UPF3B proteins are core NMD factors. UPF1 is an RNA helicase that forms a complex with the adaptor proteins UPF2 and UPF3B. UPF3B binds the exon-junction complex (EJC), a large multi-subunit complex that is recruited to RNAs just upstream of exon-exon junctions after RNA splicing. Most normal RNAs harbor a stop codon terminating the main open reading frame (ORF) in the final exon, thereby avoiding NMD because the EJC is displaced by ribosomes before translation termination. However, a select group of normal mRNAs have a downstream EJC (dEJ) and thus are targeted for decay by the NMD pathway. Other transcripts are degraded by NMD in an EJC-*independent* manner (Gehring et al., 2005). In particular, long 3’UTRs and short open reading frames upstream of the main ORF (uORFs) can elicit NMD in some contexts (Boehm et al., 2014; Kebaara and Atkin, 2009; Kertesz et al., 2006; Kurosaki and Maquat, 2013; Singh et al., 2008; Toma et al., 2015) that remain to be clearly defined (Brogna et al., 2016; Fatscher et al., 2015; Karousis et al., 2016; Kervestin and Jacobson, 2012).

While some normal RNAs directly targeted by NMD have been identified, the vast majority of NMD targets remain unknown. The finding that NMD’s reach extends beyond “quality control” strongly argues that this conserved and highly selective RNA decay pathway has roles in regulating normal cellular events. Indeed, NMD has been shown to be important for development in organisms spanning the phylogenetic scale (Hwang and Maquat, 2011; Vicente-Crespo and Palacios, 2010). In humans, mutations in the NMD gene, *UPF3*B, cause intellectual disability (Nguyen et al., 2014; Tarpey et al., 2007), and copy number variations in several NMD genes are associated with a number of neuro-developmental disorders (Nguyen et al., 2013).

Studies have demonstrated that NMD must be downregulated by neuron-expressed microRNAs to permit neural stem cells to undergo differentiation (Bruno et al., 2011; Karam and Wilkinson, 2012; Lou et al., 2014), and that feedback loops exist to sustain either high or low NMD magnitude and thus maintain the stem or differentiated cell state, respectively(Lou et al., 2014). These studies suggest that NMD plays an integral role in influencing stem cell versus differentiation decisions.

Relevant to this work, a large body of research has demonstrated that loss of individual factors that either regulate or are essential for NMD (UPF1, UPF2, UPF3A, SMG1, and SMG6) lead to early embryonic lethality in mice, as well as other organisms (Alonso and Akam, 2003; Anastasaki et al., 2011; Casadio et al., 2014; Hwang and Maquat, 2011; Li et al., 2015; Longman et al., 2007; Mao et al., 2015; McIlwain et al., 2010; Medghalchi et al., 2001; Metzstein and Krasnow, 2006; Shum et al., 2016; Silver et al., 2010; Weischenfeldt et al., 2008). Further, knock-out embryos that have been examined to date exhibit phenotypes in the pre- to peri-implantation stage in mice. For example, cultured *Upf1*-null pre-implantation (E3.5) embryos exhibited regression of the ICM and apoptosis, and, *in vivo*, exhibited a complete loss of architecture after implantation at E5.5 (Medghalchi et al., 2001). Similarly, *Smg6*-null E3.5 blastocysts fail to form the ICM after 5 days of growth in culture (Li et al., 2015), and *Upf2*-null embryos exhibit altered morphology as early as E3.5, with lethality occurring before E5.5 (Weischenfeldt et al., 2008). Loss of *Upf3a* causes morphological defects in E3.5 blastocysts, with embryonic lethality traced between E4.5 and E8.5 (Shum et al., 2016). Furthermore, consistent with a potential role for the NMD pathway at the blastocyst stage, they found that UPF3A is highly expressed in the ICM of E3.5 blastocysts but nearly undetectable in TE (Shum et al., 2016). Focused studies like this have not been conducted for the remaining NMD factors, and thus it will be interesting to determine the expression pattern of other factors during development. Further, phenotypes specific to dysregulation of pluripotency have been reported

in embryos and ESC systems devoid of NMD activity (Li et al., 2015; Lou et al., 2016; Medghalchi et al., 2001), hinting at a role for selective regulation of mRNA stability by the NMD pathway in mediating pluripotency and/or the pluripotent cell population during early embryo development.

Acknowledgments

Chapter 2 contains material as a reprint that was published in *Developmental Cell*, 2018.

Chapter 3 contains material that is being prepared for submission for publication. The dissertation author was the primary investigator and author of this material. Co-authors at this point include: Abhishek Sohni, Heidi Cook-Andersen*, and Miles F. Wilkinson* (* - authors contributed equally to this work).

Chapter 4 contains material that is being prepared for submission for publication. The dissertation author was the primary investigator and author of this material. Co-authors at this point include Cuong To, Srimeenakshi Srinivasan, Wei Zhang, Ana Lisa Yeo, Gabriel Garzo, Miles F. Wilkinson, Louise C. Laurent, and Heidi Cook-Andersen.

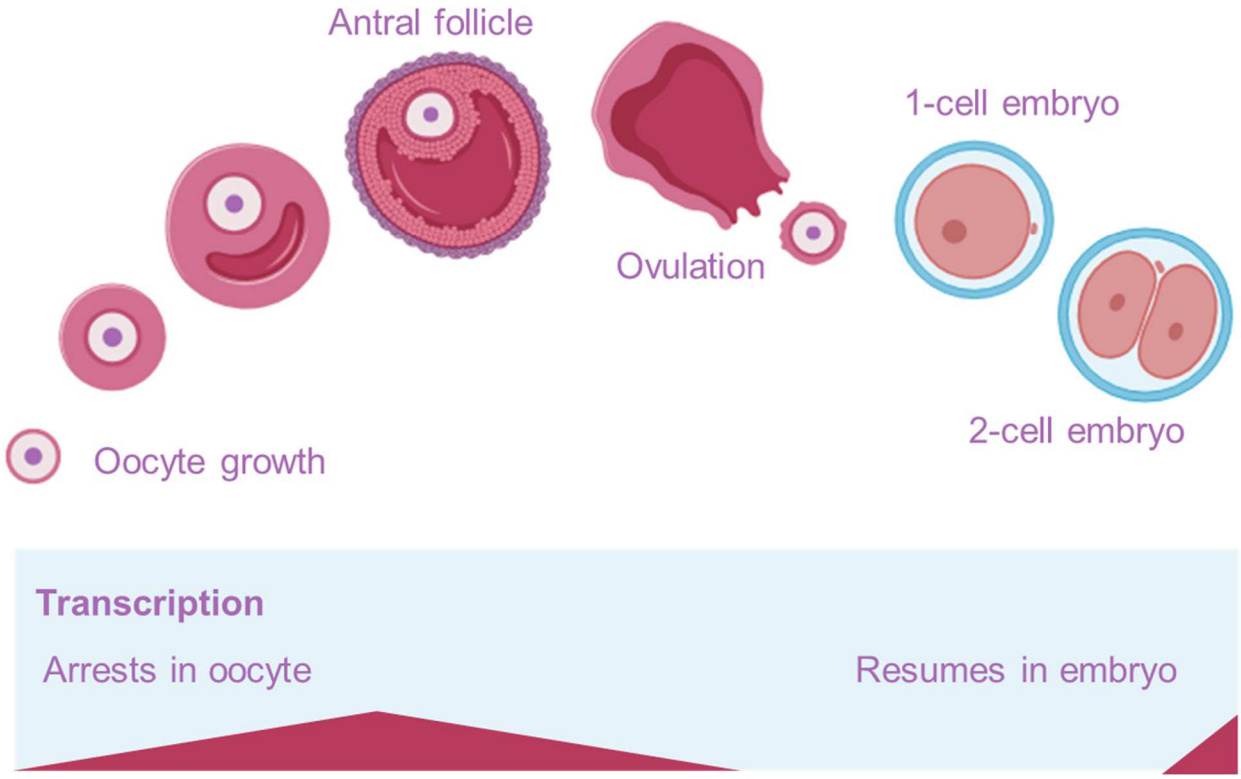


Figure 1.1 Oocyte maturation, fertilization, nuclear reprogramming, and embryo cleavage, all occur in the absence of *de novo* transcription.

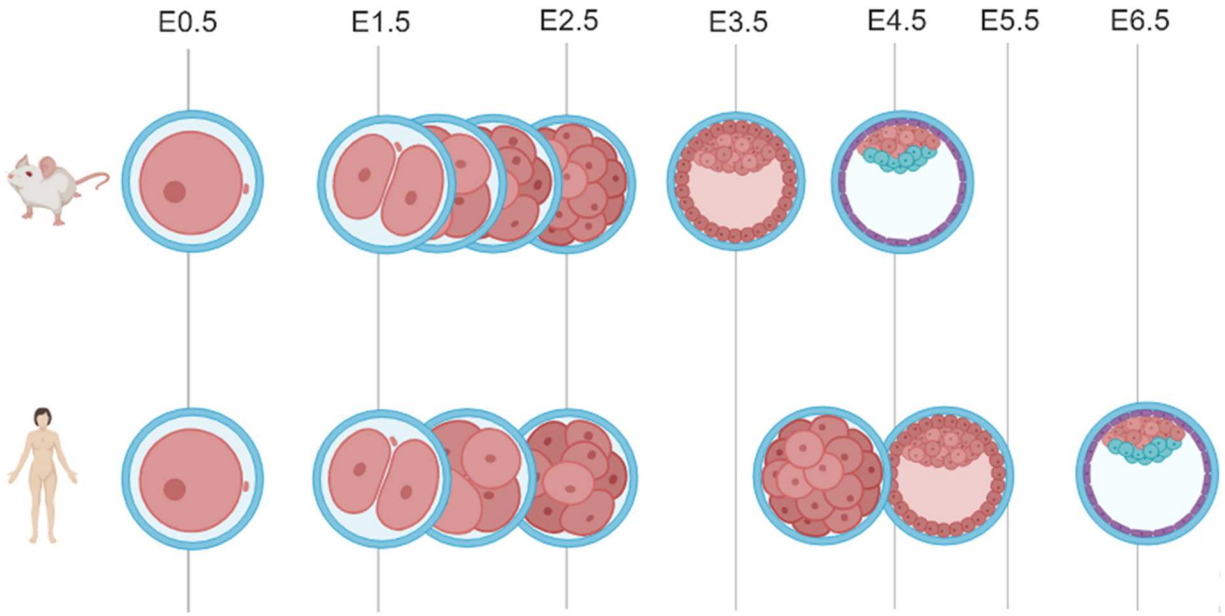


Figure 1.2 Overview of preimplantation embryo development in mouse and human. Embryo development commences on embryonic day (E)0.5 with fertilization. Transcriptional activation occurs at the 2-cell stage in mouse, and between the 4-8-cell stage in human. Blastocoele cavity formation is evident by E3.5 in mouse, and by E5.0 in human. In the mouse, all three lineages (TE, PrE, and EPI) are present by E4.5 at the time of implantation, In human, this is evident by E5.0-E6.0, and implantation begins around E7.0.

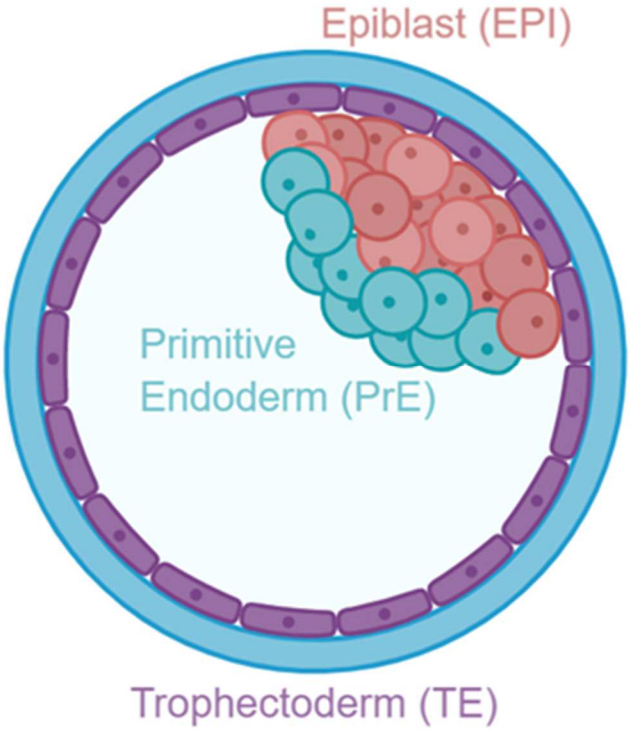


Figure 1.3 The three lineages of the blastocyst present at the time of implantation include the outer extra-embryonic trophectoderm, the extra-embryonic primitive endoderm, and the epiblast.

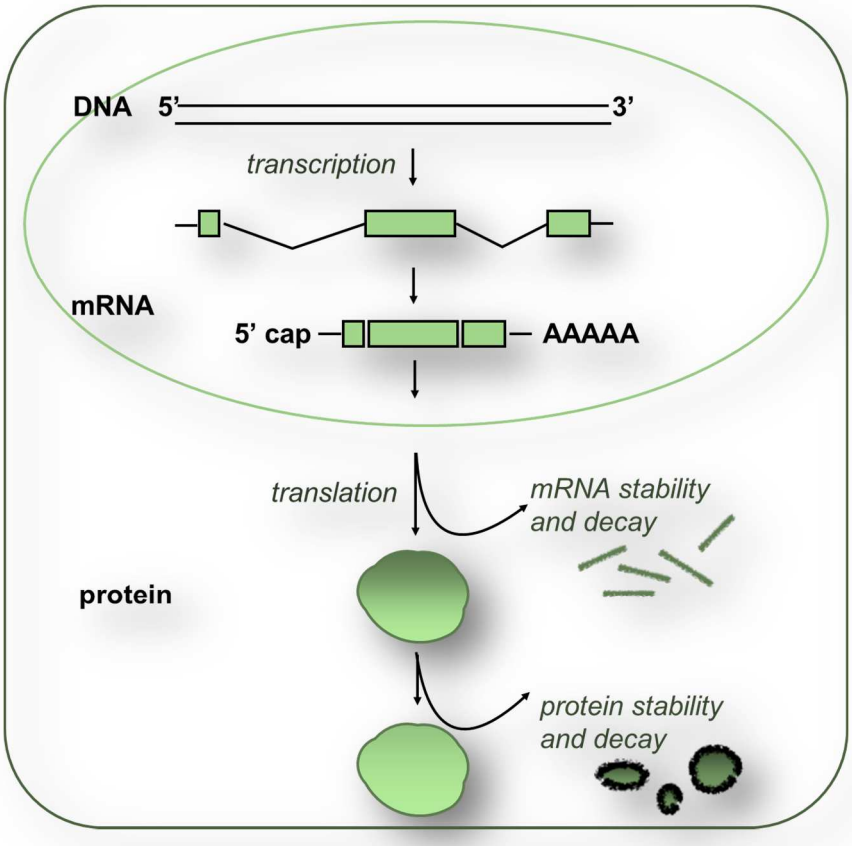


Figure 1.4. The steady-state level of an mRNA is dependent both on its rate of synthesis and its rate of decay.

References

Alonso, C.R. (2011). A complex “mRNA degradation code” controls gene expression during animal development. *Trends Genet*. 28, 78–88.

Alonso, C.R., and Akam, M. (2003). A Hox gene mutation that triggers nonsense-mediated RNA decay and affects alternative splicing during Drosophila development. *Nucleic Acids Res.* 31, 3873–3880.

Anastasaki, C., Longman, D., Capper, A., Patton, E.E., and Caceres, J.F. (2011). Dhx34 and Nbas function in the NMD pathway and are required for embryonic development in zebrafish. *Nucleic Acids Res.* 39, 3686–3694.

Andreu-Vieyra, C. V, Chen, R., Agno, J.E., Glaser, S., Anastassiadis, K., Stewart, A.F., and Matzuk, M.M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS Biol.* 8, e1000453.

Bakheet, T., Williams, B.R.G., and Khabar, K.S.A. (2006). ARED 3.0: the large and diverse AU-rich transcriptome. *Nucleic Acids Res*. 34, 111–114.

Ball, C.B., Rodriguez, K.F., Stumpo, D.J., Ribeiro-Neto, F., Korach, K.S., Blackshear, P.J., Birnbaumer, L., and Ramos, S.B. V (2014). The RNA-binding protein, ZFP36L2, influences ovulation and oocyte maturation. *PLoS One* 9, e97324.

Bedzhov, I., Graham, S.J.L., Leung, C.Y., and Zernicka-Goetz, M. (2014). Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369, 1–14.

Benjamin, D., Schmidlin, M., Min, L., Gross, B., and Moroni, C. (2006). BRF1 Protein Turnover and mRNA Decay Activity Are Regulated by Protein Kinase B at the Same Phosphorylation Sites. *Mol. Cell. Biol.* 26, 9497–9507.

Berg, D.K., Smith, C.S., Pearton, D.J., Wells, D.N., Broadhurst, R., Donnison, M., and Pfeffer, P.L. (2011). Trophectoderm Lineage Determination in Cattle. *Dev. Cell* 20, 244–255.

Bertero, A., Madrigal, P., Galli, A., Hubner, N.C., Moreno, I., Burks, D., Brown, S., Pedersen, R.A., Gaffney, D., Mendjan, S., Pauklin, S., and Vallier, L. (2015). Activin/nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev.* 29, 702–717.

Besser, D. (2004). Expression of nodal, lefty-a, and lefty-B in undifferentiated human embryonic stem cells requires activation of Smad2/3. *J. Biol. Chem.* 279, 45076–45084.

Bessonnard, S., De Mot, L., Gonze, D., Barriol, M., Dennis, C., Goldbeter, A., Dupont, G., and Chazaud, C. (2014). Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* 141, 3637–3648.

Blakeley, P., Fogarty, N.M.E., Valle, I., Wamaitha, S.E., Hu, X., Elder, K., Snell, P., Christie, L., and Robson, P. (2015). Defining the Three Cell Lineages of the Human Blastocyst by Single-Cell RNAseq. *Development* 142, 3151–3165.

Boehm, V., Haberman, N., Ottens, F., Ule, J., and Gehring, N.H. (2014). 3’ UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep.* 9, 555–568.

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

Bouniol-Baly, C., Hamraoui, L., Guibert, J., Beaujean, N., Szöllösi, M.S., and Debey, P. (1999). Differential transcriptional activity associated with chromatin configuration in fully grown mouse germinal vesicle oocytes. *Biol. Reprod.* 60, 580–587.

Brogna, S., McLeod, T., and Petric, M. (2016). The Meaning of NMD: Translate or Perish. *Trends Genet.* 32, 395–407.

Brooks, S.A., and Blackshear, P.J. (2013). Tristetraprolin (TTP): Interactions with mRNA and proteins, and current thoughts on mechanisms of action. *BBA - Gene Regul. Mech.* 1829, 666–679.

Bruno, I.G., Karam, R., Huang, L., Bhardwaj, A., Lou, C.H., Shum, E.Y., Song, H.-W., Corbett, M.A., Gifford, W.D., Gecz, J., Pfaff, S.L., and Wilkinson, M.F. (2011). Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. *Mol. Cell* 42, 500–510.

Burns, K.H., Viveiros, M.M., Ren, Y., Wang, P., DeMayo, F.J., Frail, D.E., Eppig, J.J., and Matzuk, M.M. (2003). Roles of NPM2 in chromatin and nucleolar organization in oocytes and embryos. *Science* 300, 633–636.

Cao, H., Deterding, L.J., Venable, J.D., Kennington, E.A., Yates, J.R., Tomer, K.B., and Blackshear, P.J. (2006). Identification of the anti-inflammatory protein tristetraprolin as a hyperphosphorylated protein by mass spectrometry and site-directed mutagenesis. *Biochem. J.* 394, 285–297.

Carballo, E., Cao, H., Lai, W.S., Kennington, E.A., Campbell, D., and Blackshear, P.J. (2001). Decreased sensitivity of tristetraprolin-deficient cells to p38 inhibitors suggests the involvement of tristetraprolin in the p38 signaling pathway. *J. Biol. Chem.* 276, 42580–42587.

Casadio, A., Longman, D., Hug, N., Delavaine, L., Vallejos Baier, R., Alonso, C.R., and Caceres, J.F. (2014). Identification and characterization of novel factors that act in the nonsense-mediated mRNA decay pathway in nematodes, flies and mammals. *EMBO Rep.* 16, 71–78.

Cauffman, G., De Rycke, M., Sermon, K., Liebaers, I., and Van de Velde, H. (2009). Markers that define stemness in ESC are unable to identify the totipotent cells in human preimplantation embryos. *Hum. Reprod.* 24, 63–70.

Chan, W.K., Huang, L., Gudikote, J.P., Chang, Y.F., Imam, J.S., MacLean 2nd, J.A., and Wilkinson, M.F. (2007). An alternative branch of the nonsense-mediated decay pathway. *EMBO J* 26, 1820–1830.

Chang, Y.-F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* 76, 51–74.

Chazaud, C., and Yamanaka, Y. (2016). Lineage specification in the mouse preimplantation embryo. *Development* 143, 1063–1074.

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.

Chen, A.E., Egli, D., Niakan, K., Deng, J., Akutsu, H., Yamaki, M., Cowan, C., Fitz-Gerald, C., Zhang, K., Melton, D.A., and Eggan, K. (2009). Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell* 4, 103–106.

Dumdie, J.N., Cho, K., Ramaiah, M., Skarbrevik, D., Mora-Castilla, S., Stumpo, D.J., Lykke-Andersen, J., Laurent, L.C., Blackshear, P.J., Wilkinson, M.F., and Cook-Andersen, H. (2018). Chromatin Modification and Global Transcriptional Silencing in the Oocyte Mediated by the mRNA Decay Activator ZFP36L2. *Dev. Cell* 44, 392–402.

Eckert, J.J., Velazquez, M.A., and Fleming, T.P. (2015). Cell Signaling During Mammalian Early Embryo Development.

Eppig, J.J., and Schroeder, A.C. (1989). Capacity of mouse oocytes from preantral follicles to undergo embryogenesis and development to live young after growth, maturation, and fertilization in vitro. *Biol. Reprod.* 41, 268–276.

Fatscher, T., Boehm, V., and Gehring, N.H. (2015). Mechanism, factors, and physiological role of nonsense-mediated mRNA decay. *Cell. Mol. Life Sci.*

Frankenberg, S., Gerbe, F., Bessonnard, S., Belville, C., Pouchin, P., Bardot, O., and Chazaud, C. (2011). Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Dev. Cell* 21, 1005–1013.

Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. *EMBO J.* 22, 3960–3970.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol. Cell* 20, 65–75.

Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 18, 675–685.

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5’ to 3’ mRNA decay pathways in yeast. *Mol. Cell* 12, 1439–1452.

Houseley, J., and Tollervey, D. (2009). Leading Edge The Many Pathways of RNA Degradation. *Cell* 136, 763–776.

Huang, L., and Wilkinson, M.F. (2012). Regulation of nonsense-mediated mRNA decay. *Wiley Interdiscip. Rev. RNA* 3, 807–828.

Hwang, J., and Maquat, L.E. (2011). Nonsense-mediated mRNA decay (NMD) in animal embryogenesis: to die or not to die, that is the question. *Curr. Opin. Genet. Dev.* 21, 422–430.

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

Kalista, T., Austinn Freeman, H., Behr, B., Pera, R.R., and Scott, C.T. (2011). Cell Stem Cell Forum Donation of Embryos for Human Development and Stem Cell Research. *Stem Cell* 8, 360–362.

Karam, R., and Wilkinson, M. (2012). A conserved microRNA/NMD regulatory circuit controls gene expression. *RNA Biol.* 9, 22–26.

Karousis, E.D., Nasif, S., and Mühlemann, O. (2016). Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip. Rev. RNA* 7, 661–682.

Kebaara, B.W., and Atkin, A.L. (2009). Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. *Nucleic Acids Res.* 37, 2771–2778.

Kertesz, S., Kerenyi, Z., Merai, Z., Bartos, I., Palfy, T., Barta, E., and Silhavy, D. (2006). Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic Acids Res.* 34, 6147–6157.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. *Nat. Rev. Mol. Cell Biol.* 13, 700–712.

Kuijk, E.W., van Tol, L.T.A., Van de Velde, H., Wubbolts, R., Welling, M., Geijsen, N., and Roelen, B.A.J. (2012). The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. *Development* 139, 871–882.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K.D., Yamada, R.G., Ueda, H.R., and Saitou, M. (2006). An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res*. 34, e42.

Kurosaki, T., and Maquat, L.E. (2013). Rules that govern UPF1 binding to mRNA 3' UTRs. *Proc. Natl. Acad. Sci. U. S. A.* 110, 3357–3362.

Labno, A., Tomecki, R., and Dziembowski, A. (2016). Cytoplasmic RNA decay pathways - Enzymes and mechanisms. *Biochim. Biophys. Acta - Mol. Cell Res.* 1863, 3125–3147.

Lai, W.S., Kennington, E.A., and Blackshear, P.J. (2003). Tristetraprolin and its family members can promote the cell-free deadenylation of AU-rich element-containing mRNAs by poly(A) ribonuclease. *Mol. Cell. Biol.* 23, 3798–3812.

Lelivelt, M.J., and Culbertson, M.R. (1999). Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol. Cell. Biol.* 19, 6710–6719.

Li, L., Zheng, P., and Dean, J. (2010). Maternal control of early mouse development. *Development* 137, 859–870.

Li, T., Shi, Y., Wang, P., Guachalla, L.M., Sun, B., Joerss, T., Chen, Y.-S., Groth, M., Krueger, A., Platzer, M., Yang, Y.-G., Rudolph, K.L., and Wang, Z.-Q. (2015). Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. *EMBO J.* 34, 1630–1647.

Liu, H., and Aoki, F. (2002). Transcriptional activity associated with meiotic competence in fully grown mouse GV oocytes. *Zygote* 10, 327–332.

Longman, D., Plasterk, R.H.A., Johnstone, I.L., and Cáceres, J.F. (2007). Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. *Genes Dev.* 21, 1075–1085.

Lorthongpanich, C., and Issaragrisil, S. (2015). Emerging Role of the Hippo Signaling Pathway in Position Sensing and Lineage Specification in Mammalian Preimplantation Embryos. *Biol. Reprod.* 92, 143.

Lou, C., Dumdie, J., Goetz, A., Shum, E., Brafman, D., Liao, X., Mora-Castilla, S., Ramaiah, M., Cook-Andersen, H., Laurent, L., and Wilkinson, M. (2016). Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate. *Stem Cell Reports* 6, 844–857.

Lou, C.H., Shao, A., Shum, E.Y., Espinoza, J.L., Huang, L., Karam, R., and Wilkinson, M.F. (2014). Posttranscriptional control of the stem cell and neurogenic programs by the nonsense-mediated RNA decay pathway. *Cell Rep.* 6, 748–764.

Lykke-Andersen, J., and Wagner, E. (2005). Recruitment and activation of mRNA decay enzymes by two ARE-mediated decay activation domains in the proteins TTP and BRF-1. *Genes Dev.* 19, 351–361.

Mao, H., Pilaz, L.-J., McMahon, J.J., Golzio, C., Wu, D., Shi, L., Katsanis, N., and Silver, D.L. (2015). Rbm8a haploinsufficiency disrupts embryonic cortical development resulting in microcephaly. *J. Neurosci.* 35, 7003–7018.

McIlwain, D.R., Pan, Q., Reilly, P.T., Elia, A.J., McCracken, S., Wakeham, A.C., Itie-Youten, A., Blencowe, B.J., and Mak, T.W. (2010). Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. *Proc. Natl. Acad. Sci. U. S. A.* 107, 12186–12191.

Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M., and Dietz, H.C. (2001). Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* 10, 99–105.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* 36, 1073–1078.

Metzstein, M.M., and Krasnow, M.A. (2006). Functions of the Nonsense-Mediated mRNA Decay Pathway in Drosophila Development. *PLoS Genet.* 2, e180.

Nguyen, L.S., Kim, H.-G., Rosenfeld, J.A., Shen, Y., Gusella, J.F., Lacassie, Y., Layman, L.C., Shaffer, L.G., and Gécz, J. (2013). Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders. *Hum. Mol. Genet.* 22, 1816–1825.

Nguyen, L.S., Wilkinson, M.F., and Gecz, J. (2014). Nonsense-mediated mRNA decay: inter-individual variability and human disease. *Neurosci. Biobehav. Rev.* 46 Pt 2, 175–186.

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

Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., Oleś, A.K., Araúzo-Bravo, M.J., Saitou, M., Hadjantonakis, A.-K., and Hiiragi, T. (2013). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol*. 16, 27–37.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.-K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* 135, 3081–3091.

Popp, M.W.-L., and Maquat, L.E. (2013). Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.* 47, 139–165.

Ramos, S.B. V, Stumpo, D.J., Kennington, E.A., Phillips, R.S., Bock, C.B., Ribeiro-Neto, F., and Blackshear, P.J. (2004). The CCCH tandem zinc-finger protein Zfp36l2 is crucial for female fertility and early embryonic development. *Development* 131, 4883–4893.

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

Sanduja, S., Blanco, F.F., and Dixon, D.A. (2010). The roles of TTP and BRF proteins in regulated mRNA decay. *WIREs RNA* 2, 42–57.

Schrode, N., Saiz, N., Di Talia, S., and Hadjantonakis, A.-K. (2014). GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. *Dev. Cell* 29, 454–467.

Schultz, R.M., Stein, P., and Svoboda, P. (2018). The oocyte-to-embryo transition in mouse: past, present, and future. *Biol. Reprod.* 99, 160–174.

Schweingruber, C., Rufener, S.C., Zünd, D., Yamashita, A., and Mühlemann, O. (2013). Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim. Biophys. Acta* 1829, 612–623.

Shum, E.Y., Jones, S.H., Shao, A., Dumdie, J., Krause, M.D., Chan, W.-K., Lou, C.-H., Espinoza, J.L., Song, H.-W., Phan, M.H., Ramaiah, M., Huang, L., McCarrey, J.R., Peterson, K.J., De Rooij, D.G., Cook-Andersen, H., and Wilkinson, M.F. (2016). The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. *Cell* 165, 382–395.

Silver, D.L., Watkins-Chow, D.E., Schreck, K.C., Pierfelice, T.J., Larson, D.M., Burnetti, A.J., Liaw, H.-J., Myung, K., Walsh, C.A., Gaiano, N., and Pavan, W.J. (2010). The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat. Neurosci.* 13, 551–558.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol.* 6, e111.

Smith, J.E., and Baker, K.E. (2015). Nonsense-mediated RNA decay--a switch and dial for regulating gene expression. *Bioessays* 37, 612–623.

Sorensen, R.A., and Wassarman, P.M. (1976). Relationship between growth and meiotic maturation of the mouse oocyte. *Dev. Biol.* 50, 531–536.

Stumpo, D.J., Byrd, N.A., Phillips, R.S., Ghosh, S., Maronpot, R.R., Castranio, T., Meyers, E.N., Mishina, Y., and Blackshear, P.J. (2004). Chorioallantoic fusion defects and embryonic lethality resulting from disruption of Zfp36L1, a gene encoding a CCCH tandem zinc finger protein of the Tristetraprolin family. *Mol. Cell. Biol.* 24, 6445–6455.

Stumpo, D.J., Broxmeyer, H.E., Ward, T., Cooper, S., Hangoc, G., Chung, Y.J., Shelley, W.C., Richfield, E.K., Ray, M.K., Yoder, M.C., Aplan, P.D., and Blackshear, P.J. (2009). Targeted disruption of Zfp36l2, encoding a CCCH tandem zinc finger RNA-binding protein, results in defective hematopoiesis. *Blood* 114, 2401–2410.

Tarpey, P.S., Raymond, F.L., Nguyen, L.S., Rodriguez, J., Hackett, A., Vandeleur, L., Smith, R., Shoubridge, C., Edkins, S., Stevens, C., O'Meara, S., Tofts, C., Barthorpe, S., Buck, G., Cole, J., Halliday, K., Hills, K., Jones, D., Mironenko, T., et al. (2007). Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. *Nat. Genet.* 39, 1127–1133.

Taylor, G.A., Carballo, E., Lee, D.M., Lai, W.S., Thompson, M.J., Patel, D.D., Schenkman, D.I., Gilkeson, G.S., Broxmeyer, H.E., Haynes, B.F., and Blackshear, P.J. (1996). A pathogenetic role for TNF alpha in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency. *Immunity* 4, 445–454.

Thomsen, S., Anders, S., Janga, S.C., Huber, W., and Alonso, C.R. (2010). Genome-wide analysis of mRNA decay patterns during early Drosophila development. *Genome Biol.* 11, R93.

Thoren, L.A., Nørgaard, G.A., Weischenfeldt, J., Waage, J., Jakobsen, J.S., Damgaard, I., Bergström, F.C., Blom, A.M., Borup, R., Bisgaard, H.C., and Porse, B.T. (2010). UPF2 Is a Critical Regulator of Liver Development, Function and Regeneration. *PLoS One* 5, e11650.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* 33, 201–212.

Toma, K.G., Rebbapragada, I., Durand, S., and Lykke-Andersen, J. (2015). Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. *RNA* 21, 887–897.

Tucker, M., Staples, R.R., Valencia-Sanchez, M.A., Muhlrad, D., and Parker, R. (2002). Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in Saccharomyces cerevisiae. *EMBO J.* 21, 1427–1436.

Vicente-Crespo, M., and Palacios, I.M. (2010). Nonsense-mediated mRNA decay and development: shoot the messenger to survive? *Biochem. Soc. Trans.* 38, 1500–1505.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. (2002). Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5860–5865.

Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Mönch, K., Thoren, L.A., Nielsen, F.C., Jacobsen, S.E.W., Nerlov, C., and Porse, B.T. (2008). NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev*. 22, 1381–1396.

Wickramasinghe, D., Ebert, K.M., and Albertini, D.F. (1991). Meiotic competence acquisition is associated with the appearance of M-phase characteristics in growing mouse oocytes. *Dev. Biol.* 143, 162–172.

Zuccotti, M., Giorgi Rossi, P., Martinez, A., Garagna, S., Forabosco, A., and Redi, C.A. (1998). Meiotic and developmental competence of mouse antral oocytes. *Biol. Reprod.* 58, 700–704.

Zuccotti, M., Ponce, R.H., Boiani, M., Guizzardi, S., Govoni, P., Scandroglio, R., Garagna, S., and Redi, C.A. (2002). The analysis of chromatin organisation allows selection of mouse antral oocytes competent for development to blastocyst. *Zygote* 10, 73–78.

Chapter 2: Chromatin Modification and Global Transcriptional Silencing in the Oocyte Mediated by the mRNA Decay Activator ZFP36L2.

Abstract

Global transcriptional silencing is a highly conserved mechanism central to the oocyte-to-embryo transition. We report the unexpected discovery that global transcriptional silencing in oocytes depends on an mRNA decay activator. Oocyte-specific loss of ZFP36L2 an RNA-binding protein that promotes AU-rich element-dependent mRNA decay prevents global transcriptional silencing and causes oocyte maturation and fertilization defects, as well as complete female infertility in the mouse. Single-cell RNA sequencing revealed that ZFP36L2 downregulates mRNAs encoding transcription and chromatin modification regulators, including a large group of mRNAs for histone demethylases targeting H3K4 and H3K9, which we show are bound and degraded by ZFP36L2. Oocytes lacking *Zfp36l2* fail to accumulate histone methylation at H3K4 and H3K9, marks associated with the transcriptionally silent, developmentally competent oocyte state. Our results uncover a ZFP36L2-dependent mRNA decay mechanism that acts as a developmental switch during oocyte growth, triggering wide-spread shifts in chromatin modification and global transcription.

Introduction

The carefully orchestrated transition from the fully differentiated oocyte to the totipotent embryo is arguably one of the most dynamic transitions in biology. This oocyte-to-embryo transition—fundamental to all multicellular life—occurs in the absence of *de novo* transcription. Transcription is globally silenced in the final stages of oocyte growth and does not significantly resume until the 2-cell embryo stage in mice. Therefore, successful development across the oocyte-to-embryo transition requires establishment of the proper cohort of maternal mRNA and proteins within the oocyte prior to transcriptional silencing (Li et al., 2010).

Global transcriptional silencing and the acquisition of full developmental competence are closely linked events that occur during the final stages of oocyte growth. Indeed, periovulatory oocytes (referred to hereafter as GV oocytes) that fail to undergo global transcriptional silencing have significantly reduced rates of oocyte maturation, fertilization and embryo development (Andreu-Vieyra et al., 2010; Bouniol-Baly et al., 1999; Liu and Aoki, 2002; Zuccotti et al., 2002). While some factors contributing to global transcriptional silencing have been identified (Andreu-Vieyra et al., 2010; Y.-J. Liu et al., 2012; Lowther and Mehlmann, 2015; Medvedev et al., 2011; Xia et al., 2012), the mechanisms underlying this pivotal developmental event remain poorly understood.

Global transcriptional silencing and developmental competence are closely associated with extensive chromatin modification, including changes in histone methylation, histone acetylation, and DNA methylation (Andreu-Vieyra et al., 2010; De La Fuente, 2006; Kageyama et al., 2007; Zuccotti et al., 2011). Among the best-characterized of such modifications are increases in the levels of both histone H3K9 and H3K4 trimethylation (Dahl et al., 2016a; De La Fuente, 2006; Kageyama et al., 2007; Zhang et al., 2016a). H3K9me3 is widely regarded as a marker of transcriptional repression, while H3K4me3 is generally associated with active transcription (Bannister and Kouzarides, 2011). However, recent evidence provides strong support that H3K4me3 can also act as a repressive mark in oocytes and that accumulation of H3K4me3 in the late oocyte contributes to transcriptional silencing (Dahl et al., 2016a; Zhang et al., 2016a). The factors and pathways that mediate the accumulation of these histone methylation marks in the late oocyte, and the precise functional relationships between histone methylation, transcriptional silencing and oocyte developmental competence, remain intriguing questions.

Both transcriptional silencing and developmental competence in the oocyte are also accompanied by extensive changes in chromatin structure. The chromatin pattern within the late

mammalian oocyte nucleus shifts from a diffuse configuration ("non-surrounded nucleolus" [NSN]) to a condensed structure that forms a distinct ring around the nucleolus ("surrounded nucleolus" [SN]). *In vitro* studies have demonstrated that the SN chromatin configuration strongly correlates with transcriptional silence and full oocyte developmental competence. Conversely, the NSN chromatin configuration correlates with active transcription and significantly diminished oocyte developmental competence (Bouniol-Baly et al., 1999; Zuccotti et al., 2002). However, while transcriptional silencing and chromatin condensation are closely linked temporally, these two events are genetically separable and appear to be driven by distinct pathways (Andreu-Vieyra et al., 2010; Burns et al., 2003).

Given that the oocyte-to-embryo transition occurs in the absence of transcription, it is not surprising that this transition depends upon post-transcriptional mechanisms. Several lines of evidence suggest that RNA decay is critical for multiple stages of the oocyte-to-embryo transition (Svoboda et al., 2015) although mRNA in the growing oocyte has been proposed to be unusually stable overall (Brower et al., 1981). The subject of this report—the mRNA decay activator ZFP36L2—has been implicated as having a role in the oocyte-to-embryo transition (Ball et al., 2014; Ramos, 2012; Ramos et al., 2004). ZFP36L2 is one of four members of the ZFP36 (TTP/Tis11) family in the mouse, all of which have established roles in degrading mRNAs harboring AU-rich elements (AREs) in the 3' untranslated region (UTR) (Brooks and Blackshear, 2013b; Sanduja et al., 2011). Each family member appears to have non-redundant, tissue-specific functions. Complete loss of ZFP36L2 in mice results in severe defects in hematopoiesis and death within a few weeks of birth (Stumpo et al., 2009). Female *Zfp36l2*-mutant mice that lack the 5' end of the *Zfp36l2* gene and express a 29 amino acid N-terminal truncated version of ZFP36L2, are infertile, with genetic background-specific blocks in ovulation or embryo development past the 2-cell stage (Ball et al., 2014; Ramos et al., 2004). Because this truncated protein retains the ability to trigger mRNA decay *in vitro*, it is unclear whether the

developmental arrest is a result of gain-of-function effects or reduced ARE-mediated mRNA decay activity that is only evident *in vivo* (Ramos, 2012). Thus, the precise role of ZFP36L2 and mRNA decay in the oocyte has remained unclear.

To establish whether ZFP36L2 plays a critical role in the oocyte and/or the oocyte-to-embryo transition, we generated and analyzed oocyte-specific complete *Zfp36l2* conditional knockout (cKO) mice. We found that *Zfp36l2*-cKO female mice are infertile and oocytes from these mutant mice fail to undergo normal maturation and fertilization. To investigate the underlying mechanism, we performed single-cell RNAseq analysis and found that loss of ZFP36L2 caused upregulation of scores of mRNAs encoding factors with central roles in transcriptional regulation and chromatin modification, including a large group of histone demethylases that regulate methylation at H3K4 or H3K9. We show these mRNAs are direct targets of ZFP36L2, as evidenced by mRNA decay and RNA immunoprecipitation assays, suggesting that histone methylation can be regulated by mRNA decay. Consistent with this, GV oocytes from *Zfp36l2*-cKO mice exhibited defects in H3K4 and H3K9 trimethylation and failed to undergo global transcriptional silencing, both of which are associated with the acquisition of full developmental competence in the oocyte. Together, our findings establish an essential role for ZFP36L2 in the regulation of maternal mRNAs encoding factors that control transcription and chromatin modification and provide a mechanism by which mRNA decay can bring about histone methylation, global transcriptional silencing, and the production of a developmentally competent oocyte.

Methods

Generation of conditional knockout mice

Conditional *Zfp36l2* knockout mice were generated by Xenogen Biosciences (Cranbury,

NJ) using standard embryonic stem (ES) cell targeting techniques. These mice will be described in detail in a future study (PJ Blackshear, *manuscript in preparation*). The targeting construct was designed to delete exon 2, which contains the RNA binding domain and the majority of the protein coding region. Briefly, C57BL/6 genomic DNA was used to prepare a targeting vector that contained loxP sites flanking exon 2 as well as a Neomycin (Neo) positive selection expression cassette. C57BL/6 ES cells were electroporated with the targeting vector and correctly targeted ES cells were subsequently transiently transfected with a Flp expressing plasmid to excise the Neo cassette. Blastocysts were injected with two independent clones and germline transmission was obtained by crossing chimeras with C57BL/6 Tac wild type mice. Heterozygous mice with a conditional floxed *Zfp36l2* allele were intercrossed to generate homozygous floxed mice. Deletion of *Zfp36l2* by conditional knock out was obtained by crossing the oocyte-specific *Zp3*-Cre transgenic mice with the *Zfp36l2* floxed mice (*Zfp36l2*$^{fl/fl}$) to generate *Zfp36l2*$^{fl/fl}$;Zp3-Cre males that were then maintained with *Zfp36l2*$^{fl/fl}$ females to generate *Zfp36l2*-cKO and –Fl/Fl littermate controls for experiments. For each experiment detailed below, the total number and age of mice used are clearly stated. Mice were genotyped by polymerase chain reaction (primers used for genotyping are provided in supplemental data). The C57BL/6 mice colonies were maintained in agreement with protocols approved by the Institutional Animal Care and Use Committee at the University of California, San Diego.

Primary culture of oocytes and zygotes

For experiments involving primary culture of oocytes, female mice at the ages specified in the description of each experiment, were stimulated with intraperitoneal injection 10 IU equine Chorionic Gonadotropin (eCG; A.F. Parlow, National Hormone and Pituitary Program, National Institute of Diabetes and Digestive and Kidney Diseases, USA), ovaries were harvested 44 hours after eCG injection, and GV oocytes were collected following ovarian puncture in M2 media (Sigma) containing 5 μM Milrinone. For *in vitro* oocyte maturation assessment, oocytes

were washed 5 times to remove the Milrinone (Sigma) and intact COC complexes were cultured for 18-20 hours at 37ºC in M2 media. To obtain zygotes, 5 to 8-wk-old female mice were superovulated with intraperitoneal injection 10 IU eCG followed by 10 IU of human Chorionic Gonadotropin (hCG; A.F. Parlow, National Hormone and Pituitary Program, National Institute of Diabetes and Digestive and Kidney Diseases, USA) 46 hours later. The female mice were placed with 8-12-wk-old WT males of known fertility at the time of hCG injection. The contents of the ampulla were collected only from females with a visible plug 18 hrs after hCG injection. Ampulla contents were evaluated for the presence of zygotes by microscopy following a brief digestion in hyaluronidase (Sigma). Isolated cells were cultured overnight in M16 media (Sigma) to confirm fertilization and/or fertilization failure.

Maintenance of cell lines

HeLa cells (ATCC) were grown in DMEM (Gibco) supplemented with 10% FBS (Hyclone) and 1% Penicillin/Streptomycin (Gibco), and were split using 0.25% Trypsin-EDTA (Gibco) approximately every two days, when cells reached 80% confluence. Experiments were performed on cells once they reached approximately 80% confluence.

Genotyping and DNA PCR

Ear tags were taken from mice and DNA was extracted for PCR (Viagen, Cat#102-T). Lysed ear tags were heated at 55^0C for 3-4 hours and then 85^0C for 1 hour to denature lysozymes in the buffer. PCR reactions were conducted in 25 uL volume reactions with 1uL lysed mouse DNA, 19.5 uL nuclease free water, 1 uL dNTP mixture (BioPioneer, Cat# MDM-4), 0.5 uL forward and reverse primer mixture (10uM), 2.5 uL 10x buffer (Denville, cat# CB3702-7), and 0.5 uL Taq polymerase (Denville, Cat# CB4050-2). 1.2% agarose gel with ethidium bromide was used for visualization of gene bands. Primers used are listed in Table S2.

Fertility and estrus cycle assessment

Zfp36l2-Fl/Fl (n=7), *Zfp36l2*-cKO (n=8), and WT (n=8) females were individually set up in breeding pairs with WT males of known fertility and followed for 6 months. At the time breeder cages were set up, female mice were between 8-16 weeks old. Males in each breeding pair were exchanged at 6 months of age. Maternal abdomens were monitored for swelling suggestive of pregnancy. Cages were checked daily to ensure that newly born pups were counted before potential cannibalization. To evaluate mating behavior, 6 WT male (aged 7-8 weeks) and *Zfp36l2*-cKO virgin female (aged 9-10 weeks) pairs were set up and females were evaluated each morning for 44 consecutive days for the presence or absence of a seminal vaginal plug. To assess estrus cycle length, *Zfp36l2*-cKO and *Zfp36l2*-Fl/Fl female mice (n = 3 at 15-19 weeks old for each genotype) were housed in individual cages for 25 days to obtain at least three full estrus cycles from each mouse. Females were housed individually for at least 1 month prior to the start of the experiment. Each day, a sterile pipette and 50uL of water were used to obtain a vaginal smear from each mouse. Vaginal smear slides were stained with 0.4% methyl blue and analyzed to determine the estrus cycle stage for each mouse each day using standard published criteria (Nelson et al., 1982). Estrus length was defined as the time between two consecutive estrus observations.

Ovarian follicle counts

Unstimulated ovaries were collected from singly-housed, unstimulated adult cKO and Fl/Fl females aged 15-23 weeks of age in the diestrus phase of the cycle (n = 7 for each genotype). Stimulated ovaries were isolated from sexually immature, 4-5 week old females 46 hours following intraperitoneal injection 10 IU eCG (n = 3 for each genotype). Whole ovaries from cKO and *Zfp36l2*-Fl/Fl mice were fixed in neutral buffered formalin for 16-24 hours. Whole ovaries were processed in pairs, embedded in paraffin wax and then cut in 5 um sections. All slides were Haemotoxylin and Eosin stained using standard protocols (Cardiff et al., 2014) and scanned using Zeiss Axio Scan.Z1. Every fifth section, of one ovary per mouse, was analyzed

with respect to the number and stage of follicles. Follicles were only counted if nucleus was clearly visible to avoid duplicate counting. Counts were binned into the following stage categories: Primordial (Pr), Primary (1^0), Secondary (2^0), Antral (Ant), Corpus luteum (CL), Atretic (Atr). To correct for counting only every fifth section, total ovary counts were multiplied by five as per published protocols (Flaws et al., 1997; Hirshfield and Midgley, 1978). Statistical analysis was performed using GraphPad Prism. All p-values obtained through unpaired T-tests.

Evaluation of oocyte diameter

To measure oocyte diameter, 5 week-old *Zfp36l2*-Fl/Fl and -cKO female mice (4 mice per genotype) were stimulated with intraperitoneal injection 10 IU eCG, ovaries were harvested 44 hours after eCG injection, and GV oocytes were collected following ovarian puncture in M2 media (Sigma) containing 5 μM Milrinone (Sigma). All GV oocytes completely enclosed in cumulus cells were collected, stripped of granulosa cells mechanically with pipetitng, and photographed. Oocyte diameters were determined by measurement of the images using NIH Image J software and a scale for calibration, measuring both the largest diameter the diameter perpendicular to this line, both exclusive of the zona pellucida. The oocyte diameters and distribution for each group were determined using Prism (GraphPad Software).

Ovulation, maturation and fertilization

To evaluate maturation and ovulation rates *in vivo*, 5 to 8 week-old female mice were superovulated with intraperitoneal injection 10 IU eCG followed by 10 IU of hCG 46 hours later to trigger oocyte maturation and ovulation. Oocytes released from the ovary were collected from the ampulla 15 hrs after hCG injection and cumulus cells were removed by brief hyaluronidase digestion per standard protocol and evaluated under the dissection microscope (Leica M125). The data presented is from oocytes pooled from 9 mice of each genotype in 3 separate experiments. To evaluate maturation rates *in vitro*, 5 to 8 wk-old female mice were stimulated with 10 IU of eCG by intraperitoneal injection. To isolate periovulatory, germinal vesicle (GV)

stage oocytes, the mice were sacrificed 46 hrs after eCG injection, the ovaries were harvested and follicles were collected following ovarian puncture in M2 media containing 5 μM Milrinone. Only oocytes completely enclosed in cumulus cells were collected. Oocytes were washed 5 times to remove the Milrinone and intact COC complexes were cultured for 20 hours at 37^{o}C in M2 media before assessment for polar body extrusion by microscopy. A total of 102 Fl/FL and 57 cKO oocytes were assessed from 5 mice of each genotype in 2 separate experiments. To test fertilization and obtain zygotes, 5 to 8-wk-old female mice were superovulated with intraperitoneal injection 10 IU eCG followed by 10 IU of hCG 46 hours later. The female mice were placed with 12-wk-old WT males of known fertility at the time of hCG injection. The next morning, the female mice were checked for vaginal plug. The contents of the ampulla were collected only from females with a visible plug 18 hrs after hCG injection. Ampulla contents were evaluated for the presence of zygotes by microscopy following a brief digestion in hyaluronidase (Sigma). Isolated cells were cultured overnight to confirm fertilization and/or fertilization failure. A total of 95 zygotes from 5 *Zfp36l2*-Fl/Fl females were evaluated.

Single cell library preparation

Five week-old *Zfp36l2*-Fl/Fl and -cKO female mice were stimulated with 10 IU of eCG by intraperitoneal injection. The mice were sacrificed 46 hrs after eCG injection, ovaries were harvested and GV oocytes were collected following ovarian puncture in M2 media containing 5 μM Milrinone. Only oocytes completely enclosed in cumulus cells were collected and granulosa cells were mechanically removed using pipets of decreasing diameter. The zona pellucida was removed with a brief incubation in Acid Tyrode's solution (Millipore) and the oocytes were immediately washed in PBS with 0.1% BSA. Single mouse oocytes were transferred into nuclease free 0.2ml PCR tubes and placed immediately on dry ice. In total, 26 Fl/Fl oocytes and 24 cKO oocytes were analyzed from 4 separate littermate pairs for each genotype in 4 separate experiments. As an additional means of normalization, ERCC Spike-In RNAs (Life

Technologies, Cat.no. 4456740) were added to the first set of samples, which comprised 9 Fl/Fl and 7 cKO single oocytes. The ERCC stock provided by the manufacturer was diluted 1:25,000 and 0.25 µl of the dilution was added to each oocyte before cell lysis and RNA-seq library preparation, and ERCC reads were analyzed downstream in parallel with reads from the oocyte transcriptome. RNA isolation and cDNA synthesis was carried out using the SMARTer® Ultra® Low Input RNA Kit for Sequencing - v3 (Clontech, Cat.no. 634851). Libraries were generated from the resulting cDNA using the Nextera XT DNA library preparation kit (Illumina, Cat.no. FC-131-1024) as previously described (Mora-Castilla et al., 2016).

RNA sequencing and data analysis

Approximately 21.5 million reads were generated per sample, and 58% of these reads were uniquely mapped via STAR (v2.3.0.1) (Dobin et al., 2013; Dobin and Gingeras, 2015) after trimming for low quality reads and adapter sequences via cutadapt (v1.8.1) (Martin, 2011). Processing of .sam files was accomplished with Samtools (Li et al., 2009). Counts for each gene were quantified using the python script rpkmforgenes.py (Ramsköld et al., 2009) and annotated using the Refseq mm10 genome, which had been concatenated with the ERCC Spike-In annotation. Reads were filtered, such that genes without at least one sample with at least 10 raw reads and one RPKM read were removed from the analysis, and overlapping RefSeq transcripts were collapsed giving one expression value per gene locus. The count data was normalized and differential expression was performed using the R (v.3.1.1) package scde (v.1.2.1). Briefly, scde accounts for technical noise by fitting individual error models for single cell RNA-seq measurements. This Bayesian approach estimates both the likelihood of a gene being expressed in a given cell as well as expression fold change between subpopulations (Kharchenko et al., 2014). Genes with an adjusted P value (Q-value) less than 0.05 were considered differentially expressed unless otherwise noted. Heatmaps, scatter plots, and boxplots were constructed using the R (v.3.1.1) package gplots. GO categories for heatmaps

were compiled with AmciGO (v2.3.2) (Ashburner et al., 2000a; Gene Ontology Consortium, 2014) and HIstome (Khare et al., 2012a). All functional enrichment analyses were generated using DAVID gene annotation and analysis resource (Huang et al., 2009). AU-rich elements were quantified using the ARE Score algorithm (Spasic et al., 2012). Cumulative plots in Supplemental Figures S2 and S3 were generated from custom Python scripts using ZFP36L2 RIP data from Zhang et al. (Zhang et al., 2013a), ARE Scores from Spasic et al. (Spasic et al., 2012), GO terms from AmiGO (Ashburner et al., 2000a; Gene Ontology Consortium, 2014), and single cell RNA seq data from this study with Upregulated and Downregulated transcripts defined as described above ($Q<0.05$), and log2 fold change (cKO/WT) values calculated for transcripts using the scde algorithm (Kharchenko et al., 2014).

RNA isolation and qPCR analysis

GV oocytes were isolated from Fl/Fl and cKO females aged 5 to 8 weeks in M2 media supplemented with 5 uM Milrinone following eCG stimulation, ovarian follicle puncture and mechanical removal of granulosa cells as described above. One hundred mouse oocytes of each genotype were isolated and pooled from 3-10 mice of each genotype in at least 3 separate experiments using quantitative real time PCR (qPCR). *Gapdh* or *L19* were used as an endogenous controls. RNA isolation, cDNA synthesis and qPCR reactions were set up according to the instructions in the Power SYBR Green Cells-to-Ct kit (Life Technologies, Cat. no. 4402953) per the manufacturer's protocol with the exception that lysis of the 100 oocyte samples was performed in 50 microliters of lysis buffer. The cDNA generated was used for gene expression analysis in triplicate by qPCR with custom synthesized primers and details of the primers used are listed in Table S2. qPCR was performed in at least 2 pools of oocytes from at least 10 mice of each genotype.

Run-on transcription assays

GV oocytes were isolated from Fl/Fl and cKO females aged 5 to 8 weeks in M2 media supplemented with 5 uM Milrinone following eCG stimulation, ovarian follicle puncture and mechanical removal of granulosa cells as described above. Mice were age-matched for each experiment and littermate controls were used when possible. The last 5 washes were carried out in M2 medium alone to remove the milrinone. Transcriptional activity was determined in run-on studies with 5-ethynyl-uridine (EU) using the Click-iT RNA Alexa Fluor 488 Imaging kit (Invitrogen, C10329, C10330). The oocytes were cultured in M2 medium with or without 2mM EU at 37^{o}C for 45 minutes. After the incubation in EU, the oocytes were then transferred to Acid Tyrode's solution briefly to remove the zona and washed three times in Phosphate Buffered Saline (PBS) with Bovine Serum Albumin (BSA). The oocytes were then immediately fixed in 4% Paraformaldehyde (PFA) for 30 minutes and washed three times in PBST as above. The oocytes were then permeabilized in 0.1% Triton X-100 in PBS 1 hour at room temperature and washed 3 times in PBST. The oocytes were then incubated for 30 minutes at room temperature with the Click-iT reaction cocktail prepared per the manufacturer's proctocol and washed again 3 times with PBST. DNA staining was performed using Hoechst 33342 at 1:1000 in PBST for 15 minutes immediately prior to imaging. SN was assigned to nuclei demonstrating a complete ring of chromatin around the nucleolus when visualized by confocal microscopy at 40X magnification. The fluorescent signal was detected using the 40X objective of a laser-scanning confocal microscope (Zeiss LSM780). As above, to accurately compare fluorescent intensity between oocytes, the laser power was set based upon the oocyte with the strongest signal and the same setting was used to evaluate all oocytes. A total of 140 and 80 oocytes were screened from 7 *Zfp36l2*-Fl/Fl and 10 *Zfp36l2*-cKO mice, respectively, in 2 separate experiments.

Western blotting

GV oocytes were isolated from Fl/Fl and cKO females aged 5 to 8 weeks in M2 media supplemented with 5 uM Milrinone following eCG stimulation, ovarian follicle puncture and

mechanical removal of granulosa cells as described above. An equal number of Fl/Fl and cKO GV oocytes were lysed in 25 ml of RIPA lysis buffer, incubated on ice for 1 hour and then mixed with the same volume of 2x Laemmli sample buffer (Bio-Rad, Cat. No. 1610737). After boiling for 10 minutes, total lysate was fractionated on SDS-PAGE and Western blotting was performed following transfer to Immobilon-P membranes (Millipore, Cat. No. IPVH00010) using standard protocols. Blocking was performed using Detector^{TM} Block buffer (KPL, Cat. No. 71-83-00) and detection was performed using SuperSignal^{TM} West Dura Extended Duration Substrate (ThermoFisher scientific, Cat. No. 34075). For ZFP36L2, 50 GV oocytes were loaded per lane; for H3K4me3, 70 GV oocytes were loaded per lane. Primary antibodies and dilutions used were as follows: anti-ZFP36L2 (1:2000, Sigma HPA047428); anti-b-Actin (1:5000, Sigma, Cat. No. A5316); anti-H3K4me3 (1:2000, Abcam, Cat. No. ab8580) and anti-Histone H3 (1:2500, Milipore. Cat. No. 05-928). Internal loading controls were detected on the same membrane. Detection of each target was performed at least twice using pooled oocytes from different mice.

Immunofluorescence

GV oocytes were isolated from Fl/Fl and cKO females aged 5 to 8 weeks in M2 media supplemented with 5 uM Milrinone following eCG stimulation, ovarian follicle puncture and mechanical removal of granulosa cells as described above. The zona pellucida was removed in Acid Tyrode's solution, followed by three washes in PBS with 0.1% BSA. The oocytes were fixed with 4% PFA in PBS for 1 hour at room temperature followed by three washes in PBS with 0.1% BSA and 0.1% Tween (PBST). The oocytes were then permeabilized by incubation with 1% Triton X-100 in PBS for 15 minutes at room temperature. After permeabilization, the oocytes were again washed three times in PBST and blocked with 4% normal donkey serum in PBST for 1hr at room temperature. The oocytes were incubated with primary antibodies overnight at 4^0C. All washes and incubations were performed with gentle agitation. Anti-histone H3 (tri methyl K4) antibody (Abcam, Cat. No. ab8580) was used at a dilution of 1:200, anti-trimethyl-histone H3

(Lys9) antibody (Millipore, Cat. No. 07-442) was used at a dilution of 1:50, and pan-histone H3 monoclonal antibody (ThermoFisher Scientific, Cat. No. AHO1432) was used at a dilution of 1:250. After removing the primary antibody, the oocytes were washed three times in PBST and then incubated with Alexa Fluor 555 conjugated donkey anti-rabbit IgG (H+L) secondary antibody (ThermoFisher Scientific, Cat. No. A-31572) or Alexa Fluor 488 conjugated donkey anti-mouse IgG (H+L) Highly Cross-Adsorbed secondary antibody (ThermoFisher Scientific, Cat. No. A-21202) for 1 hour at room temperature. The oocytes were then washed three times with PBST and mounted in 96 well glass bottom plates with a 50% dilution of Vectashield antifade mounting medium with DAPI (Vector Laboratories, Cat. No. H-1200). The fluorescent signal was detected using the 40X objective of a laser-scanning confocal microscope (Zeiss LSM780). To accurately compare fluorescent intensity between oocytes, the laser power was set based upon the oocyte with the strongest signal and the same setting was used to evaluate all oocytes. Signal quantification was performed using NIH Image J software; Corrected Total Cell Fluorescence (CTCF) was calculated as CTCF = Integrated density – (Area of selected region X Mean fluorescence of background readings) (Burgess et al., 2010; McCloy et al., 2014). CTCF was calculated for H3K4me3, H3K9me3, total H3, and DAPI signals; H3K4me3, H3K9me3, and total H3 were normalized to the DAPI signal for each oocyte nucleus. Significance as demonstrated in Figure 2.8B and the corresponding figure legend was calculated with an unpaired Student's parametric T-test, with error bars representing SEM. Greater than 40 oocytes pooled from 15 mice were examined in 4 separate experiments for each histone target.

RNA immunoprecipitation

Magna RIP RNA-binding protein Immunoprecipitation Kit (EMD Millipore. Cat. No. 17-700) was used for RIP assay, with some modifications from the manufacturer's protocol. In brief, 80-90% confluent HeLa cells in 6X10 cm plates were dissolved in 300 ml of lysis buffer and

supernatants were harvested as cell lysates after centrifugation at 14,000 rpm for 10 minutes. Cell lysates were pre-cleared by incubating with washed magnetic beads for 1 hour at 4°C. During pre-clearing, 5 mg of anti-ZFP36L2 antibody (Abcam, Cat. No. ab70775) was bound to pre-washed magnetic beads by incubating with rotation for 30 minutes at room temperature (RT) and 5 mg of purified rabbit IgG was bound to pre-washed magnetic beads as a negative control. Pre-cleared lysate was divided into two parts and incubated overnight at 4°C with anti-ZFP36L2 antibody-bound magnetic beads or rabbit IgG-bound magnetic beads, respectively. On the following day, beads were washed 6 times with RIP wash buffer and then treated with proteinase K for 30 minutes at 55°C. Supernatants were separated using a magnetic separator, transferred into new tubes, and then cleaned by phenol chloroform extraction. Nucleotides were isolated by ethanol precipitation and re-suspended in 12.5 ml RNAse-free water. DNA was removed by using TURBO DNA-free kit (ThermoFisher scientific, Cat. No. AM1907) and then cDNA was synthesized using superscript IV reverse transcriptase (ThermoFisher scientific, Cat. No. 18090050). Quantitative real time PCR assay was performed using SsoAdvanced Universal SyBr Green Supermix (Bio-Rad, Cat. No. 1725274). Primers used in for qPCR are shown in Table S2. Graph in Figure 2.6B represents average values of 5-6 independent experiments. For normalization, fold enrichment of each RNA in the anti-ZFP36L2 antibody-precipitated sample was determined relative to the rabbit IgG precipitated sample and values were then normalized to the value for L19. Statistical significance was calculated via two-way ANOVA, after log transformation to generate a more normal distribution of the data.

RNA decay assays

HeLa cells cultured on 6 well-plates were transfected with either 20mM of siRNA targeting human ZFP36 family members (siZFP36; guuguggaugaaguggcagcg) (Jing et al., 2005) or siRNA against Luciferase (siCtrl: cguacgcggaauacuucga) using Silentfect Lipid Reagent for RNAi (Bio Rad, Cat. No. 1703361). At 24 hours post-transfection, cells were

transfected again with the same siRNAs and incubated for another 24 hours. Then, cell culture media was exchanged with fresh media and, at 2 hours after media exchange, cells were treated with 0.5 mg/ml of Actinomycin D (Sigma) for 0 hours, 1 hour, 2 hours and 4 hours. Western blotting was performed as above using an antibody that recognizes both ZFP36L2 and ZFP36L1 (Cell Signaling Technologies Cat. No. 2119) at 1:5000 dilution. ZFP36L1 and ZFP36 were not detectable by Western blotting before or after siRNA treatment (data not shown), consistent with published data (Nagaraj et al., 2011; Paulo and Gygi, 2017). GAPDH served as loading control. Total RNAs were isolated from cells using Trizol reagent (ThermoFisher scientific, Cat. No. 15596) and qRT-PCR was performed using iScript Reverse transcription supermix (Bio-rad, Cat. No. 1708840) and SsoAdvanced Universal SyBr Green Supermix (Bio-Rad, Cat. No. 1725274). Half-lives were calculated by nonlinear regression, with an exponential fit, using GraphPad Prism Software and presented in hours. Data were plotted on a logarithmic y-axis to show a linear best-fit line. Accurate half-lives could not be calculated for mRNAs for which the regression line was near horizontal.

ARE conservation

3'UTR sequences for indicated transcripts were compiled from the NCBI Gene database. For genes with more than one transcript, that with the highest AREScore (typically the longest transcript) was used for analysis. AREScores were determined with the AREScore algorithm, using the 3'UTR fasta sequences as input (Spasic et al., 2012).

Quantification and Statistical Analysis

Statistical analysis and software used have already been detailed in the methods sections above, associated with each experiment, as well as in the figure legends. To summarize, all statistical comparisons seen in the main and supplementary figures, unless otherwise specifically stated in the detailed experimental methods above, were conducted using an unpaired T-test, where error bars represent standard error of the mean (SEM). In cases

where data did not meet the assumptions of the statistical approach (i.e., normal distribution), a logarithmic transformation of the data was first conducted, followed by the statistical test. Statistical analysis associated with cumulative distribution plots was conducted using the nonparametric Kolmogorov-Smirnov (KS) test, which is appropriate for comparing probability distributions. For RNA sequencing analysis, count data was normalized to ERCC spike-in concentration, as well as to the whole genome using Reads Per Kilobase of transcript per Million of mapped reads (RPKM). Differential expression was performed using the R (v.3.1.1) package scde (v.1.2.1), which accounts for technical noise by fitting individual error models for single cell RNA-seq measurements. This Bayesian approach estimates both the likelihood of a gene being expressed in a given cell as well as expression fold change between subpopulations, adding robustness to single-cell differential expression analysis (Kharchenko et al., 2014). Genes with an adjusted P value (Q-value) less than 0.05 were considered differentially expressed unless otherwise noted, taking into account multiple testing in genome-wide comparisons. RNA half-life was calculated by nonlinear regression, using an exponential fit; the data is plotted on a logarithmic y-axis for clarity of presentation. All statistical calculations were conducted using GraphPad Prism Software, R (v.3.1.1), and python.

Data and Software Availability

The RNA sequencing data associated with this study have been made publicly available under the GEO accession GSE96638. Additional data tables, including read counts, differential expression quantification, and calculation of transcript-associated ARE Scores (Spasic et al., 2012) are available in Supplemental Table S2. Software used for all statistical analysis includes GraphPad Prism, R (v3.1.1), and python. Specific R packages used include scde (v.1.2.1) and gplots. RNA sequencing data was trimmed of poor-quality sequences and adapter content with cutadapt (Martin, 2011), mapped with STAR (Dobin et al., 2013; Dobin and Gingeras, 2015) to the Refseq mm10 genome, and file formats were converted with Samtools (Li et al., 2009).

Reads were quantified using the python script rpkmforgenes.py (Ramsköld et al., 2009). Publicly available programs used for various computational analyses include the AREScore algorithm (Spasic et al., 2012), DAVID Gene Ontology Analysis(Huang et al., 2009), AmiGO (Ashburner et al., 2000b), and HIstome (Khare et al., 2012b). Custom Python scripts for generation of cumulative plots in Supplemental Figures S2 and S3 are available upon request.

Results

ZFP36L2 expression in oocytes is required for female fertility

To test the importance of ZFP36L2 in the oocyte-to-embryo transition, we generated conditional knockout mice in which *Zfp36l2* is deleted only in the growing oocyte. *Zfp36l2*-floxed (*Zfp36l2*^{Fl}) mice were made harboring *loxP* sites flanking exon 2, which encodes the RNA-binding domain and the majority of the coding region (467 of 484 amino acids) (Figure 2.2A). Oocyte-specific deletion was achieved by crossing *Zfp36l2*$^{Fl/Fl}$ mice (Fl/Fl or control) with mice expressing Cre recombinase driven by the *Zp3* promoter (Zp3-Cre), which is expressed only in growing oocytes at the primary follicle stage and onward (de Vries et al., 2000). Robust knockout in *Zfp36l2*$^{Fl/Fl}$-Zp3-Cre^{+} mice (*Zfp36l2*-cKO or cKO) was confirmed in GV oocytes at the DNA, RNA and protein levels (Figures 1A-B and S1A-B). Adult *Zfp36l2*-cKO females displayed normal estrus cycles (Figure 2.2C) and mating behavior (Figure 2.2D). However, crosses with wild-type males revealed that *Zfp36l2*-cKO females were completely infertile (Figure 2.1C), demonstrating that oocyte-specific expression of ZFP36L2 is required for female fertility.

ZFP36L2 is required for oocyte developmental competence

Histological analyses of ovaries from cKO and control adult females performed to investigate whether ZFP36L2 has roles in oocyte and/or follicle growth revealed that ovaries of

cKO females contained follicles at all developmental stages (Figure 2.1D). Quantification of whole ovaries from unstimulated adult cKO and control females showed overall similar numbers of follicles at each stage of growth and similar numbers of corpora lutea (Figure 2.1E), demonstrating that oocyte-specific expression of ZFP36L2 is not required for growth to the GV stage or for ovulation. Further, GV stage oocytes isolated from cKO and control females were indistinguishable from controls with respect to oocyte morphology and diameter (Figure 2.2E). However, follicle counts in cKO ovaries did demonstrate lower numbers of primary follicles and a trend toward lower numbers of follicles at later stages (Figure 2.1E). To examine this potential defect more closely, we repeated the quantification of follicles in ovaries from sexually immature females following pharmacological ovarian stimulation. Following gonadotropin challenge, cKO ovaries contained fewer growing follicles and corpora lutea relative to control ovaries (Figure 2.2F). cKO ovaries also had higher numbers of atretic oocytes compared to control ovaries, specifically at the primary and secondary stages (Figures S1F-G). Together, these findings suggest a role for ZPFP36L2 in the early preantral follicle. However, that similar numbers of oocytes grew to the GV stage and were ovulated in adult cycling mice demonstrate that the complete infertility of adult cKO females cannot be explained by a gross defect in oocyte growth, folliculogenesis, or ovulation alone.

To uncover the cause of female infertility in the *Zfp36l2*-cKO mice, we next investigated the role of ZFP36L2 in oocyte maturation and fertilization. Consistent with our histological analyses above, cKO females failed to upregulate oocyte production following gonadotropin stimulation (4 ± 3.6 vs. 23 ± 7.7 ovulated in cKO and Fl/Fl, respectively). In further analyses, cKO oocytes exhibited significantly lower rates of maturation than control oocytes, both following superovulation *in vivo* (63% in cKO vs. 85% in control, p=0.025) (Figure 2.1F) and following culture of cumulus-oocyte complexes *in vitro* (51% in cKO vs. 76% in control, p=0.04) (Figure 2.2H). Many of the cKO oocytes that reached the MII stage displayed abnormally large

polar bodies, a phenotype suggestive of a spindle defect (Figures 1F and S1H, insets). Following superovulation and mating experiments to examine fertilization, no zygotes were recovered from 8 cKO females compared to ninety-five zygotes from 5 control females, demonstrating that cKO oocytes are incompetent to undergo fertilization (Figure 2.1G). Together, these results demonstrate that oocytes lacking *Zfp36l2* exhibit abnormal oocyte maturation and a failure to undergo normal fertilization, establishing a critical role for oocyte-specific expression of ZFP36L2 in oocyte developmental competence and female fertility.

ZFP36L2 regulates mRNAs encoding factors with central roles in transcription regulation and chromatin modification

Given the well-established role of ZFP36L2 in mRNA degradation, we next sought to identify oocyte mRNAs dysregulated as a consequence of *Zfp36L2*-cKO. To this end, we performed single-cell RNAseq analyses, comparing GV oocytes from 26 control and 24 cKO females. cKO oocytes exhibited gene expression patterns distinct from control oocytes as shown by both hierarchical clustering (Figure 2.3A) and unbiased principal component analysis (Figure 2.4A). Analysis of control RNAs covering a wide range of known lengths and concentrations spiked in to the single-cell RNAseq reactions (Baker et al., 2005; External RNA Controls Consortium, 2005) showed read counts correlated with input concentrations (Figure 2.4B) and no significant differences in overall read counts in cKO vs. control oocytes (Figure 2.4C). These data provide a robust validation of our RNAseq pipeline and demonstrate that *Zfp36l2* cKO does not result in a large, nonspecific shift for a substantial fraction of the polyadenylated mRNA pool in the oocyte. Instead, our transcriptome analyses revealed significant, reproducible changes in the levels of a subset of mRNAs in the absence of ZFP36L2 (Figure 2.3A and Table S1). Of 16,614 mRNAs with detectable expression in single oocytes, we observed higher levels of 1,418 (8.5%) and lower levels of 1,201 (7.2%) transcripts in cKO

oocytes relative to controls (log_2 fold change>1; adjusted p-value (Q)<0.05), consistent with a significant role for ZFP36L2 in regulating the mRNA pool during oocyte growth.

Because ZFP36L2 is known to trigger mRNA decay, mRNAs *upregulated* by *Zfp36l2* deletion are candidates to be direct targets of ZFP36L2. Consistent with this, transcripts upregulated upon *Zfp36l2*-cKO were enriched for mRNAs that associate with ZFP36L2 in mouse erythroid cells (Zhang et al., 2013b) (Figure 2.4E). Moreover, transcripts that were both upregulated in *Zfp36l2*-cKO oocytes and bound by ZFP36L2 in mouse erythroid cells were enriched for 3'UTR AREs, relative to downregulated and unbound mRNAs, as measured by higher AREScores (Figure 2.4F), a metric shown to correlate with mRNA decay triggered by ZFP36 proteins (Spasic et al., 2012). Gene ontology (GO) analysis revealed that the two most highly enriched categories among mRNAs upregulated with *Zfp36l2* deletion were *Regulation of Transcription* (p=8.8e-6; 184 genes) and *Transcription* (p=2.4e-5; 150 genes) (Figure 2.3B). In all, four of the top six most highly enriched categories related to transcription regulation and chromatin modification. We validated a subset of upregulated mRNAs encoding transcription regulators with low, medium, and high fold-changes by RT-qPCR (Figure 2.3C). GO analysis of mRNAs downregulated in cKO oocytes, which would be expected to be regulated indirectly by ZFP36L2, encoded proteins enriched for several biological pathways not obviously relevant to oocyte development (Figure 2.4D).

ZFP36L2 is required for global transcriptional silencing during the final stages of oocyte growth

Our discovery that oocytes lacking ZFP36L2 exhibit dysregulated expression of many transcriptional regulators raised the possibility that ZFP36L2 is important for oocyte transcriptional silencing. To directly test this, we performed transcriptional run-on assays using 5-ethynyl uridine (EU) followed by fluorescence staining to examine transcription activity on a global scale. Strikingly, we found that ~80% of GV oocytes from cKO mice failed to undergo

transcriptional silencing (Figures 3A-B). In contrast, only 8% of control GV oocytes had this phenotype. This transcriptional silencing defect occurred in the absence of a gross defect in chromatin organization, with a similar number of cKO oocytes reaching the SN configuration (74% cKO v. 76% Fl/Fl) (Figure 2.5C). These observations demonstrate an essential role for the mRNA decay activator, ZFP36L2, in mediating global transcriptional silencing—independent of chromatin condensation—in the late oocyte.

mRNAs encoding histone demethylases for H3K4 and H3K9 are direct targets of ZFP36L2 decay

To identify mRNAs that ZFP36L2 might regulate to drive global transcriptional silencing, we evaluated the effect of *Zfp36l2*-cKO on groups of mRNAs in more than 1,400 narrow GO categories. This identified “histone demethylases” as a group that was significantly upregulated upon *Zfp36l2*-cKO (Figure 2.7A). We next investigated whether these are direct ZFP36L2 direct targets. Such direct target mRNAs would be expected to not only be upregulated in oocytes with *Zfp36l2*-cKO but to also (i) contain 3’UTR AREs, (ii) be bound by ZFP36L2, and (iii) be stabilized in the absence of ZFP36L2. In support, we found histone demethylase mRNAs demonstrated AU-rich 3’UTRs with high AREScores compared to mRNAs for other transcription regulators (Figure 2.7B). This group of histone demethylase mRNAs was also enriched for binding to ZFP36L2 in mouse erythroid cells (Figure 2.7C), relative to groups of mRNAs encoding other chromatin modifiers and to all other expressed transcripts as a whole.

Remarkably, 11 of the 12 histone demethylase mRNAs that were upregulated encode demethylases for histone H3K4 or H3K9 (Figure 2.6A). Examination of individual transcripts revealed the majority of these mRNAs (8 of 11) contain AU-rich 3’UTRs with AREScores of ≥4, predictive of ARE-mediated mRNA decay (Spasic et al., 2012) (Figure 2.6A). In support of the functional importance of these AU-rich 3’UTR sequences, a high AREScore is an evolutionary conserved feature of this group of histone demethylase mRNAs in mammals (Figure 2.7D). In

contrast, we identified 2 histone demethylase mRNAs that were modestly downregulated (1.5-1.7-fold) (Table S1).

To test whether histone demethylase mRNAs are bound and degraded by ZFP36L2, we turned to HeLa cells, where ZFP36L2 is the predominant ZFP36-family protein expressed (Nagaraj et al., 2011; Paulo and Gygi, 2017) and assessment of its mRNA association and degradation activity was feasible. RNA-immunoprecipitation (RIP) analysis using an antibody specific for ZFP36L2 revealed that ZFP36L2 associated with each of the histone H3K4 and H3K9 demethylase mRNAs upregulated in the absence of ZFP36L2 in the oocyte (Figure 2.6B). Depletion of ZFP36L2 using a validated siRNA targeting human ZFP36 proteins (Jing et al., 2005) resulted in stabilization of the majority of these histone demethylase mRNAs (Figures 4C-D and S3E). Together, these findings provide strong evidence that H3K4 and H3K9 histone demethylase mRNAs are direct targets of ZFP36L2-mediated RNA decay.

ZFP36L2 mediates global increases in H3K4 and H3K9 methylation in the final stages of oocyte growth

Our findings raised the question of whether ZFP36L2 acts to downregulate histone demethylase mRNAs to drive increased methylation levels at H3K4 and H3K9, changes closely associated with transcriptional silencing and chromatin in competent oocytes (De La Fuente, 2006; Kageyama et al., 2007). Immunofluorescence analyses performed in cKO and control GV oocytes to investigate this possibility revealed marked differences in the nuclei of cKO relative to control oocytes. Despite the ability of *Zfp36l2*-cKO oocytes to successfully complete the transition from the NSN to SN chromatin configuration (Figures 3C, 5A, and S4), oocytes lacking ZFP36L2 failed to upregulate both histone H3K4 and H3K9 trimethylation, while overall levels of histone H3 were not significantly affected (Figure 2.8A-B and S4). This defect was specific for SN oocytes, as levels of H3K4me3 and H3K9me3 were low in NSN oocytes both in the presence and absence of ZFP36L2. To confirm, we performed Western blot analyses, which

verified significantly lower levels of H3K4me3 in *Zfp36l2*-cKO relative to control oocytes (Figure 2.8C). We also attempted Western blot analysis with multiple antibodies against H3K9me3, but were unable to detect a signal with a feasible number of oocytes. These data indicate a critical role for ZFP36L2 in the accumulation of both H3K4 and H3K9 trimethylation that occurs in the final stages of growth in the transcriptionally silent, fully competent oocyte.

Discussion

Global transcriptional silencing is a highly conserved event in the final stages of oocyte growth tightly linked to the successful transition from oocyte to embryo (Andreu-Vieyra et al., 2010; Liu and Aoki, 2002; Zuccotti et al., 2002). Our study reveals a critical role for an mRNA decay factor, ZFP36L2, in regulating histone modification and bringing about this transcriptional silencing event (Figure 2.8D).

Our findings suggest that the role of ZFP36L2 in global transcriptional silencing in the oocyte occurs, at least in part, through a direct effect of ZFP36L2 on mRNA stability during oocyte growth. Our single-cell transcriptome analyses identified a group of mRNAs encoding histone H3K4 and H3K9 demethylases that are upregulated in the oocyte with *Zfp36l2*-cKO (Figures 2C, 4A, and S3A), contain 3'UTR AREs conserved among mammals (Figures 4A, S3B, and S3D), are bound by ZFP36L2 in both human HeLa cells (Figures 4B) and mouse erythroid cells (Figure 2.7C) (Zhang et al., 2013b), and, for the majority, are stabilized upon ZFP36L2 depletion (Figures 4C-D and S3E). Collectively, these findings strongly suggest that these histone demethylase mRNAs are direct targets of ARE-mediated decay by ZFP36L2 in the oocyte. Functional importance of downregulation of this group of H3K4 and H3K9 demethylase mRNAs by ZFP36L2 in the oocyte is suggested by our finding that loss of *Zfp36l2* resulted in a failure to accumulate increased levels of histone H3K4me3 and H3K9me3 (Figure 2.8A-C), marks with demonstrated roles in transcriptional repression (Bannister and Kouzarides, 2011;

Dahl et al., 2016a; Zhang et al., 2016a) and closely associated with both global transcriptional silencing and the developmentally competent oocyte state. In contrast, chromatin condensation – another important developmental event in the late oocyte – occurred independently of ZFP36L2 (Figures 3A, 3C, 5A-B, and S4). Together, our findings suggest a mechanism whereby ZFP36L2 acts as a developmental switch in the oocyte to downregulate a group of histone demethylase mRNAs to drive increased H3K4 and H3K9 methylation levels, global transcriptional silencing and the production of a developmentally competent oocyte (Figure 2.8D).

An important issue for the future is determining the precise role of the ZFP36L2-dependent histone H3K4 and H3K9 methylation in regulating chromatin structure and transcriptional silencing in the oocyte. While histone methylation has been implicated in the regulation of higher-order chromatin structure (Bannister and Kouzarides, 2011), our results demonstrate wild-type oocyte levels of these histone marks are not necessary for (or induced by) chromatin condensation from the NSN-to-SN configuration (Figures 3C and 5A-B). H3K9me3 is a strong candidate for a transcriptional repressive mark, as it is associated with transcriptional repression in a wide variety of cell types (Bannister and Kouzarides, 2011), and we and others observed increases in H3K9me3 levels concomitant with global transcriptional silencing in oocytes (Figure 2.8A-B) (De La Fuente, 2006; Kageyama et al., 2007). In contrast, while H3K4me3 is widely considered to be a mark of active transcription in somatic cells (Bannister and Kouzarides, 2011), several lines of evidence suggest that this mark can play the opposite role in the oocyte. H3K4me3 levels are higher in transcriptionally silent SN oocytes relative to transcriptionally active NSN oocytes (Figure 2.8A-B) (De La Fuente, 2006; Kageyama et al., 2007). Moreover, knockout of the H3K4 methyltransferase, MLL2, leads to lower levels of H3K4me3 in the face of persistent transcriptional activity (Andreu-Vieyra et al., 2010). Finally, it was recently shown that removal of H3K4me3 in transcriptionally silent oocytes by forced

overexpression of one of the histone demethylases we observed upregulated in *Zfp36l2*-cKO oocytes and regulated by ZFP36L2—KDM5B—partially restores transcriptional activity (Figures 2C and 4A-C) (Zhang et al., 2016a). Together, these observations strongly implicate both H3K9 and H3K4 methylation as marks critical for transcriptional repression in oocytes. Of note, we observed no significant effect on *Mll2* mRNA levels in *Zfp36l2*-cKO oocytes, and in contrast to histone demethylases, histone methyltransferases as a group did not stand out as potential ZFP36L2 targets (Figure 2.7A-C and Table S1). This suggests that ZFP36L2 affects the balance of histone H3 H3K4 and H3K9 methylation by targeting histone demethylases rather than histone methyltransferases and implicate mRNA decay as an important mechanism contributing to the epigenetic changes established in the late oocyte in preparation for the oocyte-to-embryo transition.

How is ZFP36L2-mediated mRNA decay regulated to allow timely transcriptional silencing? It is well established that ZFP36 family members are regulated post-translationally by phosphorylation, with each member containing upwards of ten phosphorylation sites, more than one of which have been shown to regulate mRNA decay-promoting activity (Sanduja et al., 2011). Therefore, while ZFP36 proteins are expressed in a wide range of cell types and tissues, their mRNA decay-promoting activity is likely to be dynamically regulated; the specific pathways regulating ZFP36L2 decay activity in the oocyte are not known. In addition, we note that our single-cell RNAseq analyses did not detect deadenylated mRNA, and our studies are not able to distinguish between deadenylated and degraded mRNAs. Thus, determining whether histone demethylase and/or other mRNAs targeted by ZFP36L2 in the oocyte are degraded or accumulate as deadenylated transcripts for later decay or reactivation by polyadenylation is an interesting future question.

In addition to the role for oocyte expression of ZFP36L2 in global transcriptional silencing during oocyte growth shown here, ZFP36L2 and the ARE-mediated mRNA decay

pathway likely play important roles at other stages of oocyte growth and in early embryo development as well. While oocytes lacking ZFP36L2 were able to grow and ovulate at relatively normal rates in unstimulated adult females (Figures 1D-E and S1E), stimulation of sexually immature and peri-pubertal cKO females with exogenous gonadotropin uncovered a likely additional role for ZFP36L2 in early preantral follicle growth (Figure 2.2F-G). Such a role could be in early oocyte growth or survival or in communication of the oocyte with the surrounding follicular cells, given the observed decrease in response to gonadotropin stimulation. It is also possible that decay of other ZFP36L2 mRNA targets not investigated here contribute to the defects in global transcriptional silencing, oocyte maturation and fertilization observed in *Zfp36l2*-cKO females. Roles for ZFP36L2 in oocyte maturation and embryo development past the 2-cell stage are evident from published studies with mice globally expressing a short N-terminal truncated version of ZFP36L2, although it is not known whether these roles represent a requirement for ZFP36L2 activity in the oocyte, the embryo or another cell type (Ball et al., 2014; Ramos et al., 2004). However, given that significant decay of maternal mRNA is known to occur during both oocyte maturation and the 2-cell embryo stage before the maternal-zygotic transition, it is tempting to speculate that ZFP36L2 is involved in decay at one or both of these stages.

There is a growing consensus that regulation of mRNA degradation during development might act as a developmental switch--a powerful means to switch from one program to another at pivotal developmental transitions (Svoboda et al., 2015). Our findings demonstrate such a role for an mRNA degradation pathway in establishing transcriptional silence and developmental competence in the final stages of oocyte growth. In particular, our demonstration that a post-transcriptional pathway controls the levels of multiple histone demethylase mRNAs is an unexpected finding that opens exciting avenues of investigation into the role of mRNA decay in regulating chromatin modification. The potential impact of these findings is strengthened given

our data suggesting that ARE-mediated decay of this group of mRNAs is a conserved process. Conservation of this pathway also raises the interesting possibility that defects in ZFP36L2 activity might play a role in a subset of currently unexplained causes of human female infertility. Together, that transcriptional silencing on a global scale is mediated by an mRNA decay activator acting on key regulators of transcriptional activity is an exciting possibility with broad implications for both development and transcriptional regulation. These data provide a compelling demonstration of the importance of mRNA decay at a pivotal developmental transition.

Acknowledgements

Chapter 2 contains material as a reprint of the following paper:

Dumdie JN, Cho K, Ramaiah M, Skarbrevik D, Mora-Castilla S, Stumpo DJ, Lykke-Andersen J, Laurent LC, Blackshear PJ, Wilkinson MF, Cook-Andersen H. Chromatin Modification and Global Transcriptional Silencing in the Oocyte Mediated by the mRNA Decay Activator ZFP36L2. *Developmental Cell*. 2018 Feb 5; 44(3):392-402. The dissertation author was the primary investigator and author of this material.

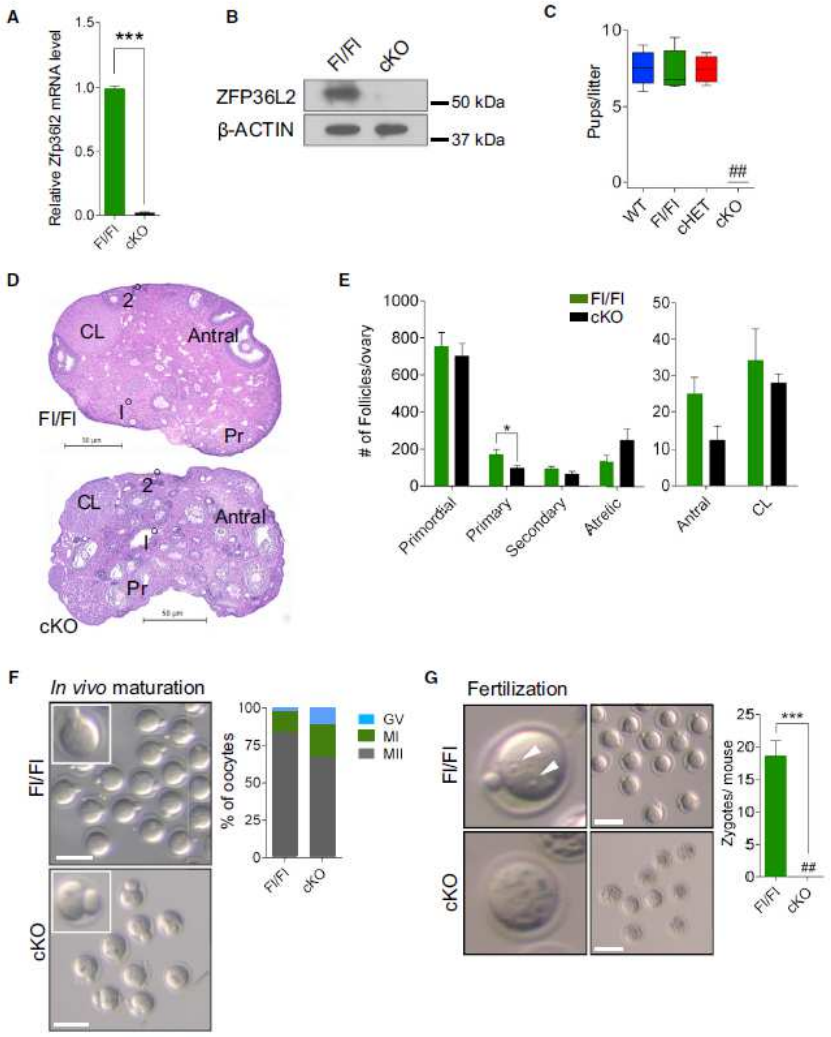


Figure 2.1 ZFP36L2 expression in oocytes is required for female fertility and oocyte developmental competence. (A) Relative level of *Zfp36l2* mRNA by qPCR analysis in cKO and Fl/Fl GV oocytes, normalized to *Gapdh*. Error bars are SEM (unpaired T-test; ***, p<0.001). (B) Western blot comparing ZFP36L2 in cKO and Fl/Fl GV oocytes. Beta-actin is the internal loading control. *(C)* Box plot showing the distribution of the number of pups/litter for cKO and controls over a six-month period. ##, no pups. *(D)* Hemotoxylin and Eosin stained sections of ovaries from unstimulated adult Fl/Fl and cKO females. Pr, primordial; 1º, primary; 2º, secondary ; Antral, antral; CL, corpus luteum. Scale bar, 50 um. *(E)* Number of follicles at each stage of development in whole ovaries of mature unstimulated Fl/Fl and cKO mice. Bars represent mean and SEM at each stage of development (unpaired T-test. *, p<0.05). *(F)* Number of GV, MI and MII oocytes isolated from cKO and Fl/Fl females following superovulation *in vivo* as a percent of total oocytes (unpaired T-test; *, p <0.05). Insets, MII oocyte enlarged to show the abnormally large polar bodies observed in many cKO oocytes. Scale bar, 100 µm. *(G)* Number of zygotes isolated from cKO and Fl/Fl females following superovulation and mating (unpaired T-test; ***, p <0.001. ##, no zygotes. Arrowheads in top left panel indicate pronuclei identified only in cKOs. Scale bar, 100 µm.

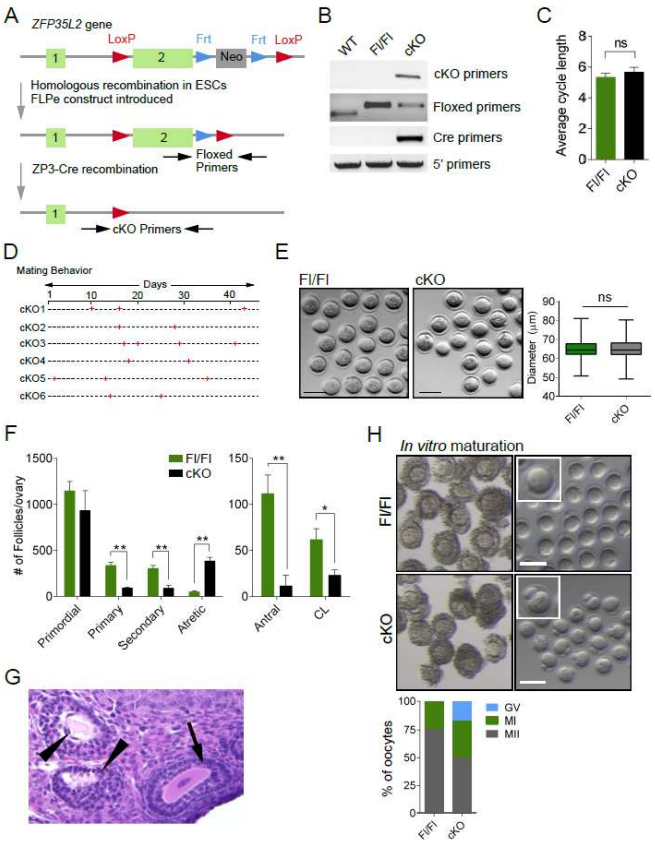


Figure 2.2 Generation of Zfp36l3-cKO mice and additional analyses of the infertile phenotype. (A) Scheme for generation of Zfp36l2 conditional KO mice. Red arrowheads depict loxP sites while blue arrowheads denote Frt sites. The location of primers used for detection of the floxed and knockout alleles are indicated. Additional details are available in Methods and primer sequences are listed in Table S2. (B) PCR analysis of genomic DNA from WT, Fl/Fl and cKO mouse GV oocytes, using primers indicated in Figure 2.2A. (C) Mean estrus cycle length is unchanged in cKO females as determined by vaginal cytology. Bars are mean and error bars are SEM of cycle length (ns, nonsignificant). (D) Zfp36l2-cKO females continue to mate as determined by the presence of seminal vaginal plugs. +, plug observed; —, no plug observed. (E) GV oocyte diameter is unaffected by ZFP36L2 cKO. The diameter of GV oocytes isolated from Fl/Fl and cKO females was measured exclusive of the zona pellucida. Scale bars, 100 µm. Box plots depict median diameters (Fl/Fl median = 65.9 µm, n=166; cKO median = 65.9 µm, n=140; p = 0.38) and distribution for each genotype. Maximum and minimum values are denoted by upper and lower bars, respectively. (F) Number of follicles at each stage of development in whole ovaries of immature Fl/Fl and cKO mice following gonadotropin stimulation. Bars represent mean and error bars are SEM for follicle numbers per ovary at each stage (unpaired T-test. *, $p<0.05$, **, $p<0.01$). (G) Oocyte atresia in cKO ovaries following gonadotropin stimulation. Ovarian cross sections from Zfp36l2-cKO females show a portion of oocytes at the primary and early secondary follicle stages appeared atretic after stimulation in cKO oocytes. Arrowheads indicate representative atretic oocytes; the arrow indicates a representative healthy oocyte at a similar stage. (H) Number of GV, MI and MII oocytes as a percent of the total oocytes obtained following in vitro culture of cumulus-oocyte complexes isolated from cKO and Fl/Fl ovaries (unpaired T-test; *, $p<0.05$). Insets, representative images of MII oocytes enlarged to show the abnormally large polar bodies observed in many of the cKO oocytes.

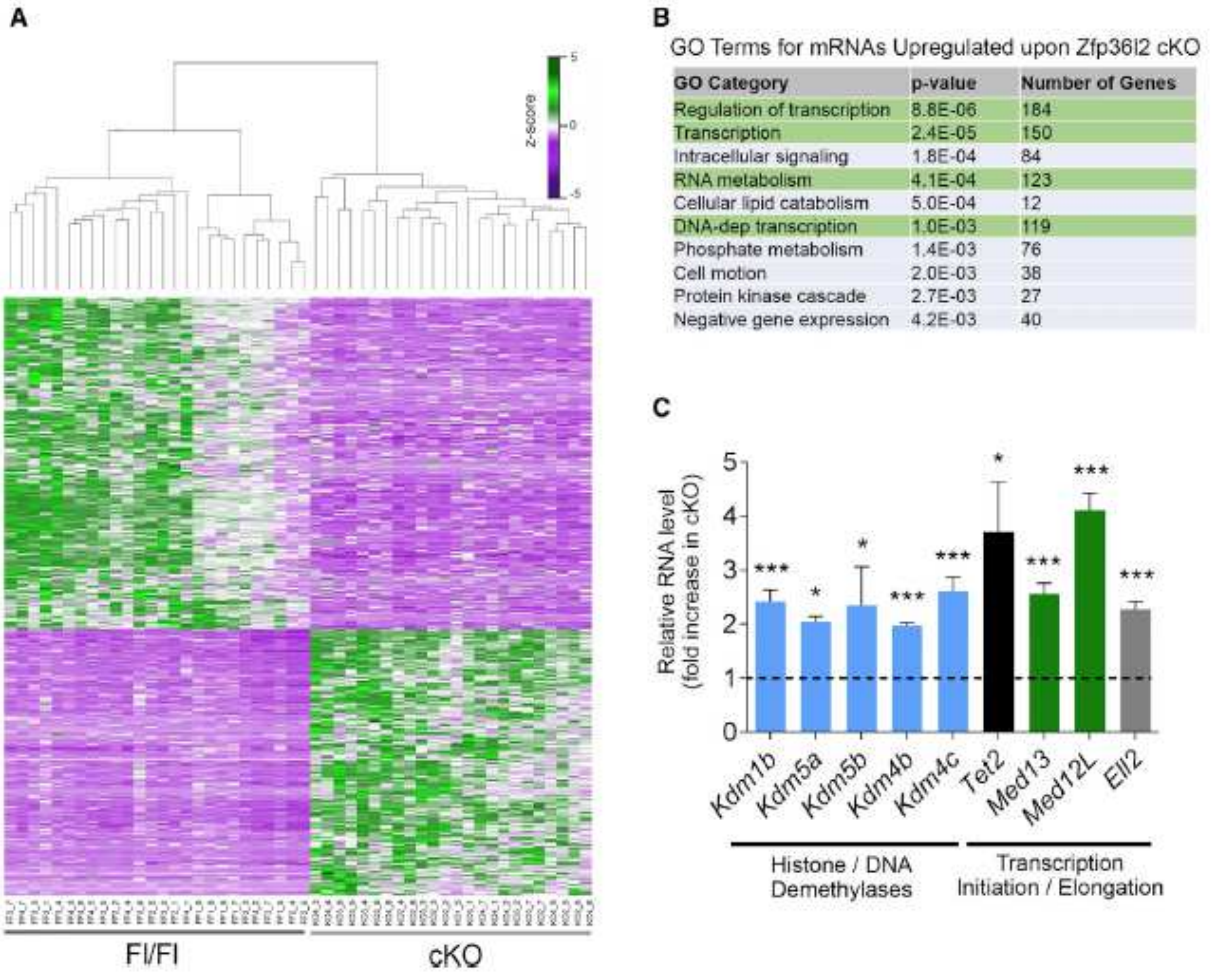


Figure 2.3 ZFP36L2 regulates mRNAs encoding factors with central roles in transcription regulation and chromatin modification. *(A)* Heat map depicting clustering and global changes in gene expression for GV oocytes from Fl/Fl and cKO females as determined by single-cell RNA sequencing. *(B)* GO analysis of transcripts upregulated in cKO oocytes show enrichment for genes involved in transcription regulation (green). Top 10 most highly enriched categories are shown. *(C)* qPCR validation for a subset of changes observed in Fl/Fl and cKO oocytes by RNAseq. RNA levels are shown relative to *L19* and Fl/Fl oocytes were normalized to 1; error bars are SEM. (unpaired T-test; *, p <0.05; ***, p<0.001).

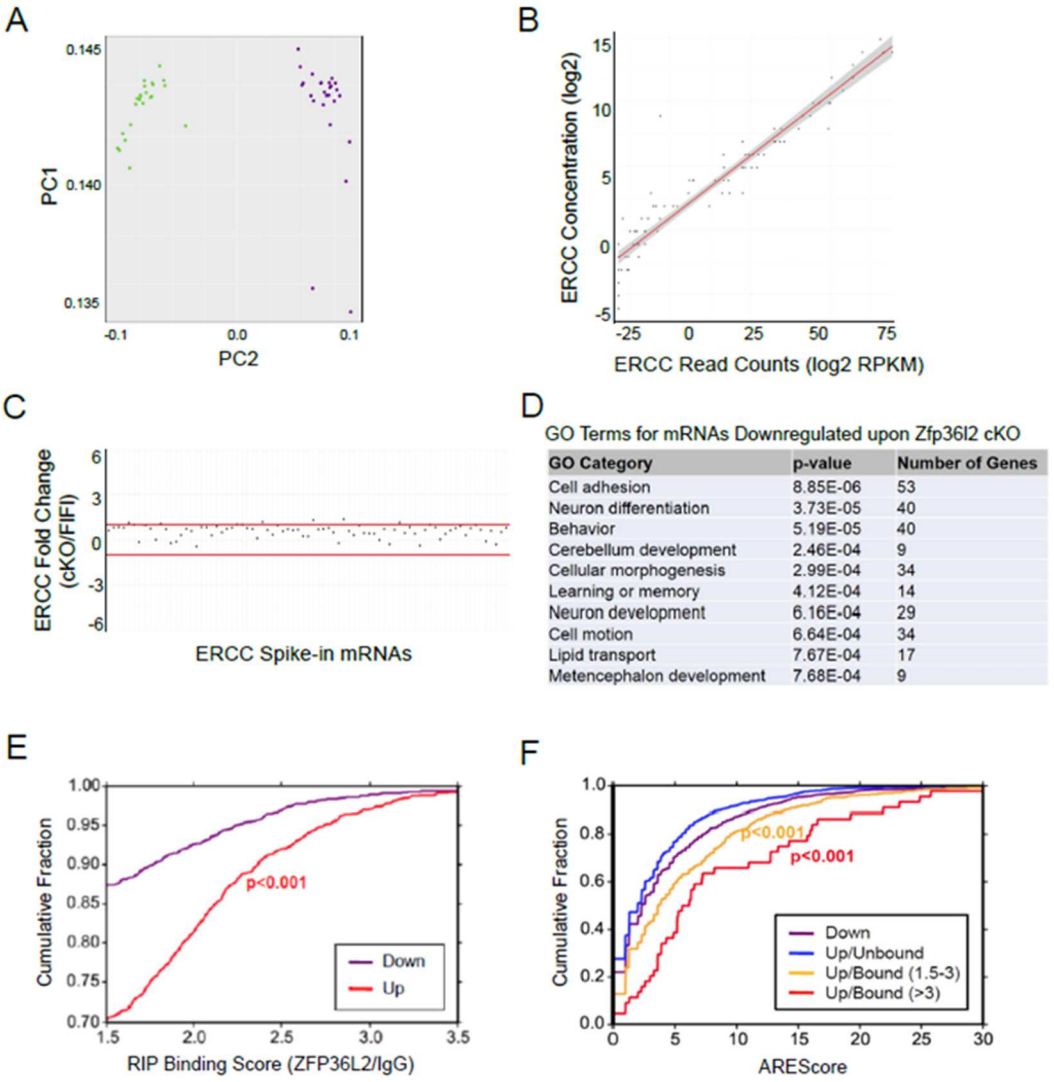


Figure 2.4 Single cell RNAseq quality assessment reveals upregulated transcripts in cKO are enriched for ZFP36L2 binding and AU-rich elements. (A) Principal component analysis (PCA) of cKO and Fl/Fl oocyte transcriptomes. Data was plotted along the first and second principal components. (B) ERCC Spike-In RNAs were added to oocytes before lysis at the start of library preparation as an additional means of normalization. Normalized ERCC reads from both cKO and Fl/Fl were plotted against ERCC mRNA input concentration using a log2 scale. ERCC RNA input concentration is linearly correlated with output normalized read counts. (C) Differential expression for each ERCC Spike-In RNA was calculated using the same methods alongside whole transcriptome data (See Methods). There was no overall shift in ERCC RNA levels between cKO and Fl/Fl oocytes. (D) GO analysis of transcripts downregulated in KO oocytes relative to Fl/Fl oocytes. Top 10 most highly enriched categories are shown. (E) Cumulative plot comparing ZFP36L2 binding in mouse erythroid progenitor cells (Zhang et al., 2013) for transcripts upregulated or downregulated (Q<0.05) in Zfp36l2-cKO oocytes. The plotted binding value refers to fold enrichment in anti-ZFP36L2 immunoprecipitation (IP) over IgG; (Zhang et al., 2013). KS-test, $p < 0.001$. (F) Cumulative plot comparing ARE Scores (Spasic et al., 2012) for transcripts downregulated (Down) upon Zfp36l2-cKO in oocytes, or upregulated and bound with the binding values indicated, or unbound by ZFP36L2 in erythroid cells (Zhang et al., 2013). p-values refer to comparison with downregulated transcripts (KS-test).

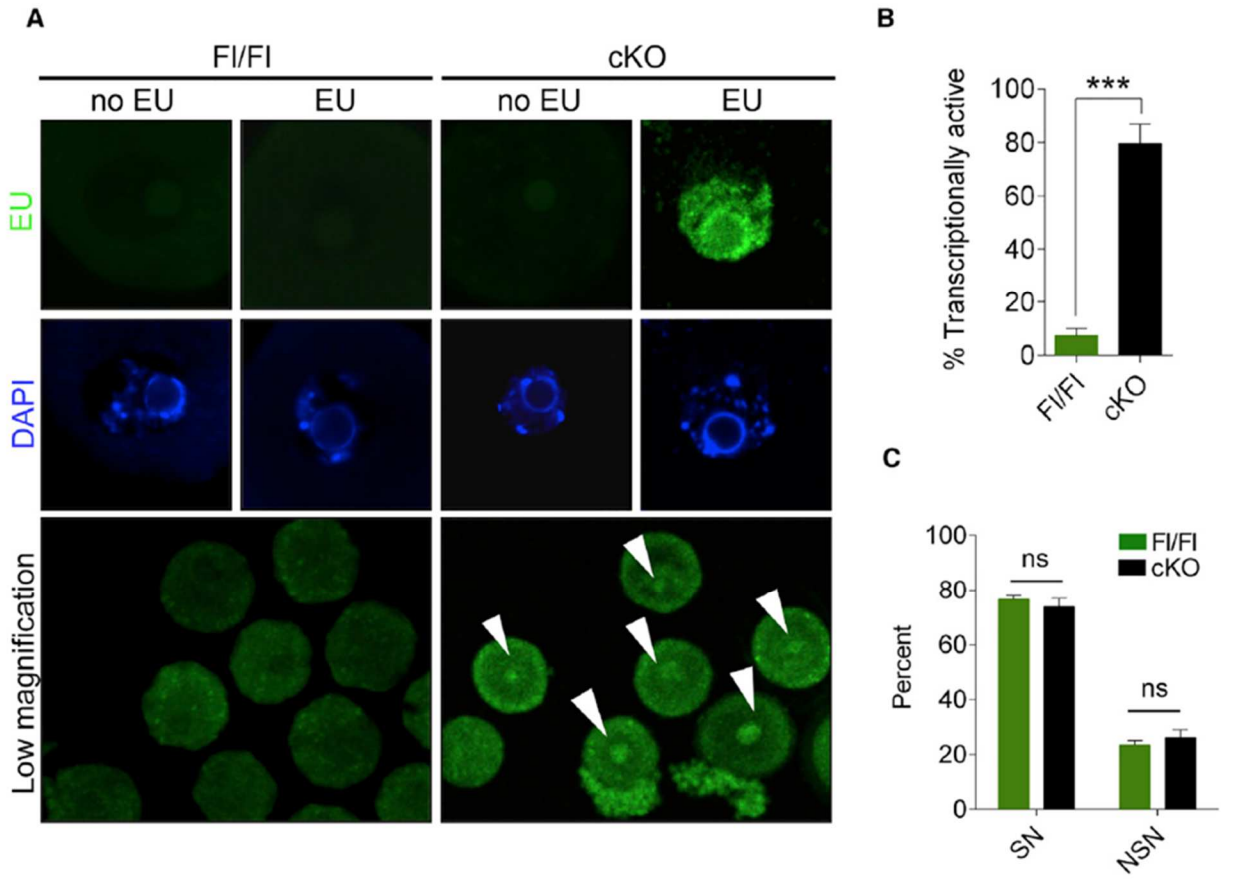


Figure 2.5 ZFP36L2 is required for global transcriptional silencing during the final stages of oocyte growth. *(A)* Run-on transcription assays performed in cKO and Fl/Fl GV oocytes by culture in the presence or absence of 5-ethynyl uridine (EU). Newly transcribed, EU-labeled cellular RNAs are detected as green fluorescence following staining; nuclei are blue (Hoechst). A lower magnification image demonstrates the range of transcriptional activity observed; arrowheads indicate transcriptionally active nuclei to the extent visible in a single confocal plane. *(B)* The percent of transcriptionally active oocytes in run-on assays from Figure 2.5A is quantified. Error bars are SEM (unpaired T-test; ***, $p<0.001$). *(C)* Chromatin organization of oocytes in the run-on assays was determined by Hoechst staining. Error bars are SEM (unpaired T-test; ns, nonsignificant).

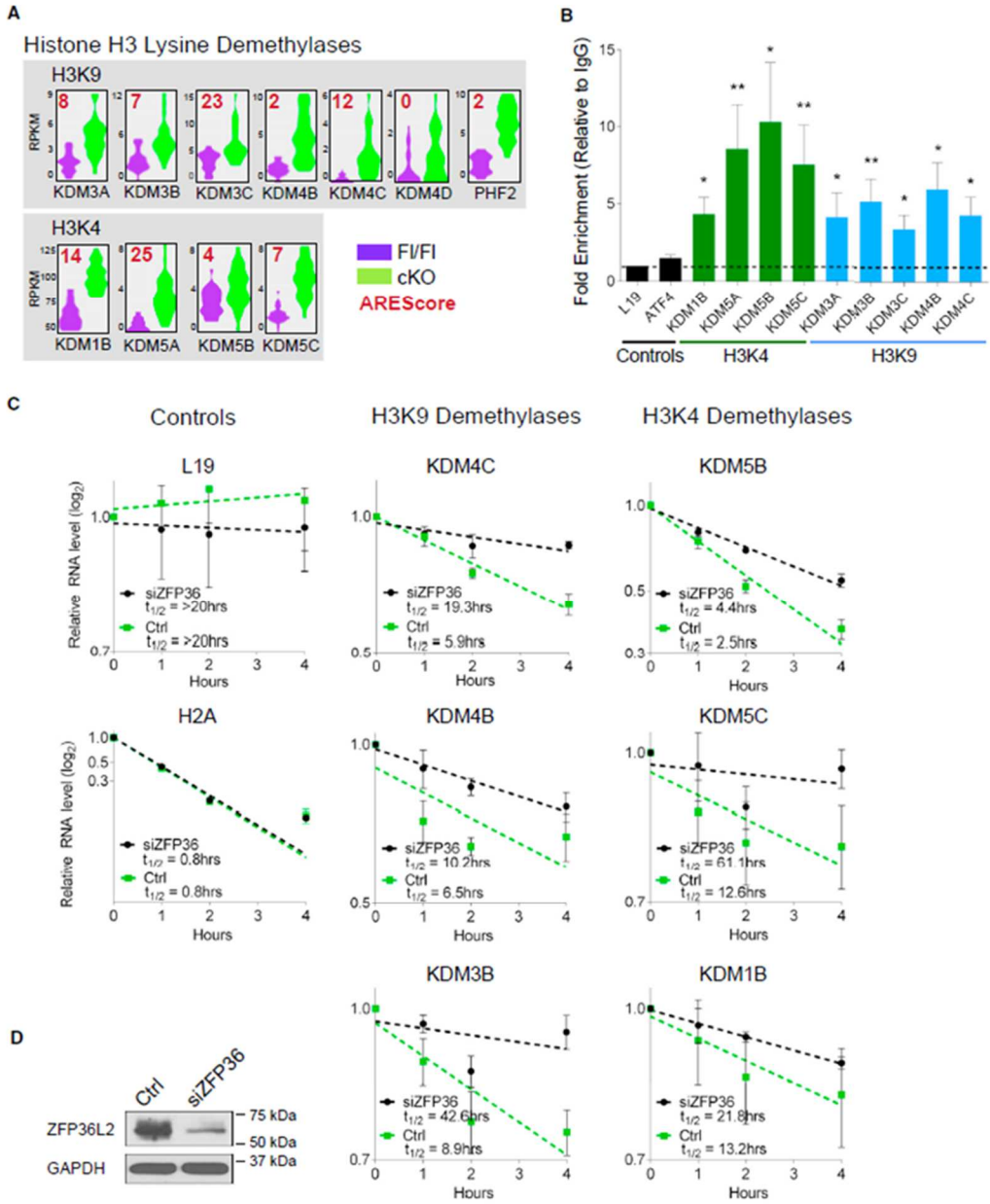
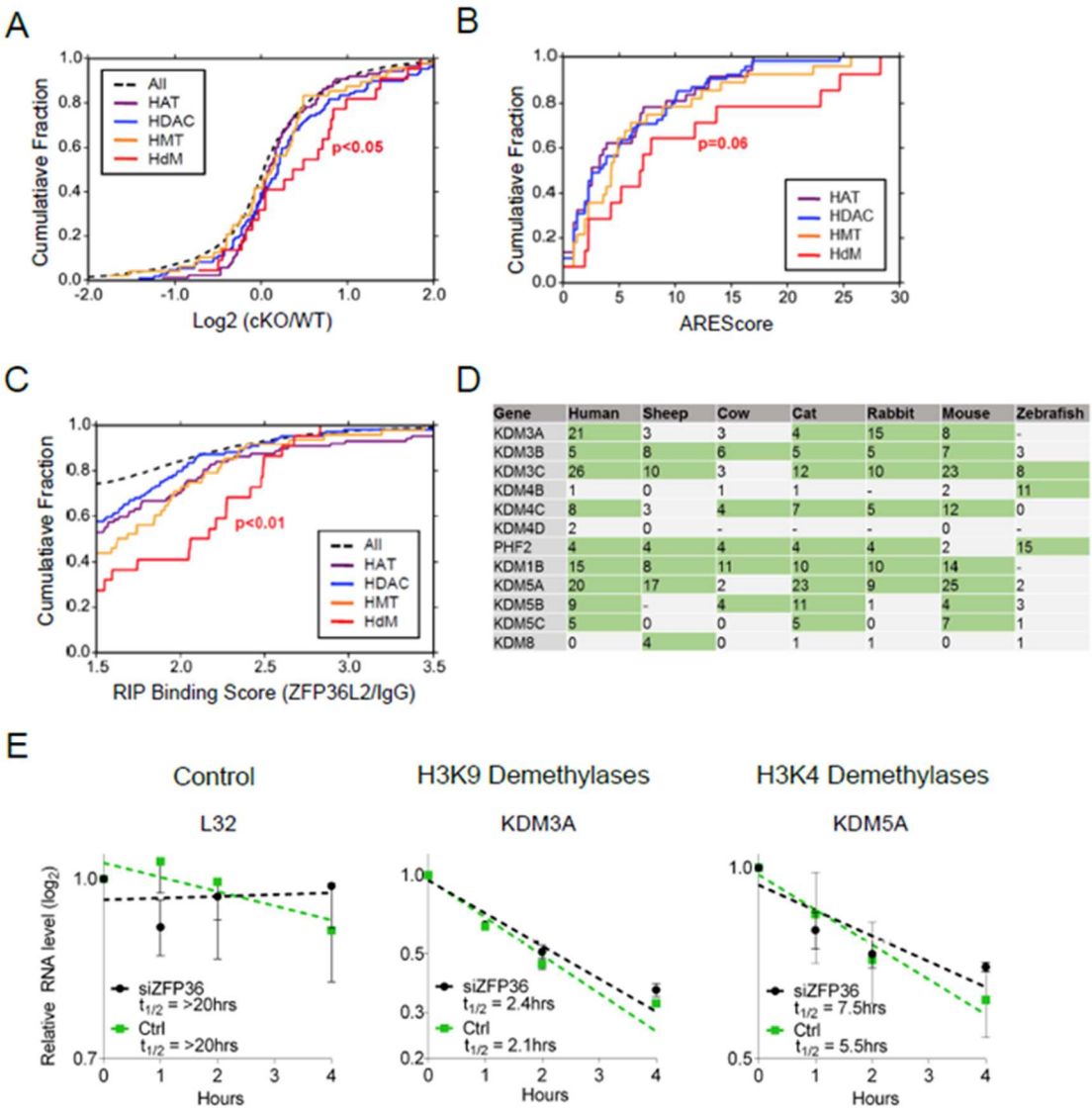


Figure 2.6 mRNAs encoding histone demethylases for H3K4 and H3K9 are direct targets of ZFP36L2 decay. *(A)* Violin plots present the distribution of normalized RNAseq reads for mRNAs encoding histone H3 lysine demethylases for H3K4 and H3K9 upregulated in cKO GV oocytes relative to Fl/Fl. AREScores for each mRNA 3'UTR are in red. *(B)* RNA immunoprecipitation assay demonstrating fold enrichment for ZFP36L2 binding for each histone demethylase mRNA in Hela cells relative to IgG control and normalized to L19. ATF4 is an additional control mRNA subjected to an mRNA decay pathway not dependent on ZFP36L2 (two-way ANOVA; *, $p < 0.05$; **, $p < 0.01$). *(C)* RNA decay assay showing increased half-lives for histone demethylase mRNAs as determined by qPCR in Hela cells following knock-down of ZFP36L2. Cells were transfected with siRNA targeting either ZFP36 proteins or luciferase as a control, then treated with Actinomycin D for indicated times. Half-life was not determined for L19 as the regression line was near horizontal. See also Figure 2.7E. *(D)* Western blot for ZFP36L2 demonstrating knock-down following siRNA transfection in HeLa cells. GAPDH is the internal loading control.



Gene	Human	Sheep	Cow	Cat	Rabbit	Mouse	Zebrafish
KDM3A	21	3	3	4	15	8	-
KDM3B	5	8	6	5	5	7	3
KDM3C	26	10	3	12	10	23	8
KDM4B	1	0	1	1	-	2	11
KDM4C	8	3	4	7	5	12	0
KDM4D	2	0	-	-	-	0	-
PHF2	4	4	4	4	4	2	15
KDM1B	15	8	11	10	10	14	-
KDM5A	20	17	2	23	9	25	2
KDM5B	9	-	4	11	1	4	3
KDM5C	5	0	0	5	0	7	1
KDM8	0	4	0	1	1	0	1

Figure 2.7 Histone demethylase mRNAs are conserved targets of ZFP36L2. (A) Lysine histone demethylase (HdM) mRNAs are enriched for upregulation upon Zfp36l2-cKO in oocytes. Cumulative plot for changes in transcript levels upon Zfp36l2-cKO (log2[cKO/WT]) for all genes (All), compared with genes with AmiGO GO terms histone demethylase (HdM), histone acetyltransferase (HAT), histone deacetylase (HDAC) or histone methyltransferase (HMT). p-value is based on KS-test, comparing with All genes. (B) Histone demethylase mRNAs are enriched for high ARESscores. Cumulative plot of ARESscores for transcripts with GO terms listed above. p-value is based on KS-test, comparing with the Transcription GO term. (C) Histone demethylase mRNAs are enriched for binding to ZFP36L2. Cumulative plot for binding to ZFP36L2 (as defined in the legend to Supplemental Figure 2.4 (Zhang et al., 2013)) for the GO categories described in panel A. p-value is based on KS-test, comparing with All genes. (D) AU-richness of histone demethylase 3'UTR sequences is conserved across mammals (quantified by ARESscore) (Spasic et al., 2012). ARESscores ≥ 4 predictive of decay by an ARE-dependent pathway are highlighted in green. (E) RNA decay assay showing increased half-lives for additional histone demethylase mRNAs as determined by qPCR in Hela cells following knockdown of ZFP36L2 as in Figure 2.6C-D.

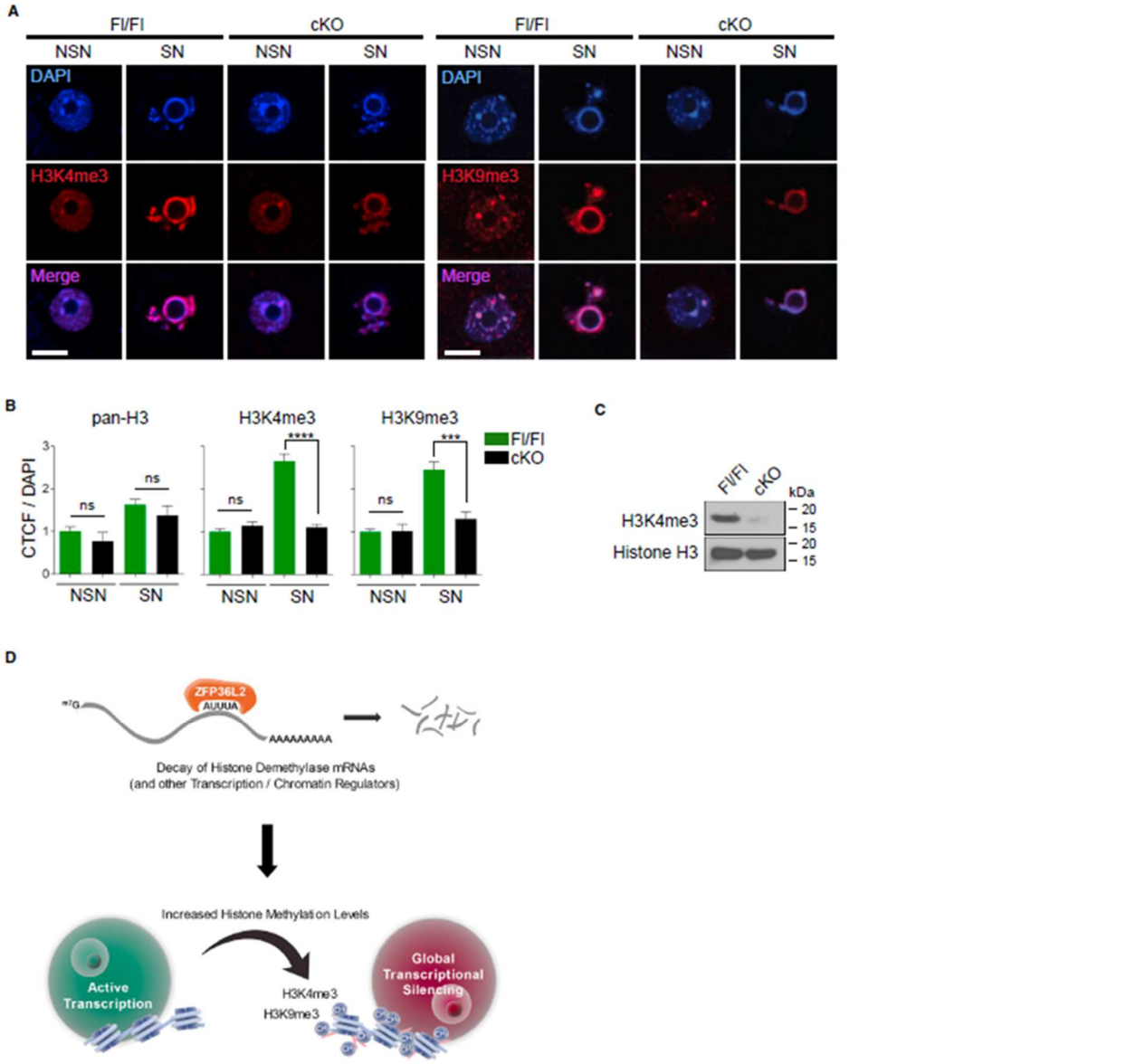


Figure 2.8 ZFP36L2 mediates global increases in H3K4 and H3K9 methylation in the final stages of oocyte growth. *(A)* Immunofluorescence of cKO and Fl/Fl GV oocytes with SN and NSN chromatin configurations to compare H3K4me3 and H3K9me3 levels. See also Figure 2.9A for pan-H3 analyses. *(B)* Quantification of immunofluorescence intensity as seen in Figure 2.8A and S4A. Corrected Total Cell Fluorescence (CTCF) was calculated for each signal then normalized to DAPI CTCF and to the respective Fl/Fl NSN intensity, which was set to 1 (unpaired T-test; ***, p<0.001; **** p<0.0001). *(C)* Western blot comparing H3K4me3 levels in cKO and Fl/Fl oocytes. Total histone H3 levels serve as internal loading control. *(D)* Model for the role of the mRNA decay activator ZFP36L2 in the oocyte. We propose ZFP36L2 acts as a "developmental switch" during oocyte growth by binding and directly degrading a group of mRNAs encoding H3K4 and H3K9 histone demethylases, leading to the accumulation of histone methylation at H3K4 and H3K9, global transcriptional silencing and the acquisition of full developmental competence.

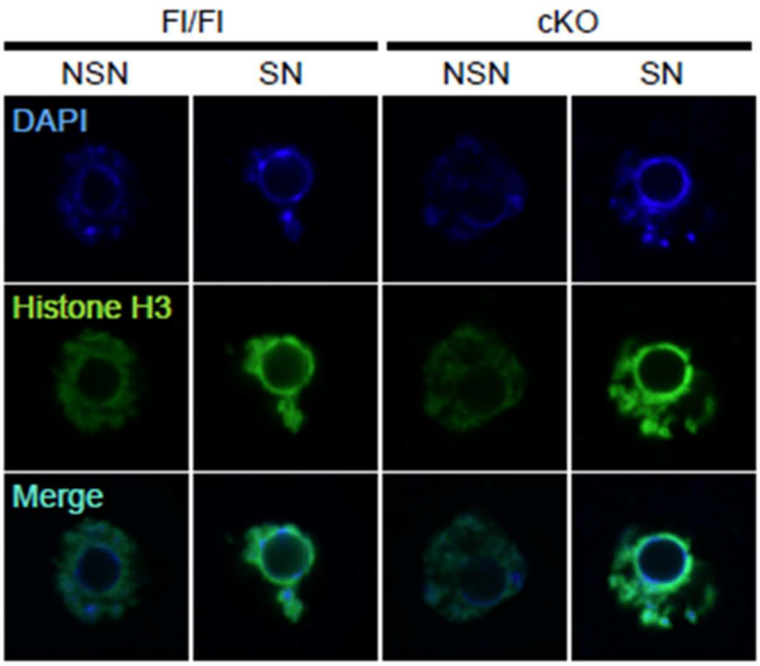


Figure 2.9. ZFP36L2 cKO does not affect total histone H3 levels. Immunofluorescence of GV oocytes from Fl/Fl and cKO mice to assay total histone H3 levels. Images presented show DAPI staining, pan-histone H3 antibody staining, and a merged image of both. cKO oocytes showed no difference in total H3 levels compared to Fl/Fl oocytes (SN nor NSN); quantification is shown in Figure 2.8B.

References

Andreu-Vieyra, C. V, Chen, R., Agno, J.E., Glaser, S., Anastassiadis, K., Stewart, A.F., and Matzuk, M.M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS Biol.* 8, e1000453.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29.

Baker, S.C., Bauer, S.R., Beyer, R.P., Brenton, J.D., Bromley, B., Burrill, J., Causton, H., Conley, M.P., Elespuru, R., Fero, M., Foy, C., Fuscoe, J., Gao, X., Gerhold, D.L., Gilles, P., Goodsaid, F., Guo, X., Hackett, J., Hockett, R.D., et al. (2005). The External RNA Controls Consortium: a progress report. *Nat. Methods* 2, 731–734.

Ball, C.B., Rodriguez, K.F., Stumpo, D.J., Ribeiro-Neto, F., Korach, K.S., Blackshear, P.J., Birnbaumer, L., and Ramos, S.B. V (2014). The RNA-binding protein, ZFP36L2, influences ovulation and oocyte maturation. *PLoS On*e 9, e97324.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395.

Bouniol-Baly, C., Hamraoui, L., Guibert, J., Beaujean, N., Szöllösi, M.S., and Debey, P. (1999). Differential transcriptional activity associated with chromatin configuration in fully grown mouse germinal vesicle oocytes. *Biol. Reprod.* 60, 580–587.

Brooks, S.A., and Blackshear, P.J. (2013). Tristetraprolin (TTP): interactions with mRNA and proteins, and current thoughts on mechanisms of action. *Biochim. Biophys. Acta* 1829, 666–679.

Brower, P.T., Gizang, E., Boreen, S.M., and Schultz, R.M. (1981). Biochemical studies of mammalian oogenesis: synthesis and stability of various classes of RNA during growth of the mouse oocyte in vitro. *Dev. Biol*. 86, 373–383.

Burgess, A., Vigneron, S., Brioudes, E., Labbé, J.-C., Lorca, T., and Castro, A. (2010). Loss of human Greatwall results in G2 arrest and multiple mitotic defects due to deregulation of the cyclin B-Cdc2/PP2A balance. *Proc. Natl. Acad. Sci.* 107, 12564–12569.

Burns, K.H., Viveiros, M.M., Ren, Y., Wang, P., DeMayo, F.J., Frail, D.E., Eppig, J.J., and Matzuk, M.M. (2003). Roles of NPM2 in chromatin and nucleolar organization in oocytes and embryos. *Science* 300, 633–636.

Cardiff, R.D., Miller, C.H., and Munn, R.J. (2014). Manual hematoxylin and eosin staining of mouse tissue sections. *Cold Spring Harb. Protoc.* 2014, 655–658.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., Preissl, S., Jermstad, I., Haugen, M.H., Suganthan, R., Bjørås, M., Hansen, K., Dalen, K.T., Fedorcsak, P., Ren, B., et al. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* 537, 548–552.

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinformatics* 51, 11.14.1-11.14.19.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

External RNA Controls Consortium (2005). Proposed methods for testing and selecting the ERCC external RNA controls. *BMC Genomics* 6, 150.

Flaws, J.A., Abbud, R., Mann, R.J., Nilson, J.H., and Hirshfield, A.N. (1997). Chronically elevated luteinizing hormone depletes primordial follicles in the mouse ovary. *Biol. Reprod.* 57, 1233–1237.

Gene Ontology Consortium (2014). Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43, D1049-56.

Hirshfield, A.N., and Midgley, A.R. (1978). Morphometric analysis of follicular development in the rat. *Biol. Reprod.* 19, 597–605.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.

Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S.-C., Gram, H., and Han, J. (2005). Involvement of MicroRNA in AU-Rich Element-Mediated mRNA Instability. *Cell* 120, 623–634.

Kageyama, S., Liu, H., Kaneko, N., Ooga, M., Nagata, M., and Aoki, F. (2007). Alterations in epigenetic modifications during oocyte growth in mice. *Reproduction* 133, 85–94.

Kharchenko, P. V, Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* 11, 740–742.

Khare, S.P., Habib, F., Sharma, R., Gadewal, N., Gupta, S., and Galande, S. (2012). HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res.* 40, D337-42.

De La Fuente, R. (2006). Chromatin modifications in the germinal vesicle (GV) of mammalian oocytes. *Dev. Biol.* 292, 1–12.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Li, L., Zheng, P., and Dean, J. (2010). Maternal control of early mouse development. *Development* 137, 859–870.

Liu, H., and Aoki, F. (2002). Transcriptional activity associated with meiotic competence in fully grown mouse GV oocytes. *Zygote* 10, 327–332.

Liu, Y.-J., Nakamura, T., and Nakano, T. (2012). Essential role of DPPA3 for chromatin condensation in mouse oocytogenesis. *Biol. Reprod.* 86, 40.

Lowther, K.M., and Mehlmann, L.M. (2015). Embryonic Poly(A)-Binding Protein Is Required During Early Stages of Mouse Oocyte Development for Chromatin Organization, Transcriptional Silencing, and Meiotic Competence. *Biol. Reprod.* 93, 43.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17, 10.

McCloy, R.A., Rogers, S., Caldon, C.E., Lorca, T., Castro, A., and Burgess, A. (2014). Partial inhibition of Cdk1 in G 2 phase overrides the SAC and decouples mitotic events. *Cell Cycle* 13, 1400–1412.

Medvedev, S., Pan, H., and Schultz, R.M. (2011). Absence of MSY2 in mouse oocytes perturbs oocyte growth and maturation, RNA stability, and the transcriptome. *Biol. Reprod*. 85, 575–583.

Mora-Castilla, S., To, C., Vaezeslami, S., Morey, R., Srinivasan, S., Dumdie, J.N., Cook-Andersen, H., Jenkins, J., and Laurent, L.C. (2016). Miniaturization Technologies for Efficient Single-Cell Library Preparation for Next-Generation Sequencing. *J. Lab. Autom.* 21(4), 557-67.

Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* 7, 548.

Nelson, J.F., Felicio, L.S., Randall, P.K., Sims, C., and Finch, C.E. (1982). A longitudinal study of estrous cyclicity in aging C57BL/6J mice: I. Cycle frequency, length and vaginal cytology. *Biol. Reprod.* 27, 327–339.

Paulo, J.A., and Gygi, S.P. (2017). Nicotine-induced protein expression profiling reveals mutually altered proteins across four human cell lines. *Proteomics* 17, 1600319.

Ramos, S.B. V (2012). Characterization of DeltaN-Zfp36l2 mutant associated with arrest of early embryonic development and female infertility. *J. Biol. Chem.* 287, 13116–13127.

Ramos, S.B. V, Stumpo, D.J., Kennington, E.A., Phillips, R.S., Bock, C.B., Ribeiro-Neto, F., and Blackshear, P.J. (2004). The CCCH tandem zinc-finger protein Zfp36l2 is crucial for female fertility and early embryonic development. *Development* 131, 4883–4893.

Ramsköld, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* 5, e1000598.

Sanduja, S., Blanco, F.F., and Dixon, D.A. (2011). The roles of TTP and BRF proteins in regulated mRNA decay. Wiley Interdiscip. Rev. *RNA* 2, 42–57.

Spasic, M., Friedel, C.C., Schott, J., Kreth, J., Leppek, K., Hofmann, S., Ozgur, S., and Stoecklin, G. (2012). Genome-wide assessment of AU-rich elements by the AREScore algorithm. *PLoS Genet.* 8, e1002433.

Stumpo, D.J., Broxmeyer, H.E., Ward, T., Cooper, S., Hangoc, G., Chung, Y.J., Shelley, W.C., Richfield, E.K., Ray, M.K., Yoder, M.C., Aplan, P.D., and Blackshear, P.J. (2009). Targeted disruption of Zfp36l2, encoding a CCCH tandem zinc finger RNA-binding protein, results in defective hematopoiesis. *Blood* 114, 2401–2410.

Svoboda, P., Franke, V., and Schultz, R.M. (2015). Sculpting the Transcriptome During the Oocyte-to-Embryo Transition in Mouse. *Curr. Top. Dev. Biol*. 113, 305–349.

de Vries, W.N., Binns, L.T., Fancher, K.S., Dean, J., Moore, R., Kemler, R., and Knowles, B.B. (2000). Expression of Cre recombinase in mouse oocytes: a means to study maternal effect genes. *Genesis* 26, 110–112.

Xia, M., He, H., Wang, Y., Liu, M., Zhou, T., Lin, M., Zhou, Z., Huo, R., Zhou, Q., and Sha, J. (2012). PCBP1 is required for maintenance of the transcriptionally silent state in fully grown mouse oocytes. *Cell Cycle* 11, 2833–2842.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., Liu, W., Kou, X., Zhao, Y., He, W., Li, C., Chen, B., Li, Y., Wang, Q., Ma, J., et al. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* 537, 553–557.

Zhang, L., Prak, L., Rayon-Estrada, V., Thiru, P., Flygare, J., Lim, B., and Lodish, H.F. (2013). ZFP36L2 is required for self-renewal of early burst-forming unit erythroid progenitors. *Nature* 499, 92–96.

Zuccotti, M., Ponce, R.H., Boiani, M., Guizzardi, S., Govoni, P., Scandroglio, R., Garagna, S., and Redi, C.A. (2002). The analysis of chromatin organisation allows selection of mouse antral oocytes competent for development to blastocyst. *Zygote* 10, 73–78.

Zuccotti, M., Bellone, M., Longo, F., Redi, C.A., and Garagna, S. (2011). Fully-mature antral mouse oocytes are transcriptionally silent but their heterochromatin maintains a transcriptional permissive histone acetylation profile. *J. Assist. Reprod. Genet.* 28, 1193–1196.

Chapter 3: Pluripotent gene expression in the inner cell mass of the mouse blastocyst and in mouse embryonic stem cells is regulated by Upf2

Abstract

The pluripotent population of cells in the blastocyst, the inner cell mass, is established in the mouse embryo approximately three days after fertilization. The inner cell mass will go on to develop into the epiblast, and undergo gastrulation to form the entire organism. It is also the cell type that can be perpetuated *in vitro* as embryonic stem cells. Here, we show that UPF2, an mRNA decay activator, is necessary for the maintenance of gene expression patterns in the mouse inner cell mass associated with pluripotency and that loss of *Upf2* leads to a decrease in the number of pluripotent cells in the blastocyst as it approaches implantation. During *in vitro* growth, the inner cell mass of Upf2-null embryos regresses, despite normal trophectoderm outgrowth, indicating a lineage-specific necessity of this RNA decay factor in embryonic development. Embryonic stem cells can be derived from *Upf2*-null blastocysts under naïve pluripotent state derivation conditions, but exhibit slowed growth and a reduced ability to maintain pluripotency in culture. RNA sequencing of embryos lacking *Upf2* reveals dramatic alterations in the gene expression of factors enriched in the inner cell mass and epiblast, including multiple markers for pluripotency, while extra-embryonic trophectoderm gene expression remains relatively normal. Finally, analysis of upstream regulators of pluripotency-associated genes that were downregulated with loss of *Upf2* revealed potential mechanisms by which *Upf2*, through RNA decay, might regulate the expression of genes necessary for maintaining pluripotency in the early embryo and in embryonic stem cells.

Introduction

Preimplantation embryo development encompasses ~4 days in the mouse and ~7 days in humans, and consists of successive rounds of cell division and differentiation, culminating in blastocyst implantation in the uterine cavity (Bedzhov et al., 2014; Eckert et al., 2015; Lorthongpanich and Issaragrisil, 2015). The first major differentiation event in the embryo—

separation of the extra-embryonic trophectoderm (TE) from the inner cell mass (ICM)—leads to the formation of a blastocyst (as early as embryonic day (E) 3.25 in mice). The ICM further differentiates into another extra-embryonic cell type, the primitive endoderm (PrE), which mostly contributes to the yolk sac, and the epiblast (EPI), which gives rise to the three germ layers that form the basis of development of all systems within the developing fetus, the endoderm, mesoderm and ectoderm.

The pluripotent ICM is also the source of embryonic stem cells (ESCs), which have been used extensively *in vitro* to identify core principles that govern our understanding of early development, as well as mechanisms underlying how cells differentiate to form all of the mature cell and tissue types of a living organism (Nichols and Smith, 2009; Weinberger et al., 2015). A large body of research in recent years has been devoted to identifying the "naïve state" or "ground state" of pluripotency, the *in vitro* cell state that most closely corresponds to the *in vivo* embryonic ICM (Nichols and Smith, 2009; Weinberger et al., 2015). However, the molecular network regulating these early stages of embryogenesis, including how the pluripotent ICM is established, what gene expression patterns define the ICM, and what is necessary for ICM proliferation and differentiation within the preimplantation blastocyst, remain poorly understood.

To date, most studies examining preimplantation development and the molecular basis of pluripotency have focused on the role of transcriptional regulation (Chazaud and Yamanaka, 2016), failing to account for the potentially important role of post-transcriptional regulation in developmental events. Given that the steady-state level of an RNA is equally dependent on its synthesis and decay, RNA turnover is an important but understudied mechanism by which gene expression is regulated during development (Alonso, 2011; Dumdie et al., 2018; Li et al., 2015; Lou et al., 2016; Medghalchi et al., 2001; Shum et al., 2016; Thoren et al., 2010; Weischenfeldt et al., 2008).

One of the most well-studied and highly conserved post-transcriptional RNA decay pathways is Nonsense-mediated RNA Decay (NMD). NMD was originally discovered by virtue of its ability to degrade aberrant RNAs harboring premature termination codons and protect cells from truncated dominant-negative proteins (Chan et al., 2009; Chang et al., 2007; Popp and Maquat, 2013; Schweingruber et al., 2013). Subsequently, it was discovered that NMD degrades subsets of normal RNAs, with loss or disruption of NMD leading to dysregulation of 5-20% of the normal transcriptome in species spanning the phylogenetic scale (Chan et al., 2007; Gatfield et al., 2003; He et al., 2003; Lelivelt and Culbertson, 1999; Mendell et al., 2004). The finding that NMD's reach extends beyond "quality control" strongly argues that this conserved and highly selective RNA decay pathway has roles in regulating normal cellular events. Indeed, there is extensive evidence that NMD is critical for development in a number of systems (Hwang and Maquat, 2011; Tarpey et al., 2007; Vicente-Crespo and Palacios, 2010). In particular, a large body of research has demonstrated that loss of individual factors that either regulate or are essential for NMD (UPF1, UPF2, UPF3A, SMG1, and SMG6) lead to early embryonic lethality in mice, as well as other organisms (Alonso and Akam, 2003; Anastasaki et al., 2011; Casadio et al., 2014; Hwang and Maquat, 2011; Li et al., 2015; Longman et al., 2007; Mao et al., 2015; McIlwain et al., 2010; Medghalchi et al., 2001; Metzstein and Krasnow, 2006; Shum et al., 2016; Silver et al., 2010; Weischenfeldt et al., 2008). Further, phenotypic defects associated with dysregulation of pluripotency have been reported in embryos and ESC systems devoid of NMD activity (Li et al., 2015; Lou et al., 2016; Medghalchi et al., 2001), hinting at a role for selective regulation of mRNA stability by the NMD pathway in regulating pluripotency during early embryo development.

Here, we characterize the embryonic lethality phenotype of embryos devoid of UPF2, a key component of the NMD pathway. We show that *Upf2*-null preimplantation blastocysts are delayed in development and exhibit lethality mostly by E5.5. We demonstrate a lineage-specific

necessity of UPF2 within pluripotent cells of the ICM, without detectable defects in extra-embryonic cell types. Compared to stage-matched controls, *Upf2*-null blastocysts exhibit regression of the ICM across implantation stages, contain fewer NANOG-positive staining cells, and demonstrate a global reduction in gene expression of pluripotency-associated mRNAs. *Upf2*-null ESC lines can be derived under naïve or "ground" state pluripotency conditions, but exhibit growth defects and, when transitioned to conventional ESC culture conditions, exhibit a propensity to differentiate *in vitro.* Thus, while *Upf2*-null ESC lines can be derived in the ground pluripotency state, *Upf2* is necessary for the transition to a more primed pluripotent state *in vitro*.

Methods

Generation of global heterozygous mice

Global knockout (heterozygous) mice were generated by crossing mice expressing Cre recombinase driven by the Ella promoter in the early preimplantation embryo to a previously published *Upf2*-floxed mouse in the C57BL/6 background (Weischenfeldt et al., 2008). The *Upf2*-floxed construct was designed to delete exons 2 and 3, and while a truncated protein is still generated, it has been shown to lack NMD activity (Weischenfeldt et al., 2008). Global heterozygous males and females were bred to generate mice for experiments, and all experiments herein were conducted on embryos obtained from global *Upf2*-heterozygous parental crosses. For each experiment detailed below, the total number and age of mice used are clearly stated. Mice were genotyped by polymerase chain reaction (primers used for genotyping are provided in supplemental data). The mice colonies were maintained in agreement with protocols approved by the Institutional Animal Care and Use Committee at the University of California, San Diego.

Genotyping and DNA PCR

Ear tags were taken from mice and DNA was extracted for PCR. Lysed ear tags were added to 0.2ug/mL Proteinase K in DirectPCR lysis buffer (Viagen, Cat #103-T), and then heated at 55C for 3-4 hours and then 85C for 30 minutes to denature lysozymes in the buffer. PCR reactions were conducted in 15 uL volume reactions with 1uL lysed mouse DNA, 10.8 uL nuclease free water, 0.6 uL dNTP mixture (BioPioneer, Cat# MDM-4), 0.6 uL forward and reverse primer mixture (10uM), 1.5 uL 10x buffer (Denville, cat# CB3702-7), and 0.5 uL Taq polymerase (Denville, Cat# CB4050-2). 1.2% agarose gel with ethidium bromide was used for visualization of gene bands.

Genotyping of single embryos was performed using a custom low-input protocol. Early postimplantation embryos were carefully dissected out of the maternal decidua and washed in 2.5% pancreatin / 0.5% trypsin in Tyrode Ringers Solution to remove external membranes and any resulting maternal tissue contamination. DNA was extracted for PCR by incubation in 10uL Blastocyst Lysis Buffer (100mM Tris-HCl, pH 8.3, 100mM KCl, 0.02% gelatin, 0.45% Tween 20, 20 mg/ml Proteinase K) for 30 minutes at 55C followed by incubation at 95C for 10 minutes to denature active lysozymes. Genotyping PCR reactions were performed using a high sensitivity taq (PrimeSTAR HS DNA Polymerase, Takara #R010A). Reactions were conducted in 20uL volume reactions with 1-3uL embryo DNA (1uL for postimplantation embryos, 3uL for preimplantation embryos), 1.6uL dNTP mixture, 2uL primer mixture (10uM), and 0.2uL PrimeSTAR Taq, with nuclease-free water to total volume. Preimplantation embryos required 38 cycles for accurate visualization of bands on a 1.2% agarose gel with ethidium bromide.

Isolation and primary culture of zygotes and isolation of blastocysts

For experiments involving the isolation and primary culture of zygotes, female mice of age 4-8 weeks were stimulated with intraperitoneal injection of 5 IU equine Chorionic Gonadotropin (eCG; Lee BioSolutions, 493-10) followed by intraperitoneal injection of 5 IU human Chorionic Gonadotropin (hCG; Sigma, C1063) 46-48 hours later. Immediately after hCG

injection, female mice were placed in the same cage as a stud male of known fertility for mating. The contents of the ampulla were collected only from females with a visible plug 18 hrs after hCG injection. Ampulla contents were evaluated for the presence of zygotes by microscopy following a brief digestion in 0.1mg/mL hyaluronidase (Sigma, H4272) dissolved in M2 media (Sigma, M7176). Zygotes were washed 4 times in M2 before being transferred to M16 (Sigma, M7292) for culture. Embryos were cultured in 20uL drops of M16, overlaid with mineral oil, 20 embryos per drop. For blastocyst experiments, zygotes were cultured at 37C, and were isolated 90-94 hours post-hCG, 100-104 hours post-hCG, and 110-114 hours post-hCG, for early-, mid-, and late- blastocyst stages, respectively. For some experiments, blastocysts were isolated from the uterine tubes of females that had been superovulated with eCG and hCG, as detailed above, and mated. E3.5 blastocysts were isolated approximately 96 hours post-hCG injection, from females with a visible plug 18 hours after hCG injection. The uterine tubes were dissected apart and M2 media was flushed through, using an insulin syringe, to expel the embryos.

Isolation of early postimplantation embryos

To obtain early postimplantation embryo counts, timed matings were set up for male and female *Upf2*-heterozygous mice. Female mice were 8-12 weeks of age, and male mice were stud males of known fertility. Females were examined for the presence of vaginal plugs following mating. Only females with observable plugs were used for postimplantation embryo dissections. On embryonic days 5.5 and 6.5, female mice were injected with 0.1ml of a 1% Chicago blue dye solution through the tail vein using a 1-ml syringe fitted with a 27-gauge needle. Mice were sacrificed approximately three minutes after the dye injection and the uterine horns were carefully removed. Implantation sites could then be visualized by the blue dye. Each embryo was separated by cutting between implantation sites along the uterine horn. The muscular uterine myometrium was peeled back to expose the decidua, and the decidua was carefully dissected back to reveal each embryo. Reichart's membrane, if still attached, was

removed by washing in 2.5% pancreatin / 0.5% trypsin in Tyrode Ringers Solution, as well as careful dissection, if necessary.

Outgrowth assay

Blastocysts were obtained from 4-8 week old female *Upf2*-heterozygous mice, as per above. To assay blastocyst attachment and outgrowth, blastocysts were individually plated in a single well of a 24-well gelatin-coated (0.1% gelatin) plate in α-MEM containing 1% Fetal Bovine Serum (FBS) and 1% Penicillin/Streptomycin. After 72 and 96 hours of incubation at 37C, images were taken for further analysis. After 96 hours, blastocyst outgrowths were washed twice with PBS and individually picked for genotyping.

Immunofluorescence

Blastocysts were isolated from *Upf2*-heterozygous females aged 4-8 weeks as per above, following superovulation. The zona pellucida was removed in Acid Tyrode's solution, followed by three washes in PBS with 0.1% BSA. Embryos were fixed with 4% PFA in PBS for 1 hour at room temperature followed by three washes in PBS with 0.1% BSA and 0.1% Tween (PBST). The embryos were then permeabilized by incubation with 1% Triton X-100 in PBS for 1 hour at room temperature. After permeabilization, the emrbyos were again washed three times in PBST and blocked with 4% normal donkey serum in PBST for 2 hours at room temperature. The embryos were incubated with primary antibodies overnight at 4^0C. All washes and incubations were performed with gentle agitation. Anti-CDX2 antibody (Thermo Scientific, MA5-14494) was used at a dilution of 1:500, anti-NANOG antibody (Invitrogen, 14-5761-80) was used at a dilution of 1:500, and anti-GATA6 (R&D Systems, AF1700) was used at a dilution of 1:500). After removing the primary antibody, the embryos were washed three times in PBST and then incubated with Alexa Fluor 647 conjugated donkey anti-rabbit IgG (H+L) secondary antibody (ThermoFisher Scientific, Cat. No. A-31573), Alexa Fluor 488 conjugated donkey anti-rat IgG (H+L) Highly Cross-Adsorbed secondary antibody (ThermoFisher Scientific, Cat. No. A-

21208), and Alexa Fluor 546 conjugated donkey anti-goat IgG (H+L) Highly Cross-Adsorbed secondary antibody (ThermoFisher Scientific, Cat. No. A-11056) for 2 hours at room temperature. The embryos were then washed three times with PBST and incubated with DAPI (Novus Biologicals, NBP2-31156) at a dilution of 1:1000 for 30 minutes. Embryos were washed with PBST three more times before being placed in individual 1ul drops in a Nunc glass bottom dish (Thermo Fisher Scientific, 150682) and overlaid with mineral oil for imaging.

Imaging methods

Embryos were imaged using a Nikon A1R confocal with a four-line LUN-V laser engine and DU4 detector, mounted on a Nikon Ti2 using a S Fluor 40x 0.9 NA objective. Images were acquired in resonant mode with bidirectional scanning and 4x line averaging, and 0.575 µm steps were used to collect Z-stacks of the entire embryo. The lasers used were 405nm (7% laser power), 488nm (5% laser power), 561nm (3% laser power), and 640nm (3% laser power). To avoid cross-talk between channels, Z-stacks were acquired of the DAPI and AlexaFluor 568 channels first, and the AlexaFluor488 and AlexaFluor 647 channels were acquired subsequently.

Automated nuclei/cell counting

To count nuclei positive for the three markers of interest, we created an automated image processing and counting routine using the General Analysis 3 tool within NIS Elements 5.11. For pre-processing, we applied local contrast, smoothing, and a rolling ball filter before a threshold was applied to generate a binary layer. The binary was then subsequently eroded and dilated, cleaned, smoothed, and touching binaries were separated and filtered for size. The resulting binaries were then counted and the records were pooled. For each image stack, the routine was validated and adjusted manually. Blastocyst total cell number was used to bin embryos into three groups (early = 20-35 cells; mid = 36-50 cells, and late = 51-65 cells), and

the percent of NANOG-, GATA6-, and CDX2-positive cells was calculated as a fraction of the total cell number.

cDNA synthesis and library preparation of individual blastocysts

In total, 19 E3.25 stage blastocysts (8 *Upf2*$^{-/-}$, 11 *Upf2*$^{+/+}$) and 15 E3.5 stage blastocysts (8 *Upf2*$^{-/-}$, 7 *Upf2*$^{+/+}$) were subjected to mRNA sequencing. As a means of normalization, ERCC Spike-In RNAs (Life Technologies, Cat.no. 4456740) were added to the samples. The ERCC stock provided by the manufacturer was diluted 1:25,000 and 0.25 µl of the dilution was added to each embryo compartment samle before cell lysis and RNA-seq library preparation, and ERCC reads were analyzed downstream in parallel with reads from the embryo transcriptome. RNA isolation and cDNA synthesis were carried out using the SMARTer® Ultra® Low Input RNA Kit for Sequencing – v4 (Clontech, Cat.no. 634888).

Genotyping of cDNA was performed using custom primers. In brief, since loxP sites surround exons 2 and 3 of the *Upf2* gene, the wild-type locus was detected using forward and reverse primers in exons 1 and 2, respectively; the knock-out locus was detected using forward and reverse primers in exons 1 and 4, respectively. Libraries were generated from the resulting cDNA (0.2ng/ul per sample) of wild-type and *Upf2*-null embryos using the Nextera XT DNA library preparation kit (Illumina, Cat.no. FC-131-1024) as previously described (Mora-Castilla et al., 2016). Indexed sequence libraries were pooled for multiplexing, normalized by MiSeq read number, and paired-end sequencing was performed on a HiSeq 4000.

Analysis of mRNA sequencing data

Reads were mapped via STAR (2.5.2a) (Dobin et al., 2013; Dobin and Gingeras, 2015) after trimming for adapter sequences. Samtools (Li et al., 2009) was used to process sam files, as well as to sort and remove PCR duplicates of bam files. Counts for each gene were quantified using the Subread package FeatureCounts using the gene level quantification in

paired-end mode (release 1.5.2) (Liao et al., 2014), and annotated using the Ensembl GRCm38 genome. Reads were filtered such that genes without at least one sample with at least 10 raw reads were removed from the analysis. Count data was normalized to TPM using the Bioconductor package edgeR (McCarthy et al., 2012; Robinson et al., 2010). Differential expression was calculated using DESeq2 (Love et al., 2014) based on a model using the negative binomial distribution. Genes with an adjusted P value less than 0.05 were considered differentially expressed unless otherwise noted. Heatmaps were constructed using the heatmap.2 function in ggplot2 (Wickham, 2016). Principle components were calculated with the prcomp function. All box plots, scatter plots, PCA plots, and heatmaps were constructed using ggplot2. All of these analyses and plot constructions were performed with RStudio, R (v3.2.2 and v3.4.0). Gene ontology analysis was performed with the online Metascape platform (Tripathi 2015 Cell host and microbe) and gene enrichment networks were visualized with Cytoscape (Shannon et al., 2003).

Inference of mRNA stability from steady-state RNA sequencing data

To infer mRNA stability from steady-state sequencing data, we employed the use of REMBRANDTS (REMoving Bias from Rna-seq ANalysis of Differential Transcript Stability), which internally uses DESeq to obtain estimates of pre-mRNA and mature mRNA abundance and estimates a gene-specific bias function. The stringency parameter used for these analyses was 0.99, which is the most stringent value possible for this parameter.

Results

Upf2 is required for mouse peri-implantation embryo viability

To test the importance of UPF2 in the early embryo, we generated a global *Upf2* heterozygous knock-out mouse line by crossing established *Upf2*-floxed (*Upf2-Fl/Fl*) mice

(Weischenfeldt et al., 2008) with mice expressing Cre recombinase driven by the *Ella* promoter, which is expressed early in preimplantation embryo development (Figure 3.2A). Knock-out was confirmed by PCR (Figure 3.2B-C), and the resulting mRNA and protein generated with loss of exons 2 and 3 have been shown to compromise NMD activity (Weischenfeldt et al., 2008). No *Upf2*-null (*Upf2*$^{-/-}$) pups were born from global heterozygous (*Upf2*$^{+/-}$ X *Upf2*$^{+/-}$) parental crossings, and *Upf2*$^{+/-}$ pups were present at a 1.5:1 ratio to wild-type pups (*Upf2*$^{+/+}$), suggesting the possibility of lethality of some heterozygous progeny (Figure 3.2B). The mean litter size obtained from global heterozygous matings was 4.6 pups.

We next determined the timing of embryonic lethality of *Upf2*-null embryos. A previous report found that loss of UPF2 leads to embryonic lethality in mice at E9.5 or earlier, but did not determine the precise timing (Weischenfeldt et al., 2008). We isolated early post-implantation embryos from *Upf2*-heterozygous crosses and found that *Upf2*-null embryos die at peri-implantation stages, with no knock-out embryos detectable at E6.5, and only two knock-out embryos present at E5.5, from a total of five breeding pairs (Figure 3.1A, 3.2F). Isolations performed at E5.5 and E6.5 produced an average of 8.2 and 8.0 implantation sites, respectively, and 1.2 and 1.0 resorption sites, respectively (Figure 3.2D-E), suggesting that some embryos may already be defective at the time of implantation and are unable to implant properly.

We next looked at earlier stages of to determine at what stage preimplantation development of *Upf2*-null embryos become defective. Preimplantation E3.5 blastocysts were isolated from the uterine horns of pregnant females subjected to heterozygous matings after superovulation. *In vivo*, *Upf2*-null blastocysts were able to mature and expand *in vivo,* but they exhibited a slight delay in development, with more *Upf2*-null blastocysts present as an early-stage blastocyst, as compared to heterozygous and wild-type littermate embryos (Figure 3.1B). We also examined blastocysts grown *in vitro* – zygotes isolated from heterozygous parental crossings after superovulation grew to expanded blastocysts. However, as observed *in vivo*, a

higher proportion of *Upf2*-null embryos stalled at the early blastocyst stage compared to control embryos (Figure 3.2G). Thus, *Upf2*-null embryos primarily die between E3.5 and E5.5, during the window of implantation, and, interestingly, some *Upf2*-null blastocysts appear to be delayed at the mid-blastocyst (~E3.5) stage.

Loss of Upf2 leads to inner cell mass regression

The finding that *Upf2*-null blastocysts tended to have a delay in development at E3.5 led us to next test their ability to initiate and progress through the next stage of development--implantation. To address this, we performed attachment and outgrowth assays, which provide a measure of the ability of the blastocyst to both attach and initiate invasion into the endometrium. Embryos were genotyped following outgrowth, with cells harvested from the TE or ICM (or both) of the blastocyst outgrowth. *Upf2*-null blastocysts, isolated from heterozygous parental crossings following superovulation, were able to attach to gelatin-coated plates at the same rate as littermate controls (data not shown). In addition, the TE differentiated into trophoblast giant cells, growing out at an area also similar to littermate control embryos at both 72 (Figure 3.2H) and 96 hours (Figure 3.1C, E) post-plating. The percent increase of trophectoderm outgrowth area from 72 to 96 hours was comparable between embryo genotypes as well, with $Upf2^{+/+}$, $Upf2^{+/-}$, and $Upf2^{-/-}$ blastocyst outgrowths increasing at a rate of 115, 110, and 95 percent, respectively. Thus, loss of *Upf2* does not seem to impair function of the TE with this assay.

However, in the same outgrowth assay, blastocysts lacking *Upf2* exhibited rapid inner cell mass regression. ICM outgrowths were graded at 72 and 96 hours based on the following numerical scale: (1) many cells forming a densely packed epithelium, (2) many cells forming a loosely packed epithelium, (3) few cells forming a loosely packed epithelium, or (4) no identifiable ICM. The average grades for $Upf2^{+/+}$, $Upf2^{+/-}$, and $Upf2^{-/-}$ ICMs were 2.1, 2.3, and 4.0 at 72 hours ($P < 0.01$; Figure 3.2I) and 2.0, 2.3, and 4.0 at 96 hours ($P < 0.01$; Figure 3.1D-E,

3.2I), respectively. Thus, all *Upf2*-null blastocyst ICMs regressed very rapidly *in vitro.* This result implied that *Upf2*-null blastocysts have an ICM-specific phenotypic defect.

Upf2 is required for expression of a key pluripotency factor in the epiblast

The discovery that *Upf2*-null blastocysts exhibited a specific ICM defect in an *in vitro* outgrowth assay led us to next assess *Upf2*-null development *in vivo* using markers known to specifically label the ICM and other early cell lineages. To this end, we performed immunofluorescence to examine the expression of the most-widely known lineage markers for each of the three lineages present in the blastocyst: NANOG (ICM/EPI), GATA6 (PrE), and CDX2 (TE). Zygotes were isolated from heterozygous parental crossings after superovulation and grown *in vitro* to the blastocyst stage. Blastocysts were fixed and stained for the three markers detailed above, and 3-D reconstructed using confocal z-stack projections. An automated image processing and counting routine, uniquely tailored to each antibody signal, allowed us to accurately count the number of NANOG-, GATA6-, and CDX2-positive cells within each embryo, in an unbiased manner. The blastocysts were genotyped following imaging. Embryos were binned into two groups (early- and mid-blastocyst), based on total cell number (see Methods). Genotypes were then compared within the early- and mid-blastocyst bins to accurately analyze marker expression in blastocysts of a comparable developmental stage. We observed a significantly lower percent of NANOG-positive cells in *Upf2*-null late blastocysts as compared to littermate controls ($P<0.05$; Figure 3.3A-B). The effect on NANOG-expressing cells was specific, as we did not observe a statistically significant difference in the percent of total CDX2-positive or GATA6-positive cells between genotypes (Figure 3.3A-B). In addition, the early blastocyst stage embryos did not demonstrate a significantly different number of NANOG-, GATA6-, or CDX2-positive cells (Figure 3.4A-B). These data supported gross morphological observations, that *Upf2*-null E3.25 blastocysts were phenotypically unaltered, but by E3.5, began to exhibit phenotypic defects. In combination with the ICM-specific defect observed when

Upf2-null blastocysts were outgrown, the lower number of cells expressing NANOG at the mid-blastocyst stage suggested a role for UPF2 in the pluripotent epiblast.

Upf2 regulates expression of mRNAs within the pluripotent ICM

To determine if other factors known to be expressed within the pluripotent epiblast were also affected in *Upf2*-null blastocysts, and at the RNA level given the known role of UPF2 in RNA decay, we performed mRNA sequencing on whole embryos isolated at E3.25 and E3.5. In total, 19 E3.25 stage blastocysts (8 *Upf2*$^{-/-}$, 11 *Upf2*$^{+/+}$) and 15 E3.5 stage blastocysts (8 *Upf2*$^{-/-}$, 7 *Upf2*$^{+/+}$) were subjected to mRNA sequencing, utilizing our previously established methods for RNAseq with low input (Dumdie et al., 2018; Mora-Castilla et al., 2016). Embryos were genotyped after cDNA synthesis (and before library preparation) by PCR (Figure 3.5A, and confirmation of *Upf2* knock-out was accomplished via read alignment at exons 2 and 3, which are the exons removed in this mutant mouse line (Figure 3.5B).

First, we wanted to determine if loss of Upf2 affects the global gene expression profile of mouse blastocysts. Plotting whole transcriptome variance profiles with dimensionality reduction (t-SNE), we observed a developmental trajectory of littermate control (wild-type) blastocysts (Figure 3.6A). While *Upf2*-null E3.25 blastocysts seemed to cluster with their control counterparts, E3.5 blastocyst transcriptome profiles resembled that of E3.25 embryos more so than E3.5 embryos. This indicated, based on global gene expression profiles, that *Upf2*-null E3.5 blastocysts exhibited a slight delay in development, consistent with morphological observations (Figure 3.1B, S1G). With the ICM-specific phenotypic defects we observed in mind, and focusing on the stage of blastocyst development before any obvious phenotypic or global transcriptomic delay (E3.25), we sought to determine if *Upf2*-null embryos possessed a gene expression profile associated with pluripotency.

Examining the expression of the well-established core set of pluripotency factors *Nanog, Oct4/Pou5f1, Sox2, Klf4,* and *c-Myc* in E3.25 embryos, we saw that all but one of these factors

were downregulated in *Upf2*-null blastocysts (Figure 3.6B). To expand upon these data, we next examined the expression of other embryonic lineage-specific markers in our RNA-seq data. To this end, we compiled mouse ICM- and TE-specific gene lists identified by Blakeley et al. (Blakeley et al., 2015), as well as EPI- and PrE-specific genes identified in early and late blastocysts by Ohnishi et al. (Ohnishi et al., 2013). At both E3.25 and E3.5, the ratio of the expression of ICM-specific and EPI-specific genes in *Upf2*-null embryos relative to control embryos decreased. This can be visualized as a deviation from one for the ratio of the normalized reads of lineage-specific genes in *Upf2*-null embryos relative to control embryos (ICM vs TE, $P < 0.01$; EPI vs PrE, $P = 0.0005$; Figure 3.7A-B). In contrast, however, the ratio of the expression of TE-specific and PrE-specific genes in *Upf2*-null embryos relative to control embryos was not significantly different. This data further supported that *Upf2*-null embryos exhibit a lineage-specifc specific defect in the pluripotent cell population of the embryo—the ICM, and subsequently the EPI.

The decrease in expression of a large set of ICM- and EPI-specific genes at the whole-embryo level could result from downregulation of gene expression in *Upf2*-null blastocysts, or alternatively, it could result from *Upf2*-null blastocysts having proportionally less ICM/EPI cells. To distinguish between these possibilities, we compiled a set of 27 housekeeping genes that have been previously verified in mouse across many tissue and cell types (Li et al., 2017). We then normalized the expression of all lineage-specific genes to the expression of the housekeeping genes, re-calculating the ratio of their expression in *Upf2*-null compared to control embryos. While ICM- and EPI- specific gene expression remained lower for *Upf2*-null E3.25 blastocysts relative to controls, the expression of these lineage-specific genes at E3.5 was approximately equal for *Upf2*-null embryos as compared to controls after normalization (Figure 3.7A-B).

This supports the conclusion that gene expression of pluripotency-associated genes is downregulated in *Upf2*-null blastocysts at E3.25, but at E3.5, the downregulation of these genes appears, instead, to represent a decrease in ICM/EPI cell number. The allocation of an embryonic cell to a particular lineage is dependent upon a complex network of signaling and transcriptional factors, and how lineage-initiating factors become segregated to their respective cell type remains an active area of research. Here, we demonstrate, in close developmental proximity, a global reduction in gene expression of ICM- and EPI-specific genes, followed by a reduction in the number of EPI cells in the blastocyst. Given that the number of PrE cells in *Upf2*-null embryos is not increasing at this transition, these cells destined to become EPI must either be delayed (slowed) in growth, or selectively undergoing apoptosis. Altogether, this data supports a hypothesis in which *Upf2* specifically modulates the expression of factors essential for establishing and/or maintaining pluripotency in the ICM/EPI cells within the early mouse blastocyst.

Loss of Upf2 leads to an inability to maintain pluripotency and slowed proliferation in vitro

Since our previous data suggested a specific role for *Upf2* within the pluripotent ICM, we next wanted to assay the ability of *Upf2*-null embryos to generate mouse embryonic stem cells (ESCs) *in vitro*. That *Upf2*-null blastocysts exhibited a delay at the mid-blastocyst stage indicated that the ICM might already possess a defect, which we might be able to assay in their ability to derive ESCs. Using conventional ESC derivation conditions (serum-containing media with LIF), only 4% of ESC lines generated were verified *Upf2*-null ESC lines (as compared to 25% as would be expected by chance) (Figure 3.8B). Because of the decreased propensity of *Upf2*-null blastocysts to generate ESCs, we attempted to derive ESCs under naïve pluripotency (2-inhibitor (2i, MEK inhibitor and GSK3 inhibitor) with LIF) conditions (Figure 3.8A). ESCs in naïve pluripotency conditions are thought to more closely resemble cells of the blastocyst ICM,

whereas conventional ESC conditions are thought to be slightly more primed towards an EPI fate. Following *Upf2*-heterozygous parental crosses, *Upf2*$^{+/+}$, *Upf2*$^{+/-}$, and *Upf2*$^{-/-}$ blastocysts were plated on feeders for ESC derivation. Of all ESC lines derived under naïve conditions, 12% were *Upf2*-null ESC lines. Thus, *Upf2*-null blastocysts could more efficiently generate ESCs under more naïve culture conditions. This data again further supported that *Upf2*-null blastocysts were already defective at the mid-blastocyst stage.

Upf2-null naïve ESC lines displayed grossly normal morphology for ESCs of this type, stained positive for alkaline phosphatase indicating they are in a pluripotent state, and expressed markers associated with naïve pluripotency, including *Nanog* and *Pou5f1/Oct4* (data not shown). However, compared to littermate control ESC lines, *Upf2*-null ESCs exhibited markedly slowed growth (Figure 3.8C), consistent with the reduced number of EPI cells in *Upf2*-null blastocysts (Figure 3.3A-B). That *Upf2*-null ESCs could not be efficiently derived in conventional ESC conditions suggested that *Upf2* is necessary for the transition from naïve to conventional pluripotency. As another test of this hypothesis, we passaged the naïve *Upf2*-null ESC lines into slightly more differentiated conditions; (1) into 2-inhibitor media with LIF but without feeders, and (2) into serum-containing media with LIF and feeders (Figure 3.8A). Attempting to transition *Upf2*-null ESCs out of naïve pluripotency with either of these conditions resulted in a morphological pattern associated with differentiation (Figure 3.8D), as well as slowed growth as observed in naïve culture conditions (Figure 3.8C). Further mechanistic studies will attempt to unravel how UPF2 regulates the proliferation and/or survival of ESCs, findings which will likely provide additional insight into the role of UPF2 in pluripotency in the embryo. That UPF2 might be necessary for the transition out of naïve pluripotency by regulating the expression of ICM- and EPI-specific genes, as well as regulating the proliferation rate of this pluripotent cell type, is an exciting possibility and will be directly tested in ongoing experiments.

Identifying Upf2-regulated mRNAs in the early embryo

In addition to assessing the expression of factors known to regulate pluripotency in the embryo, we also wanted to identify other pathways or factors that might be regulated by Upf2 and the NMD pathway in early embryonic development. To address this question, we set out to identify direct targets of UPF2 and the NMD pathway in the embryo. The NMD field has long been plagued by a lack of consensus of what defines an NMD target, along with the complicated nature that the repertoire of NMD targets varies between species and cell types. In studies that have knocked out or down an NMD factor, it is not known what proportion of the dysregulated RNAs are direct NMD targets, also known as NMD substrates. While some normal RNAs targeted by NMD have been identified via more direct experimental approaches (Hurt et al., 2013; Johansson et al., 2007; Tani et al., 2012), and via the presence of transcript features that render it likely to be degraded by NMD (such as a dEJ) (Bao et al., 2015; Lou et al., 2016; Shum et al., 2016), the vast majority of NMD substrates remain unknown. Thus, our datasets provided a unique ability to examine the expression of NMD substrates in a developmental system within a narrow time window, after removal of one of the core NMD factors (*Upf2*). Our goal was to examine differences in gene expression at each stage of blastocyst development, with a focus on the E3.25 datasets potentially containing candidate direct targets of Upf2 that might initiate the developmental delay and other defects we observed.

To this end, we first examined mRNAs upregulated in *Upf2*-null blastocysts compared to controls at the same stage of development. Given that UPF2, as a member of an RNA decay pathway, degrades its targets, transcripts upregulated with knock-out of Upf2 are candidate targets of UPF2 activity, while downregulated genes likely represent indirect, downstream effects. At the E3.25 stage, we detected 116 differentially expressed genes between *Upf2*-null and control blastocysts ($Q<0.05$). Of those, 68 were upregulated in *Upf2*-null embryos, and 48 were downregulated. Gene ontology (GO) categories enriched in upregulated genes included

those involved in the unfolded protein response (UPR) (*Atf4* and *Ddit3),* as well as nucleotide catabolism (*Itpa, Pnrc2, Exog).* In fact, NMD has been reported to mediate the UPR (Karam et al., 2015). Relaxing the statistical cutoff within this dataset (Q<0.1), produced a higher number of upregulated genes (101 genes) to analyze as potential direct targets of Upf2. GO analysis of this list produced the categories above, as well as the additional categories of cell cycle regulation, microtubule organization, and mitochondrial function. Many genes that have been identified as NMD substrates were included in this expanded list of genes upregulated with loss of Upf2, including *Atf4, Ddit3,* and *Gadd45a*. Broad GO categories of genes downregulated with loss of *Upf2* at this early blastocyst stage included organization of the extracellular matrix (*Col5a1 Ltbp3, Matn3,* and *Crtap)*, nuclear division (*Osm, Spast, Kif23, Bora,* and *Rspo1)*, and regulation of gastrulation (*Nat8, Nanog)*.

While very few genes were differentially regulated between control and *Upf2*-null blastocysts at E3.25, the repertoire of affected genes was much greater at E3.5. This is consistent with the observation that E3.25 *Upf2*-null embryos do not yet exhibit an obvious phenotype. Thus, *direct* targets likely to be initiators of the *Upf2*-null phenotype are likely present in the E3.25 datasets, whereas later timepoints likely contain clues as to how embryogenesis fails thereafter. Aligning with this hypothesis, *Upf2*-null E3.5 embryos differentially expressed 1339 genes, compared to controls (Q<0.05). Of those, 609 were upregulated. GO categories enriched in upregulated genes included those important for metabolism of ncRNAs, processing of mRNAs, histone modifications, and response to ER stress. At this same stage, 730 genes were downregulated. Strikingly, over 60 of the downregulated genes encoded for ribosomal protein (*Rpl, Rps*) mRNAs, suggesting a robust response of these *Upf2*-null embryos in downregulating ribosome production and translation. Other pathways that were downregulated included oxidative phosphorylation and other components of the mitochondria, processing of rRNAs, cell cycle regulation, and extracellular

matrix (ECM) interactions. Interestingly, knock-out or knock-down of other NMD factors has been reported to result in a decrease in adhesion or other ECM interactions, as well (L. Huang et al., 2018).

Loss of an RNA decay factor results in stabilization of its target RNAs. However, it has been shown that many stabilized RNAs are not necessarily upregulated (Tani et al., 2012). This is because the steady state level of an mRNA reflects the balance of both the rate of transcription and the rate of decay. *In vitro*, it is common to assess mRNA stability via treatment with a transcriptional inhibitor such as Actinomycin D, but this has many drawbacks. Thus, we elected to instead use a computational approach that infers RNA decay rate from steady-state RNA sequencing data and thus does not suffer from the side effects inherent in using transcriptional inhibitors. In particular, one of these methods utilizes the traditional Δexon – Δintron method of estimating RNA stability, inferring transcriptional rates from changes in abundance of intronic reads, and using exonic read counts as a measure of the steady-state mRNA abundance. Recently, Alkallas et al. defined an innate bias within this formula, defining a regression-based bias term, and demonstrated that this corrected stability measurement correlates with *in vitro* BRIC-seq stability measurements (Alkallas et al., 2017). Using this algorithm (REMBRANDTS) (Alkallas et al., 2017), we were able to define the complete repertoire of mRNAs stabilized with loss of Upf2 in the early embryo. Consistent with the accuracy of this approach, many known NMD substrates were found to be stabilized. In addition, as observed at steady state, many more mRNAs were dysregulated at the later timepoint (E3.5) as compared to the earlier stage (E3.25).

Using this approach, we quantified the number of mapped reads from our steady-state mRNA sequencing data by intron versus exon, and calculated a differential stability score (DSS), defined as the stability measurement of the *Upf2*-null samples minus the control samples. At the earlier timepoint (E3.25), we observed 1726 genes encoding mRNAs stabilized

in *Upf2*-null embryos (DSS > 0.01) (Figure 3.9A). Many of those genes (673 genes) were very highly stabilized (DSS > 0.1), as might be expected after knock-out of an mRNA decay factor. mRNAs destabilized by Upf2 at E3.25 include those involved in cell cycle regulation, DNA repair and replication, chromosome segregation, vesicle transport and assembly of organelles, and microtubule organization.

Next, we calculated DSSs for genes stabilized at E3.5. Similar to the steady-state differential expression measurements, many more genes encoding mRNAs were differentially stabilized at this timepoint, as compared to E3.25. We observed 2500 genes stabilized in *Upf2*-null embryos at this stage (DSS > 0.1) (Figure 3.9A). The GO categories of the genes encoding these stabilized mRNAs at E3.5 were similar to those at E3.25. GO categories enriched at E3.5, but not at E3.25, were “methylation”, “chromatin modification”, and “autophagy”.

Compiling transcripts stabilized in both E3.25 and E3.5 datasets in *Upf2*-null embryos relative to controls, produced a core list of 257 transcripts likely degraded by NMD in early embryogenesis (Figure 3.9B). The stability profiles of these genes indicated that NMD could be responsible for regulating a number of processes integral to cell and embryonic vitality, including the DNA damage response, cell cycle regulation and chromosomal segregation, microtubule organization, and telomere maintenance. That these integral cell functions likely require a level of robustness in terms of overall regulation that might be achieved through RNA decay by the NMD pathway is an exciting finding.

As another approach to define high-confidence NMD direct targets during embryo development, our approach was two-fold. First, we examined the lists of upregulated and stabilized mRNAs for the presence of NMD-inducing features. Transcripts targeted for degradation by NMD possess a number of features that can be detected *in silico*. The most consistent NMD-inducing feature is a downstream exon-exon junction (dEJ). Other features that elicit NMD in an EJC-independent context include long 3’UTRs and upstream open reading

frames (uORFs). In agreement with the hypothesis that RNAs stabilized upon knock-out of Upf2 were more enriched for direct targets than those simply upregulated in steady-state analysis, we found that, at E3.25, 30.2% of genes stabilized encoded an RNA with a dEJ, compared to 23.5% of genes upregulated. Similarly, genes both stabilized and upregulated at E3.25 encoded an RNA with a dEJ 36.8% of the time (19 genes, DSS > 0.01, Q > 0.05) (Figure 3.9C). This is in contrast to the overall abundance of dEJs within the mouse genome at a rate of 17%. Genes that were both stabilized and upregulated included some that have been reported as NMD direct targets, including *Ddit3*. Also in this list was *Anapc16,* a cell cycle-regulated E3 ubiquitin ligase that is required for progression through the G1 phase of the cell cycle. In fact, another NMD-deficient ESC line has been demonstrated to stall at the G1 phase of the cell cycle (Lou et al., 2016). Given our observation of slowed growth of the EPI of *Upf2*-null blastocysts and the significantly slowed proliferation rate of *Upf2*-null ESCs, it will be interesting to determine if NMD regulates ICM and ESC cell cycle progression through *Anapc16* in future experiments.

A similar relationship between mRNA stabilization and upregulation was seen at E3.5. The incidence of a dEJ within a transcript encoded by a gene upregulated upon Upf2 knock-out was 28.4%. However, in genes stabilized with knock-out of Upf2 this rose to 30.9%, and in genes both upregulated and stabilized in Upf2-null E3.5 embryos, the incidence of a dEJ was 33.7% (335 genes, DSS > 0.1, Q > 0.05) (Figure 3.9D). At E3.5, there were 113 genes that were both stabilized and upregulated and, also encoded an mRNA with a dEJ. These mRNAs encoded for proteins with known roles in histone modification and mRNA processing in general.

Defining candidate Upf2-driven regulatory networks in the early embryo

One means by which NMD could orchestrate early embryonic development is by regulating key transcription factors and/or signaling pathways. To identify whether such UPF2-based networks exist, we searched for genes regulated by UPF2 that, in turn, are known to regulate genes downstream of UPF2. The idea was to identify key regulatory factors under the

control UPF2 that regulate large numbers of downstream genes. Such regulatory circuits would amplify the effects of UPF2. One regulatory factor that fit these criteria is the transcription factor, SALL4, which is a member of the SALL subfamily of zinc finger transcription factors, and is required for proliferation of the ICM in mouse blastocysts (Sakaki-Yumoto et al., 2006). SALL4 is known to positively regulate many genes that our data indicated depend on UPF2 for maximal expression, including *Fgf4, Nanog*, and *Utf1* in E3.25 embryos, and *Dppa4, Fbox15*, and *Hmga2* in E3.5 embryos. Likewise, NANOG is a candidate to be another key factor downstream of UPF2. Factors known to regulate the expression of NANOG include SUZ12, ESRRB, RAD23B, and MED1. Surprisingly, these regulators are also stabilized with the loss of *Upf2,* making them interesting candidates to link UPF2's mRNA decay activity with the downregulation of *Nanog* observed in the absence of *Upf2.* In particular, SUZ12 is a member of the Polycomb Repressive Complex (PRC), and represses transcription through methylation of histone H3 lysine 27 (H3K27) (Pasini et al., 2007). Loss of SUZ12 leads to embryonic lethality in mice and ESCs lacking *Suz12*, upon differentiation, aberrantly express NANOG (Pasini et al., 2007). That UPF2 might regulate pluripotency in the ICM through direct regulation of the PRC is another exciting possibility.

In total, we demonstrate here that UPF2 is necessary for mouse embryo survival past the implantation stage of development. We show that loss of *Upf2* leads to specific alterations in gene expression of factors associated with pluripotency in the ICM of the mouse blastocyst, with no observable phenotype in the extra-embryonic TE. *Upf2*-null blastocysts contain fewer NANOG-positive ICM cells, a marker for pluripotent cells, and their ICM regresses in culture. Further, while *Upf2*-null ESCs can be derived under conditions that favor naïve pluripotency, they exhibit slowed growth and differentiate prematurely when encouraged to progress from an ICM-like state to a more EPI-like state. Thus, we report that UPF2, an RNA decay activator, is

necessary for pluripotent gene expression in the ICM of the mouse embryo, and further, is necessary for exit from naïve pluripotency in mESCs *in vitro.*

Discussion

Establishment of the pluripotent population of cells in the mouse embryo is critical for successful development and is associated with expression of a number of key transcription and signaling factors. Specifically, POU5F1/OCT4 becomes restricted to the ICM at the time of blastocoele formation (Chazaud and Yamanaka, 2016). Immediately thereafter, FGF4 and NANOG become restricted to the EPI, marking the population of pluripotent cells that will form the entire organism (Chazaud and Yamanaka, 2016). Our study reveals a critical role for an RNA decay activator, UPF2, in regulating the expression of these factors in the mouse blastocyst and for maintaining pluripotency *in vitro*. Similarly, these pluripotent populations of cells in the mouse embryo can be perpetuated in culture as ESCs. Depending on the signals provided, mESCs can exist in a naïve state, in which MEK signaling is inhibited (representing an "ICM-like" cell type), to a primed state, which is thought to be the *in vitro* counterpart to a mature EPI cell immediately preceding gastrulation (Marks et al., 2012; Nichols and Smith, 2009; Weinberger et al., 2015). Our *in vitro* cell culture data suggests that UPF2 is necessary for ESCs to fully transition from the naïve pluripotent state to slightly more primed states.

These findings demonstrate that embryos devoid of *Upf2* first demonstrate a global reduction in gene expression patterns associated with the ICM, the EPI, and pluripotency, and as a result, that they ultimately allocate fewer cells to the NANOG-expressing pluripotent EPI (Figure 3.3, 3.4, 3.6, 3.7). Whether this represents a failure to establish pluripotent cells, or a relative failure of these cells to proliferate, remains an interesting question for future study. *In vitro,* ESCs can be derived from *Upf2*-null embryos under naïve pluripotency conditions, but exhibit slowed growth and an inability to maintain pluripotency upon exit from the naïve state

(Figure 3.8). These data suggest the defect might be predominantly at the step in which naïve cells begin to differentiate. It is also possible that *Upf2* and the NMD pathway might be important for ICM cells and ESCs to progress through the cell cycle, as evidenced by their slowed growth (Figure 3.3, 3.8C). Transcriptomic analysis of *Upf2*-null blastocysts compared to littermate controls revealed several plausible mechanisms by which UPF2 and the NMD pathway might work to mediate the reduction in pluripotent gene expression in the early blastocyst, including indirectly through the decay of upstream regulators of pluripotency factors, or by affecting cell cycle progression. These findings provide interesting avenues for future mechanistic studies.

A number of studies have already found various connections between NMD factors and maintenance of pluripotency or differentiation *in vitro* in ESCs, as well as the necessity of these factors for embryo viability in mice and other organisms (Alonso and Akam, 2003; Anastasaki et al., 2011; Casadio et al., 2014; Hwang and Maquat, 2011; Li et al., 2015; Longman et al., 2007; Lou et al., 2016; Mao et al., 2015; McIlwain et al., 2010; Medghalchi et al., 2001; Metzstein and Krasnow, 2006; Shum et al., 2016; Silver et al., 2010; Weischenfeldt et al., 2008). The loss of UPF1, UPF3A, and SMG6 lead to peri-implantation lethality in mice, with ICM-specific phenotypes (Li et al., 2015; Medghalchi et al., 2001; Shum et al., 2016). Thus, NMD is thought to be essential for maintaining cellular homeostasis in early development, and ultimately necessary for organism survival in vertebrates. As one example, loss of *Smg6* leads to lethality at the blastocyst stage and is incompatible with ESC derivation, and while *Smg6* is dispensable in established ESCs, these ESCs exhibit impaired differentiation due to sustained expression of pluripotency markers (Li et al., 2015). The authors further demonstrate that the NMD activity of SMG6 regulates the expression of *c-Myc*, and that differentiation can be rescued by the downregulation of this key factor (Li et al., 2015). While ESCs could not be derived from *Smg6*-null blastocysts in this study under normal ESC derivation conditions, it would be interesting to

see if null lines could be derived under naïve ESC conditions, as we have done here, and if the phenotype was comparable. In human ESC studies, however, the magnitude of NMD activity has been shown to differentially regulate differentiation into one lineage versus another (Lou et al., 2016). Our data demonstrate an essential role for UPF2 in the transition from naïve to more primed ESC states *in vitro*, which coincides with the phenotypes we observed in *Upf2*-null embryos *in vivo.*

Thus, these data suggest that the NMD pathway may play a role in buffering the expression of pools of mRNAs at key developmental transitions, particularly upon exit from pluripotent states. When a cell differentiates from a 'stem cell' state to a more differentiated cell type, specific sets of mRNAs necessary for maintaining the stem cell state must be downregulated, and mRNAs important for the differentiated cell type must be upregulated. This transition likely requires a complex cellular framework to coordinate these events, and selective mRNA decay is a plausible mechanism to aid in this transition. Indeed, even in more differentiated cell types, particularly in neural development, the NMD pathway selectively modulates its activity at various stages of neural maturation (Bruno et al., 2011; Lou et al., 2016). In this study, we show that UPF2 plays an essential role at the transition from naïve to primed pluripotency in mESCs. It will be interesting to see in future studies if other NMD-deficient ESC lines can be derived under these naïve pluripotency conditions and exhibit a similar phenotype.

Transcriptomic data from *Upf2*-null blastocysts compared to control blastocysts demonstrated a global reduction in genes known to exhibit enriched expression in the ICM, in the EPI, and in pluripotent cells (Figure 3.7, 3.8). While these are not direct targets of UPF2's mRNA decay activity, we were able to identify common upstream regulators that UPF2 and the NMD pathway might work through to regulate these genes and thereby influence the ICM, the EPI, and pluripotency. Genes that will be interesting targets for future studies include those that

we show are also stabilized or upregulated with loss of *Upf2* in our embryo datasets, as these are candidates to be direct NMD targets. One of these is SUZ12, which is a member of the PRC and is necessary for differentiation of mESCs (Pasini et al., 2007). ESCs devoid of *Suz12* have no detectable H3K27me3, and when forced to differentiate in culture, fail to downregulate pluripotency factors, albeit upregulating differentiation-specific factors (Pasini et al., 2007). Our finding that *Suz12* is stabilized in *Upf2*-null blastocysts makes it an exciting target through which NMD could modulate this developmental transition. Our transcriptomic data indicates that *Esrrb* mRNA is stabilized in blastocysts devoid of *Upf2,* making it another interesting candidate for future mechanistic studies aimed at uncovering the role of NMD in mediating pluripotency and/or buffering these important developmental transitions. Several lines of evidence suggest that ESRRB is involved in pluripotency, including that ESRRB interacts with NANOG, OCT4 contains ESRRB binding sites in its 5'UTR, and knockdown of *Esrrb* in mESCs leads to premature differentiation (Zhang et al., 2008).

In summary, our data demonstrate a compelling role for a post-transcriptional pathway in regulating the establishment and/or transition between pluripotent states in the mouse blastocyst. While many studies aimed at understanding early embryo development and the propagation of ESCs focus on transcriptional regulation, we provide a likely mechanism by which selective RNA decay might help to mediate key cellular transitions during the earliest stages of development. This has broad implications for future studies aimed at understanding pluripotency *in vitro,* as well as in normal mammalian development.

Acknowledgements

Chapter 3 contains material that will be prepared for submission for publication. The dissertation author was the primary investigator and author of this material. Current co-authors

are as follows: Jennifer N Dumdie, Abhishek Sohni, Kristoffer Vitting-Seerup, Bo Porse, Heidi Cook-Andersen*, and Miles F. Wilkinson*. (* Authors contributed equally to this work)

Figure 3.1. Upf2 is necessary for outgrowth of the mouse blastocyst ICM. (A) Early postimplantation embryo isolations were performed on Upf2-heterozygous females that had been mated with Upf2-heterozygous males. Embryos were isolated from the horns of pregnant females at E5.5 and E6.5, from a total of five and four breeding pairs, respectively. No Upf2-null embryos were present at E6.5 (n=26 embryos), and only 2 Upf2-null embryos were present at E5.5 (n=34 embryos). (B) Blastocysts were flushed from the uterine horns of Upf2-heterozygous females, superovulated and bred with Upf2-heterozygous males, at E3.5. A total of 7 breeding pairs (n=63 embryos) were used for this analysis. Blastocysts were categorized as early stage or expanded, as shown by the brightfield images. Total numbers are presented as percent of blastocysts per genotype. Upf2-null embryos exhibited a slight delay in development, with a tendency towards being present as an early stage blastocyst at E3.5, as compared to littermate control embryos. (C) TE outgrowth area (um^2) after 96 hours of outgrowth is presented here by genotype. Average outgrowth area is not significantly different for Upf2-null embryos and controls ($Upf2^{+/+}$ = 30,373 +/- 2634; $Upf2^{+/-}$ = 32,794 +/- 4551; $Upf2^{-/-}$ = 33,952 +/- 2569 (mean +/- SEM)), indicating proper trophectoderm function (n=32). (D) The ICM grade after 96 hours of outgrowth was assigned blinded to genotype. The grading scale for ICM is as follows: (1) many cells forming a densely packed epithelium, (2) many cells forming a loosely packed epithelium, (3) few cells forming a loosely packed epithelium, or (4) no identifiable ICM. All Upf2-null embryos received an ICM grade of 4, indicating no identifiable ICM, as compared to normal ICM grades for littermate control embryo outgrowths (n=32).

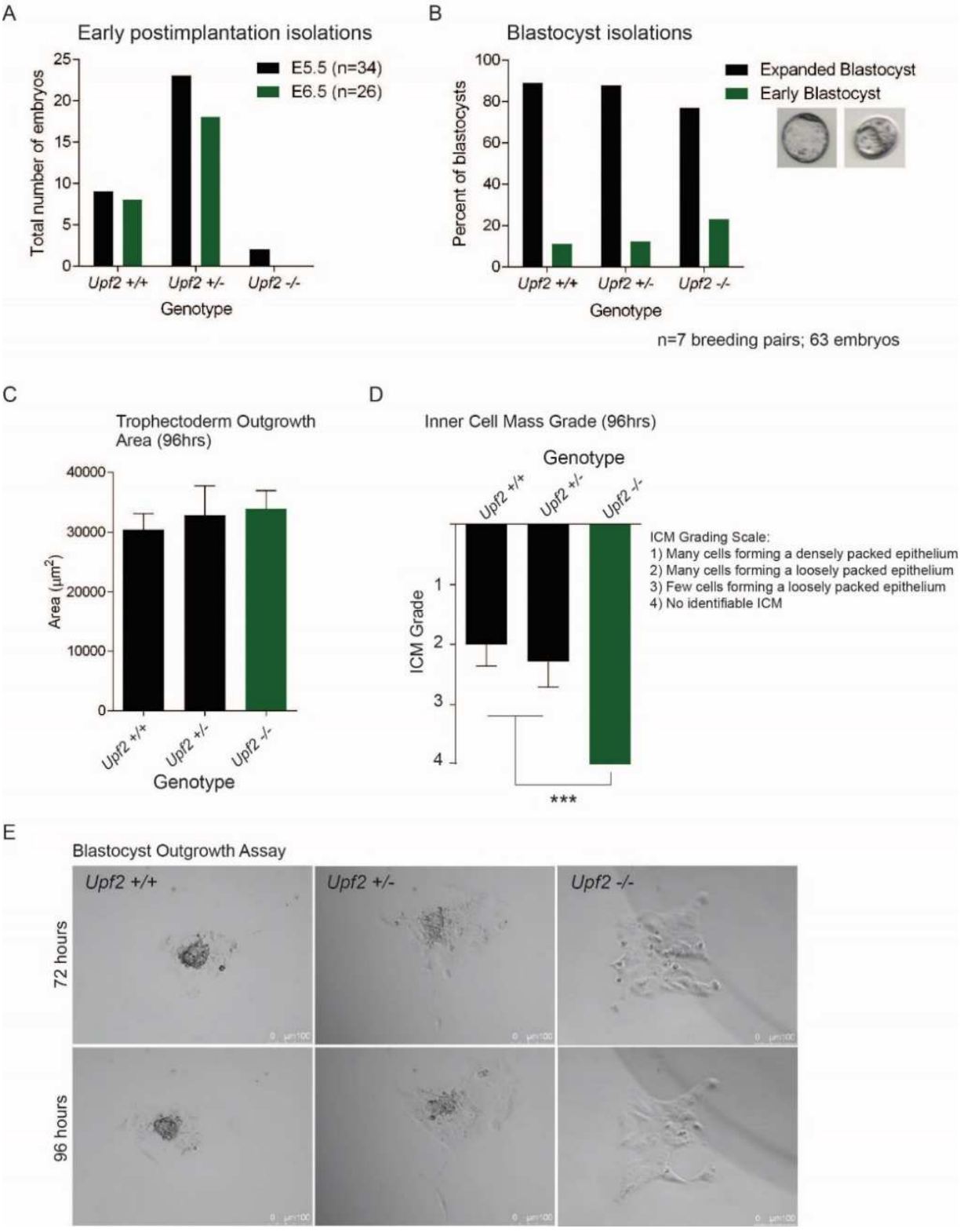
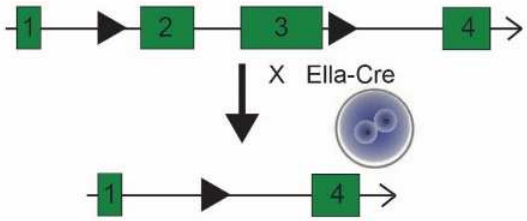

A
Early postimplantation isolations
Total number of embryos
25
20
15
10
5
0
E5.5 (n=34)
E6.5 (n=26)
Upf2 +/+
Upf2 +/-
Upf2 -/-
Genotype
B
Blastocyst isolations
Percent of blastocysts
100
80
60
40
20
0
Expanded Blastocyst
Early Blastocyst
Upf2 +/+
Upf2 +/-
Upf2 -/-
Genotype
n=7 breeding pairs; 63 embryos
C
Trophectoderm Outgrowth Area (96hrs)
Area (μm²)
40000
30000
20000
10000
0
Upf2 +/+
Upf2 +/-
Upf2 -/-
Genotype
D
Inner Cell Mass Grade (96hrs)
Genotype
Upf2 +/+
Upf2 +/-
Upf2 -/-
ICM Grade
1
2
3
4
***
ICM Grading Scale:
1) Many cells forming a densely packed epithelium
2) Many cells forming a loosely packed epithelium
3) Few cells forming a loosely packed epithelium
4) No identifiable ICM
E
Blastocyst Outgrowth Assay
Upf2 +/+
Upf2 +/-
Upf2 -/-
72 hours
96 hours

Figure 3.2. Generation of Upf2 global heterozygous mice and additional characterization of Upf2-null embryo lethality phenotype. (A) Schematic for generation of Upf2 global heterozygous mice. Upf2-floxed mice (obtained from Bo Porse laboratory), with loxP sites surrounding exons 2 and 3, were crossed with early embryo Ella-Cre mice to generate a global knock-out. (B) Table of the genotypes of live pups obtained from Upf2-heterozygous parental crosses. No Upf2-null pups were obtained from heterozygous mating pairs, indicating embryonic lethality. (C) Genotyping of individual embryos was possible using a custom DNA isolation and genotyping PCR reaction for low-input samples, as detailed in Experimental Procedures. Primer Set A (top) detects a wild-type locus and Primer Set B (bottom) detects the knockout locus. (D) The number of implantation sites detected for early postimplantation embryo isolations at E5.5 and E6.5 for Upf2-heterozygous parental crossings are shown here as a bar graph. Isolations performed at E5.5 and E6.5 produced an average of 8.2 (n=5 breeding pairs, 34 embryos) and 8.0 (n=4 breeding pairs, 26 embryos) implantation sites, respectively. (E) The number of embryo resorption sites detected for early postimplantation embryo isolations at E5.5 and E6.5 for Upf2-heterozygous parental crossings are shown here as a bar graph. Isolations performed at E5.5 and E6.5 produced an average of 1.2 (n=5 breeding pairs, 34 embryos) and 1.0 (n=4 breeding pairs, 26 embryos) implantation sites, respectively. (F) Uterine horns of Upf2-heterozygous females bred with Upf2-heterozygous males, at E5.5 (left) and E6.5 (right). Implantation sites can be visualized by the blue color (Chicago Sky Blue 6B dye). (G) Zygotes were isolated from Upf2-heterozygous females superovulated and bred with Upf2-heterozygous males, and cultured to the blastocyst stage. A total of 11 breeding pairs and 123 embryos were used for this experiment. *In vitro*-grown Upf2-null blastocysts exhibited a tendency, as *in vivo*-isolated Upf2-null blastocysts did, to be delayed in development at the blastocyst stage. Upf2-null blastocysts were more likely than their littermate control embryos, to be delayed at the early blastocyst stage, by morphological examination. (H) TE outgrowth area (um^2) after 72 hours of outgrowth is presented here by genotype. Average outgrowth area is not significantly different for Upf2-null embryos and controls ($Upf2^{+/+}$ = 27,556 +/- 2960; $Upf2^{+/-}$ = 32,242 +/- 5919 ; $Upf2^{-/-}$ = 34,596 +/- 2759 (mean +/- SEM)), indicating proper trophectoderm function (n=32). (I) The ICM grade after 96 hours of outgrowth was assigned blinded to genotype. The grading scale for ICM is as in Figure 3.1D. All Upf2-null embryos received an ICM grade of 4, indicating no identifiable ICM, as compared to normal ICM grades for littermate control embryo outgrowths (n=32).

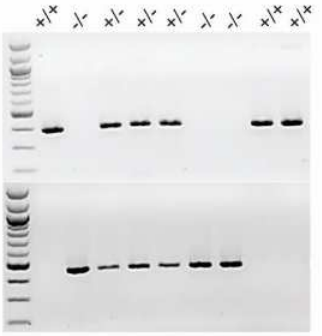
A



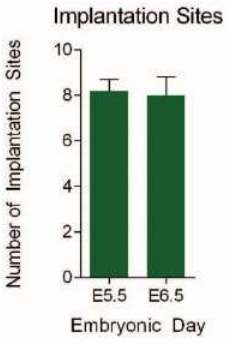
B

Genotype of pups weaned from *Upf2* +/- parental crosses (*Upf2* +/- X *Upf2* +/-)		
Upf2 +/+	*Upf2* +/-	*Upf2* -/-
36 (39%)	56 (61%)	0

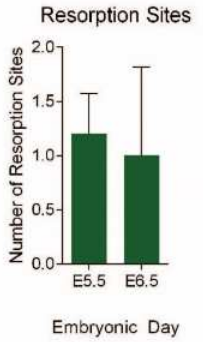
C



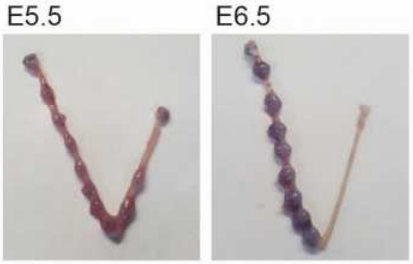
D



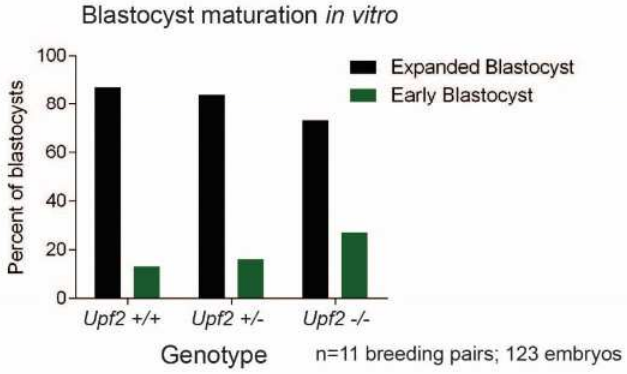
E



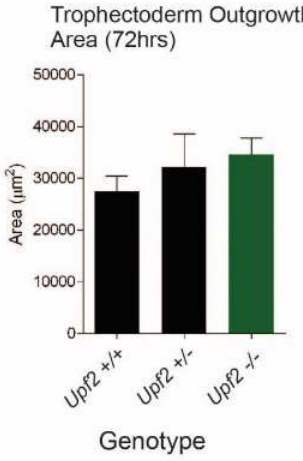
F



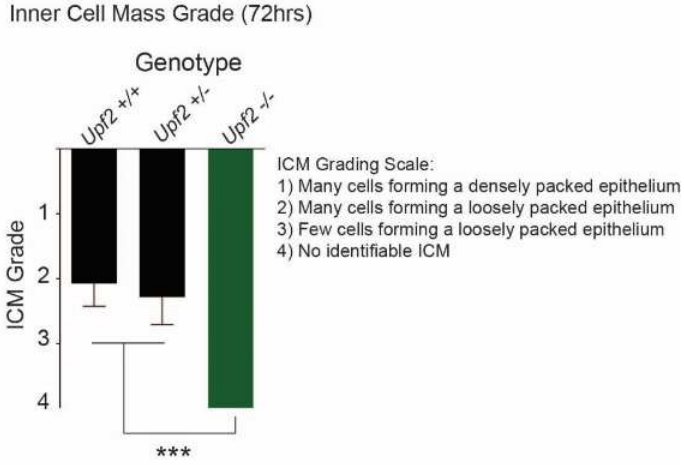
G



H



I



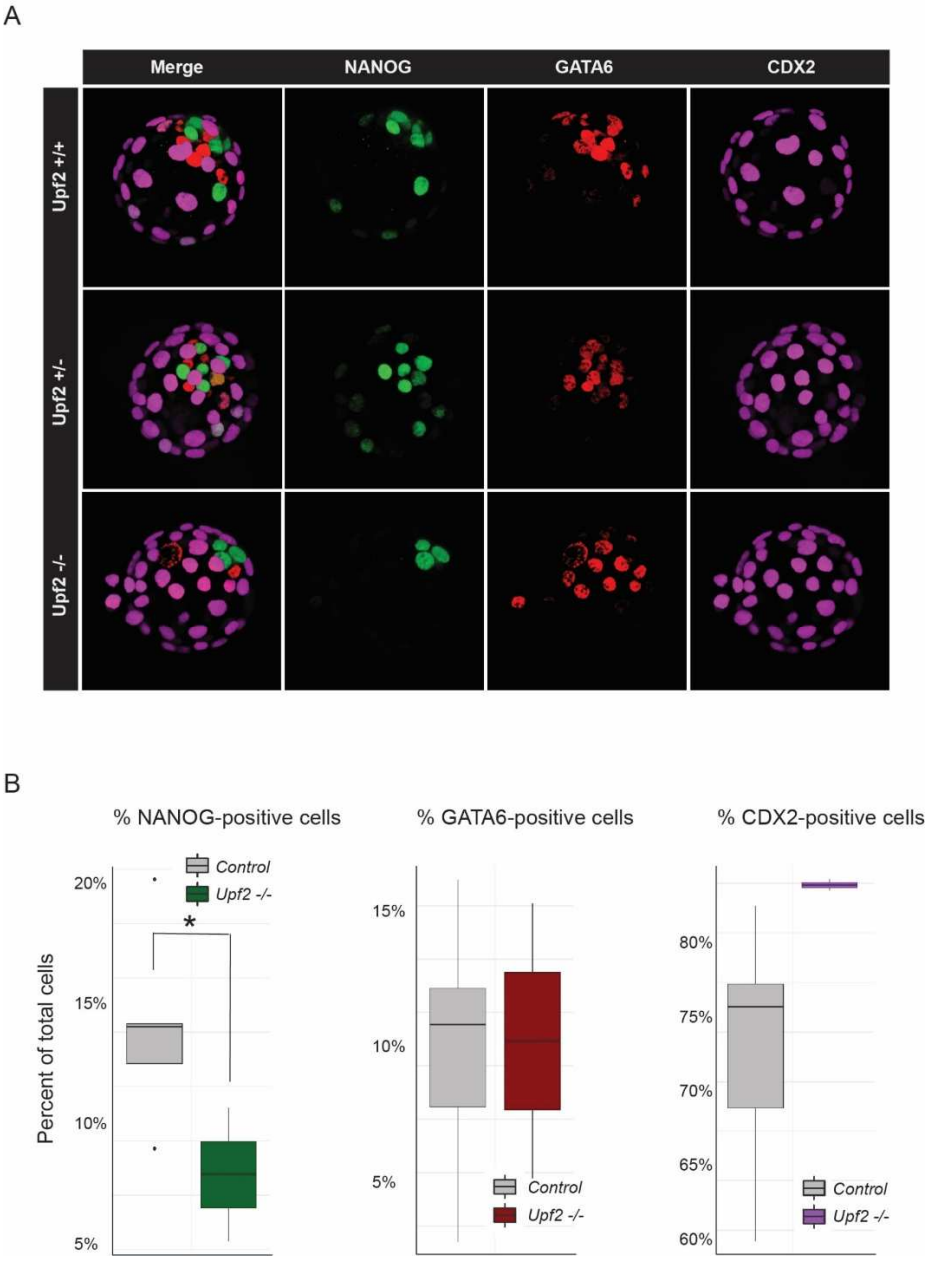


Figure 3.3. Loss of Upf2 leads to a reduction in NANOG-positive cells in the mid-blastocyst. (A) 3-dimensional (3-D) reconstructed immunofluorescence images of Upf2-null and littermate control embryos at the mid-blastocyst stage (E3.5), stained for NANOG (EPI/ICM marker), GATA6 (PE marker), and CDX2 (TE marker). Images for each channel are shown, as well as a merge of all channels. *Upf2*-null blastocysts, despite containing similar numbers of CDX2-positive, and GATA6-positive cells, contain significantly fewer NANOG-positive cells (n=7 mid-blastocyst-stage embryos analyzed). (B) Box and whisker plots of total cell counts as a percentage of total blastocyst cell number, obtained by counting nuclei positive for the three markers of interest using an automated image processing and cell counting routine. The percent of NANOG-positive cells per embryo was decreased in *Upf2*-null embryos relative to controls ($P<0.05$, Students T-test). The percent of GATA6-positive or CDX2-positive cells to total embryo cell count did not change at any of the stages of blastocyst examined.

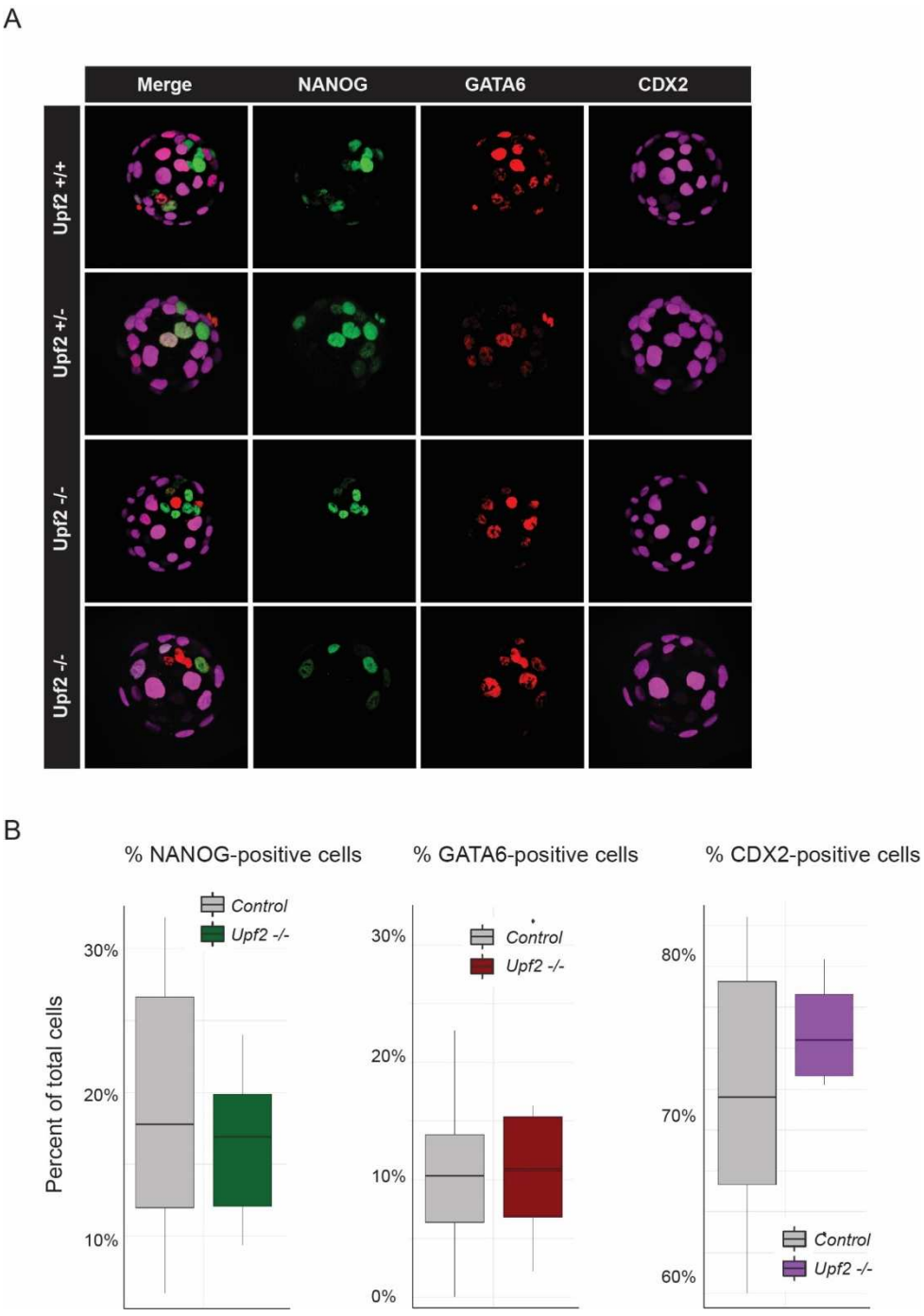


Figure 3.4. Loss of Upf2 does not affect protein expression at the early-blastocyst stage. (A) 3-D reconstructed immunofluorescence images of *Upf2*-null and littermate control embryos at the early-blastocyst stage (approximately E3.25), stained for NANOG (EPI/ICM marker), GATA6 (PE marker), and CDX2 (TE marker). Images for each channel are shown, as well as a merge of all channels. (B) *Upf2*-null E3.25 blastocysts contain similar numbers of NANOG-positive, CDX2-positive, and GATA6-positive cells (n=20 early-blastocyst staged embryos analyzed).

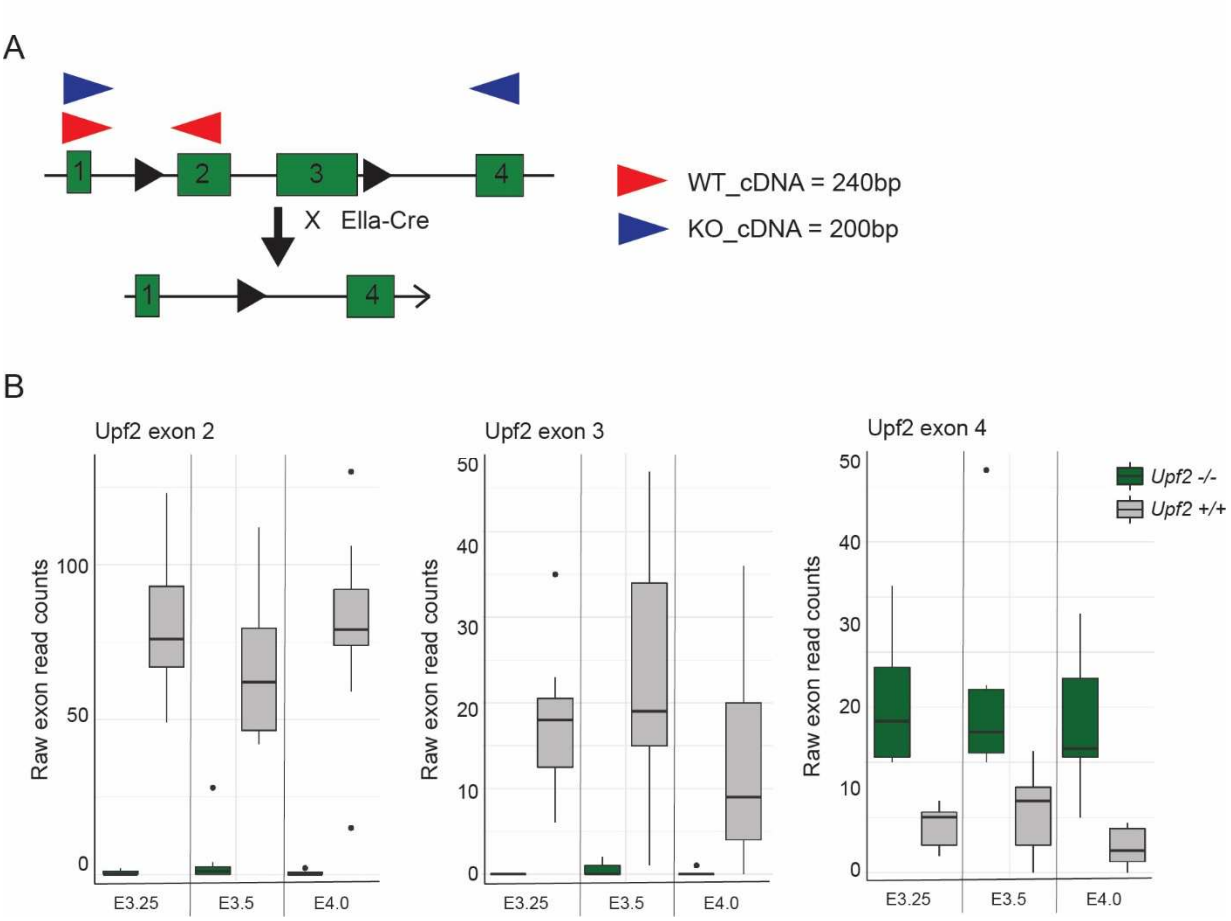


Figure 3.5. Quality control of low-input mRNA sequencing embryo datasets. (A) Schematic of custom primers designed to genotype individual embryos from purified cDNA prior to library preparation. WT_cDNA primers flank the loxP site in exons 1 and 2, and a 240bp band will be present only if that loxP site remains at a wild-type locus. KO_cDNA primers reside in exons 1 and 4, flanking both loxP sites, and a 200bp band will only be visible if knock-out has occurred. (B) Box plot of raw reads that map to exon 2, exon 3, and exon 4 of the *Upf2* gene for both Upf2-null and wild-type embryos. Exons 2 and 3 are deleted in this knock-out line, and little to no reads were observed to map to these regions in *Upf2*-null embryos. In contrast, exon 4, which is predicted to not be affected in this *Upf2*-null line, shows no decrease in the number of reads in Upf2-null compared to control embryo samples.

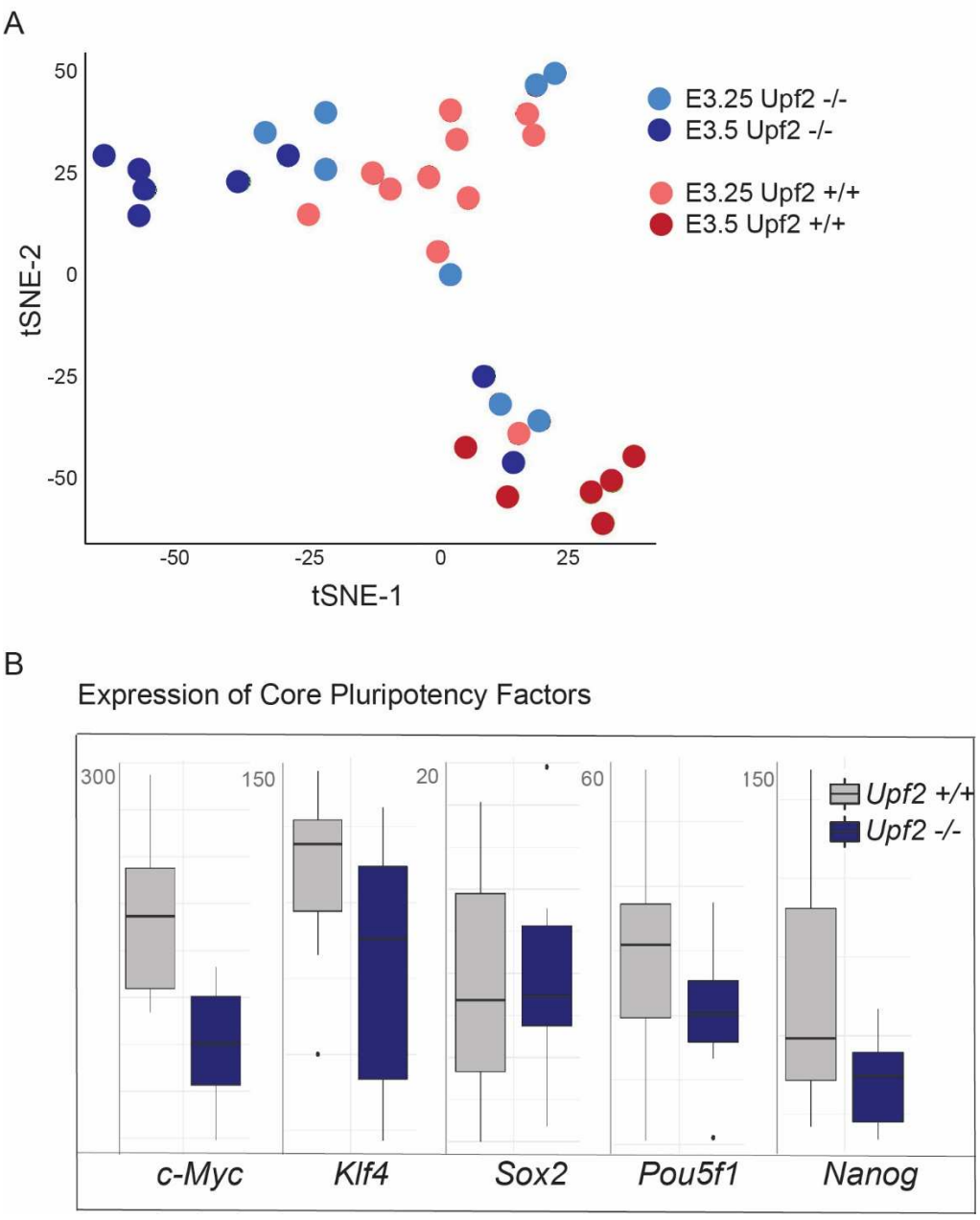


Figure 3.6. *Upf2*-null blastocysts are molecularly delayed and are reduced in their expression of pluripotency-associated mRNAs. (A) Dimensionality reduction of whole transcriptome variance profiles allowed for the visualization of low-input sample clusters, shown here as a t-SNE plot. A trajectory can be observed for control embryos, progressing from E3.25 to E3.5, shown by the arrow. While E3.25 Upf2-null embryos cluster with their age-matched control samples, the E3.5 Upf2-null embryos exhibit a delay in development, and tend to cluster with the E3.25 samples. (B) The core pluripotency factors Nanog, Pou5f1 (Oct4), Sox2, Klf4, and c-Myc are downregulated in the *Upf2*-null early blastocyst. Normalized reads (TPM) are plotted for Upf2-null and control embryos as a box and whisker plot.

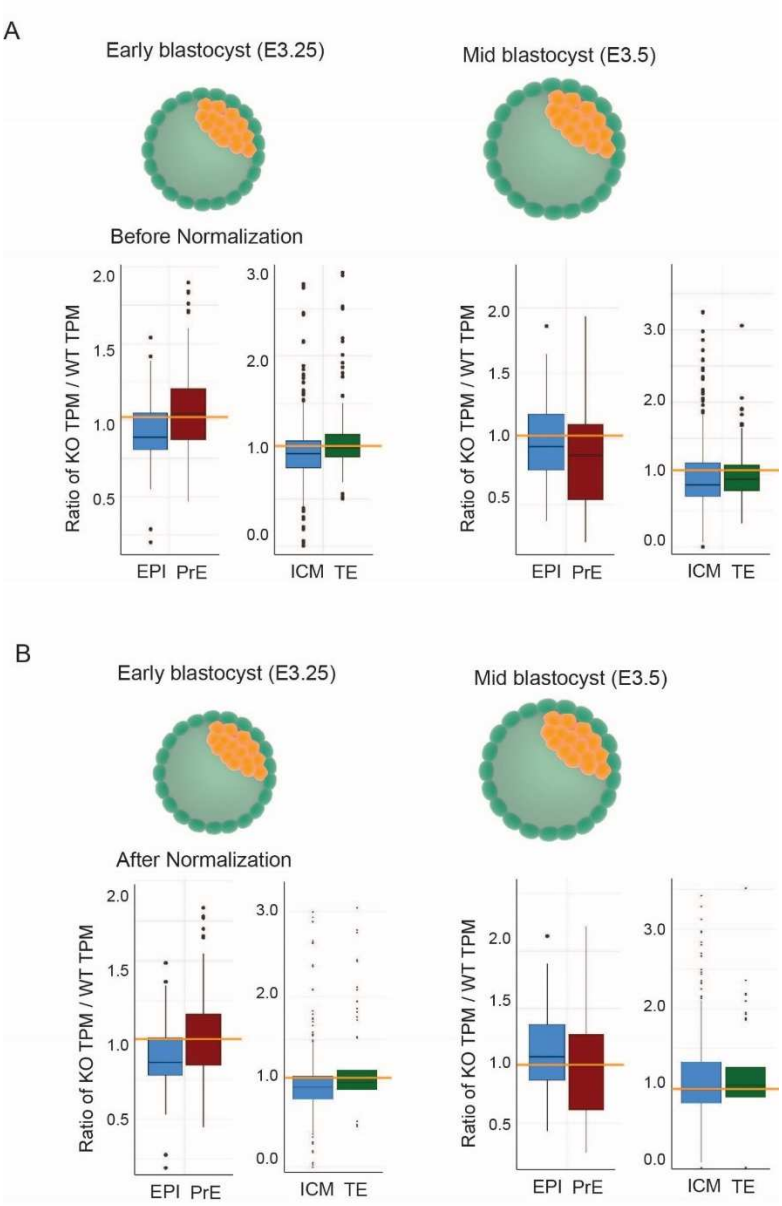
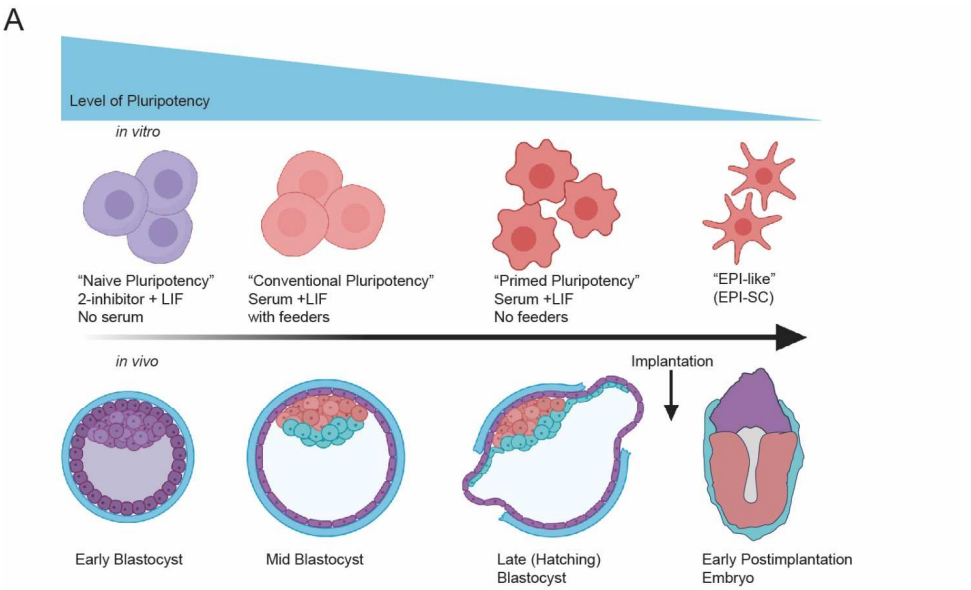
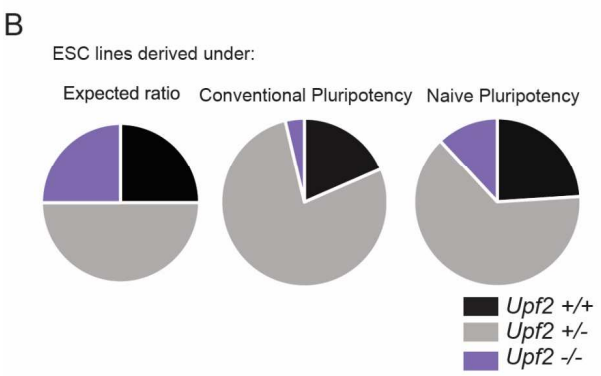
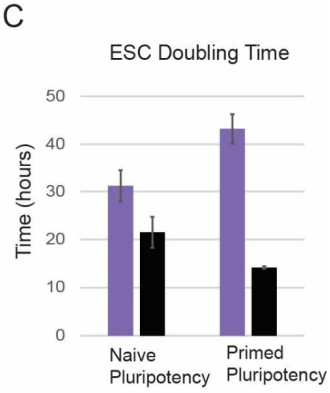


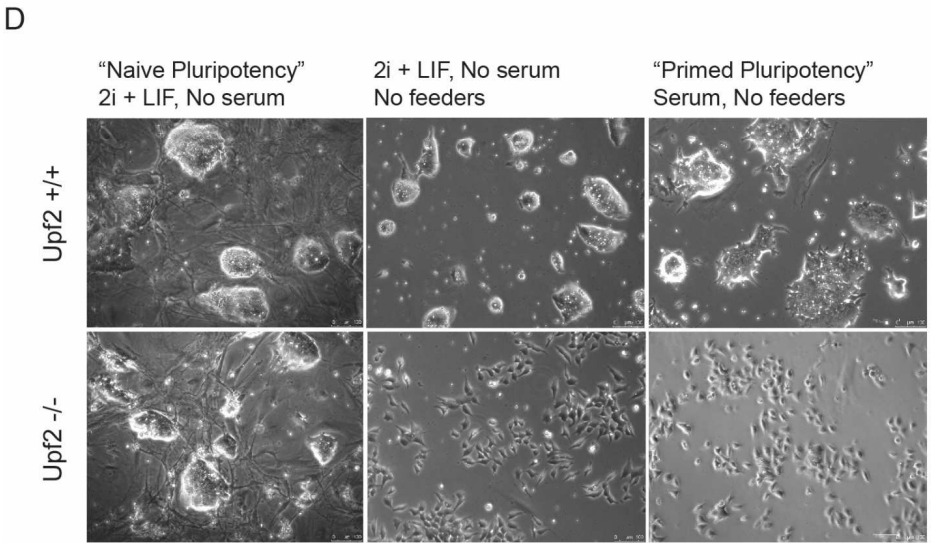
Figure 3.7. *Upf2* downregulates ICM- and EPI-specific mRNAs, resulting in a reduction in EPI cells in the mouse mid-blastocyst. (A) The ratio of the normalized reads in *Upf2*-null relative to control embryos is plotted for genes identified as either EPI- (n=50 genes) or PE-specific (n=77 genes) by Ohnishi et al. (Ohnishi et al., 2013), or ICM- (n=317 genes) or TE-specific (n=128 genes) by Blakeley et al. (Blakeley et al., 2015). A ratio of one indicates no change between Upf2-null embryos and controls. A ratio less than one indicates lower expression in Upf2-null embryos, as is observed for EPI-specific genes ($P = 0.0005$) and ICM-specific genes ($P < 0.01$) at E3.25. At E3.5, EPI- and ICM-specific genes tend to be lower in *Upf2*-null blastocysts as well. (B) After normalization of all lineage-specific genes to a panel of 27 housekeeping genes, EPI- and ICM-specific genes remain significantly lower at E3.25; however, normalization corrects the deviation in the ratio of EPI- and ICM-specific gene expression at E3.5, suggesting that the lower gene expression observed at this stage was due to a lower number of cells.

Figure 3.8. ESCs devoid of *Upf2* can be derived under naïve pluripotency conditions, but exhibit delayed growth and an inefficiency to transition to primed pluripotency. (A) Diagram of varying levels of pluripotency achieved *in vitro,* as well as their *in vivo* counterparts. Media containing 2-inhibitors with LIF and feeders represents the most immature or naïve state, and most closely resembles the ICM of the mouse blastocyst. Slightly more primed states can be achieved by removing the feeder layer, by changing the media to serum with LIF, and by then removing the feeders from the serum-containing media, respectively. (B) *Upf2*-null ESC lines can be derived at a higher rate under naïve pluripotency conditions, as compared to normal ESC derivation conditions. (C) *Upf2*-null ESCs exhibit slowed growth, as compared to littermate control ESC lines. (D) Brightfield images of *Upf2*-null ESC and control ESC lines in naïve pluripotency culture conditions and transitioned from naïve conditions into slightly more primed conditions (with 2-inhibitor and LIF, but without feeders, and in serum plus LIF but without feeders). In both cases, *Upf2*-null lines exhibit morphological changes associated with differentiation.


A
Level of Pluripotency
in vitro
"Naive Pluripotency"
2-inhibitor + LIF
No serum
"Conventional Pluripotency"
Serum +LIF
with feeders
"Primed Pluripotency"
Serum +LIF
No feeders
"EPI-like"
(EPI-SC)
in vivo
Implantation
Early Blastocyst
Mid Blastocyst
Late (Hatching)
Blastocyst
Early Postimplantation
Embryo


B
ESC lines derived under:
Expected ratio
Conventional Pluripotency
Naive Pluripotency
Upf2 +/+
Upf2 +/-
Upf2 -/-


C
ESC Doubling Time
Time (hours)
50
40
30
20
10
0
Naive
Pluripotency
Primed
Pluripotency


D
"Naive Pluripotency"
2i + LIF, No serum
2i + LIF, No serum
No feeders
"Primed Pluripotency"
Serum, No feeders
Upf2 +/+
Upf2 -/-

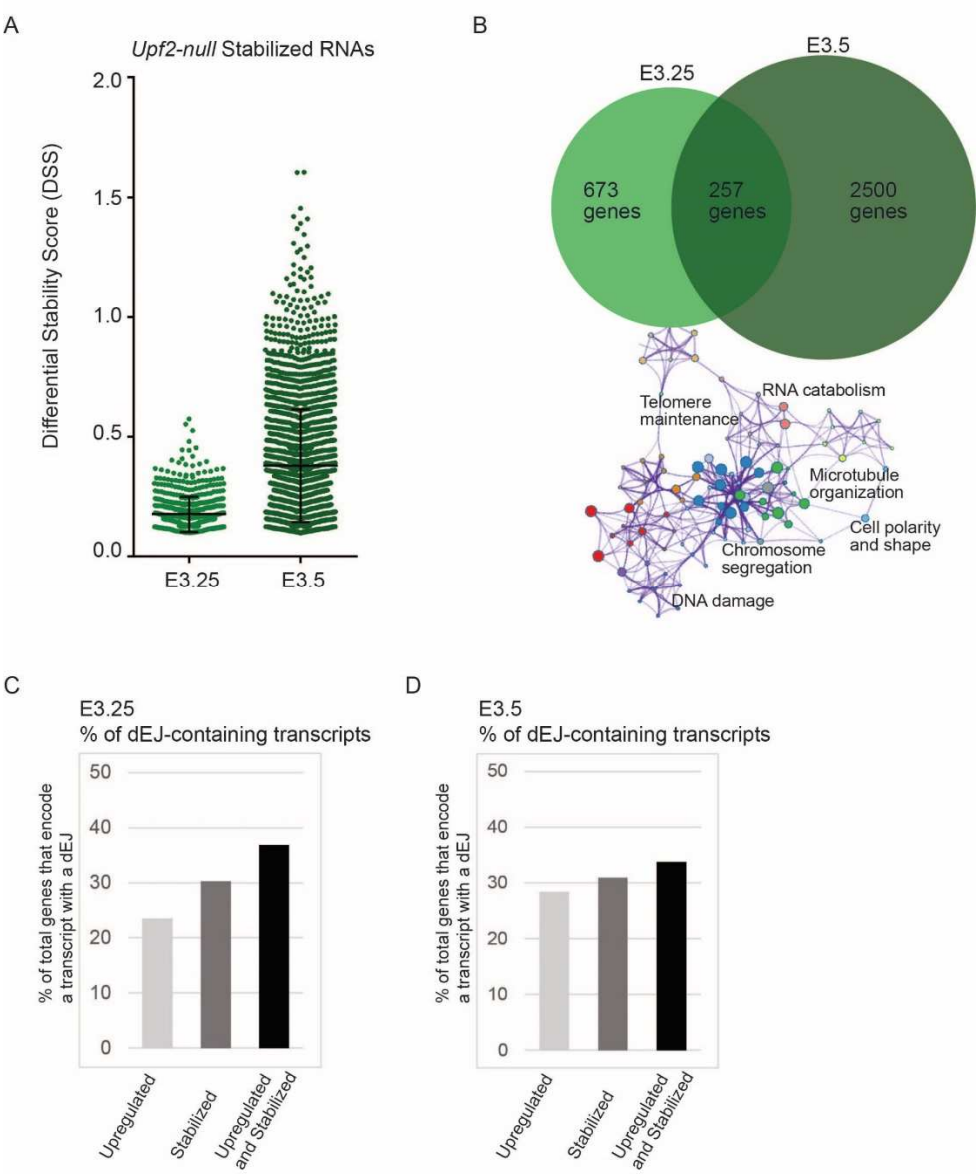


Figure 3.9. mRNAs stabilized in the absence of UPF2 in the mouse preimplantation embryo. (A) All mRNAs stabilized in *Upf2*-null blastocysts as compared to control blastocysts at E3.25 and E3.5, presented as a dot plot of the differential stability score (DSS). Many more mRNAs were differentially stabilized at the mid (E3.5) blastocyst stage, as compared to the early blastocyst stage (E3.25). (B) Venn diagram showing mRNAs stabilized in *Upf2*-null blastocysts, and how many are common to both blastocyst stages. A cytoscape network of the GO categories of commonly-stabilized mRNAs demonstrates that these likely direct target mRNAs regulate functions such as telomere maintenance, DNA damage, and RNA catabolism. (C) The percent of the total genes that encode a transcript containing a dEJ is presented as a bar graph for genes identified to be upregulated, stabilized, or both, upon loss of *Upf2* in E3.25 blastocysts. (D) As in Figure 3.9C, but at E3.5.

References

Alkallas, R., Fish, L., Goodarzi, H., and Najafabadi, H.S. (2017). Inference of RNA decay rate from transcriptional profiling highlights the regulatory programs of Alzheimer's disease. *Nat. Commun.* 8.

Alonso, C.R. (2011). A complex "mRNA degradation code" controls gene expression during animal development. *Trends Genet*. 28, 78–88.

Alonso, C.R., and Akam, M. (2003). A Hox gene mutation that triggers nonsense-mediated RNA decay and affects alternative splicing during Drosophila development. *Nucleic Acids Res.* 31, 3873–3880.

Anastasaki, C., Longman, D., Capper, A., Patton, E.E., and Caceres, J.F. (2011). Dhx34 and Nbas function in the NMD pathway and are required for embryonic development in zebrafish. *Nucleic Acids Res.* 39, 3686–3694.

Bao, J., Tang, C., Yuan, S., Porse, B.T., and Yan, W. (2015). UPF2, a nonsense-mediated mRNA decay factor, is required for prepubertal Sertoli cell development and male fertility by ensuring fidelity of the transcriptome. *Development* 142, 352–362.

Bedzhov, I., Graham, S.J.L., Leung, C.Y., and Zernicka-Goetz, M. (2014). Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369, 1–14.

Blakeley, P., Fogarty, N.M.E., Valle, I., Wamaitha, S.E., Hu, X., Elder, K., Snell, P., Christie, L., and Robson, P. (2015). Defining the Three Cell Lineages of the Human Blastocyst by Single-Cell RNAseq. *Development* 142, 3151–3165.

Bruno, I.G., Karam, R., Huang, L., Bhardwaj, A., Lou, C.H., Shum, E.Y., Song, H.-W., Corbett, M.A., Gifford, W.D., Gecz, J., Pfaff, S.L., and Wilkinson, M.F. (2011). Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. *Mol. Cell* 42, 500–510.

Casadio, A., Longman, D., Hug, N., Delavaine, L., Vallejos Baier, R., Alonso, C.R., and Caceres, J.F. (2014). Identification and characterization of novel factors that act in the nonsense-mediated mRNA decay pathway in nematodes, flies and mammals. *EMBO Rep.* 16, 71–78.

Chan, W.-K.K., Bhalla, A.D., Le Hir, H., Nguyen, L.S., Huang, L., Gecz, J., Wilkinson, M.F., and Gécz, J. (2009). A UPF3-mediated regulatory switch that maintains RNA surveillance. *Nat. Struct. Mol. Biol.* 16, 747–753.

Chan, W.K., Huang, L., Gudikote, J.P., Chang, Y.F., Imam, J.S., MacLean 2nd, J.A., and Wilkinson, M.F. (2007). An alternative branch of the nonsense-mediated decay pathway. *EMBO J* 26, 1820–1830.

Chang, Y.-F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* 76, 51–74.

Chazaud, C., and Yamanaka, Y. (2016). Lineage specification in the mouse preimplantation embryo. *Development* 143, 1063–1074.

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinformatics* 51, 11.14.1-11.14.19.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

Dumdie, J.N., Cho, K., Ramaiah, M., Skarbrevik, D., Mora-Castilla, S., Stumpo, D.J., Lykke-Andersen, J., Laurent, L.C., Blackshear, P.J., Wilkinson, M.F., and Cook-Andersen, H. (2018). Chromatin Modification and Global Transcriptional Silencing in the Oocyte Mediated by the mRNA Decay Activator ZFP36L2. *Dev. Cell* 44, 392–402.e7.

Eckert, J.J., Velazquez, M.A., and Fleming, T.P. (2015). Cell Signaling During Mammalian Early Embryo Development.

Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. *EMBO J.* 22, 3960–3970.

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol. Cell* 12, 1439–1452.

Huang, L., Shum, E.Y., Jones, S.H., Lou, C.-H., Dumdie, J., Kim, H., Roberts, A.J., Jolly, L.A., Espinoza, J.L., Skarbrevik, D.M., Phan, M.H., Cook-Andersen, H., Swerdlow, N.R., Gecz, J., and Wilkinson, M.F. (2018). A Upf3b-mutant mouse model with behavioral and neurogenesis defects. *Mol. Psychiatry* 23, 1773–1786.

Hurt, J.A., Robertson, A.D., and Burge, C.B. (2013). Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Res.* 23, 1636–1650.

Hwang, J., and Maquat, L.E. (2011). Nonsense-mediated mRNA decay (NMD) in animal embryogenesis: to die or not to die, that is the question. *Curr. Opin. Genet. Dev.* 21, 422–430.

Johansson, M.J.O., He, F., Spatrick, P., Li, C., and Jacobson, A. (2007). Association of yeast Upf1p with direct substrates of the NMD pathway. *Proc. Natl. Acad. Sci. U. S. A.* 104, 20872–20877.

Karam, R., Lou, C.-H., Kroeger, H., Huang, L., Lin, J.H., and Wilkinson, M.F. (2015). The unfolded protein response is shaped by the NMD pathway. *EMBO Rep*. 16, 599–609.

Lelivelt, M.J., and Culbertson, M.R. (1999). Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol. Cell. Biol.* 19, 6710–6719.

Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci. Rep.* 7, 4200.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Li, T., Shi, Y., Wang, P., Guachalla, L.M., Sun, B., Joerss, T., Chen, Y.-S., Groth, M., Krueger, A., Platzer, M., Yang, Y.-G., Rudolph, K.L., and Wang, Z.-Q. (2015). Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. *EMBO J.* 34, 1630–1647.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.

Longman, D., Plasterk, R.H.A., Johnstone, I.L., and Cáceres, J.F. (2007). Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. *Genes Dev.* 21, 1075–1085.

Lorthongpanich, C., and Issaragrisil, S. (2015). Emerging Role of the Hippo Signaling Pathway in Position Sensing and Lineage Specification in Mammalian Preimplantation Embryos. *Biol. Reprod.* 92, 143.

Lou, C., Dumdie, J., Goetz, A., Shum, E., Brafman, D., Liao, X., Mora-Castilla, S., Ramaiah, M., Cook-Andersen, H., Laurent, L., and Wilkinson, M. (2016). Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate. *Stem Cell Reports* 6, 844–857.

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

Mao, H., Pilaz, L.-J., McMahon, J.J., Golzio, C., Wu, D., Shi, L., Katsanis, N., and Silver, D.L. (2015). Rbm8a haploinsufficiency disrupts embryonic cortical development resulting in microcephaly. *J. Neurosci.* 35, 7003–7018.

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

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 40, 4288–4297.

McIlwain, D.R., Pan, Q., Reilly, P.T., Elia, A.J., McCracken, S., Wakeham, A.C., Itie-Youten, A., Blencowe, B.J., and Mak, T.W. (2010). Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. *Proc. Natl. Acad. Sci. U. S. A.* 107, 12186–12191.

Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M., and Dietz, H.C. (2001). Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* 10, 99–105.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* 36, 1073–1078.

Metzstein, M.M., and Krasnow, M.A. (2006). Functions of the Nonsense-Mediated mRNA Decay Pathway in Drosophila Development. *PLoS Genet.* 2, e180.

Mora-Castilla, S., To, C., Vaezeslami, S., Morey, R., Srinivasan, S., Dumdie, J.N., Cook-Andersen, H., Jenkins, J., and Laurent, L.C. (2016). Miniaturization Technologies for Efficient Single-Cell Library Preparation for Next-Generation Sequencing. *J. Lab. Autom.* 21(4), 557-67.

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

Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., Oleś, A.K., Araúzo-Bravo, M.J., Saitou, M., Hadjantonakis, A.-K., and Hiiragi, T. (2013). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol.* 16, 27–37.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* 27, 3769–3779.

Popp, M.W.-L., and Maquat, L.E. (2013). Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet*. 47, 139–165.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.

Sakaki-Yumoto, M., Kobayashi, C., Sato, A., Fujimura, S., Matsumoto, Y., Takasato, M., Kodama, T., Aburatani, H., Asashima, M., Yoshida, N., and Nishinakamura, R. (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* 133, 3005–3013.

Schweingruber, C., Rufener, S.C., Zünd, D., Yamashita, A., and Mühlemann, O. (2013). Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim. Biophys. Acta* 1829, 612–623.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504.

Shum, E.Y., Jones, S.H., Shao, A., Dumdie, J., Krause, M.D., Chan, W.-K., Lou, C.-H., Espinoza, J.L., Song, H.-W., Phan, M.H., Ramaiah, M., Huang, L., McCarrey, J.R., Peterson, K.J., De Rooij, D.G., Cook-Andersen, H., and Wilkinson, M.F. (2016). The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. *Cell* 165, 382–395.

Silver, D.L., Watkins-Chow, D.E., Schreck, K.C., Pierfelice, T.J., Larson, D.M., Burnetti, A.J., Liaw, H.-J., Myung, K., Walsh, C.A., Gaiano, N., and Pavan, W.J. (2010). The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat. Neurosci.* 13, 551–558.

Tani, H., Imamachi, N., Salam, K.A., Mizutani, R., Ijiri, K., Irie, T., Yada, T., Suzuki, Y., and Akimitsu, N. (2012). Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol.* 9, 1370–1379.

Tarpey, P.S., Raymond, F.L., Nguyen, L.S., Rodriguez, J., Hackett, A., Vandeleur, L., Smith, R., Shoubridge, C., Edkins, S., Stevens, C., O'Meara, S., Tofts, C., Barthorpe, S., Buck, G., Cole, J., Halliday, K., Hills, K., Jones, D., Mironenko, T., et al. (2007). Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. *Nat. Genet.* 39, 1127–1133.

Thoren, L.A., Nørgaard, G.A., Weischenfeldt, J., Waage, J., Jakobsen, J.S., Damgaard, I., Bergström, F.C., Blom, A.M., Borup, R., Bisgaard, H.C., and Porse, B.T. (2010). UPF2 Is a Critical Regulator of Liver Development, Function and Regeneration. *PLoS One* 5, e11650.

Vicente-Crespo, M., and Palacios, I.M. (2010). Nonsense-mediated mRNA decay and development: shoot the messenger to survive? *Biochem. Soc. Trans.* 38, 1500–1505.

Weinberger, L., Manor, Y.S., and Hanna, J.H. (2015). Stem cell states: naive to primed pluripotency. *Nat. Rev.*

Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Mönch, K., Thoren, L.A., Nielsen, F.C., Jacobsen, S.E.W., Nerlov, C., and Porse, B.T. (2008). NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev*. 22, 1381–1396.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).

Zhang, X., Zhang, J., Wang, T., Esteban, M.A., and Pei, D. (2008). Esrrb Activates Oct4 Transcription and Sustains Self-renewal and Pluripotency in Embryonic Stem Cells. *J Biol Chem* 283(51), 35825-33.

Chapter 4: Implantation and early pregnancy success of the human blastocyst is associated with distinct gene expression patterns in the inner cell mass

Abstract

Recent advances in genome-wide analysis with low input have rapidly expanded our knowledge of gene expression patterns in early human embryo development. A significant barrier to further advances, however, is a lack of qualitative analyses to identify which factors among the thousands of genes expressed might be critical for a successful pregnancy. Here, we provide an in-depth assessment of gene expression and cellular communication networks within the major compartments of the human blastocyst that are most closely associated with successful implantation. To accomplish this, we leverage cumulative clinical data in which embryos graded by morphology demonstrate significantly different pregnancy potential. Transcriptome profiling, combined with a machine learning strategy designed to deconvolute complex admixtures of cells, demonstrates that defects in the establishment and/or maintenance of the primitive endoderm lineage—following the second major embryonic differentiation event—most strongly differentiates embryos of high and low implantation potential. Delineation of cellular communication patterns further demonstrates that embryos of low implantation potential lack many of the structural and extracellular matrix interactions associated with formation of the primitive endoderm and identifies signaling pathways specifically enriched in embryos of high implantation potential. Using an unbiased machine learning approach, we demonstrate that gene expression data from the inner cell mass is the best predictor of implantation potential and we identify genes within each embryo compartment that are most closely associated with human blastocyst implantation and ongoing pregnancy. Our data supports a model in which successful implantation and ongoing pregnancy predominantly depends upon the inner cell mass and points to a previously unappreciated role for the extraembryonic primitive endoderm in the earliest stages of establishing successful pregnancy in humans.

Introduction

Successful embryo implantation is overall an inefficient process in humans for reasons that remain largely unexplained. Even in fertile couples, implantation failure accounts for 75% of all lost pregnancies and in couples undergoing treatment for infertility, lack of clinically detectible implantation accounts for up to 70% of pregnancy losses for euploid embryo transfers in *in vitro* fertilization (IVF) (Wilcox et al., 1988). A better understanding of the molecular determinants of a "good" embryo and the molecular mechanisms required for successful implantation will unquestionably lead to improvements in fertility treatments. However, these studies have been limited by technological and ethical limitations on research with human embryos, restrictions on federal funding, and the limited availability of normal embryos and associated cells for study (Kalista et al., 2011). At present, studies have not been conducted on embryos of known ploidy and quality (Blakeley et al., 2015; Petropoulos et al., 2016), limiting the functional conclusions that can be drawn from them. Here, we overcome some of these limitations. We leverage cumulative clinical data to identify groups of embryos of disparate morphology with significantly different implantation and pregnancy potential and leverage the differential gene expression analysis from euploid embryos of high and low implantation potential to identify factors likely important for implantation and early success, as well as factors that may contribute to poor outcomes.

Prior to implantation, the blastocyst has completed two successive differentiation events, forming three mature cell lineages, the TE, the primitive endoderm (PrE), and the epiblast (EPI). The first differentiation event becomes evident at cavitation, when the outer TE encloses the ICM and the blastocoele cavity is formed. In the second differentiation event, the ICM undergoes an additional round of differentiation to generate the EPI, which will undergo gastrulation and form the embryo proper, and the PrE, which forms an epithelial layer surrounding the EPI and goes on to form the yolk sac (Niakan et al., 2012). Studies in the

mouse have demonstrated that development of the blastocyst and differentiation of all three lineages requires extensive cross-talk between embryonic compartments. The Fibroblast Growth Factor (FGF) pathway is essential in mice for the establishment of the PrE (Chazaud et al., 2006; Chazaud and Yamanaka, 2016; Guo et al., 2010; Yamanaka et al., 2010) and additionally contributes to trophoblast stem cell maintenance in the polar TE (Ralston and Rossant, 2006; Tanaka et al., 1998). In the mouse blastocyst, Bone Morphogenic Protein (BMP) signaling is important for development of the TE and PrE (Graham 2014) and, from the TE, is necessary for proliferation of the ICM (Murohashi et al., 2009). However, while some of the same transcription factors and signaling pathways are commonly expressed in the human blastocyst, key differences have been observed. For instance, inhibition of FGF signaling does not affect formation of the PrE in the human blastocyst (Kuijk et al., 2012; Roode et al., 2012), and expression of both CDX2 and OCT4 (lineage-establishing factors in the mouse for the TE and ICM) are protracted, only becoming restricted to the TE and ICM, respectively, after cavitation (Cauffman et al., 2009; Chen et al., 2009). What pathways and factors act in place of these in the human embryo remains unknown.

At implantation, the extra-embryonic TE initiates contact with the maternal endometrium and differentiates to give rise to all the trophoblast cell lineages of the placenta. However, the process of embryo implantation also varies between mammalian species, with human and bovine blastocysts adhering to the endometrium via the polar trophoblast and mouse blastocysts adhering by the mural trophoblast (Lindenberg et al., 1989). Thus, despite a wealth of information on embryo development in mice and other mammals, our knowledge of the pathways and factors critical for human embryo development and implantation is severely limited.

Here, we identify patterns of gene expression and characterize the complete repertoire of cell-to-cell communication patterns within and between the main compartments of the

blastocyst in human embryos of high implantation potential, unveiling signaling pathways and structural connections most closely associated with successful implantation in humans. Surprisingly, we show that embryos of low implantation potential demonstrate a selective failure to complete or maintain formation of the PrE, suggesting a novel role for this extra-embryonic lineage in establishing developmental potential of the preimplantation human blastocyst. We demonstrate that embryos of low implantation potential lack a large number of structural and extracellular matrix interactions associated with formation of the PrE, further solidifying an essential role for this lineage in the formation of a competent human blastocyst with the potential to successfully implant in the uterus and commence a pregnancy. Using an unbiased machine learning approach, we identify gene expression profiles in the human embryo that most closely predict early ongoing pregnancy success. Our study supports a model in which implantation success and early ongoing pregnancy is ultimately more dependent on the quality of the ICM, particularly the extraembryonic PrE. Together, this data represents a rich resource for the field, expanding upon factors previously identified to be important in the human blastocyst, and enhancing our understanding of communication between cell types within the embryo. This study, and further evaluation of the factors identified within, has the potential to improve treatment for infertile couples by advancing methods for embryo culture and/or selection of the embryo with the greatest implantation potential for transfer.

Methods

Clinical outcomes assessment

After embryo transfer, the clinical pregnancy outcome was tracked and recorded as *Negative Pregnancy Test* (human chorionic gonadotrophin (β-hCG) < 5 mIU/ml, 11 days after embryo transfer)*, Biochemical Pregnancy* (β-hCG > 5 mIU/ml (although typically <100) that does not rise normally, with no evidence of ectopic pregnancy)*, Spontaneous Abortion*

(pregnancy documented by ultrasound at 6 weeks but lost before 20 weeks), or *Successful Pregnancy* (live birth).

Embryo culture and morphological assessment

Human embryos were obtained from two cohorts from the Reproductive Partners Fertility Center in La Jolla, CA, with ethical approval from the ethics board and all required patient consent measures followed.

Embryos were obtained from ICSI-fertilized oocytes. They were cultured under standard conditions as performed in the IVF Clinic (5%CO2/5%O2 in IVF media and covered in mineral oil). Each embryo underwent Prenatal Genetic Screening (PGS) before freezing using standard lab protocol. Embryos were thawed according to standard protocols and cultured in G-PGD media (Vitrolife) and allowed to expand overnight. All embryos included were graded based on morphological quality (Figure 4.2A-B) by an experienced embryologist at the time of trophectoderm biopsy and vitrification, and prenatal genetic screening (PGS) was conducted to rule out chromosomal loss or large deletions. An embryo with either an ICM or a TE that was assigned a score of A was considered “Good,” whereas a score of C for either lineage was defined as “Poor” (Figures 4.1 and 4.2A-B). Only euploid, singleton transfers were included in this retrospective analysis. Embryos transferred in multiples were only included if a single embryo could be identified by gender.

An experienced embryologist used laser dissection to separate the mural TE (mTE) from the ICM compartment. The mTE and ICM samples were individually washed in TE Wash Buffer (1mg/ml bovine serum albumin in 1XPBS), transferred to individual 0.2ml PCR tubes, and placed immediately on dry ice.

RNA sequencing library preparation

In total, six HP and eight LP expanded human blastocysts were used for this study. RNA isolation and cDNA synthesis were carried out using the SMARTer® Ultra® Low Input RNA Kit for Sequencing – v4 (Clontech, Cat.no. 634888). Libraries were generated from the resulting cDNA (0.2ng/ul per sample) using the Nextera XT DNA library preparation kit (Illumina, Cat.no. FC-131-1024) as previously described (Mora-Castilla et al., 2016). Indexed sequence libraries were pooled for multiplexing, normalized by MiSeq read number, and paired-end sequencing was performed on a HiSeq 4000.

Analysis of transcriptomic data

Approximately 18.1 million reads were generated per sample, and 70.1% of those were uniquely mapped (Figure 4.2C) via STAR (2.5.2a) (Dobin et al., 2013; Dobin and Gingeras, 2015) after trimming for adapter sequences via cutadapt (v1.10) (Martin, 2011). Samtools (Li et al., 2009) was used to process sam files, as well as to sort and remove PCR duplicates of bam files. Counts for each gene were quantified using the Subread package FeatureCounts using the gene level quantification in paired-end mode (release 1.5.2) (Liao et al., 2014), and annotated using the Ensembl GRCh38 genome. Reads were filtered such that genes without at least one sample with at least 10 raw reads were removed from the analysis. Count data was normalized to TPM using the Bioconductor package edgeR (McCarthy et al., 2012; Robinson et al., 2010). The betweenLaneNormalization function of the Bioconductor package EDASeq (Risso et al., 2011) was used to normalize the data using upper-quartile (UQ) normalization, and differential expression was calculated using DESeq2 (Love et al., 2014) based on a model using the negative binomial distribution (Figure 4.2D). No bias in library complexity or percent of uniquely mapped reads per sample was observed between ICM and TE samples, or between HP and LP embryos (Figure 4.2C). Genes with a P-value less than 0.01 were considered differentially expressed unless otherwise noted. Heatmaps were constructed using the

heatmap.2 function in ggplot2 (Wickham, 2016). Principle components were calculated with the prcomp function. All box plots, scatter plots, PCA plots, and heatmaps were constructed using ggplot2. All of these analyses and plot constructions were performed with RStudio, R (v3.2.2 and v3.4.0). Gene ontology analysis was performed with the online Metascape platform (Tripathi 2015 Cell host and microbe) and gene enrichment networks were visualized with Cytoscape (Shannon et al., 2003).

Supplementary tables of normalized read counts and calculated differential expression from Petropoulos et al. (Sophie Petropoulos et al., 2016) and Blakeley et al. (Blakeley et al., 2015) were used for the comparison of lineage-specific genes. Single-cell gene expression raw reads from Petropoulos et al. (Sophie Petropoulos et al., 2016) were used for lineage-specific marker identification, proportional composition analysis (CIBERSORT) (Newman et al., 2015), and cell compartment interaction analysis (CellPhoneDB) (Vento-Tormo et al., 2018). For these analyses, genes were filtered such that at least 50 single-cell samples must have a raw read of at least 10. CIBERSORT is a computational framework that can be used to quantify the proportion of cell types within a complex gene expression mixture. Cell types were defined using embryonic day (E)6 and E7 EPI, PrE, and polar TE samples from Petropoulos et al., and CIBERSORT was run using 1000 permutations with quantile normalization disabled. The relative level of each cell type within HP and LP ICM samples is presented, along with the Pearson correlation coefficient (R), which is generated by comparing the original mixture with the predicted input training mixture. CellPhoneDB (Vento-Tormo et al., 2018) was used to assess cellular cross-talk between cell types of the ICM and TE within our data and between the EPI, PrE, pTE, and mTE in the Petropoulos et al. dataset. This is done by deriving enriched ligand-receptor interactions between cell types from a repository of curated interactions. CellPhoneDB works by first permuting the cell type labels 1000 times to generate a null distribution for each receptor-ligand interaction; this represents all of the interactions considering

a random clustering. By then calculating what proportion of the mean values deviate from the mean, interactions enriched between two cell types can be statistically determined. Only significant interactions ($P < 0.05$) were considered in this analysis. Resulting interaction plots were created using the chordDiagram function of the circlize (v 0.4.5) package (Gu et al., 2014).

Multiple machine learning approaches were used to generate short lists of genes that could predict morphological grade given gene expression profiles. We applied a genetic algorithm to all approaches to modify and optimize individual solutions to generate a best list. Machine learning approaches used included Least Square Regression (LSR), Support Vector Machine (SVM), and Maximum Likelihood Estimation (MLE). LSR minimizes the error such that the sum of all squared error is minimized during linear regression. SVM determines a hyperplane or decision boundary that will distinctly classify the data points, and MLE is used to estimate parameters such that the probability of obtaining the observed data is maximized. All three approaches were run with a genetic algorithm with 1000 iterations, and the top ten gene lists (picked based on the smallest error in binary classification) were manually curated. First, we determined if expression of each gene list could classify HP vs LP embryos based on hierarchical clustering; we then totaled the percent of the top ten iterations for each approach that could classify embryos by clustering. This was done with complete gene expression data and with differential expression data (HP versus LP embryos; $Q<0.1$, Fold change > 1.3). For analyses with differential gene expression data as input, all gene lists that classified embryos based on hierarchical clustering were compiled to generate a “best list”, which only included genes that had been identified numerous times. By combining these “best lists” from all three algorithms, we arrived at a short list of genes predicted to classify HP versus LP embryos from gene expression data. We used gene expression data from (1) TE samples, (2) ICM samples, and from (3) TE and ICM samples combined, to determine which compartment of the embryo was more predictive in classifying HP versus LP embryos.

Results

Morphology of euploid blastocysts correlates with successful implantation and ongoing pregnancy

Preimplantation genetic screening (PGS) in combination with blastocyst morphological quality assessment, although imperfect, continues to be our most reliable and widely used method of predicting a successful pregnancy in assisted reproductive technology (Gardner et al., 2000). We retrospectively analyzed pregnancy success rates of 678 euploid blastocyst embryos following single embryo transfer in *in vitro* fertilization (IVF) cycles over a 5-year period in our IVF clinic to identify blastocysts of different pregnancy potential. Blastocysts designated as euploid by PGS with a composite grade of "good" [by standard morphological assessment of the trophectoderm (TE) and inner cell mass (ICM) (Gardner et al., 2000); also called "high-potential (HP")] resulted in a live birth in 65% of transfers (of 351 HP embryos transferred) (Figure 4.1A and 4.2A). In contrast, euploid blastocysts with a composite grade of "poor" by morphological assessment [also called "low-potential (LP")] resulted in a live birth in only 30% of transfers (of 74 LP embryos transferred; P=0.003) (Figure 4.1). This confirmed that morphological grade correlates with pregnancy success.

To look more closely at how the morphological grade of a blastocyst correlated with pregnancy success or failure, we compared the incidence of three pregnancy failure outcomes (negative pregnancy test, biochemical pregnancy, and spontaneous abortion, as defined below) between HP and LP embryo transfers. Shortly after embryo implantation occurs, syncytiotrophoblast cells of the conceptus release human Chorionic Gonadotropin (hCG), which can be confidently detected in maternal serum 6-14 days after fertilization (Makrigiannakis et al., 2017; Nepomnaschy et al., 2008; Wilcox et al., 1988). At present, the sensitivity of serum β-hCG testing limits our ability to diagnose peri-implantation failure more narrowly. Thus, we define clinical "implantation failure" as no hCG production detected by maternal serum testing 11 days

after embryo transfer, corresponding to day 16 after fertilization (β-hCG < 5 mIU/ml, Negative pregnancy test). This criteria identifies embryos for which implantation either did not begin or failed in the earliest stages before detection of hCG is reliable. Using this as a measure of clinical implantation, we found no clinically detectable evidence of embryo implantation for 49% of LP embryos, whereas only 15% of HP embryos exhibited implantation failure (P=0.002; Figure 4.1C). In contrast, other measured causes of pregnancy failure were not significantly different between HP and LP embryo groups, including biochemical pregnancy (transient detection of β-hCG, but abnormally low for a healthy pregnancy; 11% HP v. 16% LP, P=0.71) and spontaneous abortion (fetal loss at 6-20 weeks; 9% HP v. 6% LP, P=0.83). Because there was not a significant difference in the incidence of these other failure outcomes, we reasoned that the lower rates of pregnancy in LP embryos predominantly reflected lower rates of implantation. Thus, the molecular differences between these two morphological categories of euploid embryos are likely responsible for defects in establishing a successful pregnancy.

To begin to probe the relative contributions of the TE and ICM in successful implantation and pregnancy, we reanalyzed these same clinical embryo transfer data based on the individual grades assigned to the TE and ICM (A, B or C, where A is high and C is low) (Gardner et al., 2000) instead of by the composite grade as above. A low TE grade was associated with successful pregnancy in 38% of embryos transferred compared to only 16% for embryos with a low ICM grade (P=0.07; Figure 4.1D). Similarly, failed implantation occurred following transfer of 39% of embryos with a low TE grade compared to 63% of embryos with a low ICM grade (P=0.07; Figure 4.1D). Because higher grade embryos are preferentially transferred, our analysis is limited by the lower number of low grade embryos that were transferred (n=66, low TE; n=19, low ICM). However, these data demonstrate a trend and suggest that factors within the ICM play an important and early role in the successful implantation of the human embryo.

Gene expression profiles of the inner cell mass and trophectoderm from euploid human blastocysts of known high implantation potential.

To define genes most closely associated with successful implantation and pregnancy, we examined and compared the transcriptome profiles of single HP or LP embryos graded in the clinic by standard morphologic criteria as above. All embryos analyzed were euploid and at the same developmental stage, graded as hatching blastocysts on day 5-6 following fertilization, the stage immediately prior to embryo transfer and subsequent implantation (Table S1). To ensure all cells in each embryo were accurately and proportionally represented, the embryo was separated into its two major compartments – the TE compartment (comprising mural TE (mTE)) and ICM compartment (comprising EPI, PrE and polar TE (pTE)) – by laser dissection. Each compartment was sequenced and analyzed separately using established low-input RNA sequencing protocols (Figure 4.3A) (Mora-Castilla et al., 2016; Ramsköld et al., 2012). ICM and TE compartments and HP and LP embryos were easily distinguished by unbiased Principle Component Analysis (PCA) of whole transcriptome count data (Figure 4.3B). ICM and TE samples from both HP and LP embryos were enriched for the expected subset of established markers for mTE, pTE, EPI and/or PrE (Petropoulos et al., 2016) (Figure 4.3C, 4.4A-C).

Analysis of HP blastocysts provided the opportunity to determine the gene expression profile expected of “good quality” euploid embryos of high implantation and pregnancy potential. Genes identified by deep sequencing significantly overlapped with published single cell datasets (Blakeley et al., 2015; Petropoulos et al., 2016) (Figure 4.3D, S3A-B), but we were also able to significantly expand the number of genes that demonstrate enriched expression in the TE or ICM compartment of the human blastocyst (Figure 4.3D, 4.5A-B). In the TE compartment, we identified 170 genes significantly upregulated in HP TE relative to HP ICM (log2 fold change>1, Q<0.05; Table S2), of which 115 (68%) have not been previously reported as enriched in human TE (Figure 4.5A). Genes with enriched expression in HP TE include those important for

lipid and amino acid metabolism, as previously reported (Altmä et al., 2012; Gardner et al., 2015, 2001), as well as cell response to hormones, and transmembrane transport (Figure 4.5C). Similarly, in the ICM compartment, we found 677 genes significantly enriched in HP ICM relative to HP TE (log2 fold change>1, Q<0.05; Table S2), of which 497 (73%) have not been previously reported as enriched in human ICM (Figure 4.5B). These genes include some postulated to be important for EPI and PrE specification, as well as other signaling components (Blakeley et al., 2015; Petropoulos et al., 2016). Interestingly, these genes encode factors with known roles in pluripotency-related signaling, Wnt signaling, and the MAPK pathway, as well as a significant number of factors that function in cell adhesion and extracellular structure organization (Figure 4.5D). In fact, by applying MCODE, a clustering algorithm designed to predict protein-protein interaction networks from gene expression data (Bader and Hogue, 2003), inferred protein networks enriched in HP ICM included those integral to establishing a structural niche within the ICM, such as ECM organization, integrin signaling, and the formation of collagen (Figure 4.5E).

HP trophectoderm is characterized by enhanced metabolic activity

To ask what factors and pathways in the TE compartment are associated with successful implantation and pregnancy, we compared TE gene expression in HP and LP embryos. Gene expression in mTE samples was highly correlated overall (R^2 = 0.98; Figure 4.6A), as was the expression of 98 previously identified lineage-specific genes (Petropoulos et al., 2016) (R^2=0.96; Figure 4.6A). A total of 387 genes ((3.5% of the 11,050 genes with detectable expression) were differentially expressed (P<0.01); 89 genes (23%) were expressed at higher levels in HP TE while the remaining 298 differentially expressed genes (77%) were expressed at higher levels in LP TE.

Genes upregulated in HP TE were highly enriched for those encoding factors that regulate multiple aspects of embryonic metabolism. In fact, almost half (n=40) of the genes enriched in HP TE were assigned to a ‘metabolism’ category on GO analysis that has been

extensively studied in embryo development (Figure 4.7A). Our data demonstrates that HP TE is enriched for genes encoding carbohydrate and pyruvate metabolism factors, including members of the glycolysis pathway (*PFKM, LDHC*) and other carbohydrate metabolism components (*DCXR, GALE, SORD),* as well as downstream components of the electron transport chain (*PDSS1, ATP5MC1*). These findings are consistent with published data demonstrating a positive relationship between glucose uptake and successful implantation of human blastocysts (Gardner et al., 2011). In addition, we observed a number of genes encoding factors enriched in HP TE that are important for amino acid metabolism (*BCAT2, GOT1, IVD, PCBD1*). In particular, *BCAT2* and *IVD* are enzymes involved in the metabolism of leucine, which is taken up from culture medium more efficiently by embryos that will successfully develop to the blastocyst stage (Houghton et al., 2002). Other factors we identify as enriched in HP TE include those necessary for nucleotide metabolism, particularly purines (*ADSL, IMPDH2, MOCOS, PAICS*), as well as those regulating glutathione metabolism and detoxification (*GSS, GSTA4, GSTK1, HAGH, MPST*). The latter indicates a potential role of the HP TE in alleviating oxidative stress within the blastocyst. Conversely, genes enriched in LP relative to HP TE include factors important for chromatin organization, receiving and responding to chemotactic signals, and interacting with the immune system (Figure 4.7B). Why aberrant or persistent expression of these factors by mural TE is associated with decreased implantation and pregnancy success is currently unknown. Together, these findings suggest that the TE of a successful blastocyst is defined by its high metabolic and detoxification properties.

Establishment of the primitive endoderm lineage characterizes HP inner cell mass.

We next asked what factors and pathways are enriched in the ICM of HP embryos, the compartment for which our clinical analysis suggested an enhanced role in predicting successful implantation (see above). A total of 317 genes were differentially expressed (2.5% of the 12,909 genes with detectable expression) ($P<0.01$). Of those, 164 genes were expressed at a higher

level in HP ICM whereas 153 genes were more highly expressed in LP ICM. As in the mTE samples, gene expression in the HP and LP ICM samples was highly correlated overall (R^2 = 0.96; Figure 4.6B-D). However, while the expression of a previously identified panel of pTE-specific genes was similar for HP and LP pTE (R^2=0.93; Figure 4.6B) and a panel of 99 previously identified EPI-specific genes were also overall similar (R^2=0.91; Figure 4.6C), the expression of PrE-specific genes (n=82) were present at dramatically lower levels as a group in ICM from LP embryos (Sophie Petropoulos et al., 2016) (R^2=0.83; Figure 4.6D).

We investigated this further by examining the ratio of the normalized reads of all PrE-specific genes in HP ICM to that of LP ICM, which confirmed higher expression of these lineage-specific genes in HP embryos (P=1.66E-06) (Figure 4.6E). This is in contrast to the ratios of mTE- (P=0.32), pTE- (P=0.12), and EPI- (P=0.62) specific genes, which each demonstrate average ratios of approximately one. This was equally true for PrE-specific genes previously identified as both early (P=0.03) and late (P=1.63E-05) markers of PrE development (Petropoulos et al., 2016) (Figure 4.8A). We considered the possibility that, although the HP and LP embryos were at the same late blastocyst stage morphologically, LP ICM represented an earlier stage of development molecularly. However, apart from the PrE, the remaining embryo compartments demonstrated similar expression of lineage-specific genes (Figure 4.6A-C, 3E, S5A), including late stage lineage markers (Figure 4.8A). Further, PCA demonstrated that both HP and LP ICM compartments most closely segregate with published profiles of late blastocysts on Day 6 post-fertilization (Petropoulos et al., 2016) (Figure 4.8B), and plotting both HP and LP embryos on a pseudotime trajectory of human embryo development (E3-E7) placed them in the same chronological space (Figure 4.8C-D), making this possibility unlikely.

To independently reexamine whether PrE was in fact deficient in LP ICM, we used an analytical tool designed to deconvolute a mixture of cells from a complex mixture to identify the proportion of EPI and PrE cells contributing to HP and LP ICM samples (Newman et al., 2015).

Feature selection is performed via a support vector regression machine learning approach, where genes from the training dataset (single-cell human embryo dataset composed of 470 EPI, PE, and pTE cells, stages E6-7; (Petropoulos et al., 2016)) are adaptively selected to then deconvolve the input mixture (our ICM samples). Using this approach, we were able to estimate the abundances of each cell type in both HP and LP ICM samples. HP ICM samples comprised 21.9% pTE, 30.1% EPI and 48% PrE, on average (Figure 4.9E, left), compared to LP ICM samples, which comprised 27.8% pTE, 56.8% EPI, and only 15.3% PrE (R=0.76, Figure 4.9E, right). Together, these data demonstrate that the most striking difference between embryos of high or low pregnancy potential is a failure to either form or maintain the PrE cell type.

Genes significantly enriched in HP ICM relative to LP ICM included those in GO categories involved in extracellular organization and cell communication, including *adherens junctions interactions, cell junction assembly, focal adhesion*, *extracellular matrix (ECM) organization*, and *peptide secretion*, as well as those involved in intracellular organization, including *cytoskeletal organization* and *actin filament assembly*. Enrichment of cell adhesion molecules in HP ICM was particularly prominent (*CDH5, CDH11, CADM1, NECTIN3, FN1, DSCAML1, FLRT3, AMIGO2*) as were a number of genes encoding ECM and collagen biosynthesis factors (*COL4A2, P4HA1, P3H1*). These findings point to a necessity between and within cells of the ICM for cellular structural integrity (Figure 4.8E). Also enriched were genes encoding cell cycle proteins, including factors important for cell cycle progression (*TUBG1, FSD1, SGO2, CAV2*) and cell cycle checkpoints (*BUB1B, CHEK2, ZWILCH*), as well as *receptor tyrosine kinase signaling* and *response to hormones* (Figure 4.8E). GO categories enriched in LP ICM relative to HP samples included *histone modification*, components of the *NF-kB pathway*, and factors integral to *antigen presentation* (Figure 4.8F). Histone modifiers included factors important for tri-methylation at H3K4 (*KMT2C, SETD1B*) and K3K9 (*SUV39H1*), as well as histone acetyltransferases (*BRPF1, ENY2*). In addition, the gene encoding TET2, a

well-known DNA demethylase enzyme, was enriched in LP ICM samples. Why aberrant or persistent expression of these factors by the ICM is associated with decreased implantation and pregnancy success is not clear.

Cell cross-talk within and between the inner cell mass and trophectoderm in human euploid embryos of high pregnancy potential

Our transcriptome analyses identified factors mediating cell-cell communication as one of the most prominent differences in embryos of high potential for successful implantation and pregnancy relative to those with low potential. To systematically examine the cell-cell communication networks that occur within and between the main compartments of the blastocyst, we leveraged our HP embryo sequencing data and a recently published repository of interactions between ligands and their associated receptors (Vento-Tormo et al., 2018). Following 1000 permutations of the algorithm, we identified 286 significant interactions within HP blastocysts ($P<0.05$, Figure 4.9A). Of those interactions, 66 predicted directional communication (from ligand to receptor) from the ICM to the TE, 59 from the TE to the ICM, 143 between cells within the ICM, and 18 between cells within the TE (Figure 4.9A, 4.10A).

Sixty percent of the cell-cell interactions we identified involved specific cell signaling ligand-receptor pathways, with the majority originating from the ICM (i.e., the gene encoding the ligand is expressed in the ICM). Many of these signaling pathways have been shown (predominantly in non-human species) to be important in embryonic development and/or embryonic stem cell maintenance, including the BMP, EGF, Ephrin, FGF, KIT, Notch, TGF-β, and Wnt signaling pathways (Table S4) (Blakeley et al., 2015; Boroviak et al., 2015; Chazaud and Yamanaka, 2016; Graham et al., 2014; Kuijk et al., 2012; Roode et al., 2012). Also uncovered were cell-cell communication networks important in immune regulation, cell migration, and angiogenesis (Figure 4.10A), additional processes that have been demonstrated

to be important in embryonic development, as well as cell-cell interactions that have yet to be explored in the embryo. The remaining 40% of interactions uncovered involved ECM or cell-cell adhesion interactions and represented predominantly integrin-to-collagen interactions, with the large majority of these interactions arising from the ICM compartment (Figure 4.9A, 4.10A, Table S4).

To discover candidate cell-cell communication networks important for successful implantation and pregnancy, we next identified signaling-related interactions unique or enriched in HP blastocysts relative to LP embryos. In total, we discovered 115 signaling interactions unique to HP embryos. Consistent with our previous data indicating an important role for the ICM at the time of implantation, we found that LP embryos demonstrated loss of a greater number of interactions originating from the ICM than from the TE (Figure 4.9C-D). In all, significant interactions that were unique to HP embryos (and absent in LP embryos) included 24 signals from the ICM to the TE, 55 signals from ICM to ICM, 29 from TE to ICM, and 7 signals from TE to TE.

To further characterize cell-cell interaction networks within individual cell types of HP embryos, we computed interactions for single human embryo cells dissociated from blastocysts from a published dataset using the same parameters (Petropoulos et al., 2016). Using these data, we were able to infer the origin and target for 541 significant ligand-receptor interactions within the human preimplantation blastocyst at a lineage-specific level ($P<0.05$) (Figure 4.9B). Surprisingly, of all interactions identified, the cell type in which most of the interactions originated (i.e. the ligand was expressed) was the PrE (38% of all interactions). The interaction originated from the EPI in 36% of all identified interactions, and the pTE and mTE in 15% and 11% of these interactions, respectively (Figure 4.9B). Thus, as observed in our embryos of high implantation potential, the majority of these communication patterns originate from the ICM.

Combining the lineage specificity of the signaling patterns identified from a published single-cell human embryo dataset with the known ploidy and implantation potential of our datasets, we overlaid lineage-specific interactions identified with those also present in our HP embryos. In this way, we were able to infer cell type-specific communication networks present in embryos of known ploidy and implantation potential. Signaling interactions identified from this analysis included many pathways that have been previously identified as important in mouse and non-human primate embryo development, such as the BMP, FGF, TGF-β, and Wnt pathways (Figure 4.9B and 4.10B; Table S4). While these pathways are presumed to be important for human blastocyst development as well, very few studies have been able to directly examine their roles in a human system.

Expression of receptor-ligand factors involved in BMP signaling was prominent in HP embryos, including significant expression of factors for BMP2, BMP4 and BMP6 signaling. Interestingly, while most of the BMP2 and BMP6 signals were unchanged, LP embryos appeared to selectively lose BMP4 signaling originating from the ICM, particularly from the PrE (Figure 4.9C-D, 4.10B). In the mouse, BMP signaling is important for development of both the TE and PrE (Graham 2014). Bovine blastocysts treated with BMP4 have a reduced proportion of OCT4-positive cells (La Rosa et al., 2011) and human blastocysts treated with BMP4 *in vitro* demonstrated increased apoptosis (De Paepe et al., 2017), making this ligand an interesting target for future mechanistic studies in human development.

In addition, LP ICM no longer significantly expressed ligands involved in FGF signaling. FGF4 is a key regulator of PrE differentiation in mice (Chazaud and Yamanaka, 2016). However, inhibiting FGF/Erk signaling in human embryos does not affect formation of either the EPI or PrE (Kuijk et al., 2012; Roode et al., 2012). We observed here that the EPI expresses FGF2 and FGF8, and that these signals are lost in LP ICM. Non-human primate embryos rely on both FGF/ERK and Wnt signaling for correct ICM lineage specification (Boroviak et al.,

2015). In LP ICM, the majority of Wnt signaling was lost, and all PrE-expressed Wnt ligands were absent in LP ICM, as well. These signaling interactions unique or enriched in HP embryos represent candidate pathways for future studies with important roles in successful implantation and pregnancy in humans.

Most strikingly were the structural or extra-cellular matrix signals lost in LP embryos. We observed that 94% of the ECM and cell-cell adhesion interactions prominent in HP embryos were absent in LP embryos (Figure 4.9C). Further, the majority of structural interactions that were maintained were from the minor subset that originated from the TE. The majority (67%) of these structural interactions were integrin to collagen connections from the ICM to the TE, or within the ICM, and the integrin subunits most enriched in HP ICM were integrins alpha1 (ITGA1), alpha2 (ITGA2), and beta1 (ITGB1) (Figure 4.9C, 4.10C, Table S4). In the mouse blastocyst, knockout of *Itgb1* leads to defects in PrE formation and peri-implantation developmental failure (Liu et al., 2009; Moore et al., 2014). In this study and others, *Itgb1*-null embryos failed to form or maintain a PrE epithelium, and *Itgb1*-null embryoid bodies failed to form a PrE layer, instead detaching and forming aggregates (Liu et al., 2009; Moore et al., 2014). That all intra-ICM interactions involving ITGB1 were lost in LP embryos, given the importance of this factor in mediating attachment of the PrE to the underlying EPI, supports a model in which the LP ICM cannot form the PE epithelium, due in part to structural dis-integrity of the ICM or lack of an environmental niche put in place by these ECM interactions.

Unbiased machine learning demonstrates that gene expression data from the inner cell mass is more predictive of implantation potential than from the trophectoderm

To identify genes that most highly correlate with successful implantation and early ongoing pregnancy, we took an unbiased machine learning approach. To this end, we applied three machine learning algorithms to our gene expression data to test the ability to accurately

classify HP and LP blastocysts. In combination, these approaches identified genes that most highly correlate with implantation and pregnancy potential. The algorithms employed included one that utilizes feature selection (Maximum Likelihood Estimation, MLE) and two that do not perform a feature selection (Support Vector Machine, SVM and Least Square Regression, LS) (Angelov and Ekström, 2018; Guo and Wang, 2019; Heikamp and Bajorath, 2014; S. Huang et al., 2018; May et al., 2011; Zhao et al., 2011). In addition, we applied genetic algorithm to all approaches as an optimization technique to narrow the search space of interest (Ghaheri et al., 2015; Lee et al., 2018; Li et al., 2014; Manning et al., 2013) (Figure 4.11A).

When provided a list of all expressed genes from the ICM, all three machine learning approaches were able to identify a subset of genes that could accurately classify HP and LP embryos; however, none of the algorithms were able to distinguish the groups when provided a list of all expressed genes from the TE. To quantify this, for each algorithm, we picked the top ten iterations based on the smallest error in binary classification and quantified the proportion of iterations that accurately classified HP and LP embryos by hierarchical clustering (Figure 4.12A). This provided additional support that factors expressed by the ICM are more closely associated with successful implantation and pregnancy than factors from the TE.

To identify genes that most strongly define HP embryos and distinguish them from LP embryos, we optimized the search space using differential expression ($Q<0.1$) and repeated the same machine learning approaches (Figure 4.12B). Under these conditions, all approaches were able to accurately classify HP and LP embryos when provided with gene expression data from either the TE, the ICM, or both (whole embryo). For each dataset, we compiled the cumulative incidences of genes classified as predictive by any of the algorithms, ultimately ranking genes by an Incidence Score (IS), where a higher score was indicative of increased predictive ability. Genes with the highest ISs were likely identified as predictive by all three algorithms. Strikingly, 9 out of the top 10 IS-ranked genes from the whole embryo were ICM-

specific genes, and 2/3 of the genes with an IS of four or above (see Methods) were expressed in the ICM (Figure 4.11B). This result demonstrated--using an unbiased machine learning approach—that gene expression profiles of the ICM are more predictive of implantation potential than profiles obtained from the TE. Further, we were able to produce a high-confidence list of genes expressed in the whole blastocyst that is most closely associated with implantation and a successful pregnancy. Future studies in more genetically tractable models aimed at uncovering the role of these factors in embryo development and the process of implantation will shed light on how these early stages of life begin for the human embryo.

Repeating this unbiased approach, we were also able to individually query the TE and ICM and identify the genes in each of these compartments that most closely predict implantation and early pregnancy success. In the same way, we compiled cumulative incidences of genes classified as predictive and ranked genes by decreasing ISs (Figure 4.11A). In the ICM, this approach identified 55 genes highly predictive of implantation potential (IS >2). Of those, 36 genes were more highly expressed in HP ICM relative to LP ICM, i.e., potential positive predictors of success. Interestingly, the most significantly enriched GO category was ‘cell adhesion’, where CDH11, DSCAML1, and AMIGO2 were adhesion-associated genes in the ICM that most highly predict implantation. In contrast, the remaining 19 most predictive ICM genes were more highly expressed in LP ICM, acting as predictors for poor implantation outcomes. Two of these genes—*KRT8* and *KRT18*—are intermediate filaments responsible for maintaining cell structural integrity. Finally, the gene with the most positive predictive potential within the ICM was *STAT5B* (IS = 17), which, in response to cytokine or growth factor signals, acts as a transcription factor. Future studies aimed at examining the role of *STAT5B* within the ICM of the human blastocyst, along with the repertoire of genes that it might activate, could unveil important information about the preimplantation ICM and its *in vitro* counterpart, hESCs.

From the TE gene expression data, we were able to identify 54 genes highly predictive of implantation potential (IS > 2) (Figure 4.11C). Of those, the majority (47) were more highly expressed in LP TE, acting as predictors of poor implantation outcomes from TE gene expression data. GO categories most significantly enriched in this set of genes includes factors important for chromatin modification, namely histone acetylation (KAT6A) and deacetylation (RCOR1). Further, the gene most predictive of implantation success from the TE expression data was CUL4B (IS = 18), which has been demonstrated to have an essential role in the extra-embryonic trophoblast. In fact, aberrant expression of CULB4 in extra-embryonic tissue, but not in the epiblast, of mouse embryos, leads to early post-implantation lethality (Liu et al., 2012).

While extensive verification and validation experiments are needed, the genes identified in the TE that most closely correlate with successful implantation and pregnancy might aid in selecting the single embryo with the greatest pregnancy potential for transfer in IVF cycles. During PGS, a small biopsy is taken from the TE and is used to screen for chromosomal loss or large-scale deletions. Using mRNA sequencing protocols from low-input, as has been reported (Dumdie et al., 2018; Mora-Castilla et al., 2016), 1-5 cells from this biopsy could theoretically be used to screen for implantation and early pregnancy success. This type of diagnostic test would require a panel of genes that highly correlate, or even predict, clinical outcome. By implementing a diagnostic test such as this, we would be able to maximize embryo success and minimize procedure-associated risks by transferring the single best embryo. Using multiple machine learning approaches in combination, as we have done here, has allowed for the compilation of gene discovery lists most closely associated with implantation and pregnancy success in human development.

Discussion

Here, we provide a qualitative assessment of gene expression and cellular communication networks within the major compartments of the blastocyst that are most closely associated with successful implantation and ongoing pregnancy in euploid human embryos (Figure 4.3D, 4.5, 4.9, 4.10). With recent advances in sequencing with low input, our knowledge of gene expression patterns unique to human embryos, and those likely conserved, has expanded (Blakeley et al., 2015; Petropoulos et al., 2016). In addition, mounting clinical evidence has produced correlative assessments, including morphological quality and ploidy, that assist in the selection of the single best embryo for transfer to improve outcomes in IVF (Gardner et al., 2000; Irani et al., 2017). Here, we leverage retrospective clinical data to identify groups of euploid embryos of varying morphology that differ in their implantation potential to identify genes and pathways within the ICM and TE of the embryo that are highly correlated with successful implantation. Importantly, since this analysis was conducted on human embryos of known ploidy and quality, functional correlations can be drawn from the factors and pathways identified.

In the TE, we find that embryos of high implantation potential upregulate an array of mRNAs associated with metabolism of carbohydrates, amino acids, and nucleotides. These findings are consistent with published data that demonstrate a positive relationship between amino acid metabolism and blastocyst development (Houghton et al., 2002), as well as glucose uptake and human embryo implantation (Gardner et al., 2011), and further define the molecular and metabolic profile associated with a trophectoderm of high implantation potential. Whether the nutrient requirements are higher for cells of the TE compared to the ICM is unknown. It will be interesting to determine if the TE acts to provide nutrients or energy sources to the ICM and how these molecules are transferred between cells. In combination with the signaling

interactions identified herein, these data point to a dynamic and collaborative communication network of signaling pathways and energy transfer within the human blastocyst.

Surprisingly, transcriptome profiling, combined with a machine learning strategy designed to deconvolute complex admixtures of cells, demonstrated that establishment and/or maintenance of the PrE lineage most strongly differentiates embryos of high and low implantation potential (Figure 4.6, 4.8). Whether the uniform decrease in expression of PrE markers in embryos of low implantation potential represents a defect in the PrE or a delay in development is not known. However, that HP and LP embryos express similar levels of both early and late mTE, pTE, and EPI lineage markers, as well as their similar transcriptome profiles to E6-7 embryos via PCA, and their similar clustering in pseudotime space, strongly argues against an overall delay in development of embryos of low implantation potential. If a delay in development exists, it is likely an asynchronous delay in the PrE lineage specifically. (Figure 4.6, 4.8).

This data unexpectedly points to an important role for the extra-embryonic PrE in establishing a successful pregnancy. Compared to the TE and EPI, the PrE is much less studied during preimplantation embryo development, and the role of the PrE is not as well defined at these early stages leading up to implantation. After implantation, the PrE differentiates to form the visceral and parietal endoderm, which envelops the gastrulating epiblast. In the mouse, the architectural and molecular patterning of the endoderm is translated into the anterior-posterior patterning of the body axis (Lu et al., 2001; Zernicka-Goetz et al., 2002). However, we report a novel role of the PrE in establishing a human blastocyst of high implantation potential in the stages prior to implantation.

How do the events inside of the embryo, namely signals originating from the PrE, influence implantation, not making direct contact with the endometrium? While it is the pTE that ultimately attaches to the uterine endometrium, we propose that internal signals must be

communicated to convey the overall health status of the blastocyst, that it is prepared for the stages immediately preceding implantation. Communication patterns within embryos of high implantation potential, overlaid with single-cell interactions identified, revealed directional communication from both the EPI (via BMP and TGF-b signaling) and the PrE (via BMP, TGF-b, and Wnt signaling) to the pTE. Importantly, BMP and Wnt signals from the PrE were lost in embryos of low implantation potential. While we identified factors within the TE that are correlated with successful implantation as well, we propose that communication from the ICM to the TE is an additional indirect requirement for human blastocyst implantation, and that implantation itself is a complex coordination of signals within the three lineages of the blastocyst and between the TE and the endometrium.

In addition, we report the complete repertoire of candidate cell-to-cell communication patterns identified within embryos of high implantation potential, as well as those lost in embryos of decreased potential, and report that the majority of intra-embryo communication originates from the ICM. Delineation of cell-cell interactions demonstrates that embryos of low implantation potential lack many of the structural and extracellular matrix interactions associated with formation of the PrE and identifies signaling pathways specifically enriched in embryos of high implantation potential (Figure 4.9). In particular, we define a cohort of integrin-to-collagen intra-ICM structural interactions present in embryos of high implantation potential that are dynamically lost in embryos of low implantation potential. In the mouse blastocyst, loss of *Itgb1* leads to defects in formation of the PrE and associated peri-implantation lethality. In particular, embryos devoid of *Itgb1* fail to form or maintain a PrE epithelium, and resulting embryoid bodies demonstrate sloughing of the PrE, detaching and forming PrE aggregates (Liu et al., 2009; Moore et al., 2014). Whether these structural interactions must be in place for the PrE to form, or if the PrE secretes factors that initiate this structural integrity, which is then necessary for the maintenance of this lineage, is uncertain. However, the lack of structural integrity could, at least

in part, explain the morphological differences observed during grading, namely few total cells and the formation of a more loosely arranged epithelium in embryos graded as “poor”. Interestingly, this morphological assessment that has long been held as a clinical standard for embryo quality could be a visual representation of the lack of an established PrE epithelium.

Additionally, we characterized a number of signaling pathways and intra-embryo communication patterns within the human embryo with potential roles in lineage specification and/or maintenance. A combination of FGF and Wnt communication from the EPI to the PrE in human blastocysts compliments signaling patterns that have been proposed for establishment of the PrE in primates (Boroviak et al., 2015). Further, a number of signals originating from the PrE in HP embryos, including those from the BMP, TGF-b, and Wnt pathways, are selectively lost in embryos of low implantation potential. It will be important to determine the role of these pathways in the preimplantation blastocyst, and the resulting role of the PrE, at these stages. In the cascade of events that must transpire for implantation to be initiated and pregnancy to be maintained, we propose that establishment of the PrE sits at the top of that hierarchy, initiating the structural and signaling networks necessary for downstream blastocyst implantation.

Our unbiased machine learning approach, combined with our retrospective analysis of implantation and ongoing pregnancy success rates following single embryo transfer in IVF, adds to the debate of whether the ICM or the TE is more predictive of pregnancy success. Here, we report that the ICM grade is more predictive of implantation than the TE, and that gene expression data from the ICM is more predictive of implantation potential than from the TE (Figure 4.12). Utilizing a set of machine learning approaches, we additionally identify genes that are most closely associated with human embryo implantation and ongoing pregnancy (Figure 4.11). These genes provide valuable focus for future mechanistic studies in more genetically tractable models. Finally, genes found to be highly correlated with implantation in the TE might

aid in predicting clinical outcome for embryo transfer in IVF. Together, we provide a rich resource for the field through the identification of gene expression patterns and intra-embryo communication networks within the human blastocyst and most closely associated with successful implantation. Leveraging cumulative clinical data from embryos of varying pregnancy potential, we demonstrate that successful implantation predominantly depends on the ICM and we report a novel role for the extraembryonic PrE in mediating human embryo implantation.

Acknowledgements

Chapter 4 contains material that will be prepared for submission for publication. The dissertation author was the primary investigator and author of this material. Current co-authors are as follows: Jennifer N Dumdie, Cuong To, Srimeenakshi Srinivasan, Wei Zhang, Ana Lisa Yeo, Gabriel Garzo, Miles F. Wilkinson, Louise C. Laurent, and Heidi Cook-Andersen.

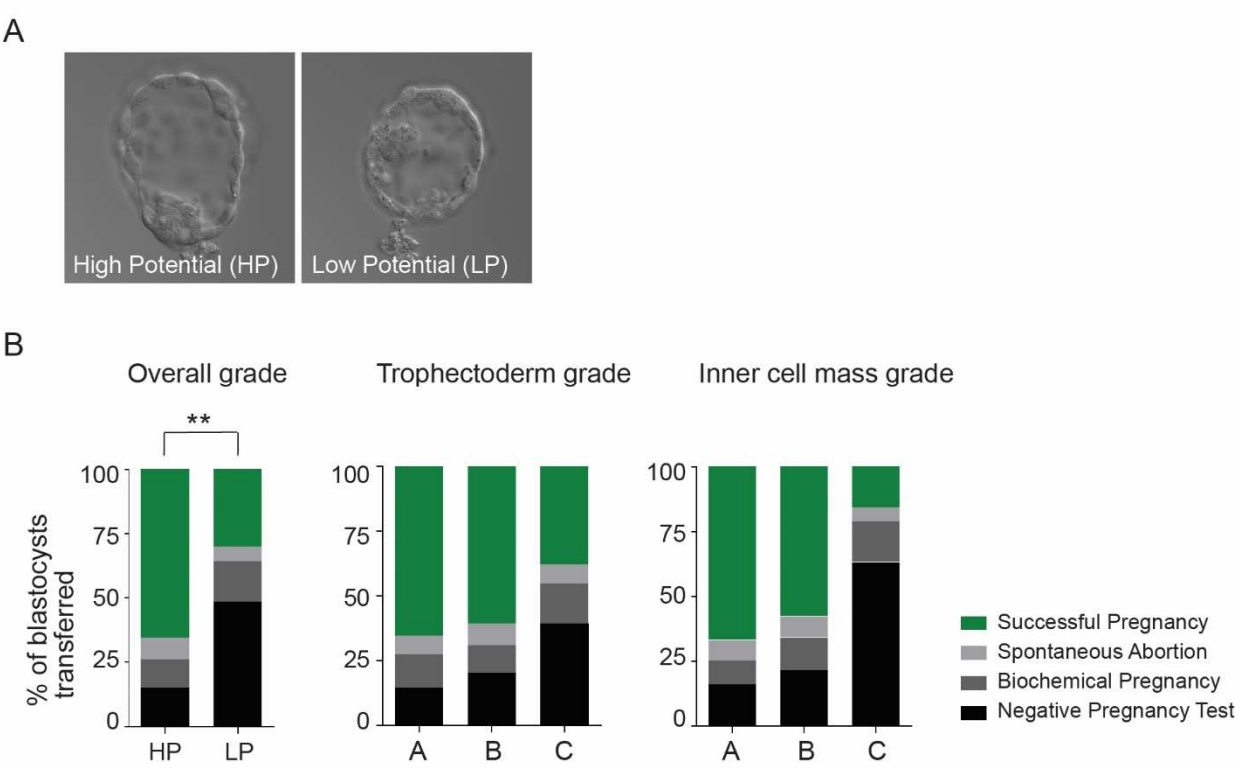


Figure 4.1. Successful implantation and early ongoing pregnancy is correlated with the morphological grade of euploid blastocysts. (A) Brightfield images of Embryonic day E5-6 human blastocysts, demonstrating varying morphological grades. Blastocysts with a grade of "good" (HP embryos) have a trophectoderm with many cells organized into a cohesive epithelium, and an inner cell mass with numerous tightly packed cells (left), while blastocysts with a grade of "poor" (LP embryos) in general display very few, larger trophectoderm cells with an inner cell mass containing few less tightly packed cells (right). (B) Clinical embryo transfer data compiled from 678 transfers from 2012 to 2017 demonstrates that euploid embryos graded as poor versus good are more likely to result in a negative pregnancy test (human chorionic gonadotrophin < 5 mIU/ml, 11 weeks after embryo transfer). Embryo transfers are plotted as a stacked bar graph as a total percent of blastocysts transferred by overall morphological grade, and outcome is classified as a successful pregnancy, spontaneous abortion, biochemical pregnancy, or a negative pregnancy test. Morphologically graded good embryo transfers resulted in a negative pregnancy test 15.1% of the time, compared to 48.6% for poor-graded embryos (P=0.002, T-test). (C) Euploid embryos transferred are plotted as stacked bar graphs individually by trophectoderm (left) and inner cell mass (right) grades. Morphological grading criteria is as in Figure 4.2A-B, such that the TE was assigned one of the following grades: A: many cells organized in a cohesive epithelium; B: several cells organized in loose epithelium; or C: few large cells. The ICM was assigned one of the following grades: A: numerous tightly packed cells; B: several loosely packed cells; or C: very few cells. ICM grade is more predictive than TE of implantation and early pregnancy loss; a C-graded TE resulted in a negative pregnancy test 39.4% of the time and a biochemical pregnancy 15.2% of the time, whereas a C-graded ICM resulted in a negative pregnancy test 63.2% of the time and a biochemical pregnancy 15.8% of the time (P=0.07, T-test).

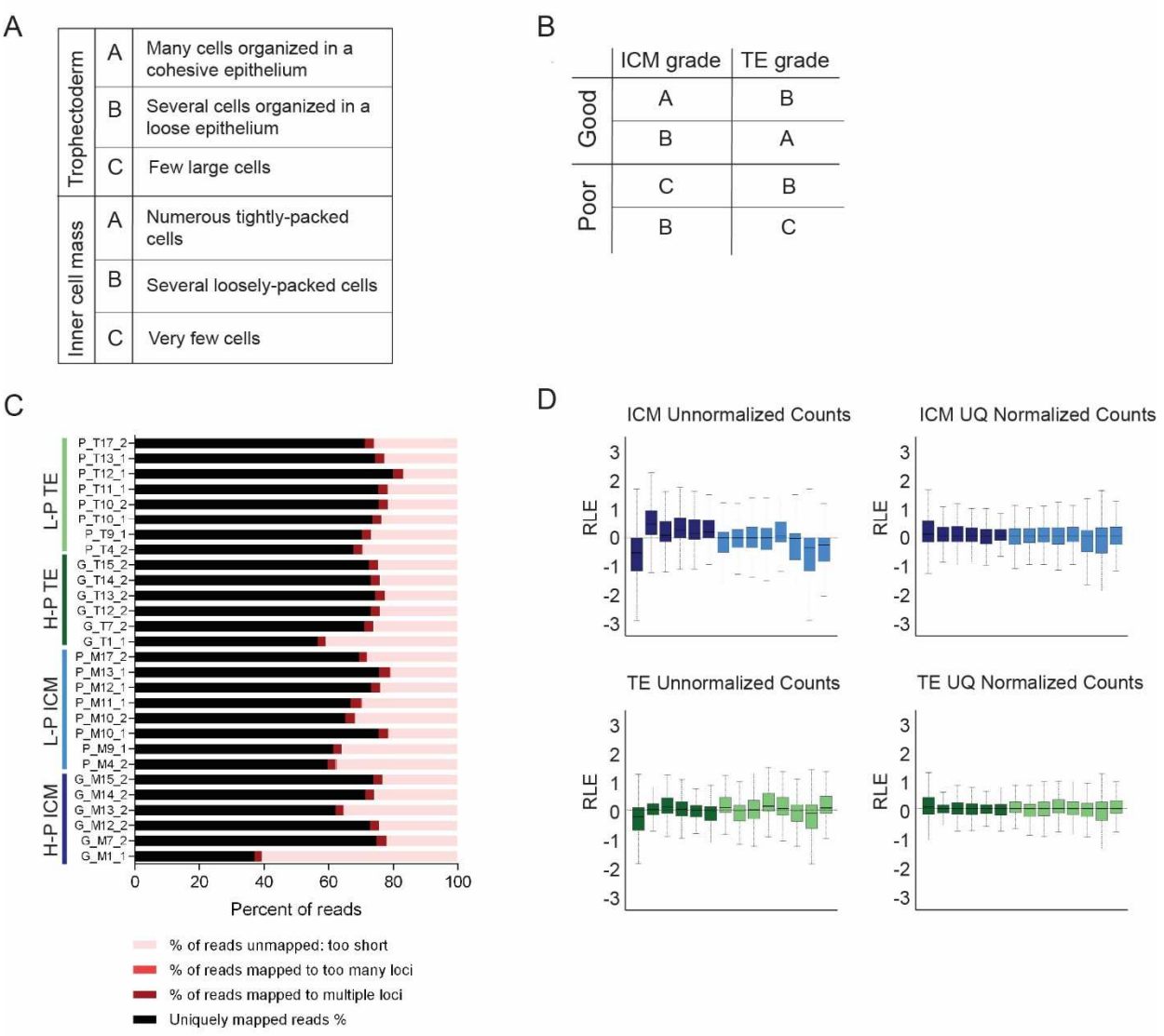
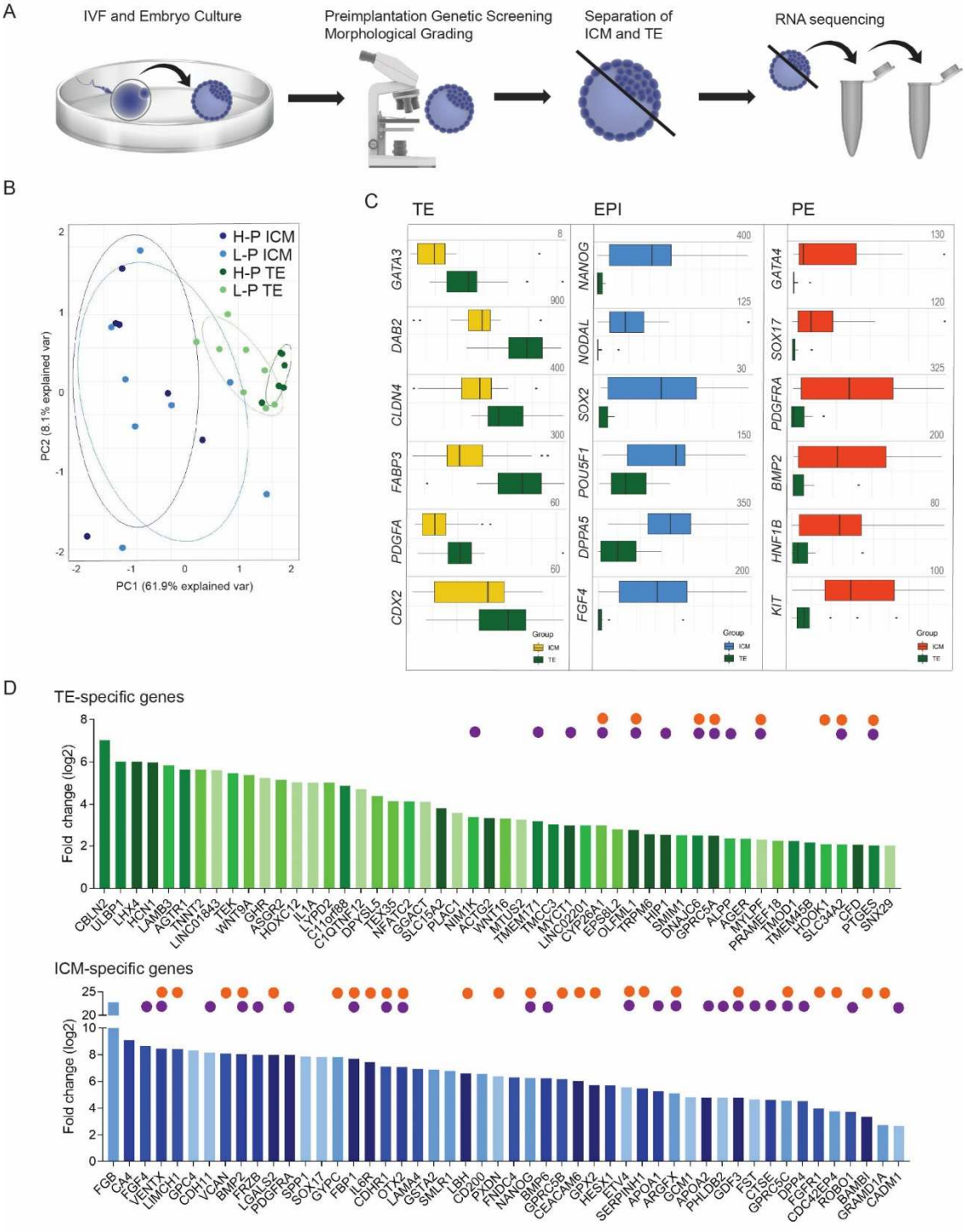


Figure 4.2. Quality control of human blastocyst inner cell mass and trophectoderm mRNA sequencing samples. (A) Table depicting the blastocyst morphological grading scale presented by Gardner et al. (Gardner et al., 2000). TE and ICM are both given a grade of either A, B, or C according to the morphological parameters herein. (B) Adapted from the original morphological grading scale by Gardner et al. (Gardner et al., 2000), the combinations of individual compartment morphological scores that result in “good” and “poor” morphological blastocyst grades. (C) For both inner cell mass and trophectoderm blastocyst samples, approximately 18.1 million reads were generated per sample, and 70.1% of those were uniquely mapped. For each sample, reads mapped or unmapped are plotted as a percent of total sequencing reads acquired. There was no difference observed between High Potential (HP) and Low-Potential (LP) samples in terms of reads uniquely mapped nor those unmapped. (D) To minimize the variability caused by highly-abundant genes, the upper quartile (UQ) normalization method was applied by estimating the scaling factors based on the 75th percentile of the gene count distribution (Dillies et al., 2013). The relative log expression (RLE; log-ratio of read count to median read count across each sample) is plotted for libraries before and after normalization. Normalized libraries exhibit median RLE values close to zero.

Figure 4.3. Transcriptome profiles of trophectoderm and inner cell mass from euploid blastocysts of high and low pregnancy potential. (A) Diagram of experimental set-up. Oocytes were fertilized via ICSI, and embryos cultured until Day 5-6 before freezing. Embryos were thawed and cultured in G-PGD media. Embryos were then graded based on morphological quality (Gardner et al., 2000) and subjected to PGS. Only euploid embryos were used (total of 6 HP and 8 LP expanded human blastocysts). The ICM and TE were separated by laser dissection and frozen in TE Cell Wash Buffer until RNA sequencing. mRNA sequencing libraries were prepared from TE and ICM samples from each embryo. (B) Variance within transcriptomic data can be visualized by Principle Component Analysis (PCA). Most of the variance between TE and ICM samples can be explained by PC1, whereas PC2 highlights variance observed within ICM samples. HP TE and ICM samples cluster the farthest apart along PC1, whereas LP TE and ICM deviate towards the center of the plot, indicating a robustness in HP embryos in the differentiation of each of these lineages. (C) Box plots of normalized reads (TPM – transcripts per million) for known lineage markers for the TE, EPI, and PE. TE samples are enriched for TE markers relative to ICM samples, and ICM samples are enriched for EPI and PE markers relative to TE samples. ICM and TE samples plotted represent the combined normalized reads of both HP and LP embryos. (D) TE- and ICM- specific genes identified by differential expression analysis between these two compartments in euploid blastocysts of high implantation potential (HP embryos). Fold enrichment within each compartment is plotted as fold change on the y-axis, and increasing statistical significance can be visualized with darker shades of blue and green for ICM- and TE-specific genes, respectively. Purple dots above genes indicate genes that have been demonstrated as lineage specific in the human blastocyst by Petropoulos et al., and orange dots above genes indicate genes that have been demonstrated as lineage specific by Blakeley et al.


A
IVF and Embryo Culture
Preimplantation Genetic Screening
Morphological Grading
Separation of
ICM and TE
RNA sequencing
B
H-P ICM
L-P ICM
H-P TE
L-P TE
PC2 (8.1% explained var)
PC1 (61.9% explained var)
C
TE
EPI
PE
GATA3
DAB2
CLDN4
FABP3
PDGFA
CDX2
NANOG
NODAL
SOX2
POU5F1
DPPA5
FGF4
GATA4
SOX17
PDGFRA
BMP2
HNF1B
KIT
Group
ICM
TE
D
TE-specific genes
Fold change (log2)
CBLN2 ULBP1 LHX4 HCN1 LAMB3 AGTR1 TNNT2 LINC01843 TEK WNT9A GHR ASGR2 HOXC12 IL1A LYPD2 C11orf88 C1QTNF12 DPYSL5 TEX35 NFATC2 GGACT SLC15A2 PLAC1 NIM1K ACTG2 WNT16 MTUS2 TMEM171 TMCC3 MYCT1 LINC02201 CYP26A1 EPS8L2 OLFML1 TRPM6 HIP1 SMIM1 DNAJC6 GPRC5A ALPP AGER MYLPF PRAMEF18 TMOD1 TMEM45B HOOK1 SLC34A2 CFD PTGES SNX29
ICM-specific genes
Fold change (log2)
FGB CA4 FGF4 VENTX LIMCH1 GPC4 CDH11 VCAN BMP2 FRZB LGALS2 PDGFRA SPP1 SOX17 GYPC FBP1 IL6R CDHR1 OTX2 LAMA4 GSTA2 SMLR1 LBH CD200 PXDN FNDC4 NANOG BMP6 GPRC5B CEACAM6 GPX2 HESX1 ETV4 SERPINH1 APOA1 ARGFX GCM1 APOA2 PHLDB2 GDF3 FST CTSE GPRC5C DPP4 FGFR1 CDC42EP4 ROBO1 BAMBI GRAMD1A CADM1

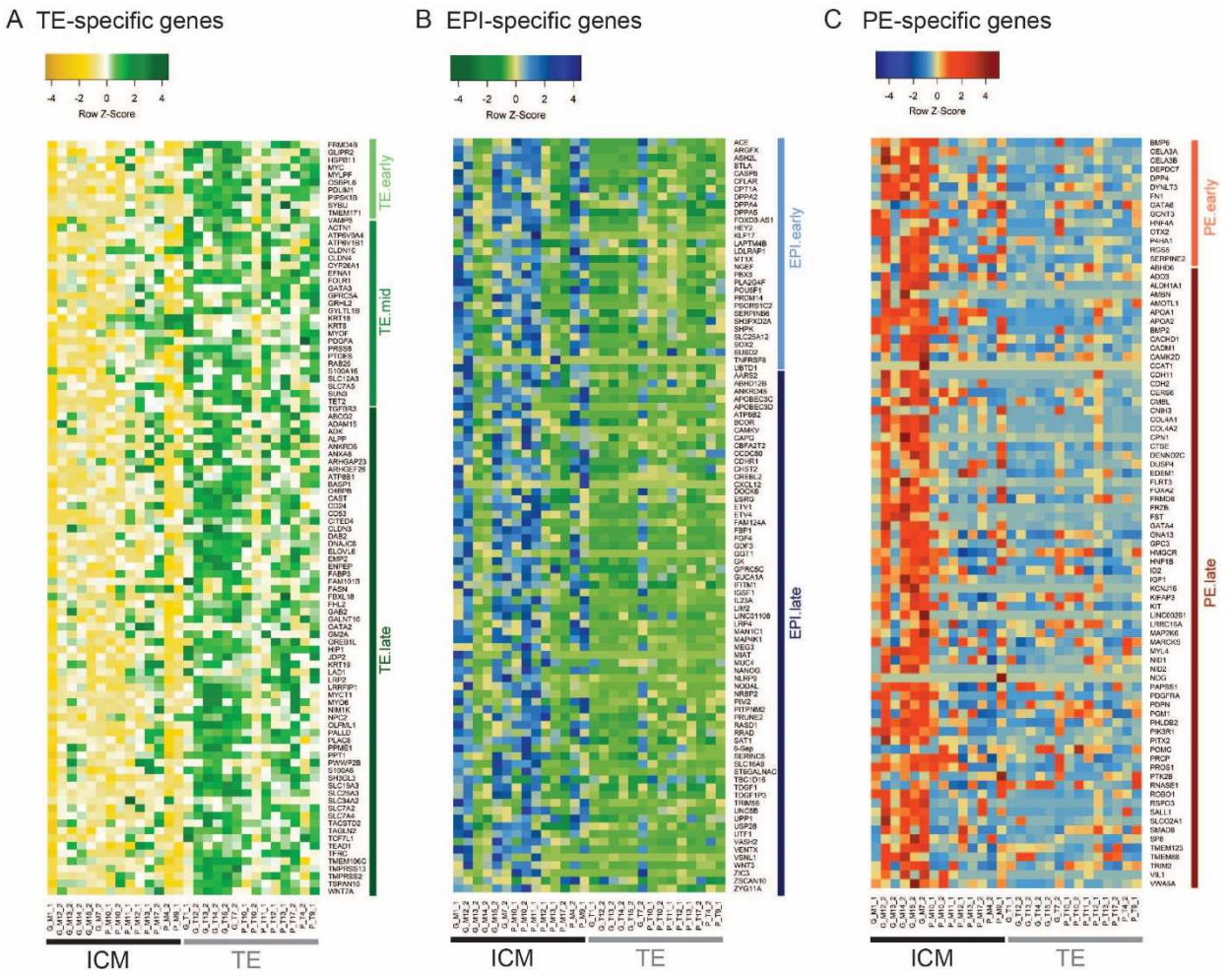


Figure 4.4. Blastocyst lineage marker expression for inner cell mass and trophectoderm samples from embryos of high and low implantation potential. (A) Full panels of TE-specific lineage markers, as defined by a recent single-cell human blastocyst RNA sequencing study (Petropoulos et al., 2016), were examined for expression in ICM and TE blastocyst samples. Gene expression data for TE marker genes identified as early, mid, and late is plotted here as a heatmap, where color is defined by z-score. TE samples demonstrate an enrichment for all TE-specific genes, as compared to ICM samples. (B) As in (A), full panels of EPI-specific lineage marker gene expression, for both early and late markers, are plotted as a heatmap. Color is defined by z-score. ICM samples demonstrate enrichment for EPI-specific markers, as compared to TE samples. (C) As in (A), early and late PrE-specific lineage markers are plotted as a heatmap, with color defined by z-score. As with EPI-specific genes, ICM samples are enriched for PrE-specific gene expression relative to TE samples.

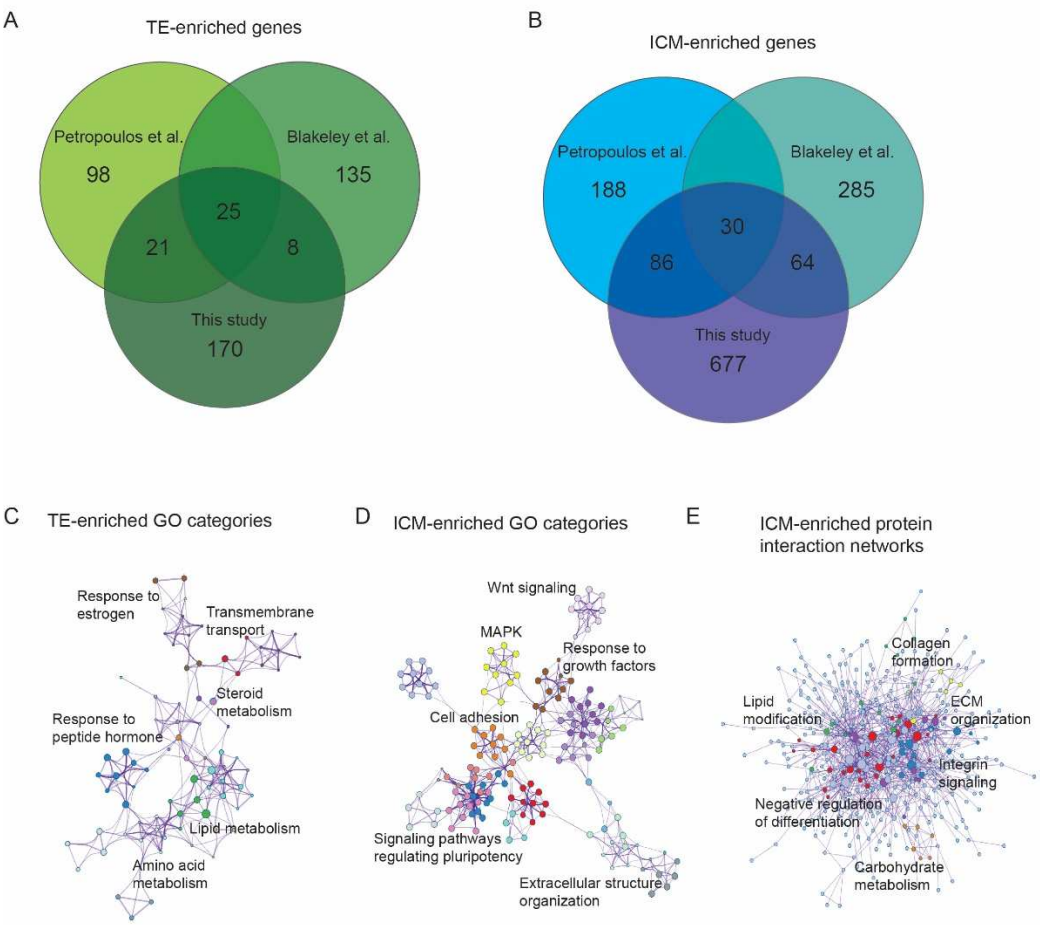
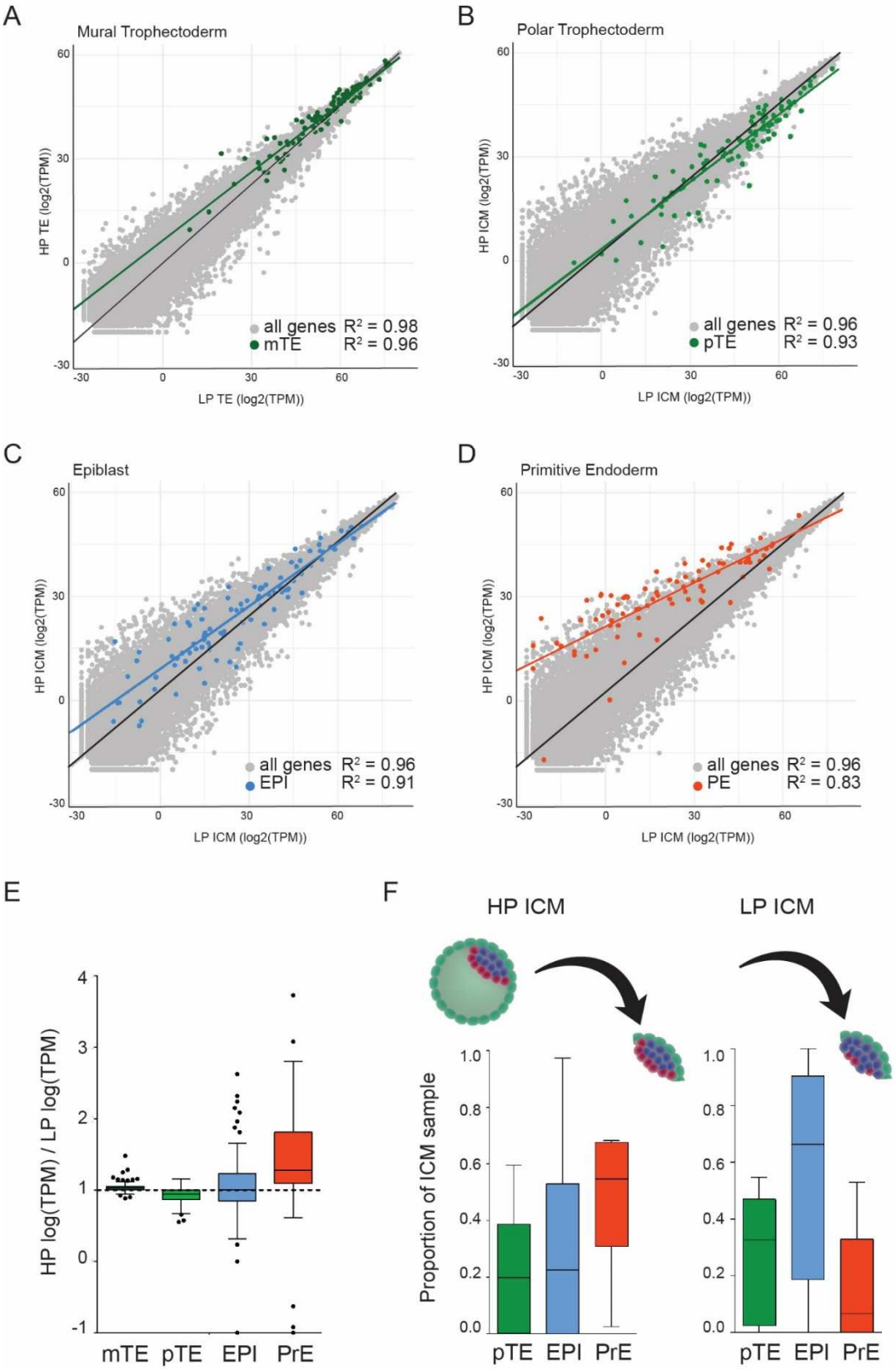


Figure 4.5. Compartment-specific gene expression patterns in euploid embryos of high implantation potential. (A) Venn diagram of TE-enriched gene expression in our data and in two other prominent human blastocyst datasets (Blakeley et al., 2015; Petropoulos et al., 2016). While many genes that we identified as being enriched in HP euploid TE samples have also been identified as TE-specific or TE-enriched in other studies (54 genes that overlap), we have additionally identified 170 genes unique to human TE in euploid blastocysts of high implantation potential. (B) Venn diagram of ICM-enriched gene expression in our data and in two other prominent human blastocyst datasets (Blakeley et al., 2015; Petropoulos et al., 2016). While many genes that we identified as being enriched in HP euploid ICM samples have also been identified as ICM-specific or ICM-enriched in other studies (180 genes that overlap), we have additionally identified 677 genes unique to human ICM in euploid blastocysts of high implantation potential. (C) Gene ontology (GO) categories enriched in TE-specific genes in HP euploid embryos include those important for metabolism of lipids, steroids, and amino acids, as well as response to hormones and transmembrane transport. (D) GO categories enriched in ICM-specific genes in HP euploid embryos include signaling pathways, including those important for regulating pluripotency, Wnt, and MAPK. Also enriched were the categories *Extracellular structure organization* and *Cell adhesion.* (E) Protein interaction networks identified by MCODE to be enriched from ICM-specific gene expression patterns include *ECM organization, Collagen formation,* and *Integrin signaling,* as well as *Negative regulation of differentiation* and *Carbohydrate metabolism.*

Figure 4.6. LP ICM samples demonstrate a relative failure to complete or maintain differentiation of the primitive endoderm lineage. (A) Scatter plot of the summed log of normalized reads (TPM) for TE samples, where HP TE samples are plotted on the y-axis, and LP TE samples are plotted on the x-axis. All genes expressed in TE samples are shown in grey, with the line of best fit for these samples displayed in black. Mural TE (mTE) gene expression data for genes defined as TE-specific by Petropoulos et al. are plotted in green, with the line of best fit displayed in green, as well. The coefficient of determination for each comparison is displayed as R-sq (all TE genes, R-sq = 0.98; mTE lineage-specific genes, R-sq = 0.96). mTE lineage-specific gene expression does not deviate far from expression of all genes expressed in TE samples. (B) As in (A), the summed log of the normalized reads (TPM) for ICM samples is shown as a scatter plot. All genes expressed in the ICM are shown in grey, with the line of best fit in black. The expression of genes identified as TE-specific by Petropoulos et al. in ICM samples is plotted in light green, and represents polar TE (pTE)-specific gene expression. The line of best fit is shown in light green. The coefficient of determination for each comparison is displayed as R-sq (all ICM genes, R-sq = 0.96; pTE lineage-specific genes, R-sq = 0.93). pTE lineage-specific gene expression does not deviate far from expression of all genes expressed in ICM samples. (C) The expression of EPI-specific genes, as in (A) and (B), is plotted in blue with the expression of all genes expressed in ICM samples in grey. All ICM genes, R-sq = 0.96; EPI lineage-specific genes, R-sq = 0.91). EPI-specific gene expression coincides with expression of all genes in the ICM. (D) The expression of PrE-specific genes, as in (A), (B), and (C), is plotted in orange with the expression of all genes expressed in ICM samples in grey. All ICM genes, R-sq = 0.96; PrE lineage-specific genes, R-sq = 0.83). PrE-specific genes are more highly expressed in HP ICM samples and more lowly expressed in LP ICM samples. (E) The ratio of the log of normalized reads (TPM) of HP versus LP blastocysts for all genes identified as lineage-specific by Petropoulos et al. is presented here as a box plot. mTE-, pTE-, EPI-, and PrE-specific genes are presented in dark green, light green, blue, and red, respectively. Ratio values around one (dotted line) indicate no change between HP and LP embryonic samples. The mean of the ratio of the expression of mTE-, pTE-, and EPI-specific genes is close to one (mTE mean = 1.04, SEM = 0.01; pTE mean =0.93, SEM=0.01; EPI mean = 1.2, SEM = 0.18), the ratio of PrE-specific gene expression deviates toward higher expression in HP ICM samples (PrE mean = 1.6, SEM = 0.23). (F) The proportion of each cell type contributing to HP (left) and LP (right) ICM samples was predicted using an analytical tool (Newman et al., 2015), run with 1000 permutations. HP ICM samples were on average classified as 21.9% pTE, 30.1% EPI, and 48% PrE (correlation between trained mixture file and input mixture file, Pearson correlation R=0.63). LP ICM samples were classified as 27.8% pTE, 56.8% EPI, and 15.3% PrE (R=0.76), demonstrating significantly less PrE contribution than HP ICM. Blastocyst models depict representative contributions of each lineage within the ICM sample collected by laser dissection.


A
Mural Trophectoderm
HP TE (log2(TPM))
LP TE (log2(TPM))
all genes R² = 0.98
mTE R² = 0.96
B
Polar Trophectoderm
HP ICM (log2(TPM))
LP ICM (log2(TPM))
all genes R² = 0.96
pTE R² = 0.93
C
Epiblast
HP ICM (log2(TPM))
LP ICM (log2(TPM))
all genes R² = 0.96
EPI R² = 0.91
D
Primitive Endoderm
HP ICM (log2(TPM))
LP ICM (log2(TPM))
all genes R² = 0.96
PE R² = 0.83
E
HP log(TPM) / LP log(TPM)
mTE pTE EPI PrE
F
HP ICM
LP ICM
Proportion of ICM sample
pTE EPI PrE
pTE EPI PrE

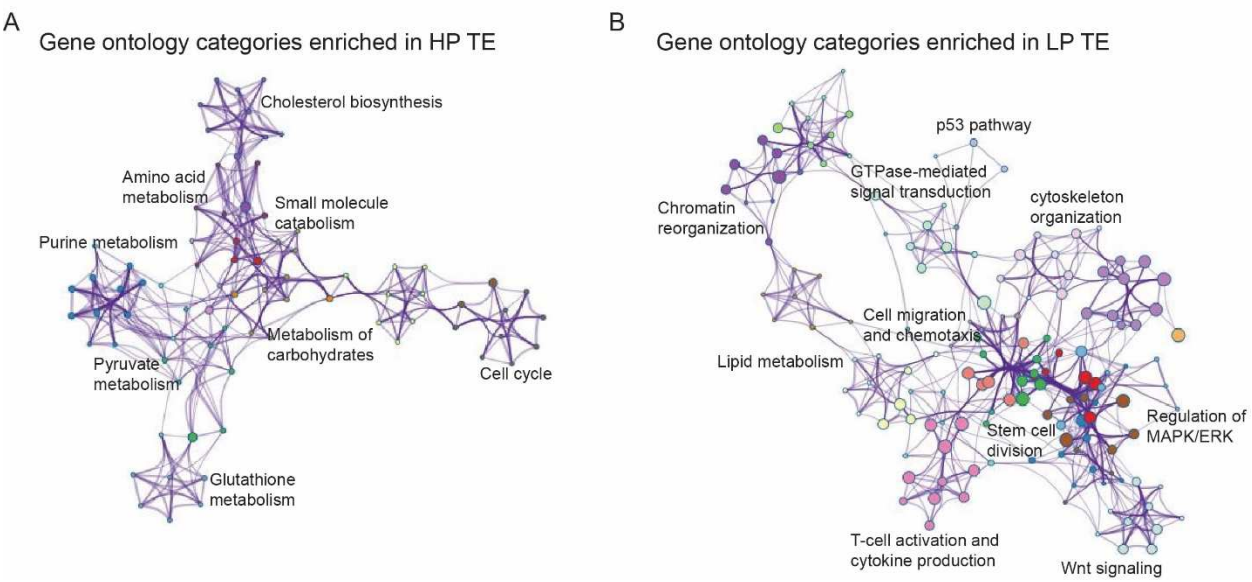


Figure 4.7. HP trophectoderm is metabolically active. (A) Gene ontology (GO) analysis for genes enriched in HP TE relative to LP TE is presented as a network. Categories upregulated in TE of high implantation potential include those important for carbohydrate metabolism, as well as the breakdown of amino acids known to correlate with embryo developmental potential. Also enriched are factors important for detoxification, including glutathione metabolism. (B) GO analysis for genes enriched in LP TE relative to HP TE. Enriched gene categories are presented as a network, and include pluripotency factors, immune and chemotactic modulators, and regulation of Wnt and MAPK/ERK signals.

Figure 4.8. HP embryos lack a defined primitive endoderm despite lack of evidence of developmental delay. (A) The ratio of the log of normalized reads (TPM) of HP versus LP blastocysts for genes identified as early, mid, or late lineage-specific by Petropoulos et al. is presented here as a box plot. mTE-, pTE-, EPI-, and PrE-specific genes are presented in dark green, light green, blue, and red, respectively. Ratio values around one (dotted line) indicate no change between HP and LP embryonic samples. Only genes identified as PrE-specific were significantly different between HP and LP embryo samples (mTE.early P=0.62, mTE.mid P=0.97, mTE.late P=0.30, pTE.early P=0.64, pTE.mid P=0.17, pTE.late P=0.31, EPI.early P=0.99, EPI.late P=0.55, PrE.early P=0.03, PrE.late P=1.63E-05). (B) Transcriptome-wide variance between E5, 6, and 7 single-cell ICM samples from Petropoulos et al. is calculated with PCA and plotted here. These cells exhibit a linear progression from earlier to later stages of development. HP and LP ICM samples are plotted with this data, and cluster together, near E6 samples. This indicates that neither HP nor LP ICM samples represent different stages of development. (C) A single-cell developmental trajectory for human embryo cells (E3 – E7) is plotted in pseudotime. HP and LP ICM samples cluster in chronological order with approximately E6-7 embryo cells. (D) As in (C), HP and LP TE samples cluster with E6-7 embryo cells in pseudotime space. (E) GO categories enriched in HP ICM include those related to cell-cell adhesion and integrin organization, as well as amino acid and nucleotide metabolism and cell cycle regulation. (F) GO categories enriched in LP ICM, indicative of potential causes of early implantation failure, include altered histone methylation patterns, as well as immune interaction and response to nutrient deficiencies.

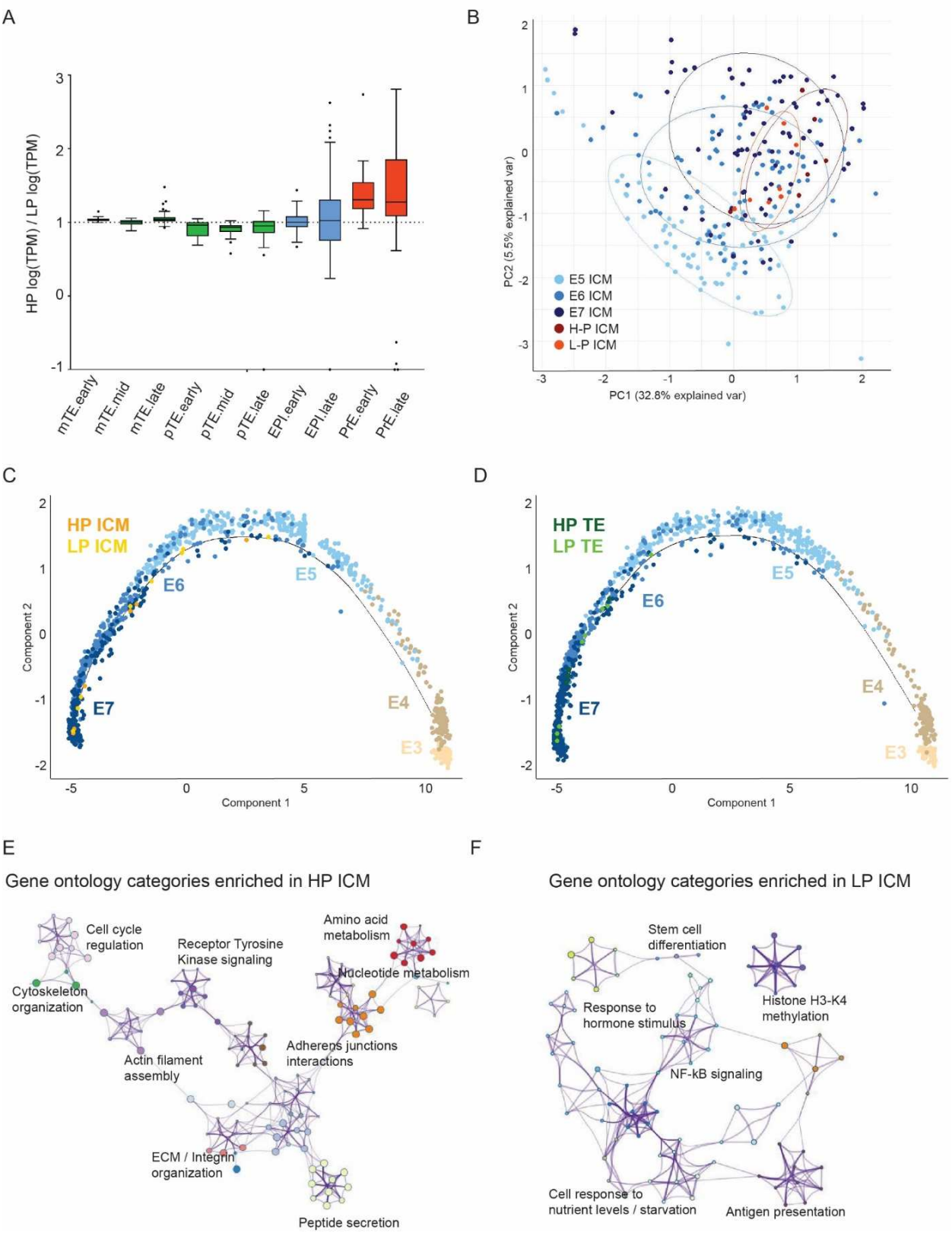

A
HP log(TPM) / LP log(TPM)
mTE.early
mTE.mid
mTE.late
pTE.early
pTE.mid
pTE.late
EPI.early
EPI.late
PrE.early
PrE.late
B
PC2 (5.5% explained var)
PC1 (32.8% explained var)
E5 ICM
E6 ICM
E7 ICM
H-P ICM
L-P ICM
C
HP ICM
LP ICM
E5
E6
E7
E4
E3
Component 1
Component 2
D
HP TE
LP TE
E5
E6
E7
E4
E3
Component 1
Component 2
E
Gene ontology categories enriched in HP ICM
Cell cycle regulation
Receptor Tyrosine Kinase signaling
Amino acid metabolism
Nucleotide metabolism
Cytoskeleton organization
Actin filament assembly
Adherens junctions interactions
ECM / Integrin organization
Peptide secretion
F
Gene ontology categories enriched in LP ICM
Stem cell differentiation
Histone H3-K4 methylation
Response to hormone stimulus
NF-kB signaling
Cell response to nutrient levels / starvation
Antigen presentation

Figure 4.9. Cell-cell communication patterns identified in human embryos of high implantation potential. (A) Using a ligand-receptor repository, we determined significant ($P<0.05$) cell-to-cell interactions within euploid embryos of high implantation potential. Interactions between the ICM and TE, as well as within each compartment, are plotted here, where cross-talk originating from the ICM (ligand expressed in the ICM) is plotted in yellow, and cross-talk originating from the TE (ligand expressed in the TE) is plotted in green. Interaction lines are directional, with the indented line representing the ligand (or "from") direction, and the non-indented line representing the receptor (or "to") direction. Gene bar sizes are proportional to the number of interactions that gene is involved with. (B) Interactions identified as significant within a single-cell human embryo dataset in which the cell types were known, that were also present in our HP embryo interaction set, are plotted here, allowing us to infer potential cell subset interactions within the human embryo. As in (A), interaction lines are directional, such that the line is indented at the ligand end. Interactions originating from the EPI, PE, pTE, and mTE, are blue, orange, light green, and dark green, respectively. (C) Examining significant interactions that were lost in LP relative to HP embryos provided insight into the mechanisms likely important for successful human embryo implantation and early ongoing pregnancy. Here, structural and signaling pathway interactions are plotted for HP embryos in grey, and significant interactions maintained in LP embryos are plotted with red interaction lines. The vast majority of structural interactions were lost in LP embryos, indicating a general loss of architectural stability. (D) Cell signaling interactions for the BMP, TGF-β, FGF, Notch, and Wnt pathways identified in euploid embryos of high implantation potential are shown here, with directionality. Communication patterns unique to HP embryos are shown in grey, whereas those that are maintained in LP embryos are in red.

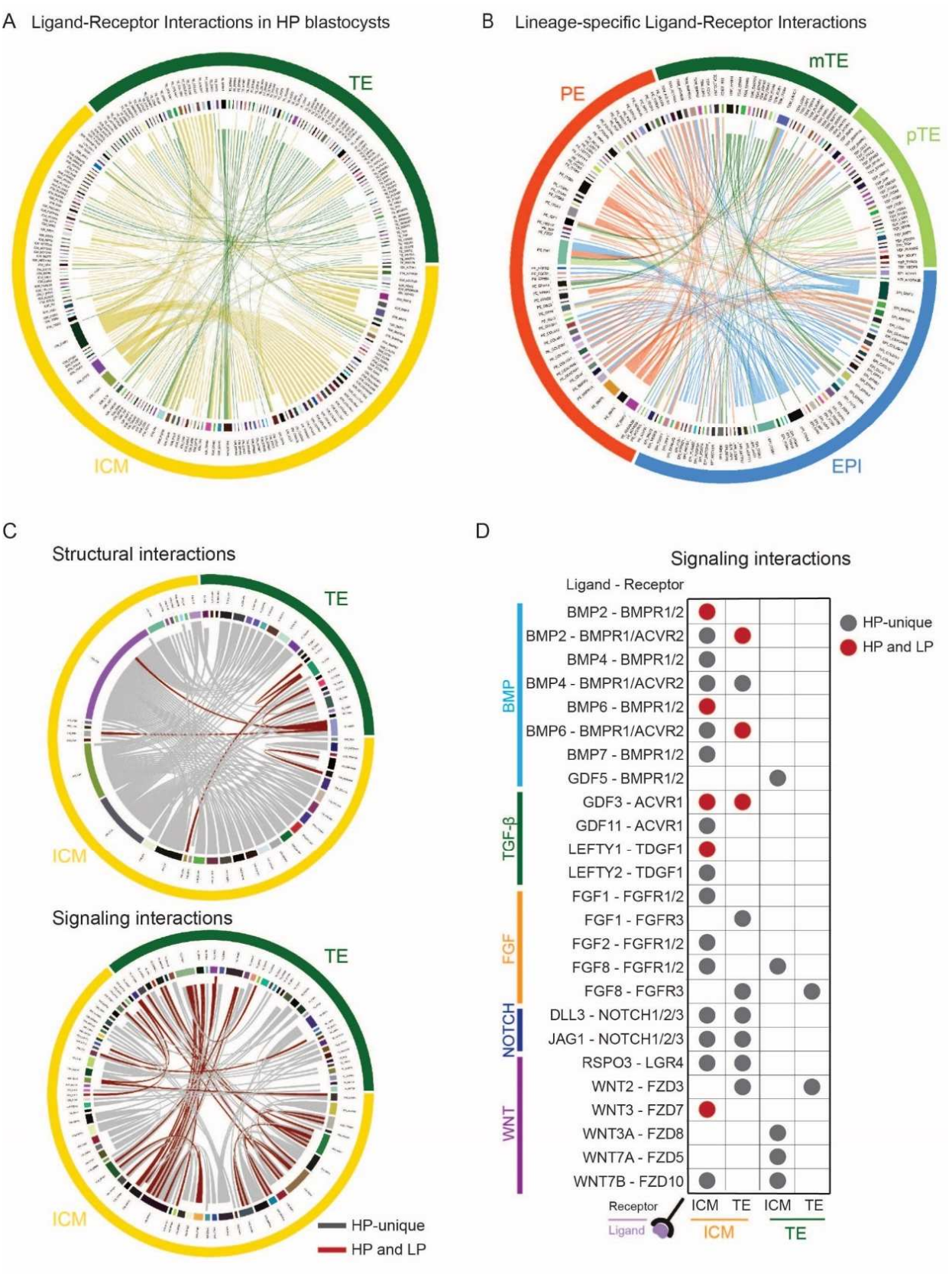
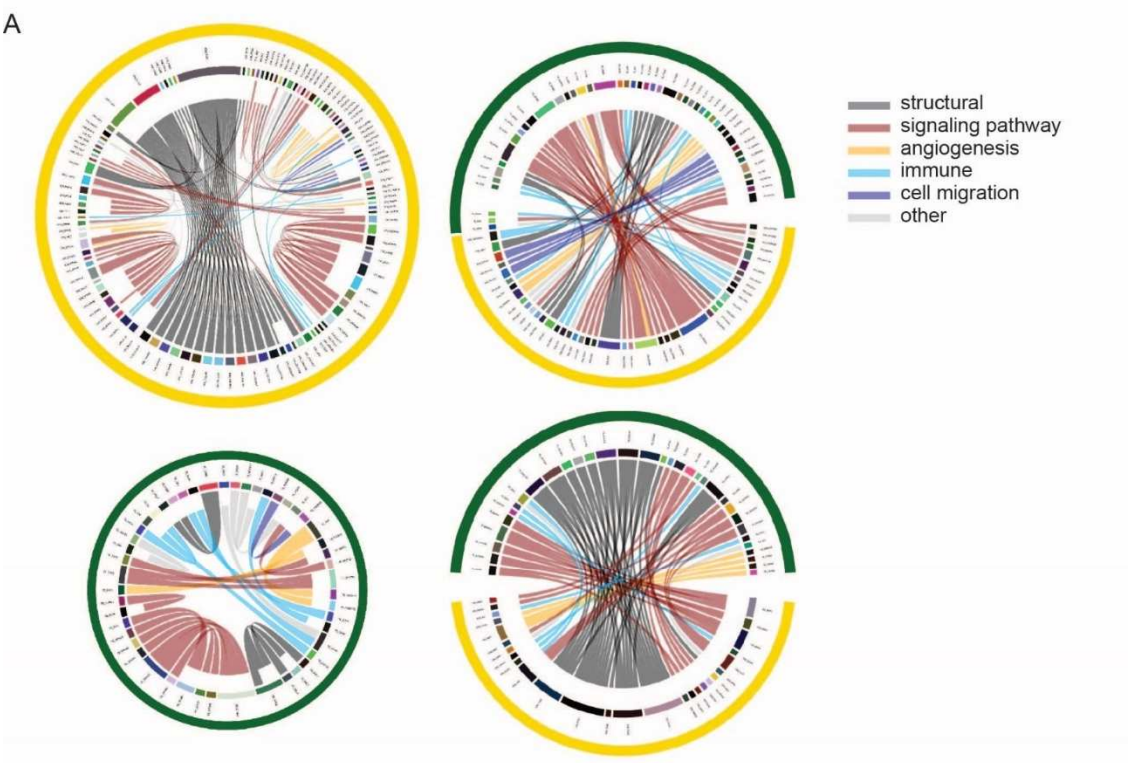
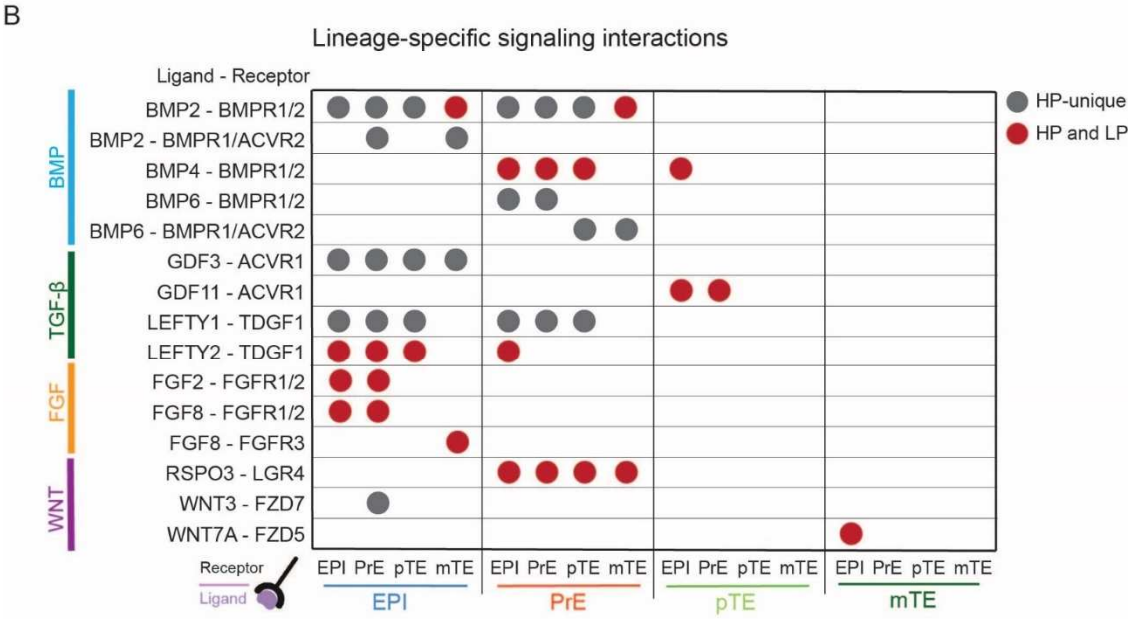
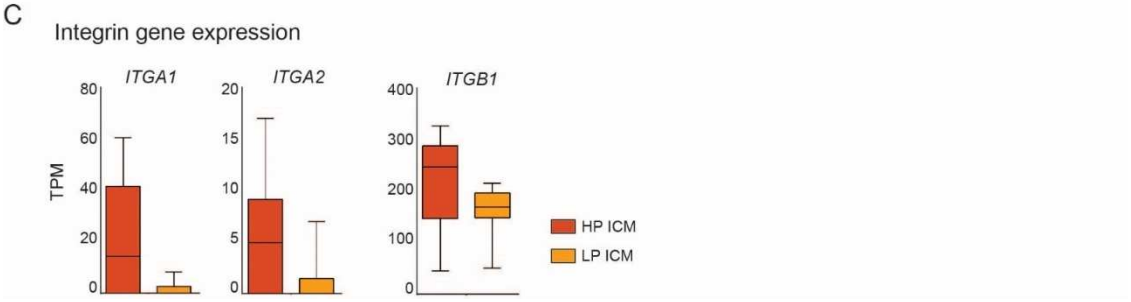

A Ligand-Receptor Interactions in HP blastocysts
TE
ICM
B Lineage-specific Ligand-Receptor Interactions
mTE
pTE
PE
EPI
C
Structural interactions
TE
ICM
Signaling interactions
TE
ICM
HP-unique
HP and LP
D
Signaling interactions
Ligand - Receptor
BMP2 - BMPR1/2
BMP2 - BMPR1/ACVR2
BMP4 - BMPR1/2
BMP4 - BMPR1/ACVR2
BMP6 - BMPR1/2
BMP6 - BMPR1/ACVR2
BMP7 - BMPR1/2
GDF5 - BMPR1/2
GDF3 - ACVR1
GDF11 - ACVR1
LEFTY1 - TDGF1
LEFTY2 - TDGF1
FGF1 - FGFR1/2
FGF1 - FGFR3
FGF2 - FGFR1/2
FGF8 - FGFR1/2
FGF8 - FGFR3
DLL3 - NOTCH1/2/3
JAG1 - NOTCH1/2/3
RSPO3 - LGR4
WNT2 - FZD3
WNT3 - FZD7
WNT3A - FZD8
WNT7A - FZD5
WNT7B - FZD10
BMP
TGF-β
FGF
NOTCH
WNT
HP-unique
HP and LP
Receptor
Ligand
ICM
TE
ICM
TE
ICM
TE

Figure 4.10. Cell-cell communication patterns identified as lineage-specific. (A) Main categories of cell compartment interactions identified in HP human embryos are presented here, with cell compartments separated into four interaction plots. Genes expressed in the ICM are surrounded by a yellow arc, whereas genes expressed in the TE are surrounded by a green arc. Broad categories of interactions identified include “structural”, “signaling pathway”, “angiogenesis”, “immune”, “cell migration”, and “other”, and each can be visualized by different colored interaction lines, as shown in the key. Intra-ICM interactions are shown on the top left, with indented interaction lines at the side of the ligand. TE-to-ICM, ICM-to-TE, and TE-to-TE interactions are plotted in the top right plot, bottom right plot, and bottom left plot, respectively. (B) Interactions identified as significant within a single-cell human embryo dataset in which the cell types were known, that were also present in our HP embryo interaction set, are plotted here, allowing us to infer potential cell subset interactions within the human embryo. This dot plot represents a subset of the interactions shown in Figure 4.9B – only signaling pathway interactions. Interactions maintained in LP embryos are shown in red, whereas signaling interactions unique to HP embryos are shown in grey. (C) Normalized read counts (TPM) for three of the main integrins (*ITGA1, ITGA2, ITGB1*) identified as important for embryonic structural stability within the literature and by cell compartment interaction analysis are plotted as box and whisker plots. All of them demonstrate lower expression in LP relative to HP ICM.


A
structural
signaling pathway
angiogenesis
immune
cell migration
other


B
Lineage-specific signaling interactions
Ligand - Receptor
BMP2 - BMPR1/2
BMP2 - BMPR1/ACVR2
BMP4 - BMPR1/2
BMP6 - BMPR1/2
BMP6 - BMPR1/ACVR2
GDF3 - ACVR1
GDF11 - ACVR1
LEFTY1 - TDGF1
LEFTY2 - TDGF1
FGF2 - FGFR1/2
FGF8 - FGFR1/2
FGF8 - FGFR3
RSPO3 - LGR4
WNT3 - FZD7
WNT7A - FZD5
BMP
TGF-β
FGF
WNT
HP-unique
HP and LP
Receptor
Ligand
EPI PrE pTE mTE
EPI
PrE
pTE
mTE


C
Integrin gene expression
ITGA1
ITGA2
ITGB1
TPM
HP ICM
LP ICM

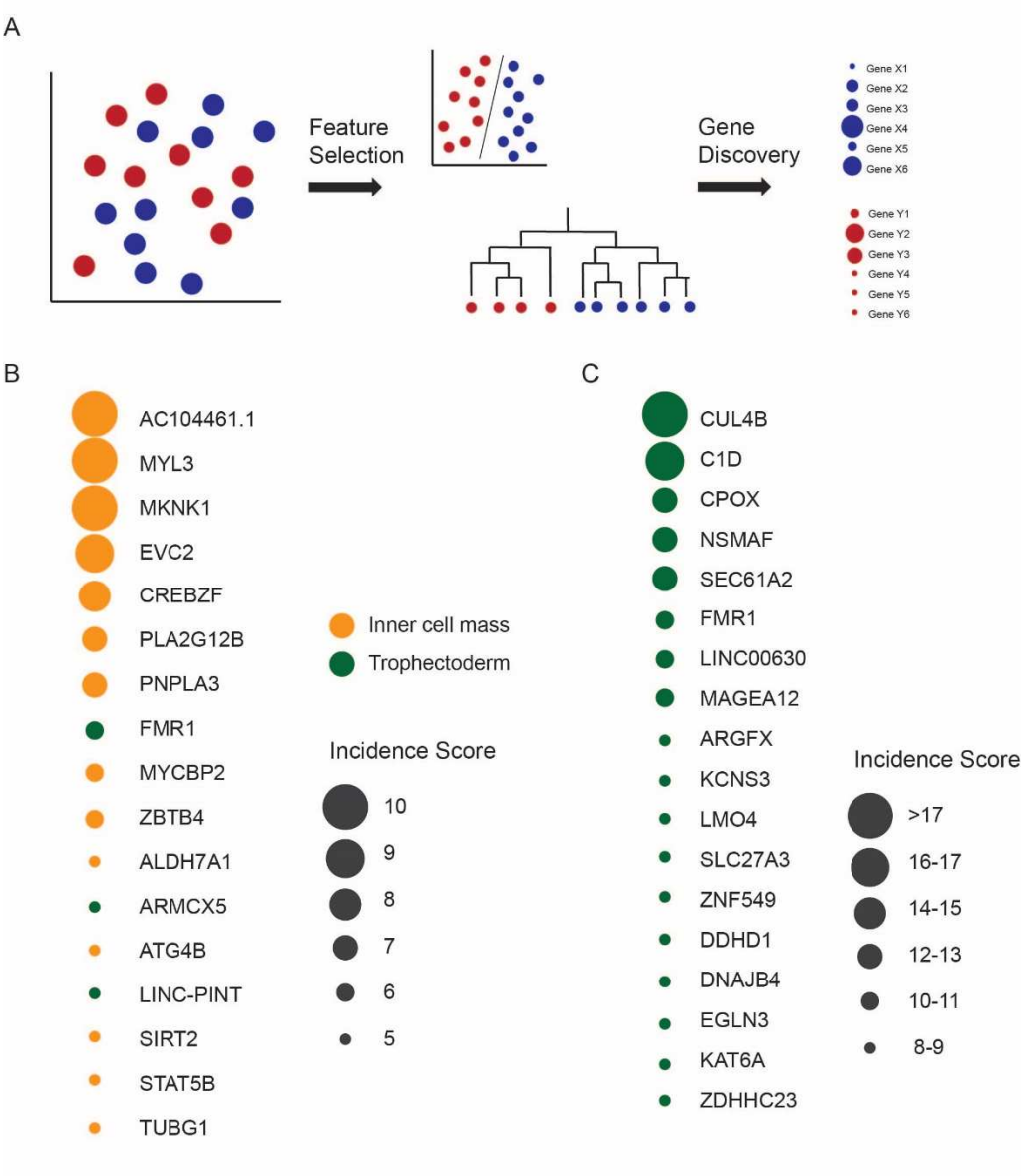


Figure 4.11. Machine learning identifies genes in both the inner cell mass and the trophectoderm that most highly correlate with implantation of euploid human blastocysts. (A) Schematic of our machine learning strategy for determining genes with the most predictive potential for implantation and early pregnancy success. Three algorithms were used (MLE, SVM, and LS), and genetic algorithm was applied as an optimization technique with feature selection. For each algorithm, the top ten iterations were chosen based on the smallest error in binary classification and only those that classified HP versus LP by hierarchical clustering (HC) were included. Final gene lists were compiled from the cumulative incidences of genes classified as predictive and assigned Incidence Scores (ISs). (B) Genes identified as most predictive from whole embryo (ICM and TE datasets combined) are presented here as a dot plot, in order of decreasing IS and colored by the compartment that the gene was expressed in (yellow for ICM and green for TE). (C) Genes identified as most predictive in the TE compartment are presented here as a dot plot in order of decreasing IS.

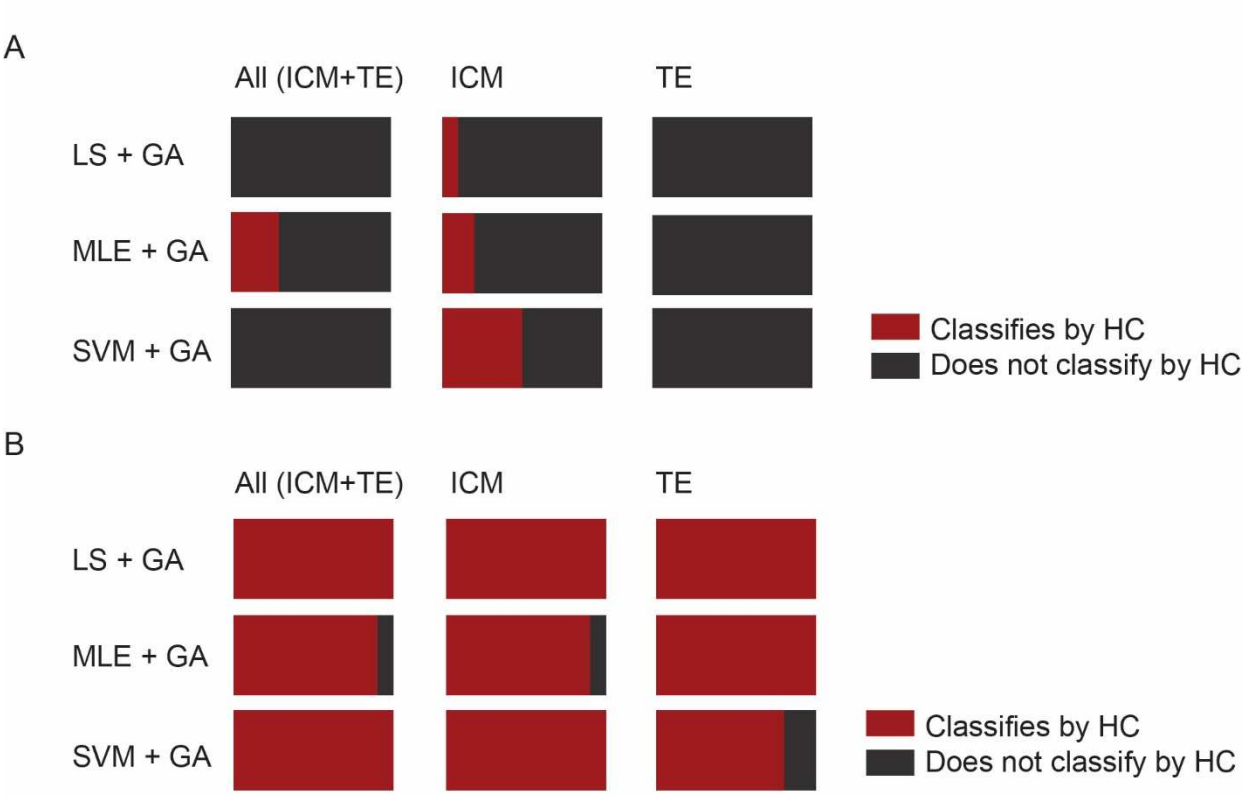


Figure 4.12. Machine learning demonstrates a higher predictive ability for the inner cell mass in determining implantation potential. (A) Three machine learning algorithms (LS, MLE, and SVM) were used in combination with a genetic algorithm (GA) on three datasets – whole embryo (All = ICM + TE), ICM, and TE – to determine genes that are most closely correlated with implantation of euploid human blastocysts. For each algorithm, the top ten iterations were chosen based on the smallest error in binary classification. The proportion of the top ten iterations that successful classified HP and LP embryo samples by hierarchical clustering (HC) is presented here, where red indicates that the iteration passed by HC, and dark grey indicates that it did not. For ICM datasets, at least a portion of the top ten iterations successfully classified HP versus LP ICM by HC for each algorithm. For the TE datasets, none of the algorithms were able to successfully classify HP versus LP TE. (B) As in (A), but the search space was optimized using differential expression analysis (Q<0.1), providing usable gene lists for all categories.

References

Altmä, S., ri Reimand, J., Hovatta, O., Zhang, P., Kere, J., Laisk, T., Saare, M., Peters, M., Vilo, J., Stavreus-Evers, A., and Salumets, A. (2012). Research Resource: Interactome of Human Embryo Implantation: Identification of Gene Expression Pathways, Regulation, and Integrated Regulatory Networks. J.R 14.

Angelov, A.G., and Ekström, M. (2018). Maximum likelihood estimation for survey data with informative interval censoring. *AStA Adv. Stat. Anal.* 1–20.

Bader, G.D., and Hogue, C.W. V (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4, 2.

Blakeley, P., Fogarty, N.M.E., Valle, I., Wamaitha, S.E., Hu, X., Elder, K., Snell, P., Christie, L., and Robson, P. (2015). Defining the Three Cell Lineages of the Human Blastocyst by Single-Cell RNAseq. *Development* 142, 3151–3165.

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

Cauffman, G., De Rycke, M., Sermon, K., Liebaers, I., and Van de Velde, H. (2009). Markers that define stemness in ESC are unable to identify the totipotent cells in human preimplantation embryos. *Hum. Reprod.* 24, 63–70.

Chazaud, C., and Yamanaka, Y. (2016). Lineage specification in the mouse preimplantation embryo. *Development* 143, 1063–1074.

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.

Chen, A.E., Egli, D., Niakan, K., Deng, J., Akutsu, H., Yamaki, M., Cowan, C., Fitz-Gerald, C., Zhang, K., Melton, D.A., and Eggan, K. (2009). Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell* 4, 103–106.

Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, N.S., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., and Jaffrézic, F. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.* 14, 671–683.

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinformatics* 51, 11.14.1-11.14.19.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

Dumdie, J.N., Cho, K., Ramaiah, M., Skarbrevik, D., Mora-Castilla, S., Stumpo, D.J., Lykke-Andersen, J., Laurent, L.C., Blackshear, P.J., Wilkinson, M.F., and Cook-Andersen, H. (2018). Chromatin Modification and Global Transcriptional Silencing in the Oocyte Mediated by the mRNA Decay Activator ZFP36L2. *Dev. Cell* 44, 392–402.e7.

Gardner, D.K., Phil, D., Lane, M., Stevens, J., Schlenker, T., and Schoolcraft, W.B. (2000). Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertil. Steril.* 73, 1155–1158.

Gardner, D.K., Phil, D., Lane, M., Stevens, J., and Schoolcraft, W.B. (2001). Noninvasive assessment of human embryo nutrient consumption as a measure of developmental potential. *Fertil Steril.* 76, 1175–1180.

Gardner, D.K., Wale, P.L., Collins, R., and Lane, M. (2011). Glucose consumption of single post-compaction human embryos is predictive of embryo sex and live birth outcome. *Hum. Reprod*. 26, 1981–1986.

Gardner, D.K., Meseguer, M., Rubio, C., and Treff, N.R. (2015). Diagnosis of human preimplantation embryo viability. *Hum. Reprod. Update* 21, 727–747.

Ghaheri, A., Shoar, S., Naderan, M., and Hoseini, S.S. (2015). The Applications of Genetic Algorithms in Medicine. *Oman Med. J.* 30, 406–416.

Graham, S.J.L., Wicher, K.B., Jedrusik, A., Guo, G., Herath, W., Robson, P., and Zernicka-Goetz, M. (2014). BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. *Nat. Commun*. 5, 5667.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812.

Guo, H., and Wang, W. (2019). Granular support vector machine: a review. *Artif. Intell. Rev.* 51, 19–32.

Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 18, 675–685.

Heikamp, K., and Bajorath, J. (2014). Support vector machines for drug discovery. *Expert Opin. Drug Discov.* 9, 93–104.

Houghton, F.D., Hawkhead, J.A., Humpherson, P.G., Hogg, J.E., Balen, A.H., Rutherford, A.J., and Leese, H.J. (2002). Non-invasive amino acid turnover predicts human embryo developmental capacity. *Hum. Reprod.* 17, 999–1005.

Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y., and Xu, W. (2018). Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics* 15, 41–51.

Kalista, T., Austinn Freeman, H., Behr, B., Pera, R.R., and Scott, C.T. (2011). Cell Stem Cell Forum Donation of Embryos for Human Development and Stem Cell Research. *Stem Cell* 8, 360–362.

Kuijk, E.W., van Tol, L.T.A., Van de Velde, H., Wubbolts, R., Welling, M., Geijsen, N., and Roelen, B.A.J. (2012). The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. *Development* 139, 871–882.

Lee, N.K., Li, X., and Wang, D. (2018). A comprehensive survey on genetic algorithms for DNA motif prediction. *Inf. Sci.* 466, 25–43.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Li, S., Kang, L., and Zhao, X.-M. (2014). A survey on evolutionary algorithm based hybrid intelligence in bioinformatics. *Biomed Res. Int.* 362738.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.

Liu, J., He, X., Corbett, S.A., Lowry, S.F., Graham, A.M., Fassler, R., and Li, S. (2009). Integrins are required for the differentiation of visceral endoderm. *J. Cell Sci*. 122, 233–242.

Liu, L., Yin, Y., Li, Y., Prevedel, L., Lacy, E.H., Ma, L., and Zhou, P. (2012). Cul4b is essential for mouse embryonic development 1258 Essential role of the CUL4B ubiquitin ligase in extra-embryonic tissue development during mouse embryogenesis. *Nat. Publ. Gr*. 22, 1258–1269.

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

Makrigiannakis, A., Vrekoussis, T., Zoumakis, E., Kalantaridou, S.N., and Jeschke, U. (2017). The Role of HCG in Implantation: A Mini-Review of Molecular and Clinical Evidence. *Int. J. Mol. Sci.* 18.

Manning, T., Sleator, R.D., and Walsh, P. (2013). Naturally selecting solutions. *Bioengineered* 4, 266–278.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17, 10.

May, R.C., Ibrahim, J.G., and Chu, H. (2011). Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits. *Stat. Med.* 30, 2551–2561.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 40, 4288–4297.

Moore, R., Tao, W., Smith, E.R., and Xu, X.-X. (2014). The Primitive Endoderm Segregates from the Epiblast in beta-1 Integrin-Deficient Early Mouse Embryos. *Mol Cell Biol.* 34(3), 560-72.

Mora-Castilla, S., To, C., Vaezeslami, S., Morey, R., Srinivasan, S., Dumdie, J.N., Cook-Andersen, H., Jenkins, J., and Laurent, L.C. (2016). Miniaturization Technologies for Efficient Single-Cell Library Preparation for Next-Generation Sequencing. *J. Lab. Autom.* 21(4):557-67.

Murohashi, M., Nakamura, T., Tanaka, S., Ichise, T., Yoshida, N., Yamamoto, T., Shibuya, M., Schlessinger, J., and Gotoh, N. (2009). An FGF4-FRS2α-Cdx2 Axis in Trophoblast Stem Cells Induces BMP4 to Regulate Proper Growth of Early Mouse Embryos. *Stem Cells* 28, N/A-N/A.

Nepomnaschy, P.A., Weinberg, C.R., Wilcox, A.J., and Baird, D.D. (2008). Urinary hCG patterns during the week following implantation. *Hum. Reprod.* 23, 271–277.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457.

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

De Paepe, C., Aberkane, A., Dewandre, D., Essahib, W., Sermon, K., Geens, M., Verheyen, G., Tournaye, H., and Van de Velde, H. (2017). BMP4 plays a role in apoptosis during human preimplantation development. *Mol. Reprod. Dev.* 86, 53–62.

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

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., Schroth, G.P., and Sandberg, R. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* 30, 777–782.

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC Bioinformatics* 12, 480.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.

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

La Rosa, I., Camargo, L.S., Pereira, M.M., Fernandez-Martin, R., Paz, D.A., and Salamone, D.F. (2011). Effects of bone morphogenic protein 4 (BMP4) and its inhibitor, Noggin, on in vitro maturation and culture of bovine preimplantation embryos. *Reprod. Biol. Endocrinol*. 9, 18.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504.

Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.-E., Stephenson, E., Polański, K., Goncalves, A., Gardner, L., Holmqvist, S., Henriksson, J., Zou, A., Sharkey, A.M., Millar, B., Innes, B., Wood, L., Wilbrey-Clark, A., et al. (2018). Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* 563, 347–353.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).

Wilcox, A.J., Weinberg, C.R., O'Connor, J.F., Baird, D.D., Schlatterer, J.P., Canfield, R.E., Armstrong, E.G., and Nisula, B.C. (1988). Incidence of Early Loss of Pregnancy. *N. Engl. J. Med*. 319, 189–194.

Yamanaka, Y., Lanner, F., and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* 137, 715–724.

Zhao, W., Boyett, J.M., Kocak, M., Ellison, D.W., and Wu, Y. (2011). Maximum likelihood estimation of reviewers' acumen in central review setting: categorical data. *Theor. Biol. Med. Model.* 8, 3.

Chapter 5

Discussion

Global transcriptional silencing in the mouse oocyte is mediated by ZFP36L2

Here, we demonstrate a critical role for the RNA decay activator ZFP36L2 in regulating histone modification and bringing about transcriptional silencing in the mouse oocyte. ZFP36L2 acts, at least in part, to stabilize mRNAs encoding histone demethylase (KDM) mRNAs, resulting in the accumulation of H3K4me3 and H3K9me3, as well as global transcriptional silencing. More recent studies suggest that these histone methylation marks are at least partially reversed in the embryo at the time of transcriptional activation (zygotic genome activation or ZGA) (Dahl et al., 2016; Zhang et al., 2016). An obvious next step would be to determine the role of demethylation at H3K4 and H3K9 in mediating transcriptional re-activation in the embryo, and if this is accomplished via the selective upregulation of KDM mRNAs. Even more interesting would be if KDM mRNA upregulation is accomplished in part through reduced activity of ZFP36L2. The fact that re-activation of transcription must happen in the absence of new transcripts being formed makes epigenetic marks an even more plausible mechanism by which this transition might be mediated. In fact, studies have demonstrated that marks present at H3K4 and H3K9 in the oocyte at the time of global transcriptional silencing are retained during reprogramming and might contribute to the epigenetic profile of the embryo (Dahl et al., 2016; Zhang et al., 2016). That these epigenetic marks might actually label the first genes to be transcribed at ZGA is an exciting and plausible scenario.

In addition to a role for ZFP36L2 in mediating global transcriptional silencing, this factor likely plays important roles at other stages of oocyte growth. For example, our data suggested a role for ZFP36L2 in mediating preantral follicle growth, as evidenced by follicle counts in sexually immature cKO females following stimulation with exogenous gonadotropin (Figure 2.2F-G). These stages of oocyte growth are difficult to study. However, it will be interesting to determine if ZFP36L2 plays a role in survival of the oocyte at these very early stages of growth, potentially by affecting the ability of the oocyte to respond to gonadotropin stimulation or to communicate with surrounding supporting cells. These studies will require knock-out of

ZFP36L2 at earlier stages of oocyte growth, likely at the primordial follicle, and *in vitro* follicle culture methods to survey the effect of loss of this factor on oocyte and follicle growth and maturation.

Finally, it will be interesting to determine how ZFP36L2 itself is regulated in the oocyte. If its activity is needed for global transcriptional silencing and a reduction in activity is necessary for selective upregulation of KDM mRNAs associated with reactivation of transcription in the embryo, it presumably must be regulated in a complex and robust manner. Members of the ZFP36 family are known to be regulated post-translationally by phosphorylation and contain upwards of ten phosphorylation sites with demonstrated roles in mediating decay activity (Sanduja et al., 2011). In particular, ZFP36 (TTP) has been shown to be a downstream target of phosphorylation through a number of major signaling pathways, including ERK MAPK, p38 MAPK, JNK, GSK3β, PKA, PKB/AKT, and PKC (Sanduja et al., 2011). Thus, many signals originating from the oocyte itself, or from external cues (from surrounding supporting cells, for example) could be translated into phosphorylation or dephosphorylation to reduce or promote mRNA decay, respectively

Pluripotent gene expression in the inner cell mass of the mouse blastocyst and in mouse embryonic stem cells is regulated by Upf2

These data in Chapter 3 suggest a critical role for the RNA decay activator UPF2 in regulating the expression of ICM- and EPI-specific genes, many of which have been demonstrated to mediate pluripotency both *in vivo* and *in vitro,* in the mouse blastocyst (Blakeley et al., 2015; Chazaud and Yamanaka, 2016; Nichols and Smith, 2009; Ohnishi et al., 2013). In response to the downregulation of a large set of these pluripotency-associated genes in *Upf2*-null early blastocysts, a reduction in the total cell number of EPI cells in *Upf2*-null E3.5 blastocysts is observed (Figure 3.3 and 3.7). This reduction in cell number could be explained by a reduction in proliferation of this cell type, which is reflected in the reduced proliferation rate

of ESCs derived from *Upf2*-null blastocysts (Figure 3.8). Human ESCs with reduced levels of *Upf1,* another key factor of the NMD pathway, have been demonstrated to stall in G1 of the cell cycle (Lou et al., 2016). We found *Anapc16*, which controls progression through G1, to be both stabilized and upregulated in the absence of *Upf2* in the mouse blastocyst. Since the phenotypic defects we observed seemed to be specific for the pluripotent population of cells within the blastocyst, with no obvious defects demonstrated in extraembryonic cell types, it will be interesting to see, (i) if UPF2 and the NMD pathway help mediate progression through the cell cycle at G1, and (ii) if this role is specific for pluripotent cells or if it can be generalized to extraembryonic, or even other differentiated cell types.

At E3.25, *Upf2*-null blastocysts already displayed a reduction in the expression of many genes that have been found to be ICM- and EPI-specific, including core pluripotency factors such as *Nanog, Pou5f1 (Oct4), c-Myc, Klf4,* and *Fgf4* (Figure 3.6B and 3.7). Since UPF2 is an RNA decay factor, this downregulation would either need to be an indirect effect of UPF2's mRNA decay activity or the effect of another function of UPF2 that is independent of NMD. If the downregulation of pluripotency-associated mRNAs is indirect, UPF2 and the NMD pathway likely function to degrade an mRNA (or set of mRNAs) that encode proteins, such as transcription factors, that positively regulate the gene expression of these mRNAs. By definition, such "intermediate regulatory factors" would be encoded by mRNAs that are upregulated and/or stabilized in *Upf2*-null embryos. Some of the factors that fit this description are proteins involved in pluripotency, including SALL4, SUZ12, ESRRB, RAD23B, and MED1. A particularly interesting candidate is SALL4, a transcription factor required for proliferation of the ICM in the mouse blastocyst (Sakaki-Yumoto et al., 2006). That these mRNAs were also stabilized by the loss of *Upf2* in the blastocyst makes them likely direct targets of UPF2 and NMD. Future studies will directly test this possibility *in vitro.* In addition, NMD factors are known to have cellular roles in additional to their RNA decay activity, such as telomere maintenance (Azzalin et al., 2007).

Future studies will be needed to determine whether SALL4 and the other candidates act as intermediate regulatory factors downstream of NMD.

Here, we show that *Upf2*-null ESCs derived under naïve pluripotency conditions, upon transition to slightly more primed states, undergo spontaneous differentiation as evidenced by morphological assessment and alkaline phosphatase staining (data not shown). Loss of other NMD factors has been shown to dictate stem versus differentiation decisions in other cell types, in both human and mouse systems (Bruno et al., 2011; Li et al., 2015; Lou et al., 2016). Thus, another possibility is that UPF2 and the NMD pathway act as a cellular buffer in stem cells, allowing them to robustly transition from pluripotent to more differentiated states, by degrading pools of RNAs important for this transition. The process of differentiation requires a coordinated series of actions within a cell, to downregulate pools of mRNAs important for the more primitive state and to upregulate mRNAs associated with the differentiated cell type. One hypothesis is that selective mRNA decay helps to orchestrate these developmental transitions. As more studies directly examine the role of RNA decay at the transition from stem to differentiated states, this will hopefully become clearer.

Gene expression patterns associated with human blastocyst implantation

This study, to our knowledge, is the first transcriptomic comparison of embryonic lineages within euploid human blastocysts of clinically variant implantation potential. It represents a rich resource for the field, characterizing gene expression patterns and cell-to-cell communication networks present in human embryos of high implantation potential, as well as those lost in embryos of clinically decreased pregnancy potential. We define clear sets of genes and signaling patterns that are most closely correlated with implantation potential of human embryos of known ploidy and quality. These genes represent the starting point for future studies aimed at understanding human embryo implantation in more tractable model organisms and in

cell culture systems and will likely shed light on human-specific processes necessary for a successful pregnancy.

Most surprising was that loss of the PrE lineage best predicts embryos of low implantation potential. We demonstrate that, although HP and LP ICM samples have similar expression of mTE-, pTE-, and EPI-specific lineage marker genes, that HP ICM samples express PrE-specific genes at a much higher level (Figures 4.6 and 4.8). This is true for both early and late markers of the PrE lineage, and not at the expense of the maturation of the embryo as a whole, or a detectable defect in the maturation of any other lineage (Figures 4.6 and 4.8), suggesting a lineage-specific defect of the PrE in LP embryos. In addition, the identification of cell-cell communication patterns revealed that the structural and ECM interactions that characterize HP embryos are essentially lost in embryos of low implantation potential. Even more interesting is that mouse blastocysts devoid of this structural integrity fail to form a PrE epithelium (Liu et al., 2009; Moore et al., 2014). Whether an intra-embryo structural network or the PrE cellular program must be put in place first for the other to be carried out is uncertain. However, the lack of structural integrity could, at least in part, explain the differences observed during morphological assessment and could be a visual indication of an unestablished PrE.

An interesting future question is if this PrE-specific defect could be rescued, either by extended culture, allowing this lineage to “catch up”, or by adjusting IVF culture conditions, likely with media containing factors important for PrE proliferation and/or maturation. With respect to the former, it is unclear if more time in culture would affect the development of the PrE or if it would simply narrow the window of implantation, maturing the TE and EPI compartments to a state incompatible with implantation itself. In addition, the morphological indications of a “poor” embryo (namely less tightly packed cells) are not associated with an early state of development, indicating that blastocysts of poor morphology are not simply delayed, but that they are missing key molecular signals necessary for successful growth, maturation, and subsequently,

implantation. A more plausible approach might be to adjust culture conditions of the embryo, following *in vitro* fertilization, such that all the factors necessary for correct PrE establishment are present. Some signaling components likely important, as we have shown here, could be a combination of BMP, FGF, and Wnt stimulation (Figures 4.9 and 4.10). That selective addition of a few signals necessary for the maturation of the PrE lineage could enhance implantation rates for the many embryos graded as “poor” in IVF clinics is an exciting possibility.

Using an unbiased machine learning approach, we identified a number of genes, both in the ICM and the TE, that most closely correlate with implantation potential (Figure 4.11). Future studies in hESC-based systems, as well as in traditional mouse models, will shed light on the role of these factors in human pre-implantation development and potentially on their role in implantation itself. While the function of these genes will be useful for our general understanding of human-specific development and for mechanistic studies moving forward, we have additionally defined lists of genes highly correlated with implantation potential in the TE as candidates for clinical diagnosis of embryo quality and potential. In theory, just 1-5 cells from a biopsy already taken during PGS could be used to choose the single best embryo for transfer, enhancing success rates and decreasing complications associated with multiple transfers. Future studies with more samples might directly test the gene panels identified here in verification and validation steps. Further, TE biopsies sequenced from the same embryo undergoing embryo transfer and followed for the duration of the pregnancy will provide the most direct assessment of genes correlated with implantation. These panels, using the approach we have defined here, may one day replace the standard blastocyst morphological assessment for choosing the single best embryo for transfer that has been in place for decades.

References

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318, 798–801.

Blakeley, P., Fogarty, N.M.E., Valle, I., Wamaitha, S.E., Hu, X., Elder, K., Snell, P., Christie, L., and Robson, P. (2015). Defining the Three Cell Lineages of the Human Blastocyst by Single-Cell RNAseq. *Development* 142, 3151–3165.

Bruno, I.G., Karam, R., Huang, L., Bhardwaj, A., Lou, C.H., Shum, E.Y., Song, H.-W., Corbett, M.A., Gifford, W.D., Gecz, J., Pfaff, S.L., and Wilkinson, M.F. (2011). Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. *Mol. Cell* 42, 500–510.

Chazaud, C., and Yamanaka, Y. (2016). Lineage specification in the mouse preimplantation embryo. *Development* 143, 1063–1074.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., Preissl, S., Jermstad, I., Haugen, M.H., Suganthan, R., Bjørås, M., Hansen, K., Dalen, K.T., Fedorcsak, P., Ren, B., et al. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* 537, 548–552.

Li, T., Shi, Y., Wang, P., Guachalla, L.M., Sun, B., Joerss, T., Chen, Y.-S., Groth, M., Krueger, A., Platzer, M., Yang, Y.-G., Rudolph, K.L., and Wang, Z.-Q. (2015). Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. *EMBO J.* 34, 1630–1647.

Liu, J., He, X., Corbett, S.A., Lowry, S.F., Graham, A.M., Fassler, R., and Li, S. (2009). Integrins are required for the differentiation of visceral endoderm. *J. Cell Sci.* 122, 233–242.

Lou, C., Dumdie, J., Goetz, A., Shum, E., Brafman, D., Liao, X., Mora-Castilla, S., Ramaiah, M., Cook-Andersen, H., Laurent, L., and Wilkinson, M. (2016). Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate. *Stem Cell Reports* 6, 844–857.

Moore, R., Tao, W., Smith, E.R., and Xu, X.-X. (2014). The Primitive Endoderm Segregates from the Epiblast in beta-1 Integrin-Deficient Early Mouse Embryos. *Mol Cell Biol.* 34(3), 560-72.

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

Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., Oleś, A.K., Araúzo-Bravo, M.J., Saitou, M., Hadjantonakis, A.-K., and Hiiragi, T. (2013). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol.* 16, 27–37.

Sakaki-Yumoto, M., Kobayashi, C., Sato, A., Fujimura, S., Matsumoto, Y., Takasato, M., Kodama, T., Aburatani, H., Asashima, M., Yoshida, N., and Nishinakamura, R. (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* 133, 3005–3013.

Sanduja, S., Blanco, F.F., and Dixon, D.A. (2011). The roles of TTP and BRF proteins in regulated mRNA decay. *Wiley Interdiscip. Rev. RNA* 2, 42–57.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., Liu, W., Kou, X., Zhao, Y., He, W., Li, C., Chen, B., Li, Y., Wang, Q., Ma, J., Yin, Q., Kee, K.,

Meng, A., Gao, S., Xu, F., Na, J., Xie, W. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature 537, 553–557.