UNIVERSITY OF CALIFORNIA,
IRVINE

Molecular Mechanisms Regulating Extraembryonic Endoderm Lineage Potential

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in

Biological Sciences

by

Paula Duyen Anh Pham

Dissertation Committee:
Professor Ken W.Y. Cho, Chair
Professor Kyoko Yokomori
Associate Professor Wenqi Wang

2022

DEDICATION

In loving memory of my grandparents and my cat Chloe.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

LIST OF ABBREVIATIONS

AVE, anterior visceral endoderm

ChIP, chromatin immunoprecipitation

CREs, Cis-regulatory elements

cXEN or XEN, chemical-induced / extraembryonic endoderm stem cells

DVE, distal visceral endoderm

E, embryonic day

EC, Embryonal carcinoma

Epi, Epiblast

E-P, Enhancer-promoter

EMT, epithelial to mesenchymal transition

emVE, embryonic visceral endoderm

ExE, extraembryonic ectoderm

ExEn, Extraembryonic Endoderm

exVE, extraembryonic visceral endoderm

GRN, gene regulatory network

GO, gene ontology

ICM, Inner cell mass

KO, knock out

mESCs, mouse embryonic stem cells

(MID) Hi-C, minimized interruption and double-digestion

mTE, mural trophectoderm

Oct4, Pou5f1

PE, Parietal Endoderm

PROTAC, Proteolysis targeting chimeric

PrE, Primitive Endoderm

PYS, Parietal yolk sac

RA, Retinoic acid

RM, Reichert's Membrane

TE, Trophectoderm or Trophoblast

TF, transcription factor

TPM, transcript per million

VE, Visceral Endoderm

VYS, Visceral yolk sac

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor and mentor, Dr. Ken W. Y. Cho, for his guidance, and support for me to pursue interdisciplinary research projects in this lab. He provided the necessary foundation by providing me the research opportunities and resources to train me into a better scientist. One major lesson I have learned from Dr. Cho is how to focus and narrow down to tackle a research problem and to consider various models and caveats to my work. I am grateful for his mentorship to show me how to lead a project independently.

I would like to thank my dissertation committee Dr. Wenqi Wang, and Dr. Kyoko Yokomori for their helpful research input, discussion, and support. I also would like to thank my advancement committee Dr. Grant Macgregor and Dr. Zeba Wunderlich for their inputs in my research direction and emphasizing the importance of formulating tractable research aims and goals. I am also thankful to Dr. Thomas F. Schilling for general support.

The text of Appendix II in this dissertation is a reprint of the material as it appears in "Ma, N., Mochel, N.R.d., Pham, P.D. *et al.* Label-free assessment of pre-implantation embryo quality by the Fluorescence Lifetime Imaging Microscopy (FLIM)-phasor approach. *Sci Rep* **9**, 13206 (2019). https://doi.org/10.1038/s41598-019-48107-2", used with permission from SpringerNature. The co-authors listed in this publication are Ning Ma, Soledad Reyes De Mochel,, Michelle Digman, Tommy Yoo, and Ken W.Y. Cho. The co-authors Michelle Digman and Ken W.Y. Cho listed in this publication directed and supervised research which forms the basis for the thesis/dissertation. (Note: An official permission and copyright statement from the publisher is attached to this manuscript and it can be found on page 173).

VITA

Paula Duyen Anh Pham

Education

2016-2022 Ph.D. in Biological Sciences, University of California, Irvine
Dissertation Title: 'Molecular Mechanisms Regulating Extraembryonic Endoderm Stem Cell Lineage Potential'

2016-2021 M.S. in Biological Sciences, University of California, Irvine

2014-2016 B.S. in Biochemistry and Cell Biology, University of California, San Diego

Research Positions

2017-2022 Graduate Student Researcher, University of California, Irvine

2015-2016 Undergraduate Researcher, University of California, San Diego

2015-2016 Research Intern, The Scripps Research Institute

Publications

Pham, P.D.[†], Lu, H.[†] , Han, H., Zhou, J.J., Madan, A., Wang W., Murre, C., & Cho, K.W.Y. Transcription factor network instructs lineage divergence and plasticity in extraembryonic endoderm **(Manuscript Submitted).**

Lu, H.[†], **Pham, P.D.**[†] , Han, H., Zhou, J.J., Wang W., Murre, C., & Cho, K.W.Y. Enhancer Cis Elements Instruct Enhancer-Promoter Interaction, Enhancer-Promoter Insulation and Compartmental Segregation. **(Manuscript In Preparation).**

Zhou, J.J., **Pham, P.D.**, Han, H., Wang, W., Cho, K.W.Y. Foxh1 engages in chromatin regulation revealed by protein interactome analyses. *Development Growth and Differentiation* (July 2022).

Ma, N.[†], Mochel, N.R.d.[†], **Pham, P.D.**[†], Yoo, T.Y., Cho, K.W.Y., Digman, M.A. Label-free assessment of pre-implantation embryo quality by the Fluorescence Lifetime Imaging Microscopy (FLIM)-phasor approach. Sci Rep 9, 13206 (2019). [†]**Co-First, Equal Contribution**

Professional Teaching Experience:

o Head Teaching Assistant, *Genetics*, UCI, Fall 2021

o Teaching Assistant, *Eukaryotic Human Genetics*, UCI, Fall 2020, Fall 2019

o Teaching Assistant, *Developmental Biology*, UCI, Winter 2020, Summer 2019, Winter 2019, Summer 2018, Winter 2018

o Administrative Teaching Assistant, *DNA to Organisms*, UCI, Fall 2018

o Undergraduate Instructional Assistant, *Molecular Biology*, UCSD, Summer 2016, Winter 2016

o Undergraduate Instructional Assistant, *Metabolic Biochemistry*, UCSD, Summer 2015

o Undergraduate Instructional Assistant, *Organismic and Evolutionary Biology*, UCSD, Fall 2015, Spring 2015

ABSTRACT OF THE DISSERTATION

Molecular Mechanisms Regulating Extraembryonic Endoderm Stem Cell Lineage Potential

by

Paula Duyen Anh Pham

Doctor of Philosophy in Biological Sciences

University of California Irvine, 2022

Professor Ken W. Y. Cho, Chair

The separation of embryonic and extraembryonic lineages is a process that has evolved in amniotes to adapt for fetus survival on land. One of the major roles of the extraembryonic tissues is to provide nutritive support sourced from the maternal environment, since development occurs *in utero*. The initial differentiation of the extraembryonic lineages occurs first prior to embryonic cell differentiation to prioritize cell survival. Thus past studies have extensively focused on understanding these lineage segregation mechanisms which take place in three major waves of lineage differentiation events. The key commonality of these events begins with a bipotent progenitor that co-expresses lineage-specific transcription factors (TFs), which is resolved by mutual inhibition of TFs and key signaling pathways. During the first lineage breaking event, differential Hippo signaling activation and cross-antagonism of POUF51(OCT4) (ICM TF) and CDX2 (TE TF) forms the trophectoderm (TE) and the inner cell mass (ICM). Likewise, in the second lineage breaking event, differential Fgf signaling activation and cross-antagonism of NANOG (Epi TF) and GATA6 (PrE TF) regulates the ICM bifurcation to the epiblast (Epi) and primitive endoderm (PrE). Next, the extraembryonic trophectoderm differentiate to facilitate the implantation process. Concurrently, during this rapid and dynamic event, the PrE differentiates

into the parietal endoderm (PE) and the visceral endoderm (VE). However, less is known about the regulation of these extraembryonic endoderm (ExEn) lineages partly due to inaccessibility of these cells during implantation. Therefore, the molecular basis of PrE cell fate divergence to PE or VE and the concerted actions of TFs that moderate PE- versus VE-cell gene regulatory programs remains to be elucidated.

To investigate the regulative behavior of ExEn cells, we analyzed scRNA-seq datasets of E4.5 mouse embryos. We delineated PrE, PE and VE cell states and resolved TFs associated with PE or VE. Comparative motif analysis of the enhancer repertoires of VE cells and PE cells suggests that GATA6, SOX17, and FOXA2 are core TFs in the ExEn gene regulatory network. To test this model, we performed transcriptome analyses on cXEN cells using degron tagged GATA6 and SOX17 combining with *FoxA2* knockout revealed that PE development requires positive GATA6 and SOX17 inputs, whereas VE development requires FOXA2 to activate VE gene program suppressed by GATA6 or SOX17. Next, we found that BMP signaling cues instruct the PE-VE lineage decision. These data reveal a core gene regulatory module that underpins PE and VE cell fate choice. Lastly, we compare GATA4 and GATA6 functional activities in PE cells, which suggest they have overlapping functions in ExEn development. Overall, this dissertation will present a characterization of the ExEn gene regulatory network (GRN), while also proposing new models and regulators that govern ExEn development.

CHAPTER 1:

Transcription factors and signaling pathways regulating extraembryonic endoderm development

Introduction:

The Molecular Events Leading to the Formation of the Primitive Endoderm

Developing mammalian embryos have evolved to form *in utero* to ensure safe fetal growth and maturation through its interactions with extraembryonic tissues. Upon the successful fertilization of the oocyte by sperm, the mouse zygote undergoes multiple rounds of cell cleavages to prepare for the first lineage segregation event beginning at embryonic day 2.5 (E2.5) (Figure 1.1) (Mihajlović and Bruce, 2017). Notably, differentiation of the first extraembryonic tissue, the trophectoderm (TE), begins by E2.75 (Figure 1.1) (Mihajlović and Bruce, 2017). A major morphological change can be observed by E3.5 where a fluid filled cavity, known as the blastocoel, forms (Figure 1.1) (Manejwala et al., 1989; Mihajlović and Bruce, 2017; Wiley, 1984). At this time, the TE becomes specified as an epithelial layer enveloping the bipotent inner cell mass (ICM) (Figure 1.1) (Mihajlović and Bruce, 2017). The main function of the TE at this time is to influx Na+ and fluid to form a blastocoel cavity marking the formation of the early blastocyst (Manejwala et al., 1989; Mihajlović and Bruce, 2017; Wiley, 1984). Molecular frameworks of these dynamic events have been carefully dissected through the use of *in vivo* and *in vitro* experimentations.
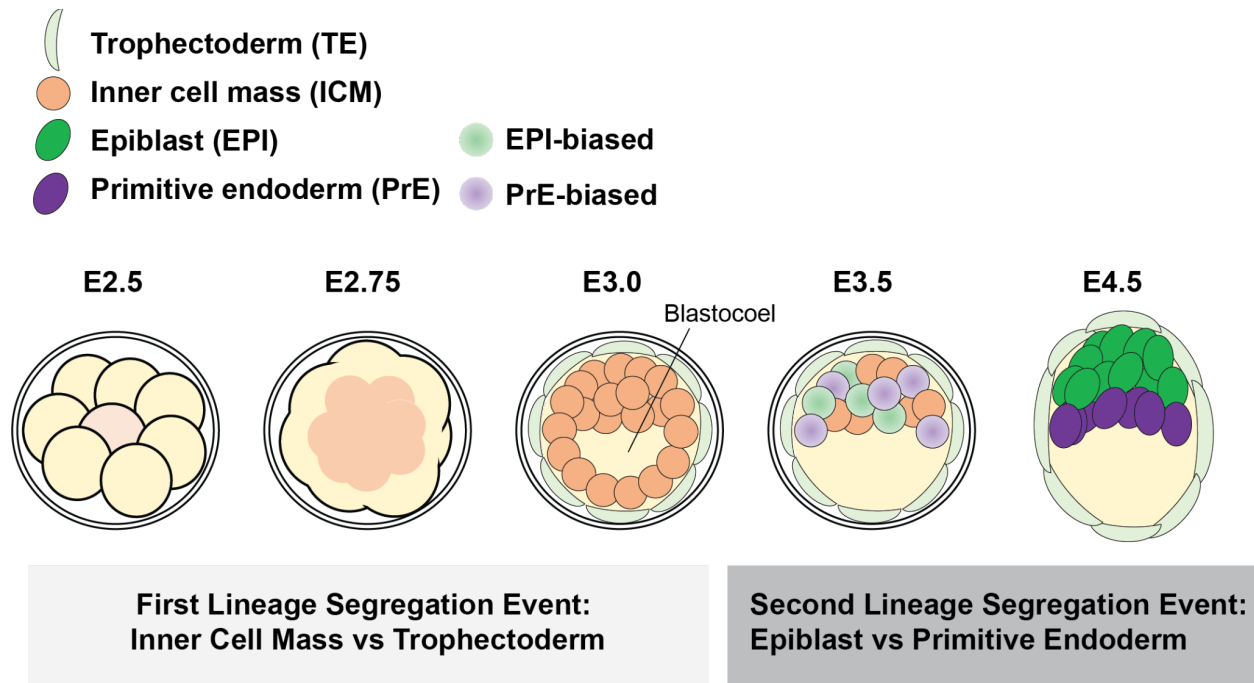
Figure 1.1. Preimplantation embryo development stages. Around the 8-cell stage (E2.5), the embryo initiates compaction by E2.75 to form distinct inside and outside cell positions. Next, the first lineage segregation event proceeds. By E3.0 (early blastocyst stage), the inner cell mass and trophectoderm cells are specified. Next, starting at E3.5 (mid blastocyst stage), the inner cell mass cells undergo the second lineage segregation event. The final stage of preimplantation embryo development, E4.5 (mature blastocyst stage), there are now two distinct tissues which are specified as the epiblast and primitive endoderm.

First Lineage Segregation Event: Inner Cell Mass vs Trophectoderm

The first cell lineage segregation requires antagonistic interactions of two master transcription factors (TFs), *Pou5f1* (*Oct4*) (positive ICM regulator) and *Cdx2* (positive TE regulator). The model of OCT4-CDX2 antagonism was first discovered through the study of mouse mutant embryos (Nichols et al., 1998). The initially characterized mutant embryos formed blastocysts, but the cells comprising the would-be ICM displayed the TE-like identity (Nichols et al., 1998). Further elaboration of this relationship was examined using a conditional, doxycycline *Oct4* knockout system. Mouse embryonic stem cells (mESCs) that lack OCT4 lost the expression of embryonic pluripotent markers (e.g., *Sox2*, *Rex-1*, and *Otx-1*) and ectopically activated *Cdx2* and obtained TE-like features (Niwa et al., 2000). Such phenotypic observations led to the proposal that OCT4 was an active repressor for the TE gene regulatory network (GRN). Likewise, opposing regulatory roles can be attributed to the master regulator for TE cell fate, *Cdx2* (Jedrusik et al., 2015; Strumpf et al., 2005). Strong parallels can be seen among *Cdx2* deficient embryos, where ectopic expression of embryonic pluripotency markers (e.g. *Nanog* and *Oct4*) is detected in the trophectoderm structures (Jedrusik et al., 2015; Strumpf et al., 2005). Additionally, overexpression of *Cdx2* in mESCs transdifferentiated the cells from the embryonic pluripotent cell state to the TE cell state (Blij et al., 2015; Huang et al., 2017; Nishiyama et al., 2009; Niwa et al., 2005). Collectively, these observations support the notion that CDX2 antagonizes OCT4 regulated GRN.

Given the molecular evidence supporting for the CDX2 and OCT4 antagonism, the association of these TFs to their cis-regulatory elements (CREs) was directly examined by ChIP-qPCR and ChIP-seq approaches (Chen et al., 2009; Huang et al., 2017; Niwa et al., 2005; Xiong et al., 2022). The previous ChIP experiments have proposed several regulatory molecular

mechanisms that allow for the activation and maintenance of the TE or embryonic pluripotent cell states. First, self-activation of *Cdx2* or *Oct4* expressions was shown through ChIP experiments demonstrating that they can bind to their own promoters. Next, cross-repression of *Cdx2* or *Oct4* expressions was shown through their reciprocal binding on each other promoters (Chen et al., 2009; Huang et al., 2017; Niwa et al., 2005; Xiong et al., 2022). Lastly, genome-wide analyses of CDX2 and OCT4 occupancy revealed that CDX2 and OCT4 can also reciprocally bind at each other's lineage specific genes associated with the embryonic pluripotent and TE cell identity (Chen et al., 2009; Huang et al., 2017; Niwa et al., 2005; Xiong et al., 2022). Collectively, the binding profiles of CDX2 or OCT4 and loss-of-function studies of *Cdx2* or *Oct4* conducted via *in vivo* and *in vitro* model systems demonstrate that these factors activate their GRNs while repressing each other GRNs. Thus, OCT4 and CDX2 were mutually antagonistic at the level of cis-regulatory interaction. These master TFs were co-expressed prior to the ICM-TE lineage commitment, and their protein expression levels are refined during this developmental period to produce the early blastocyst by E3.5 (Figure 1.1) (Dietrich and Hiiragi, 2007; Holmes et al., 2017; Niwa et al., 2005; Rossant and Tam, 2009). Blastocyst development is robust as most mouse embryos form proper blastocysts at about the same time after fertilization. The robust patterning mechanism to produce the blastocyst structure raises the question about the upstream signals instructing the lineage segregation of the totipotent cell to the TE and ICM cell fates.

Multiple signaling inputs regulate the TE and ICM bifurcation, which comprises a complex network of signaling crosstalks (Mihajlović and Bruce, 2017). The defining signaling pathway that confers the major effect for TE and ICM lineage segregation is Hippo signaling (Nishioka et al., 2009). Differential Hippo signaling inputs were activated upon initial

positioning and polarity of the cells which occurs around the E2.5 or the 8-cell stage (Hirate et al., 2013; Hirate et al., 2015). These embryos undergo the first morphological change, known as "compaction", where the cells increase cell-to-cell adhesion contacts (Mihajlović and Bruce, 2017). Although the inducing signal that initiates compaction has yet to be identified, this key morphological event sets up a distinct outer cell morphology where cells become polarized (Mihajlović and Bruce, 2017). These cell polarization events are guided by intracellular processes that form the apical domain, which are regulated by polarity complex, PAR3-PAR6-aPKC (Alarcon, 2010; Hirate et al., 2013; Hirate et al., 2015; Kono et al., 2014; Plusa et al., 2005). The polarity complex localizes a key regulator for Hippo signaling to the apical domain known as *Amot* (Hirate et al., 2013; Leung and Zernicka-Goetz, 2013). Further cell-to-cell contacts in the basolateral domains mediated by CDH1 form the epithelial-like outer layer, which surround the apolar inner cells (Cockburn et al., 2013; Stephenson et al., 2010). These key differences influence the cellular characteristics of blastocyst cells - the inner apolar (ICM) cells where AMOT is localized at its adherens junctions and the outer TE cells at its apical domains (Hirate et al., 2013; Leung and Zernicka-Goetz, 2013). These morphological changes induce the embryo cells to differentially respond to Hippo signaling cues, thereby driving the key events for ICM and TE lineage segregation.

Hippo signaling is turned off when the apically localized AMOT prevents the phosphorylation of YAP1, a key transcriptional cofactor for TEAD4 TF (Hirate et al., 2013; Home et al., 2012; Kaneko and DePamphilis, 2013; Leung and Zernicka-Goetz, 2013; Nishioka et al., 2009). This allows for YAP1 to become localized to the nucleus where it is now readily available to form the YAP1/TEAD4 transcriptional complex to activate the TE gene program (Home et al., 2012; Kaneko and DePamphilis, 2013; Nishioka et al., 2009). In the reciprocal

case, activation of Hippo signaling occurs in the apolar inner cells where AMOT is localized at adherens junctions (Home et al., 2012). AMOT-(associated adherens junctions) acts with the LATS1 kinase to phosphorylate YAP1 to prevent its nuclear translocation (Hirate et al., 2013; Leung and Zernicka-Goetz, 2013). Therefore, Hippo signaling cue serves as a key signal where transcriptional complexes of YAP/TEAD become selectively activated in the outer cells to fuel *Cdx2* expression for TE cell fate commitment, while inner cells lack YAP/TEAD activation. In sum, the co-expressed CDX2-OCT4 cell state in the totipotent cell can be resolved through a combination of signal integration and interaction of CDX2 and OCT4 that regulate TE and ICM GRNs.

Second Lineage Segregation Event: Epiblast vs Primitive Endoderm

Following the establishment of the TE and ICM, the next wave of lineage differentiation occurs in ICM bifurcation to epiblast (EPI) and primitive endoderm (PrE) to form the mature blastocyst by E4.5 (Figure 1.1) (Bedzhov et al., 2014a). A similar regulatory paradigm discussed earlier also guides this second lineage differentiation event. Around E3.5 stage, ICM cells initially co-expresses lineage-specific TFs, NANOG (EPI TF) and GATA6 (PrE TF) in a "salt-and-pepper"-like fashion, where some ICM cells are high in NANOG expression relative to GATA6 and vice versa (Bedzhov et al., 2014a). The ICM cells that have high NANOG expression are known as EPI-biased cells, likewise, the cells that have high GATA6 expression are known as PrE-biased cells (Bedzhov et al., 2014a). These co-expressed cell states are further sorted out by E4.5 where cell fate commitment is complete and the EPI cells comprise a mass of cells that are lined by the epithelial PrE (Figure 1.1) (Bedzhov et al., 2014a).

Multiple regulatory principles have been dissected to demonstrate various mechanisms that can influence EPI versus PrE cell fate choice. The most simplest model is the mutual

repression of *Nanog* and *Gata6* through regulating each other's GRNs (Chickarmane and Peterson, 2008; Schröter et al., 2015). One part of this model can be explained through NANOG affecting the expression of *Gata6* and vice versa (Schröter et al., 2015). This simple TF model was explored by mouse mutants of *Nanog*, which were able to undergo the first lineage differentiation wave regulated by OCT4-CDX2, but not the second wave where *Nanog* deficient embryos failed to specify an epiblast (Messerschmidt and Kemler, 2010; Mitsui et al., 2003). Interestingly, the ICM cells underwent apoptosis, and consequently, the PrE cells were not formed (Messerschmidt and Kemler, 2010; Mitsui et al., 2003). However, a closer examination of these mutant embryos showed that the remaining, surviving ICM cells had defaulted to the PrE cell state as revealed by the expression of GATA4 (Messerschmidt and Kemler, 2010). This observation suggests that a non-cell autonomous mechanism, which was later revealed to be FGF4 ligand secretion from EPI-biased cells, is present to promote the PrE cell fate where the EPI-biased cells prevents the apoptosis of PrE precursors (Kang et al., 2013; Messerschmidt and Kemler, 2010). Additionally, *in vitro* embryo explant cultures of these *Nanog* deficient embryos showed complete conversion to PrE or parietal endoderm-like cell identity based on morphology analysis (Messerschmidt and Kemler, 2010; Mitsui et al., 2003). Taken together, the role of NANOG is to actively repress the PrE differentiation program in the bipotent ICM cells and promote the PrE cell fate through a non-cell autonomous mechanism mediated by FGF4 secretion from EPI-biased cells to presumptive PrE cells (Messerschmidt and Kemler, 2010). Moreover, *in vitro* TF binding studies using CUT&RUN showed that NANOG binds to *Gata6* CREs and GATA6 binds to CREs of *Nanog*. This binding process is akin to what was observed in the OCT4-CDX2 binding model (Thompson et al., 2022).

Additional supporting studies pointed to the key role of FGF signaling positively promoting the PrE cell gene program (Bessonnard et al., 2014; Chazaud et al., 2006; Hamilton et al., 2019; Kang et al., 2013; Kang et al., 2017; Molotkov et al., 2017; Yamanaka et al., 2010). Application of high amounts of FGF4 to early embryos (around E2.75 to E3.0 cell stage) can direct all ICM cells to adopt the PrE cell fate suggesting that the PrE gene program is tightly controlled by FGF signaling (Kang et al., 2013). Additionally, FGF4 ligand mutants can be rescued by providing external sources of FGF4, demonstrating the requirement of FGF4 signaling to maintain the PrE gene program (Kang et al., 2013). However, FGF signaling does not induce *Gata6* expression upon its first detection at E2.5, suggesting the involvement of another unidentified signal for initial activation of *Gata6* (Kang et al., 2013). The role of FGF is to promote maintenance and survival of the PrE lineage, while GATA6 is considered the primary specification factor for the PrE (Cai et al., 2008; Kang et al., 2013; Koutsourakis et al., 1999; Schrode et al., 2014). Evidence to demonstrate that GATA6 is the primary regulator of PrE cell fate was demonstrated through mouse embryos deficient for *Gata6* lack a PrE, resembling that of *Fgf4* knockout mutants (Cai et al., 2008; Koutsourakis et al., 1999; Schrode et al., 2014). Furthermore, in the absence of GATA6, the cells had defaulted to the epiblast cell fate evidenced by a pan-expression of NANOG (Cai et al., 2008; Koutsourakis et al., 1999; Schrode et al., 2014).

To elaborate on the downstream signaling pathway of FGF signaling, an *in vitro* cell-based system was designed to tune ERK activity (the downstream signaling effector for FGF signaling) (Hamilton et al., 2019). Their results demonstrate that ERK activity affects the differential association of the transcriptional machinery with pluripotent specific enhancers that promote the EPI gene program versus MAPK-associated enhancers that promote the PrE gene

program (Hamilton et al., 2019). However, ERK activity did not influence key pluripotent TF binding (e.g., EP300, ESRRB, and SOX2) suggesting that FGF signaling influences EPI versus PrE stochastic gene expression dynamics by modulating activities of lineage-specific enhancers (Hamilton et al., 2019). To conclude, an intricate tristable network of NANOG, GATA6, and FGF signaling inputs shapes ICM cells to gradually form PrE and EPI structures (Bessonnard et al., 2014; Hamilton et al., 2019). This brings us to the third wave of lineage differentiation (Figure 1.2).

Efforts to study this third wave (Figure 1.2) have been hampered since the embryo initiates implantation, where it burrows into the uterus, making it difficult for *in vivo* observations and isolation (Bedzhov et al., 2014a). Here I focus primarily on differentiation of PrE cells during the time of implantation. The differentiation models for trophectoderm cells during implantation have been extensively discussed elsewhere (Aplin and Kimber, 2004; Bedzhov et al., 2014a; Red-Horse et al., 2004; Soncin et al., 2015). In this next section, I will present an overview of the ExEn lineages, followed by potential models of PrE differentiation from pre- to peri-implantation stages. Afterwards, I will discuss the core functions of their derivative lineages and their differentiation mechanisms by their respective TFs and signaling molecules.
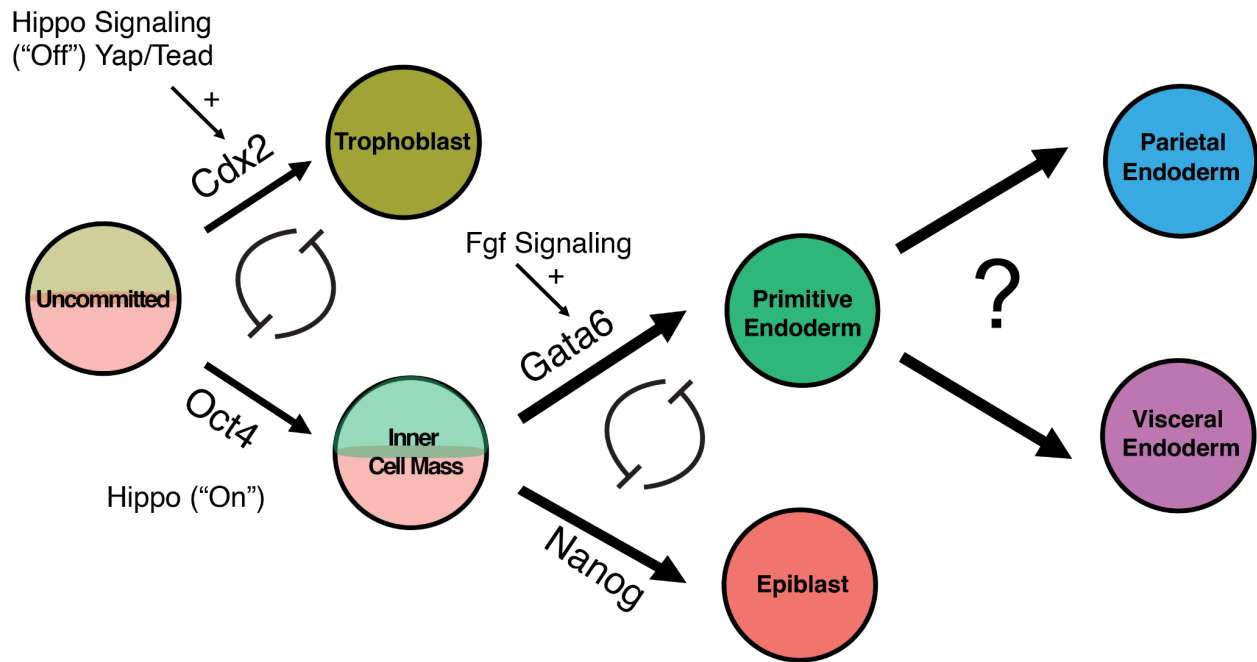
Figure 1.2. Summary of signaling and transcription factors that regulate the lineage segregation events from pre to peri-implantation development. The diagram represents the major lineage segregation events following the fertilization of the oocyte by sperm. The first two major lineage differentiation events relate to 1) ICM/TE segregation followed by 2) EPI/PrE. The third major lineage differentiation event involves both TE and PrE differentiation. Here we focus primarily on the PrE lineage differentiation to PE/VE for the work presented in this dissertation. Abbreviations: Inner cell mass (ICM), Trophectoderm/Trophoblast (TE), Primitive endoderm (PrE), Epiblast (EPI), Parietal Endoderm (PE), Visceral Endoderm (VE).

The "Black Box" of Embryo Development: What is known about the PrE differentiation during the implanting blastocyst stage?

Overview of Extraembryonic Endoderm Development

Mammalian embryonic development has evolved to use extraembryonic tissues which serve essential support functions during implantation and through gestation. These temporary organs are necessary for the growth of the embryo proper by facilitating proper nutrient supply. Of the various extraembryonic lineages that can arise from the early preimplantation stages, the primitive endoderm (PrE) gives rise to two sub-lineages, the parietal endoderm (PE), and the visceral endoderm (VE) (Figure 1.3). PE and VE are extraembryonic membranes that envelop the embryo proper for nutrient transport and support implantation before differentiating into the parietal and visceral yolk sac for supporting gestation (Figure 1.3).

Both PE and VE have distinctive characteristics that differentiate these two sub-lineages where the VE is epithelial while the PE is mesenchylmal (Filimonow and Fuente, 2021; Rossant, 1995). Additionally, the VE, which differentiates further to the embryonic VE (emVE) and extraembryonic VE (exVE) (Figure 1.3), has been found to have additional functions in patterning of the epiblast by acting as an inductive tissue for anterior-posterior axis establishment and differentiation of the hematopoietic lineages for the yolk sac vasculature (Filimonow and Fuente, 2021). Interestingly, it has also been shown that the VE can be incorporated into the definitive endoderm showing that these extraembryonic cells can contribute to the embryo proper (Kwon et al., 2008; Nowotschin et al., 2019).

Establishment of the PrE and subsequent differentiation of these sub-lineages relies on extrinsic signaling and downstream molecular cues which is not well-characterized. Isolation of these cells from embryonic cells *in utero* can be difficult since differentiation of these lineages

occur during implantation and are often present in limited quantities. Features that define the PrE

and its derivatives from embryonic lineages are the expression of key endodermal TFs (e.g.

Gata6, *FoxA2*, *Sox17*, *Hnf4a*, etc.) (Wamaitha et al., 2015). However, the relative contribution of

these TFs to the endodermal cell state is not clear since perturbation of these TFs affects the cell

state of the progenitor PrE cells.
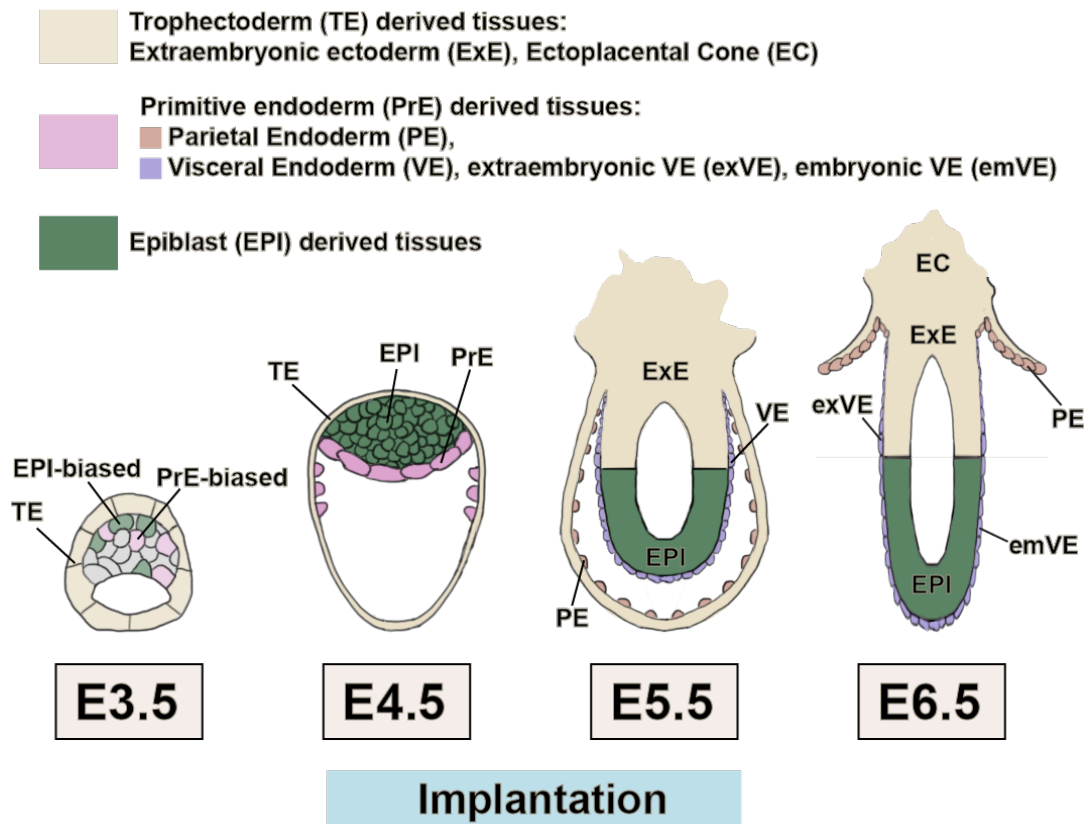
Figure 1.3. Formation of primitive endoderm derived tissues from E3.5 to E6.5.

Establishment of epiblast and primitive endoderm tissues occurs between E3.5 to E4.5. The primitive endoderm differentiates as the blastocyst prepares for implantation between E4.5 to E5.5 to form the parietal endoderm and visceral endoderm. The parietal endoderm lineage is maintained and lined along the trophectoderm derived cells while the visceral endoderm differentiates further to form the extraembryonic visceral endoderm (which lines the epiblast first, then the trophectoderm derived cells by E5.5) and the embryonic visceral endoderm (which lines the epiblast cells). Abbreviations: TE (Trophectoderm), EPI (Epiblast), PrE (Primitive Endoderm), PE (Parietal Endoderm), VE (Visceral Endoderm), exVE (Extraembryonic Visceral Endoderm), emVE (Embryonic Visceral Endoderm), EC (Ectoplacental Cone), ExE (Extraembryonic Ectoderm).

Specification of the Parietal Endoderm and its Functions

Functional Roles of the Parietal Endoderm

The first appearance of the PE was indicated by cells delaminating from the epithelial PrE layer overlaying the epiblast (Filimonow and Fuente, 2021). This morphological event marks the first epithelial to mesenchymal transition (EMT) event to occur during embryo development as these PE cells migrate along the mural TE to prepare for implantation (Filimonow and Fuente, 2021). Earlier ultrastructural studies were done on pre- to post-implantation embryos to dissect the general functions of PE upon its initial emergence in the blastocyst to mid-gestation (Adamson and Ayers, 1979; COOPER et al., 1981; Lane et al., 1983; Salamat et al., 1993; Salamat et al., 1995; Semoff et al., 1982). The PE main function is to form Reichert's Membrane (RM) which refers to a specialized basement membrane structure that structurally supports the embryo and acts as a permeable filter for nutrient and gas exchange between fetal cells and uterine tissues (Adamson and Ayers, 1979; COOPER et al., 1981; Lane et al., 1983; Salamat et al., 1993; Salamat et al., 1995; Semoff et al., 1982; Ueda et al., 2020).

Upon E4.5 blastocyst formation, early signs of potential PE differentiation were evidenced by a weak detection of possible extracellular matrix (ECM) material (Salamat et al., 1995). The ECM components became more prominent during the peri-implantation period where a thick structural compartment cushioned between the mural TE and the migratory PE cells were found (Salamat et al., 1995). Biochemical experiments further categorized this thick extracellular matrix meshwork as a specialized basement membrane enriched in laminin, collagen, and glycoproteins by comparative staining of early embryo stages and assessing parietal tissues in mid-gestation (Adamson and Ayers, 1979; Clark et al., 1975; COOPER et al., 1981; Hogan et al., 1980; Hogan et al., 1982; Salamat et al., 1995; Smith and Strickland, 1981; Wallingford et al.,

2013). Next, the appearance of RM was linked to PE's ability to produce these proteins

demonstrated through isolated *in vitro* experiments using PE/PE-like cells (Adler et al., 1990;

Clark et al., 1975; Fowler et al., 1990; Hogan et al., 1982; Howe and Solter, 1980; Oshima,

1981; Stary et al., 2005; Stary et al., 2006). Additionally, the RM is continuously remodeled as

the embryo continues to grow (Behrendtsen and Werb, 1997). With the continued growth of the

embryo, the PE secretes ECM proteins alongside plasminogen and metalloproteinase proteins to

remodel and expand the RM during the egg-to-cylinder transition (Behrendtsen and Werb, 1997;

Cheng and Grabel, 1997; Smith and Strickland, 1982; Snyder et al., 1992; Strickland et al.,

1976).

The function of PE and RM has been overlooked since its functions can be dispensable

for embryonic lineage differentiation as demonstrated by *ex vivo* differentiation of mouse

embryonic stem cells (mESCs) (Keller, 2005). However, recent research further points to the RM

as an important membrane for embryo morphogenesis within the *in utero* environment (Miner et

al., 2004; Ueda et al., 2020). For instance, upon implantation on the abembryonic pole of the

embryo, the embryo undergoes an egg-to-cylinder elongation through its proximo-distal axis, in

the post-implantation stages (Ueda et al., 2020). However, in the *in utero* environment, high and

variable intrauterine pressures are caused by naturally occurring contractions from the uterine

horn (Ueda et al., 2020). These contractions expose the embryo to periodic mechanical forces

that affect its morphogenesis (Ueda et al., 2020). Therefore, the presence of the RM shields and

protects the embryo from external forces to allow for proper egg-to-cylinder elongation to ensure

normal gastrulation (Miner et al., 2004; Ueda et al., 2020).

Indeed, embryos defective in key extracellular matrix components excreted by the PE

cells, are malformed and had a high level of peri-implantation lethality rate, likely due to embryo

resorption (Miner et al., 2004). To summarize, the RM continues its structural support functions and nutritional functions by acting as a filter for nutrient and gas exchange with the maternal tissues until the placenta, which is derived from the trophoblast cells, can be fully formed. Upon placenta formation, the PE becomes the parietal yolk sac (PYS) and continues to maintain the RM by encasing the embryo through gestation. Future studies may include examining the overall contribution of the mechanical cues by RM on extraembryonic structure morphogenesis.

Specification of the Parietal Endoderm: Transcription Factors and Signaling Pathways

Little is known about the transcription factors regulating parietal endoderm specification. However, immunohistochemistry analyses have reported loss of OCT4 expression, but high levels of GATA6, SOX7, and SOX17 are maintained preferentially in PE cells (Wallingford et al., 2013). While the expression of GATA6 is lost in the VE cells gradually, GATA4 is expressed comparably in both PE and VE cell types (Cai et al., 2008). Expression of SOX17 and SOX7 is higher in the PE in general, with variable expression levels in the different subtypes of the VE (e.g. emVE versus exVE) (Artus et al., 2011). Therefore, most of the knowledge about the PE cell state regulation is inferred only from the expression of key transcription factors. Recent single cell analyses comparing transcriptomes of the progenitor cell to the emerging cell type has nominated *Myc* as a potential PE regulator (Qiu et al., 2022). Further interrogation and targeting of TFs will aid in uncovering the function of PE TFs instructing PE cell fate.

The instructive signals promoting PE differentiation have primarily been explored through mouse knockout phenotypes and the use of F9 teratocarcinoma or embryonal carcinoma (EC) cell lines (Strickland et al., 1980). Therefore, key signaling pathways of PE are yet to be revealed in other ExEn stem cell lines that are more closely related to its *in vivo* counterparts such as XEN cells or PrESCs (Niakan et al., 2013; Ohinata et al., 2022). Here, I will review key

pathways promoting PE-like characteristics based on relevant mouse knockout studies and EC cell lines.

Signaling pathways responsible for the initiation of PE have been proposed through induction of a monolayer culture of ECs with retinoic acid (RA) and cAMP (Goldstein et al., 1990; Harris and Childs, 2002; Strickland et al., 1980). RA is a general endoderm signaling molecule (Kelly and Drysdale, 2015), and promotes PrE or VE cell identity depending on EC cell culture manipulation (Hogan et al., 1981; Strickland and Mahdavi, 1978). Addition of cAMP with RA produces PE characteristics (Goldstein et al., 1990; Harris and Childs, 2002; Strickland et al., 1980). Upstream signaling molecules such as PTHrP increases cAMP levels in EC cells suggesting that PTHrP signaling is a key signaling pathway for PE identity (Stolpe et al., 1993; Verheijen et al., 1999a). Indeed, embryos deficient in PTHrP receptor type 1 show defects in PE growth *in vivo*, while *in vitro* differentiation to PE-cells in the absence of the PTHrP receptor prevents PE differentiation in ESC cells (Verheijen et al., 1999a). Collectively, this data suggests that PTHrP signal relay is critical for PE differentiation, growth and maintenance.

One of the possible downstream effectors of PTHrP signaling may be related to CREB protein as previously proposed through the EC system (Masson et al., 1993). Interestingly, we also report CREB3L2 to be highly enriched in the PE cells, which may be a TF candidate for this proposed pathway (Figure 2.5C). PTHrP signaling can activate a cascade of biological processes, though this relay is often cell-type specific, mediated through activating common signaling effectors of other signaling pathways (e.g. *PKC*, *Erk1/2*, *RhoA* etc.), cross-activation of other signaling pathways such as WNT or BMP, or activation of developmental transcription factors (e.g. *Runx1* in chrondrocytes or *Msx2* in mammary gland cells) (Cheloha et al., 2015; McCauley and Martin, 2012). Several questions have yet to be answered about other specific pathways that

PTHrP signaling activates in PE cells besides raising cAMP levels. One finding is that PTHrP

signaling does not activate these pathways via RAS/ERK signaling effector (Verheijen et al.,

1999b). However, RAS/ERK signaling is temporarily activated to bring the EC cells into the

PrE-like state, while RAS is inhibited upon addition of cAMP or PTHrP hormone with RA

(Verheijen et al., 1999b). Other signaling pathways such as RHO/ROCK signaling, promote and

control PE characteristics by proper focal adhesion complex formation and migration via myosin

ii phosphorylation turnover to prevent abnormal growth (LaMonica et al., 2009; Mills et al.,

2009).

Specification of the Visceral Endoderm and its Functions

Extraembryonic VE vs Embryonic VE Specification during Pre-to-Peri-implantation

The PrE cells that line the epiblast become the future VE (Filimonow and Fuente, 2021). As the

peri-implantation period proceeds, the egg cylinder elongates and the VE differentiates into two

cell types, namely, the extraembryonic VE (exVE) and the embryonic VE (emVE), which are

established by E5.5 stage (Figure 1.3) (Filimonow and Fuente, 2021). The exVE wraps around

the extraembryonic ectoderm (ExE) structure that overlays the epiblast cells (Figure 1.3)

(Filimonow and Fuente, 2021). In contrast, the emVE lines the epiblast associated cells (Figure

1.3) (Filimonow and Fuente, 2021).

Due to the locations of these cell types, emVE and exVE have different properties despite

being quite similar. Through a combination of *in vivo* and *in vitro* experiments, it has been

proposed that exVE continuously receives BMP signals to maintain its identity, in contrast,

emVE receives NODAL signals (Artus et al., 2012; Coucouvanis and Martin, 1999; Julio et al.,

2011; Mesnard et al., 2006; Paca et al., 2012). Although exVE and emVE are distinct, it is still

speculative on the origin of these two cell types that appear by E5.5, particularly whether exVE is a default VE state and emVE is divergent from exVE.

Hints to these models are suggested from recent transcriptional studies where exVE has a similar basal transcriptome to emVE (Nowotschin et al., 2019). In contrast, emVE was found to have a more divergent, specific gene expression signature (Nowotschin et al., 2019). Other evidence suggesting that exVE is the default lineage was shown through *in vitro* evidence by forming embryoid bodies and co-incubating it with BMP signals which induces a exVE identity (Artus et al., 2012; Coucouvanis and Martin, 1999; Paca et al., 2012). These assays induced the formation of a VE-like cell layer along with cavity formation, a morphological process that depends on the signals produced by VE cells (Coucouvanis and Martin, 1999). While these events are *in vitro* based, they parallel the initial cavitation of the egg cylinder *in vivo* (Coucouvanis and Martin, 1999). Additionally, evidence of high BMP sources have been found in the epiblast in the preimplantation embryo stages, whose signals later become localized in the ExE (Figure 1.4) (Ben-Haim et al., 2006; Coucouvanis and Martin, 1999; Graham et al., 2014; Mochel et al., 2015; Sozen et al., 2021; Waldrip et al., 1998).

Correspondingly, several BMP signaling mutants have markedly reduced embryo size with death around gastrulation which suggests the importance of BMP in the early embryo stages (Beppu et al., 2000; Chu et al., 2004; Lawson et al., 1999; Sirard et al., 1998; Yang et al., 1998). Consistent with this, recent improvements in performing *in vitro* cultures past the preimplantation stages have made it possible to culture embryos to mimic postimplantation development (Bedzhov et al., 2014b). However, this culture system has limitations in the study of certain extraembryonic tissues since the mural TE side (lining the blastocoel cavity) of the embryo is dissected out in this culture system (Bedzhov et al., 2014b). Experiments using this

new culture system demonstrated the requirement of BMP signals at the pre-to-peri-implantation stages, where both embryonic and extraembryonic tissues (limited to EPI, ExE, and VE cell types) were markedly affected (Sozen et al., 2021). These experiments suggest that BMP signaling is functional at this point in time (Sozen et al., 2021). However, there have also been indications of the presence of NODAL and/or ACTIVIN signals during the preimplantation stages (Papanayotou and Collignon, 2014). Detection of the early NODAL/ACTIVIN signaling activity has been associated with the marking of future emVE cells rather than initial specification of the VE (Papanayotou and Collignon, 2014; Takaoka et al., 2011; Takaoka et al., 2017). The question of whether NODAL signals are present prior to implantation is still up for debate because of conflicting results (Kumar et al., 2015).

However, *Nodal* mutants and *Smad2/3* double knockouts show general VE formation but fail to form the VE subvariant or emVE (Mesnard et al., 2006; Senft et al., 2018). One attractive model that may explain the discrepancy is to assume a model based on balancing NODAL and BMP signals during pre- to peri-implantation (Figure 1.4). Based on the spatial expression patterns of *Nodal* and *BMP* ligands, expression levels of *Nodal* increases in the EPI cells from E4.5 to E5.5, while *Bmp4* ligand is prominently present in the EPI cells at E4.5 but by E5.5 *Bmp4* is expressed preferentially in the extraembryonic ectoderm (ExE) cells (Figure 1.4) (Brennan et al., 2001; Lawson et al., 1999; Mochel et al., 2015; Takaoka et al., 2006; Winnier et al., 1995). Therefore, the ligand expression patterns implies a model where increased secretion of NODAL signals from the EPI cells combined with the elongation of the egg cylinder distances the paracrine BMP4 secreted from the ExE while maximizing NODAL signals for the specification of the emVE over exVE cell fate. Besides the control of NODAL versus BMP signaling dosage exposure, these two signaling pathways have a strong mutual antagonism at the

molecular level. Both NODAL and BMP signaling activate SMAD transcriptional effectors unique to their own pathway. NODAL activates SMAD2/3 while BMP activates SMAD1/5/9 (Candia et al., 1997; Cho and Blitz, 1998). However, the transcription activation of these SMAD effectors of the NODAL and BMP signaling pathways depends on its complex formation with SMAD4 cofactor which is shared between the two signaling pathways (Candia et al., 1997; Cho and Blitz, 1998). Thus, SMAD4 is a rate limiting factor for both NODAL and BMP signaling activation (Candia et al., 1997; Cho and Blitz, 1998). Additionally, other antagonistic mechanisms between NODAL and BMP signaling may also occur at the extracellular level. For instance, NODAL can heterodimerize with BMP ligands (e.g. BMP7) where they can cross-inhibit each other's activities (Cho and Blitz, 1998; Yeo and Whitman, 2001). Thus, the relative contribution of NODAL signals versus BMP signaling in the initial VE specification and differentiation has yet to be fully dissected. Collectively, these pieces of evidence indicate that BMP signals are present earlier on to promote VE differentiation, and additional contributing signals by NODAL/ACTIVIN signaling produces the divergent emVE cell state (Figure 1.4).
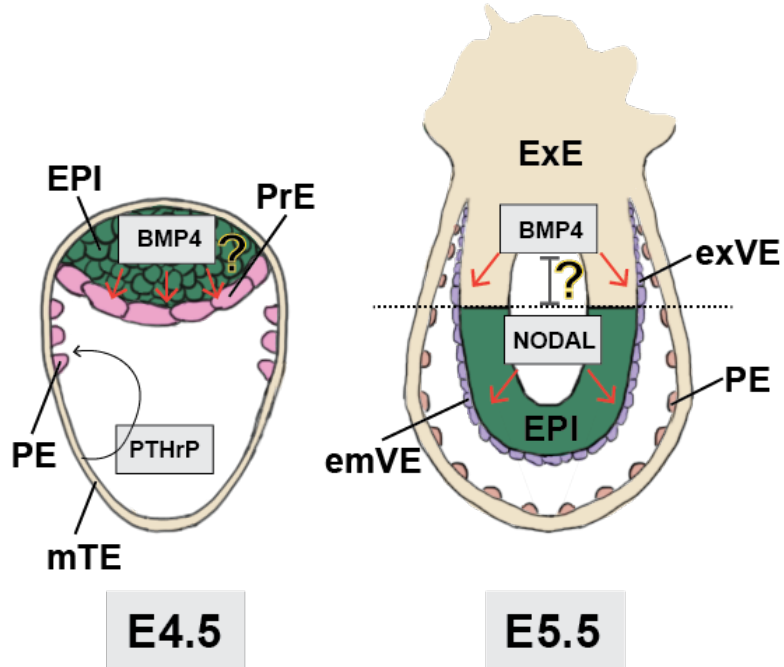
Figure 1.4. Proposed model for VE specification and differentiation. BMP4 from the epiblast cells signals to the adjacent primitive endoderm cells to promote VE identity. Primitive endoderm cells further away from these signals undergo PE differentiation by receiving PTHrP signals. Differentiation of emVE cells occurs by receiving NODAL signaling inputs from the epiblast cells. The BMP4 and NODAL signaling inputs are tuned through a potential of various antagonistic mechanisms which can include regulation at its transcriptional SMAD effectors or at the extracellular level.

Visceral Endoderm: Nutritional Support Functions

The VE cells developed specialized organelles known as apical vacuoles which mediate endocytosis pathways and have lysocytic properties (Assmat et al., 2005; Ichimura et al., 1994; Kawamura et al., 2012; Sun-Wada and Wada, 2022; Wada, 2014; Wada et al., 2014). Mouse knockout for genes (e.g. *Rab7*, *Cubn*, *Atp6v0c*, *Rab11a*, *Slc40a1*) associated with apical vacuole formation and/or participate in the endocytosis pathways have early VE defects that either lead to peri-implantation lethality or failed gastrulation (Donovan et al., 2005; Kawamura et al., 2012; Kawamura et al., 2020; Perea-Gomez et al., 2019; Sun-Wada et al., 2021; Yu et al., 2014). The abundant apical vacuoles allow for the VE to efficiently facilitate biosynthesis, processing, and transport of nutrient, hormone, apolipoproteins, and other molecules from the maternal environment to support embryonic cell development (Assmat et al., 2005; Ichimura et al., 1994; Kawamura et al., 2012; Sun-Wada and Wada, 2022; Wada, 2014; Wada et al., 2014). Additionally, these apical vacuoles serve as general centers for controlling the activity of signal transduction pathways by processing of ligands and receptors which have been extensively reviewed elsewhere (Sorkin and Zastrow, 2009). However, little is known about how these organelles participate in modulating signaling in VE cells. One example demonstrated RAB7 (a small GTP-binding proteins that forms apical vacuoles in VE cells) regulates the spatial distribution of *Dkk1*, a key Wnt antagonist important for formation of the initial anterior-posterior (AP) axis, by endocytosis of *Dkk1* to influence localized availability of *Dkk1* (Kawamura et al., 2020). Future studies are needed to examine how VE-associated apical vacuoles regulate trafficking of other WNT signaling components along with TGFβ, BMP, or FGF signaling components to influence gastrulation processes. (Tam and Loebel, 2007). Thus,

the VE has diverse functions in nutritional support, and early embryo patterning and differentiation during peri-implantation to gastrulation stages.

exVE and emVE Inductive Signaling Roles in Embryo Patterning

As development progresses, the fetal cells prepare to set up the anterior posterior axis (Tam and Loebel, 2007). These events highly depend on specification of emVE cells, which is determined by a NODAL signaling gradient sourced from epiblast cells and expression of Nodal signaling agonists from VE cells, to produce distinct and regionalized VE subtypes (Tam and Loebel, 2007). The origin and morphogenetic migration of these emVE subtypes have been a subject of debate. This review will focus on the functions and specification of these VE subtypes, since the origin, migration and regionalization of these emVE cells have recently been extensively discussed in several, detailed reviews (Robertson, 2014; Stower and Srinivas, 2014; Takaoka, 2014).

Briefly, upon VE and PE lineage divergence by E5.5, the distal visceral endoderm (DVE) differentiates in the distal region of the embryo, thus marking the proximo-distal axis (Figure 1.5) (Filimonow and Fuente, 2021; Tam and Loebel, 2007). Concomitantly, anterior visceral endoderm (AVE) caudally located to DVE is a distinct population that migrates together with DVE to mark the anterior side of the embryo (Figure 1.5) (Filimonow and Fuente, 2021; Tam and Loebel, 2007). Upon migration, AVE expresses *Hex*, *Lefty1*, *Cer*, and *Dkk1* but heterogeneously during migration where *Dkk1* expression is found highest at the embryonic and extraembryonic boundary (Figure 1.5) (Hoshino et al., 2015). This illustrates that AVE is distinct from posterior VE by presence of *Dkk1*, a Wnt agonist, which prevents ectopic anterior primitive streak formation for proper gastrulation (Figure 1.5) (Hoshino et al., 2015). The primitive streak is locally induced on the posterior side by VE through expression of *Wnt3* ligand promoted by

BMP signals from the extraembryonic ectoderm (ExE) (Figure 1.5) (Gadue et al., 2006; Liu et al., 1999; Yoon et al., 2015). To conclude, the emVE beginning from E5.5 to E6.5 prior to gastrulation becomes further regionalized and shaped, and this process is dependent on its position in the egg cylinder marking the axis to prepare for gastrulation (Figure 1.5) (Filimonow and Fuente, 2021; Tam and Loebel, 2007). Beyond patterning the epiblast cells, there is additional evidence showing that the emVE can be incorporated into the future gut tube, which demonstrates that the separation of extraembryonic and embryonic origin is not strictly segregated (Kwon et al., 2008; Nowotschin et al., 2019). However, whether these emVE cells persist in the gut tube temporarily or propagate further is still debated.
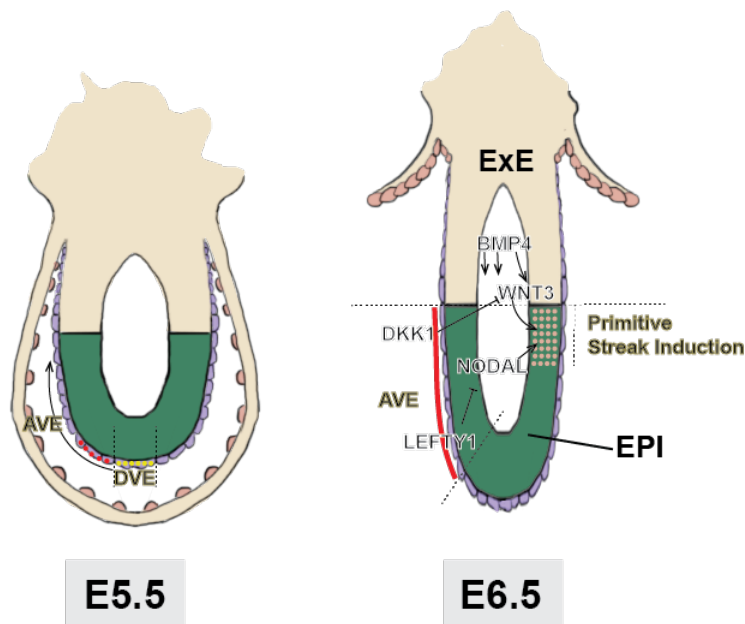
Figure 1.5. Diagram of primitive streak induction by embryonic visceral endoderm. AVE is induced and migrates upwards from E5.5 to E6.5. The AVE expresses *Lefty1* (NODAL inhibitor) and *Dkk1* (WNT inhibitor). *Dkk1* is expressed more highly at the embryonic and extraembryonic boundary. BMP4 is secreted from the ExE to activate *Wnt3*. These asymmetrical signaling gradients allow for the primitive streak induction at the posterior end. Note: *Hex* and *Cer* are also expressed in the AVE which is not annotated in this figure. Abbreviations: AVE (Anterior visceral endoderm), DVE (Distal visceral endoderm), EPI (Epiblast), ExE (Extraembryonic ectoderm)

Upon primitive streak induction, a subset of these cells differentiate to mesoderm cells that migrate and ingress into the extraembryonic components, thus becoming the extraembryonic mesoderm (Tam and Loebel, 2007). The exVE induces the first hematopoietic derived lineages or blood islands from the extraembryonic mesoderm through secretion of *Vegfa* and *Ihh* ligands (Artus et al., 2012; Belaoussoff et al., 1998; Daane and Downs, 2011; Damert et al., 2002; Dyer et al., 2001; Ferkowicz and Yoder, 2005; Pierre et al., 2009). To support this, evidence from *in vitro* BMP4 induced VE-like cells co-cultured with primitive erythroid cells can induce hematopoietic lineage and vasculogenesis differentiation (Artus et al., 2012). However, there have also been speculative debates on alternative sources of the blood lineages particularly from the exVE derived cells since blood cells were found localized within the exVE layers (Rodriguez and Downs, 2017). Originally, early experiments examining whether VE cells can contribute to hematopoietic lineages were limited due to assay limitations in lineage tracing (Downs, 2020). Improvements on *ex vivo* embryo culture and cell fate tracing assays further clarified the hematopoietic potential of VE cells (Rodriguez and Downs, 2017).

The exVE cells are in contact with the extraembryonic mesoderm as it forms (Filimonow and Fuente, 2021; Rodriguez and Downs, 2017). The extraembryonic mesoderm through its interactions with the exVE further differentiates to form two distinct regions, namely the blood island and the future allantois (Rodriguez and Downs, 2017). This suggests that the exVE cells that line the mesoderm may also be regionalized similar to the emVE as discussed earlier (Rodriguez and Downs, 2017). Indeed, grafting allantois associated exVE cells (AX-exVE) from donor *Rosa26-lacZ* reporter embryos to host embryos demonstrated that the AX-exVE have a hematopoietic potential observed through budding hematopoietic-like cells and positive *Runx1* expression in yolk sac associated cells (Rodriguez and Downs, 2017). However, the signal(s) that

is responsible for inducing the exVE (adjacent to the allantois) differentiation to hematopoietic-like cells have yet to be identified. Collectively, this suggests that the exVE may delineate to become placenta blood cells.

Transcription factors regulating VE Development

Several TFs were found to be important for VE development through analyzing phenotypes of mouse mutants. Majority of VE associated factors found were defects in VE tissue morphology and gastrulation failures. One major VE TF identified, *Hnf1b*, is considered a critical VE specifier (Barbacci et al., 1999; Coffinier et al., 1999). Mouse embryos deficient in *Hnf1b* displayed a highly disorganized VE with PE-like cell differentiation (Barbacci et al., 1999; Coffinier et al., 1999). Further examination using embryo explants demonstrated that the loss of *Hnf1b* was unable to induce the key early and late VE markers (Barbacci et al., 1999; Coffinier et al., 1999). These observations suggest that *Hnf1b* is essential and required for initiation of the VE GRN cascade by perhaps regulating other VE factors (Barbacci et al., 1999; Coffinier et al., 1999). For instance, *Hnf4a* was a key TF found to be induced by *Hnf1b*. VE structural integrity remained mostly intact in *Hnf4a* deficient embryos relative to *Hnf1b* knockout embryos. Further examination of gene expression markers (e.g. *Ttr*, *Apo-AI*, *Apo-AIV*, *Apo-B*, *Afp*) suggests that *Hnf4a* functions to activate nutritional pathways to support the embryonic cells (Chen et al., 1994; Duncan et al., 1997). Additionally, *Hnf4a* knockout embryos also had gastrulation defects, however, the contribution of Hnf4a to gastrulation is not clear. Although primitive streak induction proceeds in *Hnf4a* knockout embryos, the first event that initiates gastrulation, it is delayed by around one day (Chen et al., 1994; Duncan et al., 1997). Taken together, in VE cells, *Hnf4a* may also have a role in activation of proper gastrulation signals.

28

Another prominent VE TF that was identified is *FoxA2* (*Hnf3β*) (Dufort et al., 1998). In *FoxA2* knockout embryos, "VE" cells are present, but these "VE" cells are mis-aligned along the embryonic and extraembryonic boundary, and accompanied by abnormal accumulation of PrE/PE cells distally at the end of the embryo (Dufort et al., 1998). These observations suggest that some aspects of the VE structural integrity were maintained perhaps by the presence of *Hnf1b* due to redundant TF functions in VE development. Additionally, *FoxA2* knockout embryos were found to be smaller which may imply that FOXA2 functions to activate nutritional pathways to support embryonic cell growth. Lastly, *FoxA2* knockout embryos failed gastrulation, where primitive streak induction failed, suggesting that FOXA2 activates signaling related pathways for patterning the embryo (Dufort et al., 1998).

Several other TFs associated with VE do not function to specify the VE, but their expressions are pattern regionally along the VE tissue, specifically the emVE, which encapsulates the epiblast cells. These VE-associated TFs (e.g. *Hex*, *Otx2*, *Eomes*, and *Lhx1*) function mark the anterior-posterior axis for proper gastrulation to proceed. For instance, *Hex*, which is expressed in the PrE, marks the future DVE and is later expressed in the AVE (Paca et al., 2012; Perea-Gomez et al., 2004; Thomas et al., 1998). The signals that regulate *Hex* regionalization are not clear, though, it has been found to be induced by BMP signaling, which may allude to its early expression in the PrE (Paca et al., 2012). How *Hex* becomes restricted to the DVE and AVE has yet to be understood. Interestingly, *Hex* deficient embryos initiate gastrulation normally, but future anterior tissues were malformed (Barbera et al., 2000). The requirement of *Hex* in regulating proper anterior morphogenesis was suggested to arise from the refinement of *Hex* spatial expression from the AVE to the midline anterior definitive endoderm (Barbera et al., 2000). However, current evidence mainly characterizes *Hex* as a cell marker for

the anterior region, but the direct transcriptional roles that *Hex* imparts within its VE-associated cell types is unclear. Though complementary experiments performed in xenopus embryos suggest a molecular mechanism by which HEX acts as a transcriptional suppressor to promote WNT signaling activity by repression of TLE4 (Brickman et al., 2000; Zamparini et al., 2006). These molecular pathways imply that *Hex* may function in the AVE to fine tune WNT activation levels, by coordinating its actions with *Dkk1* (a WNT signaling agonist), at the anterior side of the embryo.

While molecular causes for the spatial expression of *Hex* in the DVE and AVE have yet to be dissected, other TFs such as, *Otx2*, *Eomes*, and *Lhx1* are regionally expressed in the AVE (Hoshino et al., 2015). The expression of patterns of *Otx2*, *Eomes*, and *Lhx1* is regulated by NODAL signals emanating from the epiblast cells. First, dose-controlled NODAL signaling (as discussed earlier) received by the emVE cells activates SMAD effectors (SMAD2/3) which leads to *Eomes* expression. The presence of EOMES, in turn, activates *Lhx1* in the AVE (Costello et al., 2015). Additionally, FOXA2 is required for the activation of *Otx2* in the DVE and AVE where OTX2 serves as a required transcriptional activator for *Dkk1*, which plays a key role in A-P axis formation as discussed earlier (Kimura-Yoshida et al., 2005; Kimura-Yoshida et al., 2007; Paca et al., 2012; Thomas et al., 1998). Although, how FOXA2 acts regionally to activate *Otx2* is not clear. However, using *in vitro* complementary experiments, LHX1 can regulate *Otx2* and *FoxA2* expression and form a protein-protein interaction complex demonstrated through *in vitro* ChIP and mass spectrometry experiments (Costello et al., 2015). Collectively, these pieces of evidence infers a regulatory model where 1. NODAL-activated SMAD2/3 effectors activate *Eomes, 2.* EOMES activates *Lhx1* 3. LHX1 enhances *FoxA2* expression in the AVE relative to

other VE cells 4. Increased FOXA2 levels activates *Otx2*, and lastly 5. LHX1, OTX2, and FOXA2 form a tripartite complex to regulate AVE functions.

To conclude, there has been a great deal of work that has elucidated PE and VE tissue development and its biological functions during early mouse embryogenesis. However, little is known still about the lineage differentiation events that govern PrE to PE or VE cell fate commitment. In the following chapters, I will focus on elucidating molecular mechanisms that orchestrate the differentiation of the PrE to PE or VE cell fates.

For their contributions in Chapter 1, thanks to: Patrick Pham (Artwork for the embryo drawings).

CHAPTER 2:

Transcription factor network instructs lineage divergence and plasticity in extraembryonic

endoderm

2.1 Introduction

The amnion, chorion, yolk sac, and allantois are critical extraembryonic membranes present in mammals, and proper extraembryonic tissues are essential for nutritive support and survival of fetus as well as assisting the patterning of the embryonic axes (Ferner and Mess, 2011; Filimonow and Fuente, 2021; Sheng and Foley, 2012). The origin of extraembryonic structures can be traced back to the blastocyst stage of embryonic development when trophectoderm (TE) cells and the inner cell mass (ICM) cells develop from the totipotent progenitors that co-express CDX2 and POU5F1 (OCT4) (Dietrich and Hiiragi, 2007; Holmes et al., 2017; Niwa et al., 2005). The status quo of the progenitor cell state is broken by Hippo signaling, leading to the commitment of TE or ICM cell fate that exclusively expresses CDX2 or OCT4 in respective cells (Leung and Zernicka-Goetz, 2013; Nishioka et al., 2009; Rayon et al., 2014). The outer layer of trophectoderm cells gives rise to the chorion, whereas ICM cells develop to establish the embryo proper (Filimonow and Fuente, 2021). The ICM cells maintain their fate by co-expressing NANOG and GATA6, which are subsequently expressed in a 'salt and pepper'-like manner comprising epiblast (EPI) and PrE cells (Bessonnard et al., 2014; Chazaud et al., 2006; Plusa et al., 2008; Saiz et al., 2016; Schrode et al., 2014; Xenopoulos et al., 2015). FGF signaling promotes ICM cells to adopt GATA6 positive PrE cells, whereas NANOG positive EPI cells arise by repression of FGF signaling (Bessonnard et al., 2014; Chazaud et al., 2006; Hamilton et al., 2019; Kang et al., 2017; Molotkov et al., 2017; Yamanaka et al., 2010). Thus, during the emergence of the two major extraembryonic cell types, bipotential progenitor cells utilize extracellular signals (developmental cues) to segregate into two distinct cell states determined by master transcription factors (TFs).

During the late blastocyst stage, PrE cells differentiate to become parietal endoderm (PE) and visceral endoderm (VE) cells (Filimonow and Fuente, 2021). VE and PE are major subtypes of the extraembryonic endoderm (ExEn) (Filimonow and Fuente, 2021). Previous studies reported that VE cells in post-implantation embryos can undergo transdifferentiation to PE cells (Casanova and Grabel, 1988; Cockroft and Gardner, 1987; Ninomiya et al., 2005; Yagi et al., 2016), while a PE-like extraembryonic endoderm (XEN) cell line acquires VE identity in response to BMP signaling (Artus et al., 2012; Paca et al., 2012; Sozen et al., 2018; Zhang et al., 2019). This plasticity in mature PE/VE cells suggests that differentiating PrE cells are in a bipotential state that can adopt PE or VE cell fates, and that plausibly the PE-VE cell fate divergence choice is instructed by dynamic interactions involving signaling pathway(s) and ExEn TF networks, analogous to the regulatory paradigm observed in the TE-ICM, and EPI-PrE cell fate decision processes. Past work implied the critical roles of GATA and SOX TFs in this process (Artus et al., 2011; Cai et al., 2008; Kanai-Azuma et al., 2002; Koutsourakis et al., 1999; Lim et al., 2008; McDonald et al., 2014; Morris et al., 2010; Niakan et al., 2010; Qu et al., 2008; Schrode et al., 2014; Shimoda et al., 2007; Wamaitha et al., 2015). However, the molecular mechanism underlying this decision-making process has remained obscure largely due to the challenge in accessing scarce material of pre- to peri-implantation embryos. Therefore, characterization of differentiating PrE cells starting at E4.5 has been poorly defined. Additionally, interpretation of genetic knockout studies involving *Gata4*, *Gata6*, *Sox7*, and *Sox17* using mouse models is often difficult due to the compounded phenotypes across tissue types that are exacerbated by overlapping functions caused by TF redundancy (Artus et al., 2011; Bowles et al., 2000; Cai et al., 2008; Kanai-Azuma et al., 2002; Koutsourakis et al., 1999; Kuo et

al., 1997; Molkentin et al., 1997; Schrode et al., 2014; Wat et al., 2012). Our goal is to fill the knowledge gap by uncovering the critical TF regulatory network that operates during this cell-fate-decision process and identifying potential signaling cues that control the generation of PE and VE cells.

Here we focused on analyzing the transcriptionally defined PrE, PE, and VE cell states based on scRNA-seq dataset of E4.5 mouse embryos (Cheng et al., 2019; Mohammed et al., 2017; Nowotschin et al., 2019; Qiu et al., 2022). We found that PrE cells were bipotential and have the property of co-expressing TFs that control both PE- and VE-associated gene regulatory modules. Comparative analysis of PE and VE cell type-specific enhancer landscapes revealed that a subset of active enhancers present in VE cells was highly accessible in PE cells. DNA motif enrichment analysis suggested that these enhancers were regulated by interaction involving FOXA2, which was preferentially expressed in VE cells as well as GATA6 and SOX17 that were preferentially expressed in PE cells. To directly address the roles of these TFs in PE and VE specification, we developed XEN cells harboring a dTAG-inducible FKBP12^{F36V} degron system (Nabet et al., 2018) that can rapidly degrade GATA6 and SOX17 in combination with *FoxA2* knockout XEN cells. This system allowed us to unveil the critical roles of GATA6, SOX17, and FOXA2 in PE and VE development and to identify target genes that were regulated by these TFs. Additionally, we show that BMP signaling is a critical developmental cue for PE-VE cell fate decisions.

2.2 Results

2.2a Gene Signatures Specifying Parietal Endoderm (PE) and Visceral Endoderm (VE) Cell Fates

To investigate the cell states within the extraembryonic endoderm (ExEn) lineage at embryonic day 4.5 (E4.5), we integrated published scRNA-seq datasets of E4.5 embryos (Cheng et al., 2019; Mohammed et al., 2017; Nowotschin et al., 2019; Qiu et al., 2022) followed by cell clustering analysis. We detected TE, EPI, and PrE (ExEn) lineages as reported previously (Figure 2.1A). Furthermore, the integrated dataset with a greater number of single cells (311 cells) enabled us to detect three subpopulations within the ExEn lineages (Figure 2.1A). Expression levels of *Pou5f1* (*Oct4*)*, Lama1*, and *Ttr* (markers of PrE, PE, and VE) (Kwon and Hadjantonakis, 2009; Niimi et al., 2004; Ohinata et al., 2022; Wallingford et al., 2013) in these individual cells revealed that the three subpopulations were characterized with $Oct4^{hi}Lama1^{med}Ttr^{med}$, $Oct4^{med}Lama1^{hi}Ttr^{lo}$, and $Oct4^{lo}Lama1^{lo}Ttr^{hi}$, which corresponds to the PrE, PE, and VE cell states (Figure 2.1B and 2.1C).
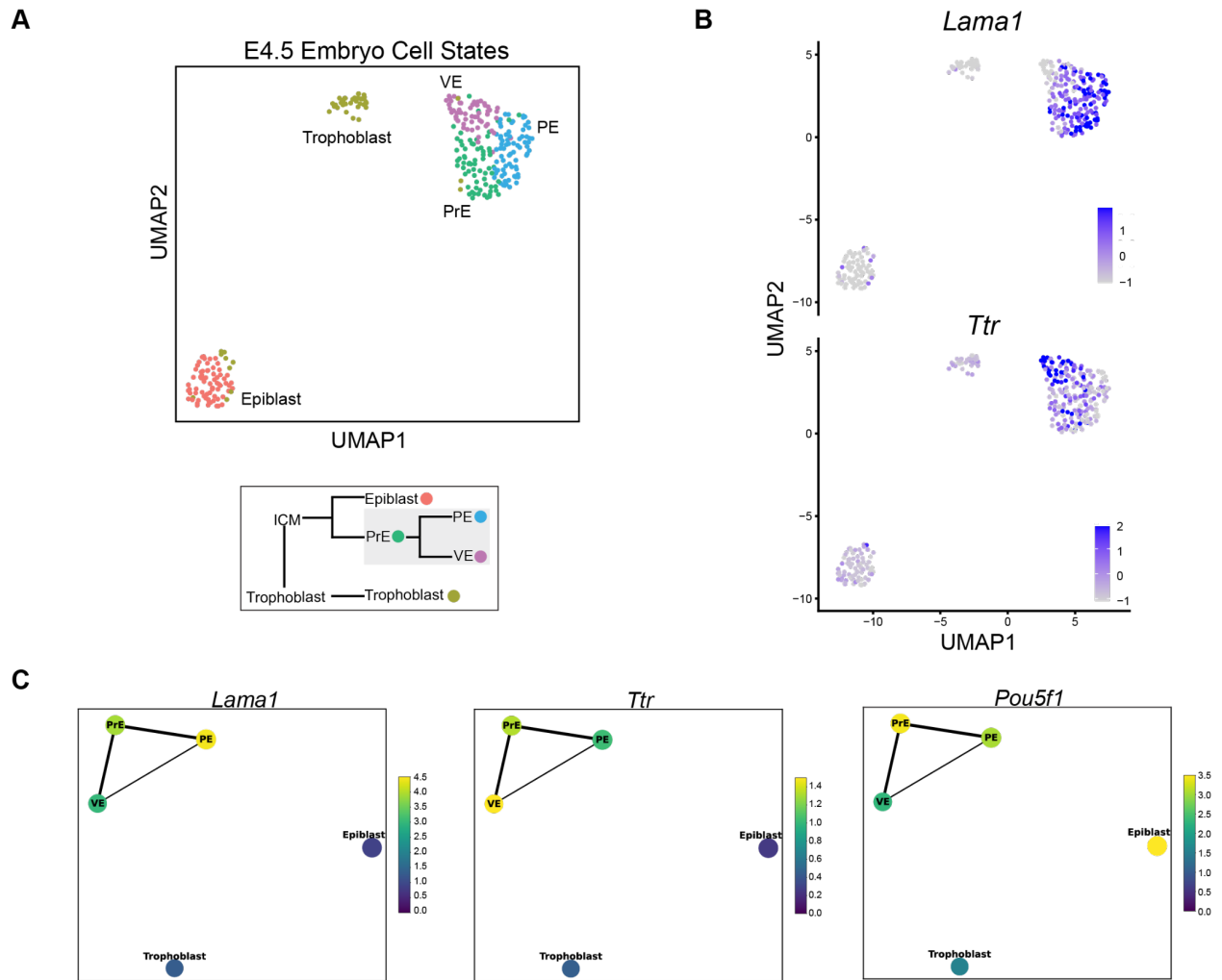
Figure 2.1. Single-cell clustering and identification of the extraembryonic endoderm lineages in E4.5 embryo. (A) UMAP embedding for the integrated scRNA-seq data of the E4.5 embryo from Nowotschin et al. (Nowotschin et al., 2019) and Qiu et al. (Qiu et al., 2022) Epiblast, trophoblast, and extraembryonic endoderm subpopulations were color coded. Lineage differentiation paths between E3.5 and E4.5 were displayed in a tree plot. (B) Expression of *Ttr* or *Lama1* for the E4.5 single cells were imposed in UMAP embedding. (C) PAGA graph displaying the connections among cell states found in E4.5 scRNA-seq data. Expression of marker genes for parietal endoderm (PE), visceral endoderm (VE), and primitive endoderm (PrE)-*Lama1*, *Ttr*, and *Pou5f1 (Oct4)*-were imposed.

Single cell trajectory inference reconstructed from E4.5 and E5.5 embryos supported the marker-based assignment of these cell states (Figure 2.2A and 2.2B). The candidate PrE cell subpopulation was positioned at the root of ExEn lineage bifurcations, which diverged to the PE and VE cell subpopulations (Figure 2.2A and 2.2B). These cells further differentiated to form E5.5 PE and E5.5 embryonic VE (emVE) and extraembryonic VE (exVE) cells (Figure 2.2A and 2.2B). To identify transcriptional signatures that distinguish PE and VE cell states at the E4.5 stage, we performed high-dimensional weighted gene correlation network analysis (hdWGCNA) (Langfelder and Horvath, 2008; Morabito et al., 2021). The resulting PE and VE gene modules were preferentially expressed in respective PE and VE cells at both E4.5 and E5.5 stages (Figure 2.2C). Gene ontology (GO) analysis on the mouse anatomical ontology database (Hayamizu et al., 2013) showed that these modules were related to the PE or VE tissue terms (Figure 2.2D).
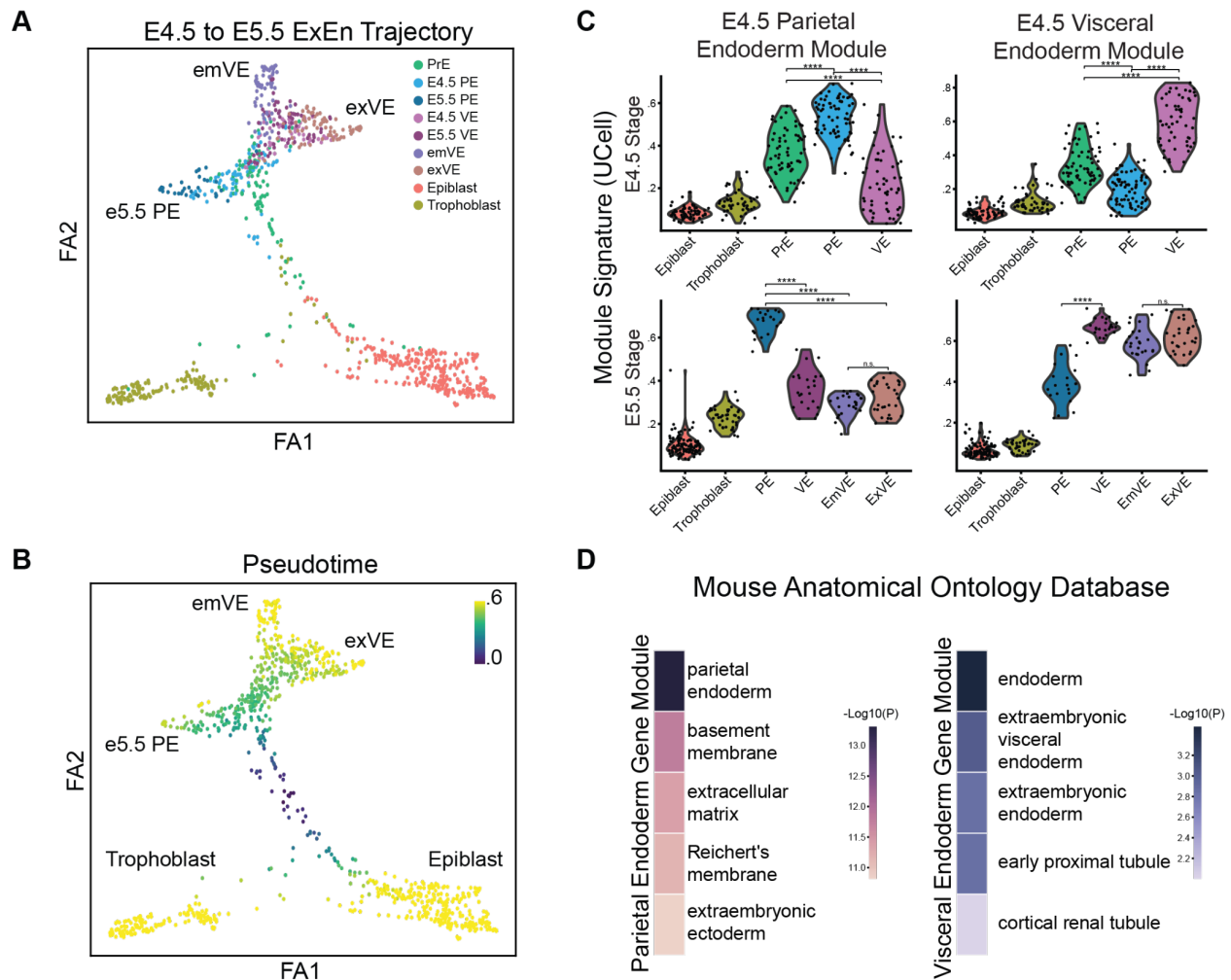
Figure 2.2. Characterization of cell states within the extraembryonic endoderm lineage in E4.5 embryo. (A) Force-directed graph showing the trajectory of E4.5 and E5.5 single cells. (B) Force-directed graph embedding E4.5 and E5.5 single cells as in Figure 2.2A with color encoding for pseudotime calculated by SCANPY. (C) Violin plots displaying the transcriptional signature (calculated by UCell) of cell states in E4.5 and E5.5 scRNA-seq datasets for E4.5 parietal endoderm module and E4.5 visceral endoderm module identified by hdWGCNA (top 50 genes). n.s. denotes non-significant, **** denotes p<0.0001 (Wilcoxon Test). (D) Enriched anatomical terms of PE and VE gene module in Figure 2.2D (kME > 0.5) from the mouse anatomical ontology database.

Moreover, KEGG pathway analysis (Kanehisa and Goto, 2000) for the PE or VE gene modules showed enrichment for biological functions related to PE or VE cells (Filimonow and Fuente, 2021) (Figure 2.3A and 2.3B). Next, we constructed molecular interaction networks for both gene modules in order to annotate the PE- or VE-associated genes to the enriched KEGG pathways (Figure 2.3A and 2.3B). Notably, we found that PE module was enriched for *Myc*-mediated self-renewal pathways (Figure 2.3A, see details below), whereas VE module was enriched for gene families such as *V-ATPases*, *Rab*, *Slc*, and other genes related to VE's histotrophic functions (Donovan et al., 2005; Filimonow and Fuente, 2021; Kawamura et al., 2012; Kawamura et al., 2020; Meehan et al., 1984; Perea-Gomez et al., 2019; Sun-Wada and Wada, 2022; Sun-Wada et al., 2021) (Figure 2.3B, see details below). Taken together, we uncovered distinct transcriptional programs associated with PE and VE cell states that were diverging in E4.5 embryos.
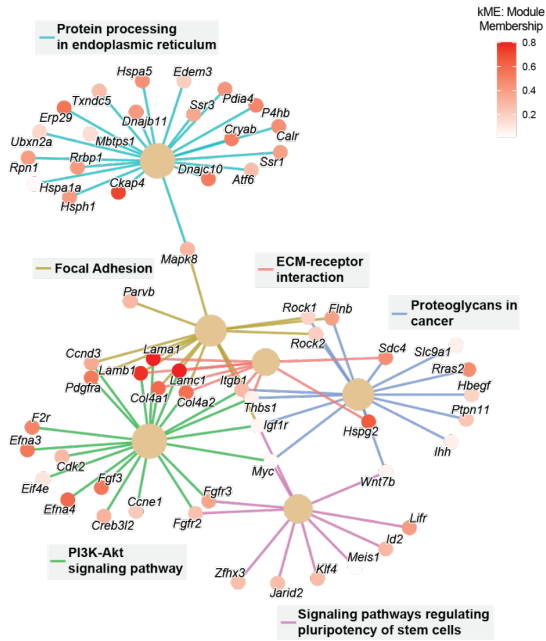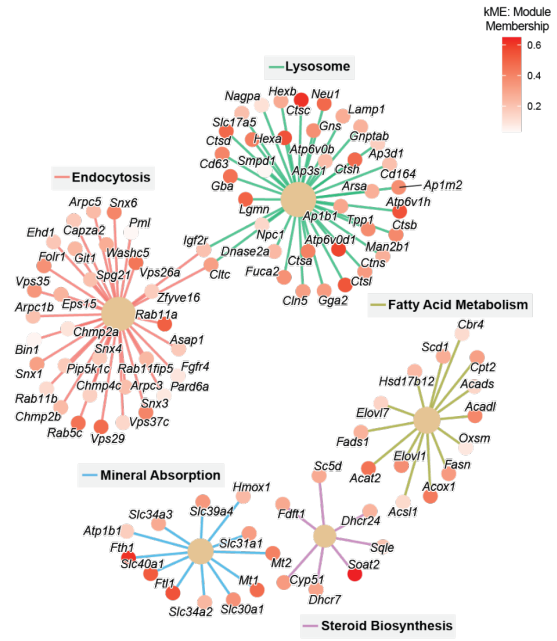
Figure 2.3. Gene network of parietal endoderm or visceral endoderm biological pathways.

(A and B) Networks connecting the E4.5 parietal endoderm module genes (A) and E4.5 visceral

endoderm module genes (B) with enriched KEGG pathways. Individual genes were color coded according to their membership score (kME) relative to the respective gene modules.

2.2b Primitive Endodermal Cells (PrE) Express TFs Linked to PE and VE Differentiation

To investigate the transcriptional state of PrE progenitors, we evaluated the transcriptional signatures of PE and VE gene modules operational in ExEn cells at the E4.5 stage. A scatter plot (Figure 2.4A) showed that the transcriptional signature of PrE cells resided in the central area partially overlapping the gene module clusters of PE and VE cells, indicating that transcriptional signature of PrE shared that of PE and VE gene modules. When the gene expression levels of top 5% of PE and VE module genes were compared, the gene expression profiles of PrE cells showed strong correlation with those of both PE and VE cells (Figure 2.4B).
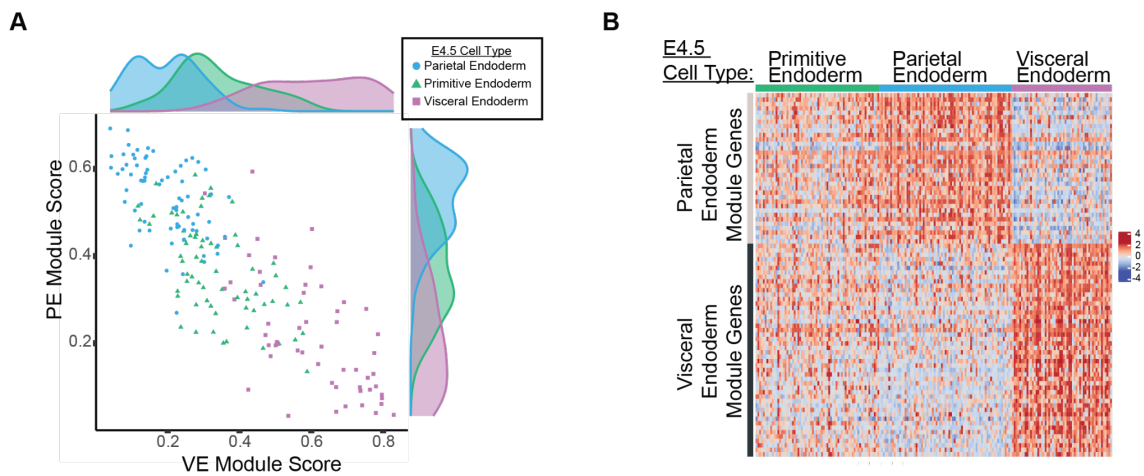
Figure 2.4. Primitive endoderm cells co-expresses parietal endoderm and visceral endoderm associated genes. (A) Scatter plot projecting the single cells of extraembryonic endoderm lineages in E4.5 embryo to the transcriptional signatures of E4.5 parietal endoderm

and visceral endoderm gene modules. Marginal distributions were plotted next to the corresponding axis. (B) Heatmap showing the expression profiles of top 5% correlated genes (kME > 0) in the parietal endoderm and visceral endoderm module in the single cells of the E4.5 extraembryonic endoderm lineages.

Since TFs are known to control gene expression programs, we then examined TFs that were highly correlated with PE or VE modules (Figure 2.5A and 2.5B). We identified TF sets that were selectively expressed in single cells (hdWGCNA analysis) (Langfelder and Horvath, 2008; Morabito et al., 2021) of PE and VE to characterize PE and VE cell types (Figure 2.5A and 2.5B). Next, Figure 2.5C and 2.5D showed the average gene expression profiles of the top correlated TF encoding genes in PE and VE cells (Table 2.1). These TFs were preferentially expressed in PE and VE cells, respectively. Interestingly, many of the PE- or VE-associated TFs were already expressed at lower levels in PrE cells (Figure 2.5C and 2.5D). These findings

indicate that PrE cells were in a plastic state, in which both PE and VE gene programs were

already in play.
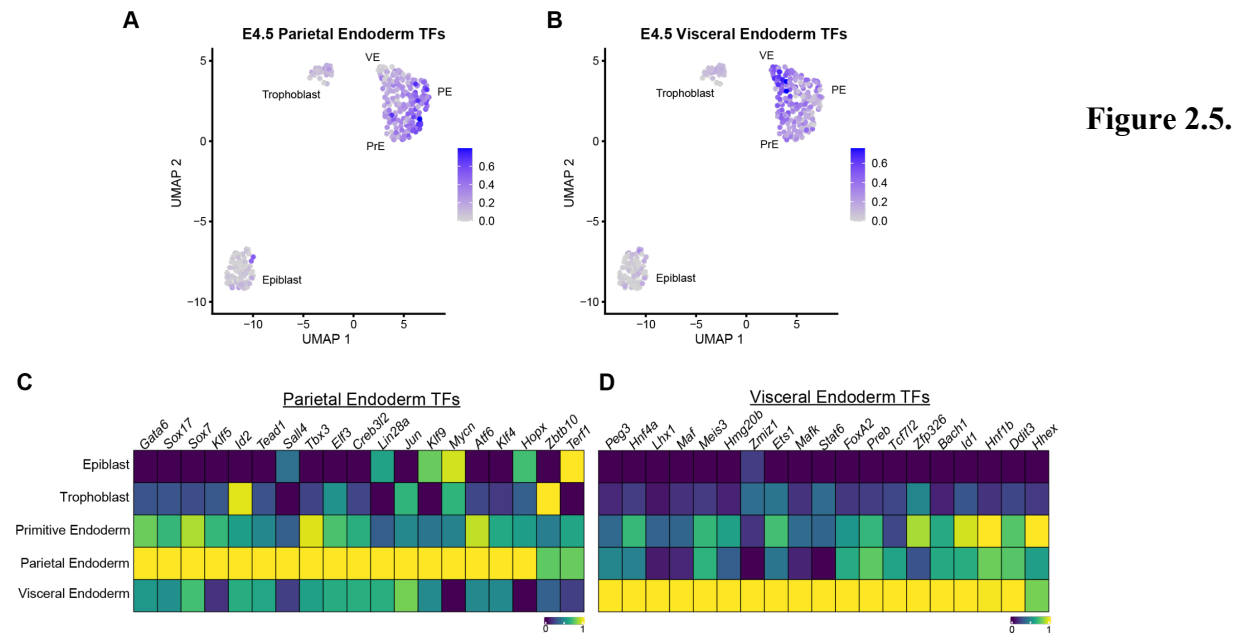
Identification of parietal endoderm and visceral endoderm transcription factors. (A and B)

Transcription signature of combined E4.5 parietal endoderm (A) or visceral endoderm (B)

transcription factors (by UCell) found among E4.5 single cells embedded in UMAP space. (C

and D) Average gene expression levels of the top correlated transcription factors in the parietal

endoderm (C) or visceral endoderm (D) modules for the five cell states present in E4.5 embryos.

Table 2.1. Parietal endoderm and visceral endoderm cell-type biased expressed transcription factor modules associated with Figure 2.5 (C) and (D)

Cell Type	Cell-Type Biased Expressed Transcription Factor
Parietal Endoderm	*Gata6, Sox7, Sox17, Klf5, Tead1, Sall4, Tbx3, Elf3, Creb3l2, Lin28a, Jun, Zbtb10, Terf1, Klf9, Mycn, Atf6, Klf4, Hopx, Klf6, Pbrm1, Satb2, Gm17067, Tead4, D630045j12rik, Zfhx3, Zfp26, Sall1, Cenpa, L3mbtl3, Ahctf1, Sp8, Gm14139, Fos, Snai1, Zfp395, Hmga2, Zfp239, Mynn, Zfp101, Adnp2, Zmat3, Zfp429, Zfp518b, Jarid2, Lyar, Myc, Csrnp1, Baz2b, Meis1, Mta3, Gm3604, Zfp809, Zscan10, Zfp280d, Tcf4, Fosb, Tfdp2, Zfp60, Elk3, Nfil3, Zfp574, 2610021a01rik, Rora, Phtf1, Zbtb6, Foxf2, Pou5f1, Zfp46, Esrrb, Zfp345, Zfp719, Epas1, Rbpj, Mybl2, Zfp9, Zfp759*
Visceral Endoderm	*Peg3, Hnf4a, Lhx1, Hhex, Maf, Hnf1b, Eomes, Meis3, Hmg20b, Zmiz1, Mafk, Stat6, Foxa2, Preb, Tcf7l2, Zfp326, Bach1, Ddit3, Blzf1, Stat3, Tcf23, Nfe2l2, Ets1, Bbx, Rit1, Creb3l1, Zfp516, Zic5, Lrrfip1, Maf1, Sub1, Creb3, Tulp1, Zfp445, Zfp35, Etv4, Tcf15, Klf10, Junb, Smad7, Arid3a, Hes6, Pura, Gm14403, Pitx1, Aff1, Klf15, Vezf1, E2f1, Pitx2, Zfp865, Zfp68, Pias2, Zfp956, Zfp141, Gm14399, E2f8, Hmga1, Ppard, Gm14406, Tfcp2l1, Cbfb, Nr2f6, Gtf2ird2, Otx2, Arid4a, Zfp945, Irx1, 2810021j22rik, Zfp422, Arid4b, Zfp954, Zfp184, Zfp58, Zfp385a, Zfp949, Gm14401, 2700081o15rik, Rcor3, Prdm1, Zfp874b, En1, Zbtb48, Zfp595, Sox12, Zfp933, Zfp940, Zscan20, Zfp623, Zfp423, Zfhx2, Zfp760, Crebl2, Plag1, Stat5a, Zfp53, Dbp, Gbx2, Casz1, Zfp946, Zfp30, Nfatc2, 2610008e11rik, Sall2, Creb3l3, Zfp52, Zfp319, Mecom, Nfic, Aw146154, Twist1, Zbtb32, Etv2, En2, Irf1*

2.2c Regulation of VE Cell Fate Mediated by *Gata6* and *Sox17*

To interrogate how 'cell-type-specific' TFs identified in Figure 2.5C and 2.5D instruct the PE or

VE gene programs, we generated cXEN cell lines (PE-like extraembryonic cells derived from

mESCs) (Cho et al., 2012; Niakan et al., 2013) and compared the enhancer landscapes between

cXEN cells and VE cells from E6.5 embryos. Derived cXEN cells were PDGFRA positive

(Figure 2.6), which is a hallmark of PrE/PE extraembryonic endoderm cells (Artus et al., 2010;

Cho et al., 2012; Plusa et al., 2008).
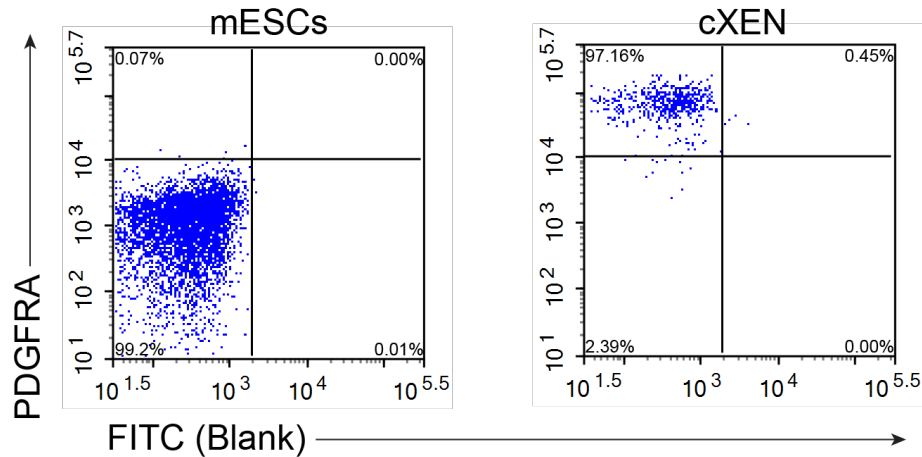
Figure 2.6. XEN cell establishment from mESCs. Flow cytometric analysis of PDGFRA

stained mESCs and cXEN cells.

ATAC-seq and H3K27ac CUT&RUN were performed using cXEN cells and compared

to matching epigenetic data sets (GSE125318) from E6.5 VE cells (Xiang et al., 2020). A

tornado plot displayed the accessibility of enhancers based on ATAC-seq and H3K27ac

CUT&RUN (Figure 2.7A). Active enhancers specific for PE or VE cells were linked to nearby

genes. PE and VE enhancers were upregulated in the corresponding cell types at the E4.5 and

E5.5 stages (Figure 2.7B).
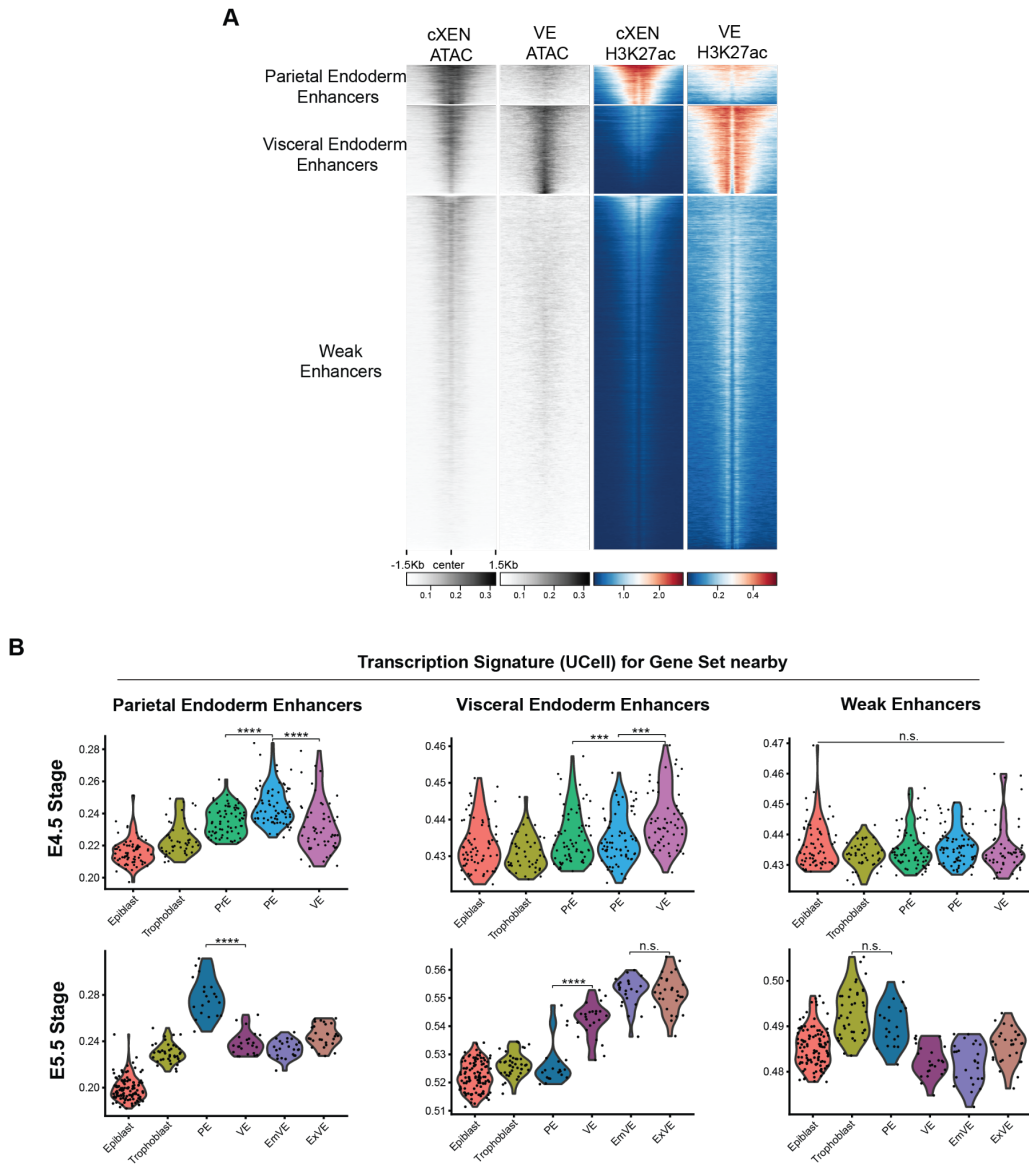
Figure 2.7. Identification of parietal endoderm and visceral endoderm enhancers.

(A) Tornado plots displaying signals of chromatin accessibility and H3K27ac mark in cXEN and E6.5 VE cells centered at H3K27ac enriched regions. (B) Violin plots displaying the transcriptional signature (calculated by UCell) of assigned cell states for non-overlapping gene sets nearby parietal endoderm, visceral endoderm, and weak enhancers in E4.5 and E5.5 scRNA-seq datasets as described in Figure 2E. n.s. denotes non-significant, **** denotes p<0.0001, *** denotes p<0.001 (Wilcoxon Test).

PE enhancers were enriched for GATA, SOX, KLF, TEAD, and JUN TF motifs, which

coincided with the list of TFs highly expressed in the single-cell PE module (Figure 2.5C and

2.8A, Table 2.1). Unexpectedly, VE enhancers were highly enriched for the PE-associated

GATA and SOX motifs (Figure 2.5C and 2.8A, Table 2.1), as well as FOX, HNF1, HNF4, and

LHX motifs, whose members (FOXA2, HNF1b, HNF4a, and LHX1) were also found in the VE

module (Figure 2.5D and 2.8A, Table 2.1). Such a motif composition suggests that VE enhancers

are co-regulated by both PE- and VE-associated TFs.

A

Parietal Endoderm Enhancer Motifs

PWM	Motif Family	p-value
	GATA	1e-92
	SOX	1e-78
	KLF	1e-40
	TEAD	1e-26
	JUN	1e-23
	FOX	1e-21

B

Visceral Endoderm Enhancer Motifs

PWM	Motif Family	p-value
	GATA	1e-129
	FOX	1e-92
	SOX	1e-81
	HNF1B	1e-66
	HNF4A	1e-59
	LHX	1e-20
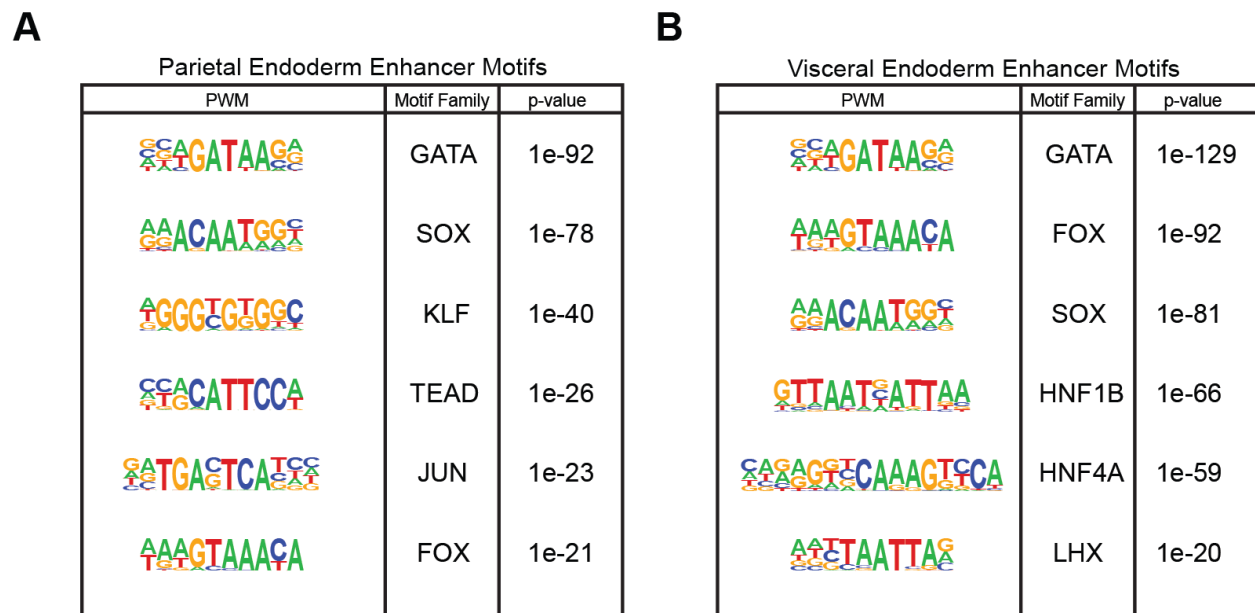
Figure 2.8. Identification of parietal endoderm and visceral endoderm enhancers. (A and B)

Enriched transcription factor binding motifs at parietal endoderm (A) and visceral endoderm

enhancers (B) identified in (Figure 2.7A).

Surprisingly, we found that VE enhancers were accessible, albeit inactive, in PE-like

cXEN cells (Figure 2.7A), indicative of TF occupancies. Therefore, we examined the bindings of

GATA6, SOX17, and FOXA2 (GS&F), which happened to be the most enriched motifs for VE

enhancers, in cXEN cells by CUT&RUN (Figure 2.9A and 2.9B). CUT&RUN data revealed that

GATA6 and SOX17 were colocalized with FOXA2 at a subset of VE enhancers (30.3%) (Figure

2.9A and 2.9B). Hereafter, these were termed as GS&F-co-bound VE enhancers (Figure 2.9A

and 2.9B). These GS&F-cobound enhancers possessed higher levels of H3K27ac and H3K4me1

marks and chromatin accessibilities than the GS&F-unbound enhancers, thus considered as latent

enhancers in cXEN cells (Figure 2.9A-C).
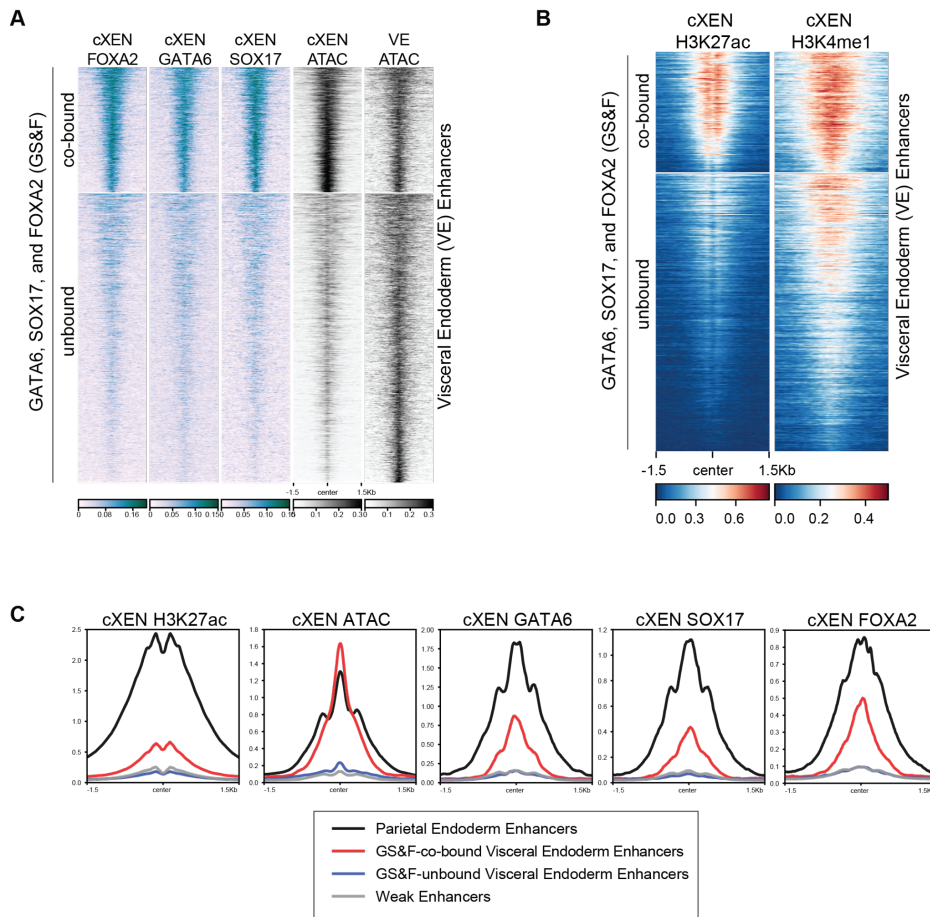
Figure 2.9.

Epigenetic states at

extraembryonic

endoderm

enhancers. (A)

Tornado plots

displaying signals of

chromatin

accessibilities and

GATA6, SOX17, and

FOXA2 bound

regions around

visceral endoderm enhancers identified in Figure 2.7A. (B) Tornado plots displaying signals of

50

H3K27ac and H3K4me1 at the visceral endoderm enhancers as clustered in Figure 2.9A. (C)

Meta profiles of the average enrichment for H3K27ac or ATAC signals, FOXA2, GATA6, or

SOX17 signals at the parietal endoderm enhancers, GS&F-co-bound visceral endoderm

enhancers, GS&F-unbound visceral endoderm enhancers, and weak enhancers.

Additionally, the nearby genes of GS&F-cobound VE enhancers were expressed higher

in extraembryonic VE (exVE) (Figure 2.10A, top panel), indicating that the exVE may be the

default cell fate of VE cells as previously proposed (Nowotschin et al., 2019). To characterize

highly specific targets of GS&F-cobound VE enhancers, we compared the bulk RNA-seq data

between E6.5 VE and E6.5 epiblast cells and removed the genes that were not differentially

expressed. The resulting genes, denoted as GS&F-cobound VE enhancer targets, were also

preferentially expressed in VE cells (Figure 2.10B, Table 2.2). Taken together, we propose that

the VE transcriptional program is jointly regulated by *Gata6, Sox17,* and *FoxA2*.
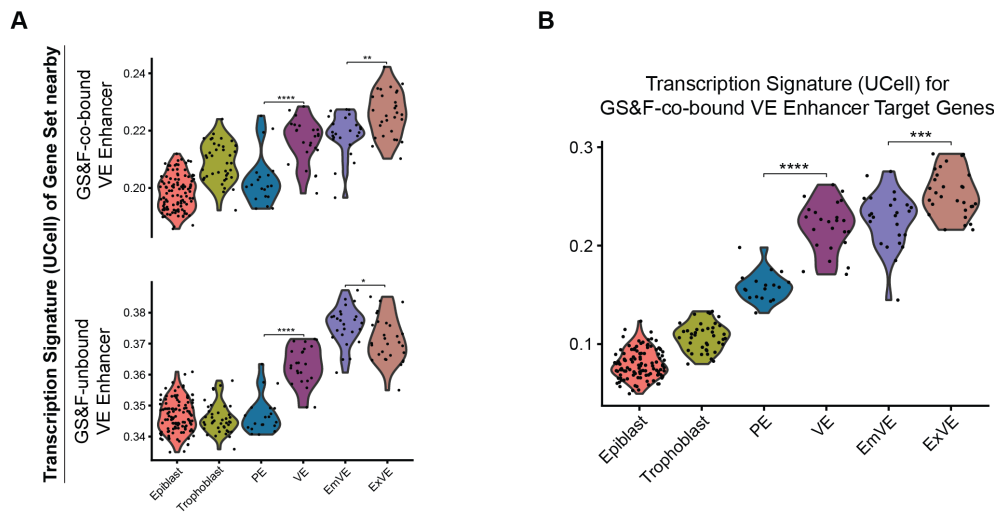
Figure 2.10. Correlation of GS&F-co-bound VE enhancer or GS&F-bound associated

genes with single cell expression states. (A) Violin plots displaying the transcriptional signature

(calculated by UCell) of assigned the cell states for genes nearby GS&F-co-bound visceral endoderm enhancers and GS&F-unbound visceral endoderm enhancers in E5.5 scRNA-seq datasets. **** denotes p<0.0001, ** denotes p<0.01, * denotes p<0.05 (Wilcoxon Test). (B) Violin plot displaying transcriptional signatures of various cell states based on E5.5 scRNA-seq data of differentially expressed genes between E6.5 visceral endoderm and E6.5 epiblast that overlap with the nearby genes of GS&F-co-bound visceral endoderm enhancer identified in (2.9A). *** denotes p<0.001 (Wilcoxon Test).

Table 2.2. GS&F-co-bound VE Enhancer Target Genes associated with Figure 2.10 (B)

GS&F-co-bound VE Enhancer Target Genes
Ebp, Creb3l3, Rbms2, Jcad, Scpep1, Dqx1, Frzb, Rab4a, Abhd3, Pbxip1, Chrnb1, Gcsh, Usp11, Zfp185, Nrg4, Pam, Mt2, Ggcx, 2510002D24Rik, Vps8, Gfi1, Ambp, Entpd3, Praf2, Smim14, Dgkk, Prss16, Soat2, Septin8, Lamp2, Amdhd2, Glrx, Rpn1, Zfp57, Eif2ak2, Gata3, Tmem120a, Adora2b, Mfsd4b1, Triqk, Ano9, Hykk, Txndc11, Ttr, Dennd5b, Fam161a, Slc39a3, Gpr39, Asns, Xkr8, Haus7, Zfp42, Xylb, Cdx2, Vasn, Slc44a3, Klhl2, Med12, Decr2, Fam3a, Pafah2, Nagpa, Gpr157, Abhd16a, Snx24, C130074G19Rik, Prr13, Ttc38, Leap2, Pld2, Ihh, Ccl25, Pxmp4, Slc12a6, Fzd6, Atp6v0b, Ssh2, Zfp52, Golm1, C1galt1c1, Atp8a1, Pxn, Pdgfra, Creld2, Rrbp1, Capn6, Borcs7, Slc46a3, Pxdc1, Cdc42bpb, Frmd4b, Tfeb, Scp2, Nkain4, Sel1l3, Lgals9, Ap1g2, Rdh10, Osbpl8, Lrp10, Ptafr, Dusp3, Leng1, Pxk, Fundc1, Slc38a9, Tmem125, Mkrn2os, Agrn, Slc13a4, Rarres1, Rwdd2b, Thpo, Gdpd1, H19, Cited1, Aplp2, Usp43, Rnf128, Slc39a4, Pitpna, Tanc2, Rp2, Mast2, Ppp3ca, Slc36a1, Dop1a, Slc8b1, Adh7, Cast, Tle6, As3mt, Twsg1, Dop1b, Mical1, Socs5, C2cd2l, Zfp202, Mfsd4b3-ps, Gpr155, Fkbp15, Gaa, Proc, Agtrap, Zhx3, Camsap3, Rflnb, Hgsnat, Aph1c, Hexb, Magt1, Bco1, Fbxl21, Slc52a2, Gas2l1, Tnip2, Trp53inp2, Marchf2, Renbp, Mfsd11, Dnmbp, Esyt1, Prss8, Mtarc2, Pvr, Mt1, Irak2, Dynlt3, Dpp4, Tor2a, Scamp5, Gng12, Coq8a, Lama5, B4galt3, Gba, Acadvl, F2, Synj1, Tmem219

2.2d Combinatorial Action of GATA6, SOX17 and FOXA2 Instructs VE Enhancer

To gain better understanding of the GS&F-cobound VE enhancers in (PE-like) cXEN cells, we compared the accessibility profiles between the latent GS&F-cobound VE enhancers and the active PE enhancers using ATAC-seq data in cXEN cells (Figure 2.11). A Vplot displaying ATAC-seq fragment distribution displayed that ATAC-seq fragment distribution centered at the active PE enhancers were highly open across the entire enhancer platform demonstrating an active TF recruitment (Figure 2.11, left panel). In contrast, TF occupancy was apparent at the central regions of GS&F-cobound VE enhancers (Figure 2.11, right panel) as revealed by a well-protected region by ATAC-seq.
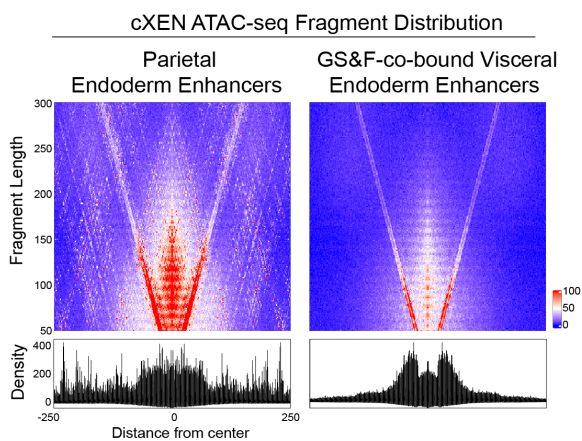
Figure 2.11. Comparison of parietal endoderm to GS&F-co-bound visceral endoderm enhancer accessibility profiles. Vplots displaying the fragment distribution for cXEN ATAC peaks centered at the parietal endoderm enhancers or the GS&F-co-bound visceral endoderm enhancers.

Column-wise summations were plotted at the bottom of the Vplots.

How do GS&F-cobound VE enhancers assume a latent state in PE cells, but become active in VE cells? To address this question, we employed TOBIAS (Transcription factor Occupancy prediction By Investigation of ATAC-seq Signal) software (Bentsen et al., 2020) to

analyze the differential TF footprinting profiles by comparing ATAC-seq data, in this case, the

GS&F-cobound VE enhancers between VE and PE-associated cell types (Figure 2.12).
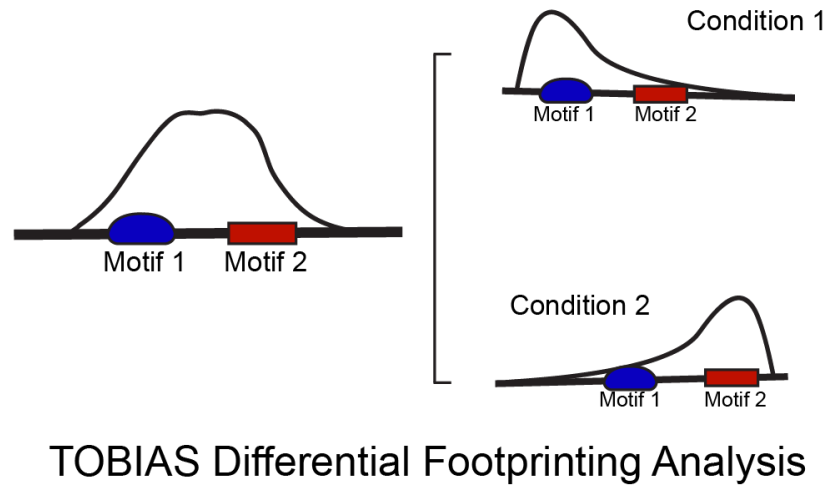
Figure 2.12. Schematic diagram illustrating differential footprinting analysis using TOBIAS software.

This analysis revealed that the accessibility of FOXA2, HNF1B, and HNF4A motifs in

both E6.5 VE and (PE-like) cXEN cells was unbiased as the differential footprinting scores of

these motifs were not significant (Figure 2.13A). This suggested that the VE-associated TFs

(FOXA2, HNF1B, HNF4A) maintained the accessibilities of GS&F-cobound VE enhancers

regardless of their states being latent or active. On the contrary, differential footprinting scores

for GATA6, SOX17, TEAD, and JUN/FOS were lower in the active enhancer state in E6.5 VE

(relative to (PE-like) cXEN) cells (Figure 2.13A). Indeed, lower expression levels of *Gata6* and

Sox17 in VE (relative to (PE-like) cXEN) cells (Figure 2.5C and 2.13B) coincided with the

timing of GS&F-cobound VE enhancer activation. This was also accompanied by a concomitant

loss of GATA6 and SOX17 footprinting as well as gain of (VE-associated) LHX footprinting in

E6.5 VE cells (Figure 2.13A and 2.13B). Collectively, these data denote a VE gene regulatory

circuit composed of (PE-associated) GATA6 and SOX17 repressing the VE regulatory program,

while requiring positive inputs from (VE-associated) FOXA2, HNF1B, and HNF4A (Figure

2.13C).
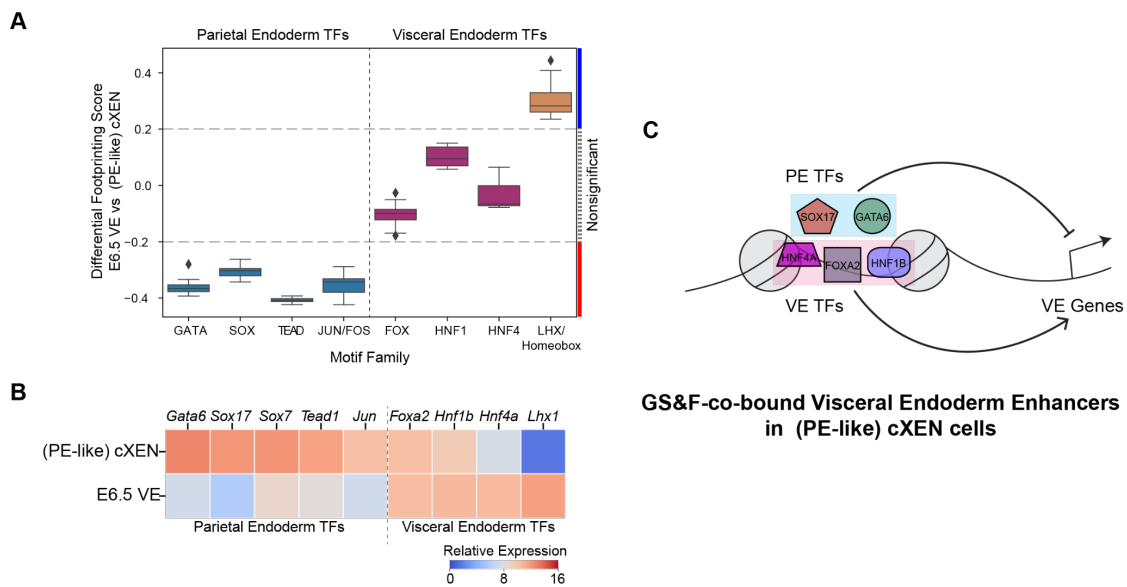
Figure 2.13. Differential transcription factor (TF) footprinting at the GATA6, SOX17, and

FOXA2 (GS&F) co-bound VE enhancers. (A) Differential footprinting scores for motifs

enriched at GS&F-co-bound VE enhancers between cXEN ATAC-seq and E6.5 visceral

endoderm ATAC-seq. Similar motifs (PWM matrixes) obtained from the JASPAR and

HOCOMOCO databases were clustered into motif families and significance was assigned by

TOBIAS, which was visually represented as a boxplot per TF family. Binding scores (y-axis)

represent footprint scores calculated by TOBIAS to indicate preferential binding in PE-like

(cXEN) cells (i.e. <0 on the y-axis) and preferential binding in E6.5 VE cells (i.e. >0 on the y-

axis). Log10(p-values), for each PWM matrix can be found in Table S8. "Significant" differentially bound TFs (indicated by the blue and red color coded bars to the right side of the graph) assigned to PE-like (cXEN) or E6.5 VE cells were defined by default thresholds set by the TOBIAS software by cutting-off the top 5% changed TFs in each direction. (B) Relative gene expression profiles calculated by DESeq2 median of ratios for TFs belonging to the TF families in (A) from cXEN and E6.5 visceral endoderm RNA-seq datasets. (C) Summary of proposed model: Either GATA6 or SOX17 represses VE gene program while FOXA2, HNF1B, or HNF4A maintains and promotes the VE gene program by associating with GS&F-co-bound enhancers in (PE-like) cXEN cells.

2.2e GATA6 and SOX17 Promote PE Development via Repression of VE Gene Program and Activation of *Mycn*

GATA6 and SOX17 appear to have a dual function - positively promoting the PE cell state, while repressing the VE cell state. To examine the roles of GATA6 and SOX17 in PE cells, cXEN cells were subjected to targeted protein degradation using a dTAG-inducible FKBP12^{F36V} degron system (Nabet et al., 2018). A DNA template encoding the dTAG-inducible FKBP12^{F36V} degron was inserted into the *Gata6* or *Sox17* locus in cXEN cells using CRISPR/Cas9-guided gene editing. The dTAG-inducible FKBP12^{F36V} degron cassette included a mNeonGreen reporter gene to follow the degradation efficacy of GATA6 and SOX17 fusion proteins in cXEN cells (Figure 2.4A and 2.S4A). We generated homozygous inserted knock-in *Gata6*-FKBP12^{F36V} and *Sox17*-FKBP12^{F36V} cXEN cell lines (Figure 2.14A and 2.14B). Effectiveness of the degron

system was examined by treating these cells with dTAG^V-1, followed by flow cytometric analysis (Figure 2.14C). The degron tagged GATA6 and SOX17 proteins were rapidly degraded upon dTAG^V-1 treatment (Nabet et al., 2020) as examined by disappearance of green fluorescent by flow cytometric analysis and by western blot analysis (Figure 2.14C and 2.14D).
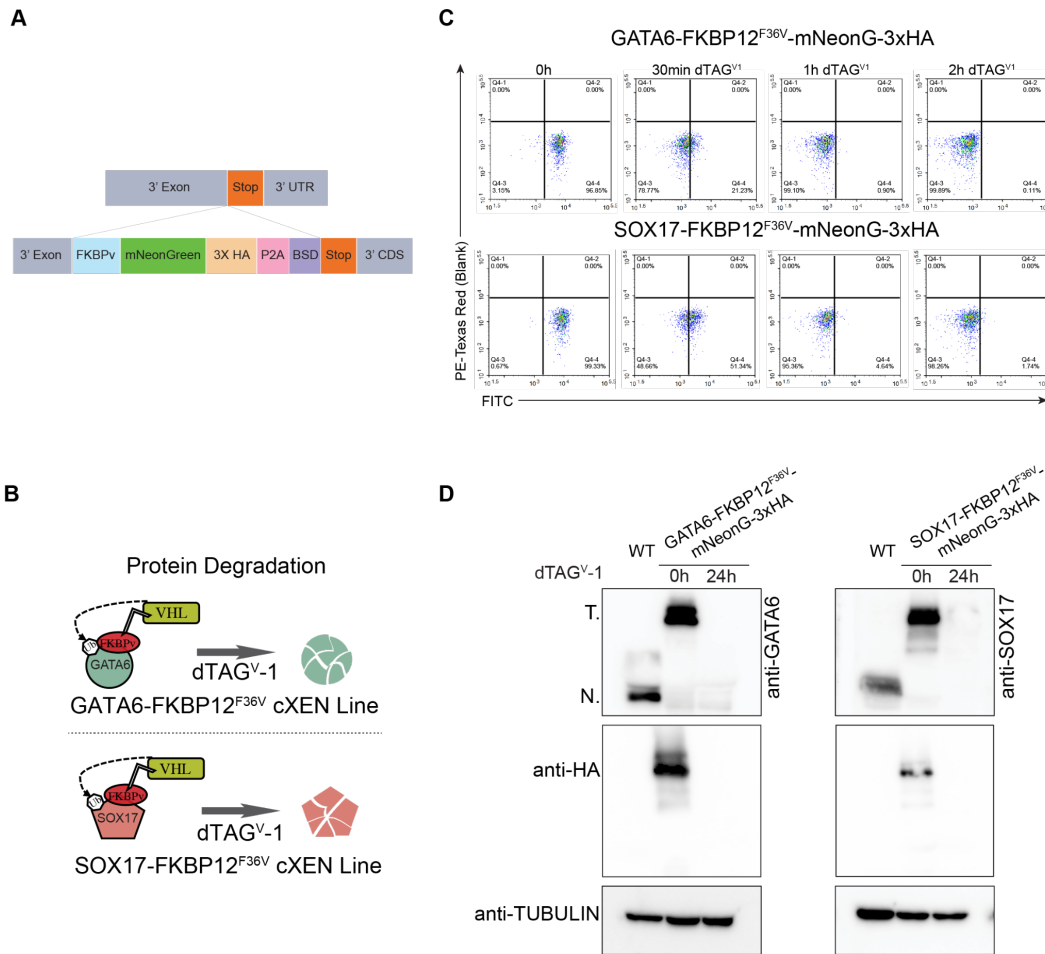
Figure 2.14. Generation of *Gata6*-FKBP12^{F36V} and *Sox17*-FKBP12^{F36V}. (A) Schematic illustration of the FKBP12^{F36V} cassette used to target the C-terminus of the endogenous *Gata6* or *Sox17* locus. (B) Schematic illustrating the strategy for acute degradation of GATA6 or SOX17 using the dTAG inducible FKBP12^{F36V} degron system. (C and D) Flow cytometric analysis (C)

and western blots (D) after acute degradation of GATA6 or SOX17 in homozygously inserted

cXEN clones. T: tagged-sized protein, N: native-sized protein.

Next, we performed time-course (0-96 hours) RNA-seq experiments to monitor the

transcriptomic alterations after GATA6/SOX17 depletion in cXEN cells. Heatmap clustering

analysis of RNA-seq samples showed major transcriptomic changes occurring after 48 hours of

GATA6 or SOX17 depletion (Figure 2.15A). Interestingly, this clustering analysis also revealed

that the transcriptomes under GATA6 and SOX17 depletion shared a certain level of similarity

(Figure 2.15A), suggesting that these two TFs regulate common pathways in cXEN cells. Next,

we generated a rank ordered gene list based on expression changes that occurred 72 hours after

the dTAG^V-1 treatment. Initially, gene set enrichment analysis (GSEA) (Subramanian et al.,

2005) uncovered that GATA6/SOX17 upregulated the VE gene module identified from scRNA-

seq data and the GS&F-cobound VE enhancer targets (Figure 2.15B, Table 2.2 and 2.3).
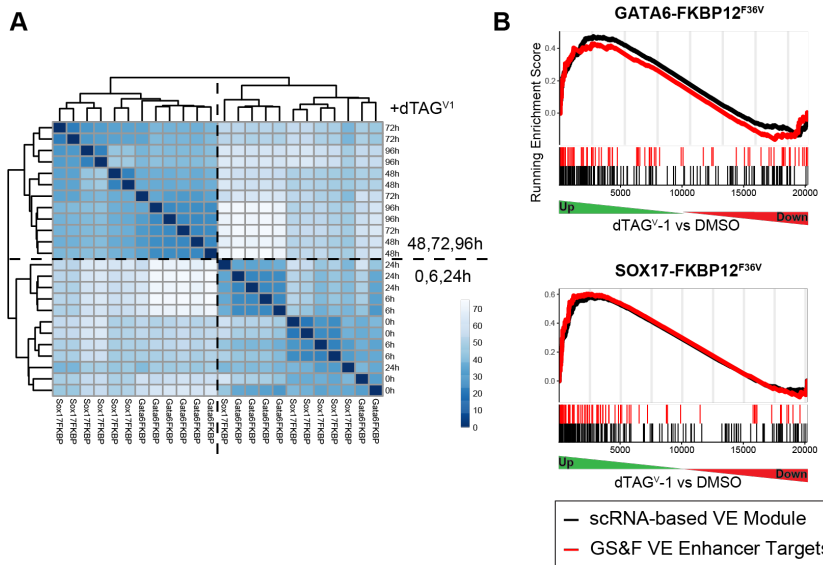
Figure 2.15. GATA6 and SOX17 regulate common pathways related to VE cell identity. (A) Hierarchical clustering heatmap of the euclidean distance matrix displaying the similarities and dissimilarities among the time

course RNA-seq samples during GATA6 or SOX17 depletion in cXEN cells. Color bar

represents euclidean distance values. (B) Barcode plot showing gene set enrichment analyses (GSEA) on gene sets (legend on the right) against ranked genes according to expression changes in 72h dTAGV-1 versus DMSO treated *Gata6*-FKBP12[F36V] or *Sox17*-FKBP12[F36V] cXEN cells.

Table 2.3. GS&F-co-bound VE Enhancer Target Genes associated with Figure 2.15 (B), and Figure 2.23 (C)

scRNA-based VE Module
Sigirr, Sycp3, Tmem37, Uap1l1, Soat2, Cotl1, Cyp26a1, Slc9a3r1, Pga5, Fth1, Ctsc, Atp6v1d, Clic6, Commd4, Atp6v0d1, Lhx1, Agpat3, Slc16a13, Fgf5, Gcat, Gcat, Has2, Cldn6, As3mt, Cldn7, Cstb, Ftl1, Ftl1-ps1, Slc7a7, Atp6v1h, Gpr137b, Atp6v0b, Ctsl, Ttr, Cndp2, Dapk2, Plin3, Slc40a1, Gramd1b, Ifi30, Acot13, Vamp8, Clic1, Atp6v1e1, Gnpda1, Aqp8, Rab11a, Lgmn, Gng12, Adk, Atp8a1, Atp6v1g1, Lratd2, Gabarap, Crb3, Apom, Ctsd, Neu1, Rdh10, Ctsh, Rhoc, Mt1, Vps29, Mfsd12, H19, Plekha2, Lrrc8d, Sirt2, Gba, Rab5c, Slc43a2, Acat2, Mttp, Cd59a, Maf, Srgn, Vps26a, Vegfb, Blvrb, Pcyt2

To corroborate the affected gene programs after GATA6 and SOX17 depletion, we performed Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath, 2008) and the resulting heatmap was grouped into four gene modules, which we denoted as: GATA6 and SOX17 (GS) co-upregulated genes, GATA6-only upregulated genes, SOX17-only upregulated genes, and GS co-downregulated genes (Figure 2.16A). Noticeably, the marker gene for VE cells, *Ttr*, appeared in the GS co-upregulated module (Figure 2.16A). GO analysis demonstrated that GS co-upregulated module was enriched for terms related to VE biological functions including nutritional supports (e.g., endocytosis and lipid metabolism), vasculogenesis (e.g., vasculature development and blood vessel migration), and BMP signaling

(Filimonow and Fuente, 2021) (Figure 2.16B). Also, *Gata6*-only and *Sox17*-only upregulated

modules were enriched for these terms but with different emphases (Figure 2.16B). In sum,

Gata6 and *Sox17* appeared to have distinct functions in VE differentiation, and work together to

suppress the VE transcriptional program in PE-like cXEN cells.
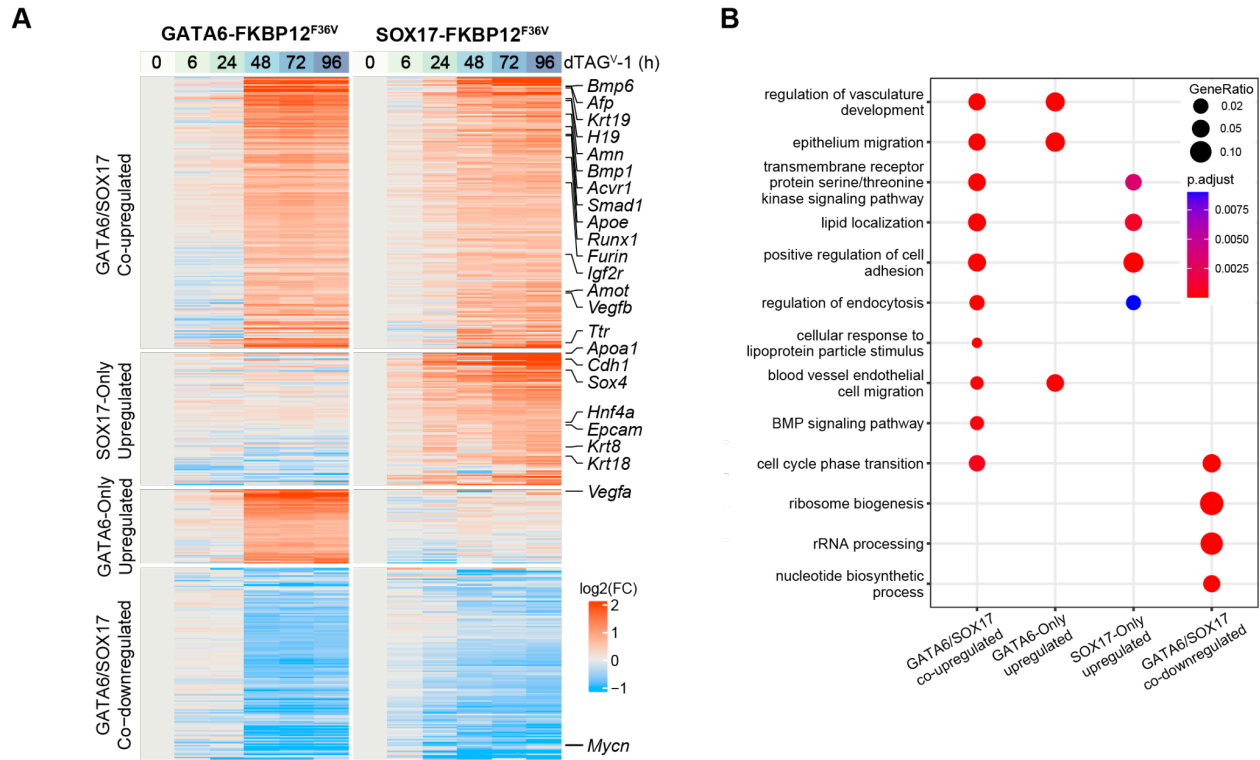
Figure 2.16. Characterization of GATA6 and SOX17 affected pathways. (A) Heatmap

displaying fold changes (log2 normalized TPM versus 0h) of genes from gene modules identified

by WGCNA during the course of GATA6 and SOX17 depletion. Selected genes were annotated.

GATA6/SOX17 Co-upregulated # of genes: 960, GATA6/SOX17 Co-downregulated # of genes:

681, SOX17 Upregulated # of genes: 469, GATA6 Upregulated # of genes: 265. (B) Aggregated

functional profiles of gene ontology biological processes enriched in gene modules in (A)

generated by clusterProfiler. Significance is color coded by p-adjusted values. Gene ratio reflects

the number of genes associated with the gene ontology grouping over the total number of input

genes

Finally, GO-term analysis of GS co-downregulated module showed enrichment of terms

such as cell cycle regulation and biosynthesis (Figure 2.16B). These terms matched the functions

of *Myc*-mediated pathways, which were linked to activities associated with the PE function

revealed by the single-cell KEGG pathway analysis (Figure 2.3A). To further characterize the

role of *Myc* in PE cells, we examined whether Myc target genes were also affected using the rank

ordered gene list as described earlier. GSEA confirmed that 'hallmarks of *Myc* targets' (from the

Molecular Signature Database, MsigDB) (Liberzon et al., 2015) were negatively regulated when

GATA6 or SOX17 was depleted (Figure 2.17B). Interestingly, the expression of *Myc* was

downregulated in SOX17-depleted condition, but the expression of *Mycn* was decreased in both

GATA6- and SOX17-depleted cXEN cells (Figure 2.17A). It is worth noting that Mycn was

previously identified as a PE-associated TF based on scRNA-seq clustering analysis (Figure

2.5C, Table 2.1). Collectively, these data indicate that *Gata6* and *Sox17* reinforce PE-cell

identity by simultaneously promoting self-renewal property via activation of *Mycn* and *Myc,*

while repressing the VE gene regulatory program.

Figure 2.17. Characterization of GATA6 and SOX17 MYC/MYCN-mediated pathways.

(A) Heatmap displaying log2 fold changes of *Myc* and *Mycn* transcripts in normalized TPMs of

GATA6- or SOX17-depleted cXEN cells relative to the 0h DMSO control. (B) Barcode plot

showing gene set enrichment analyses (GSEA) on gene sets (legend on the right) against ranked

genes according to expression changes in 72h dTAGV-1 versus DMSO treated *Gata6*-

FKBP12[F36V] or *Sox17*-FKBP12[F36V] cXEN cells.

2.2f FOXA2 Maintains VE Gene Program Activation Potential and Represses PE-associated Mycn

FOXA2 is a well-known pioneer TF regulating definitive and visceral endoderm development

(Ang and Rossant, 1994; Burtscher and Lickert, 2009; Cirillo et al., 2002; Dufort et al., 1998;

Weinstein et al., 1994). Here, our current data suggests that *FoxA2* also functions in ExEn

lineage bifurcation to PE or VE cell fate, based on the modest expression in PrE cells and abundant expression of FoxA2 in VE cells (Figure 2.5D), the presence of FOX motif enrichment in accessible VE chromatin regions (Figure 2.8B), and the physical binding of FOXA2

62
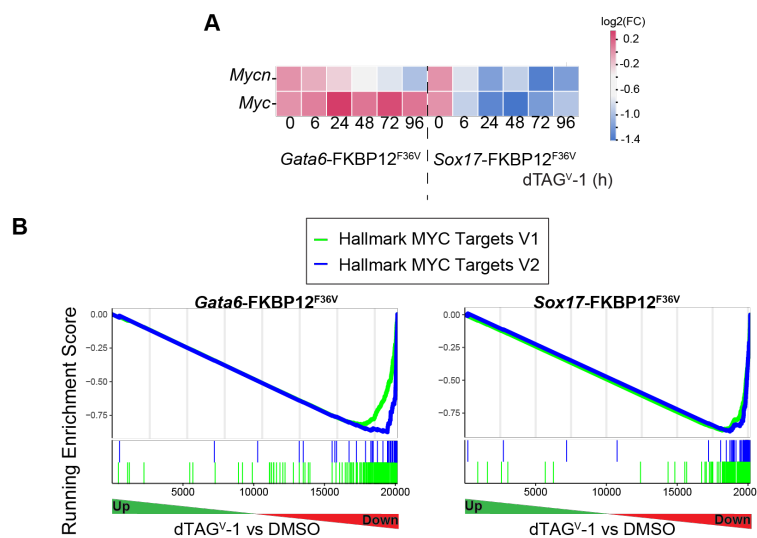
to the enhancers of cXEN cells (Figure 2.9A and 2.9C). These pieces of evidence suggest that

FoxA2 plays a fundamental role during ExEn lineage divergence to PE or VE cell fate. We

examined how the VE gene regulatory program in cXEN cells was affected by *FoxA2* using

FoxA2 knock-out (KO) cell lines. The DNA binding domain in the DNA sequence of *FoxA2*

gene was deleted using a standard CRISPR/Cas9 targeting method (Figure 2.18A) in *Gata6*-

FKBP12^{F36V} or *Sox17*-FKBP12^{F36V} cXEN cells. These modified cXEN cells lacked the full

length FOXA2 as confirmed by western blot (Figure 2.18B) and transcriptomic analysis o*f*

FoxA2 mutant cell lines (Figure 2.18C).

Figure 2.18. Generation of *FoxA2*-KO lines. (A) Schematic illustrating the strategy for acute

degradation of GATA6 or SOX17 (Figure 2.14B) in combination with *FoxA2* DNA binding

A

Gata6 or *Sox17*-FKBP12^F36V   *FoxA2*-KO

+dTAG^V-1

B

FOXA2 KO

WT FOXA2 / SOX17-FKBP12^F36V / GATA6-FKBP12^F36V

kDA

70

FOXA2→

37

anti-TUBULIN
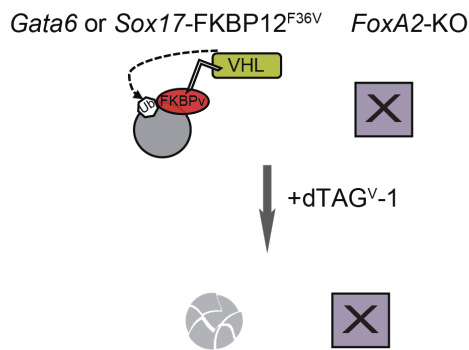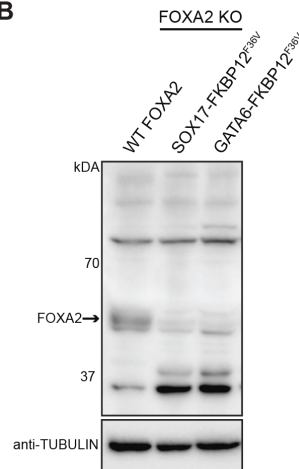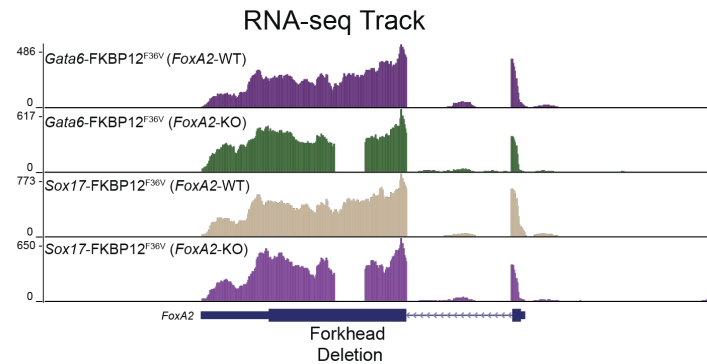
domain deletion. (B) Western blot examining FOXA2 expression in various cell lines indicated on the top. (C) Genome browser tracks displaying RNA-seq reads in the *FoxA2* locus.

C

RNA-seq Track

486 *Gata6*-FKBP12^F36V (*FoxA2*-WT)
0
617 *Gata6*-FKBP12^F36V (*FoxA2*-KO)
0
773 *Sox17*-FKBP12^F36V (*FoxA2*-WT)
0
650 *Sox17*-FKBP12^F36V (*FoxA2*-KO)
0

FoxA2

Forkhead Deletion

We first examined the effect of FOXA2 on extraembryonic lineage development. RNA-seq data collected from both FoxA2-KO and WT cXEN cells were subjected to Differentially Expressed Genes (DEGs) analysis (Love et al., 2014). Several VE-associated genes such as *Ttr* and *Hnf4a* were downregulated, while *Mycn* was upregulated (Figure 2.19A). GSEA analysis was performed on rank ordered gene list based on expression changes comparing FOXA2 knockout to WT cXEN cells to determine whether FOXA2 function as an activator or repressor on GS&F-cobound enhancers. We found that the gene regulatory direction of FOXA2 was opposite to that of GATA6 or SOX17 (Figure 2.19B), suggesting that FOXA2 has an activating effect and is required for sustaining the basal expression levels of the GATA6/SOX17-repressed VE gene program in cXEN cells.
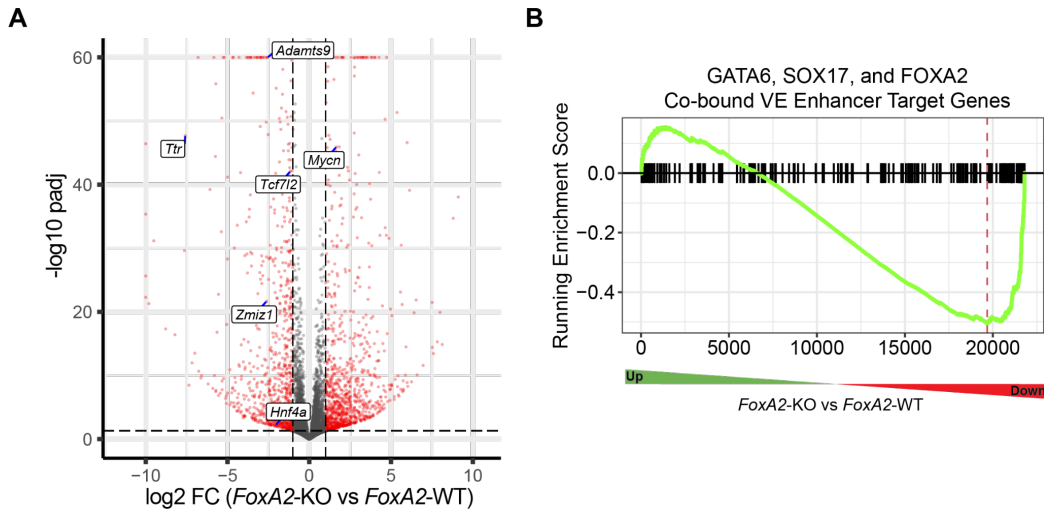
Figure 2.19. Transcriptomic analysis on *FoxA2*-KO cXEN cells. (A) Volcano plot comparing the transcriptomes of *FoxA2*-KO versus *FoxA2*-WT cXEN cells. (B) Barcode plot showing gene set enrichment analyses (GSEA) on GS&F-co-bound VE enhancer targets (Figure 2.15B) against ranked genes according to expression changes in *FoxA2*-KO versus *FoxA2*-WT cXEN cells.

Next, we treated *Gata6*-FKBP12^{F36V}/*FoxA2*-KO and *Sox17*-FKBP12^{F36V}/*FoxA2*-KO cXEN cells with and without dTAG^V-1 and performed RNA-seq analyses to examine the effect of FOXA2 on GATA6 and SOX17 depleted cXEN cells. Figure 2.20A and 2.20B shows how the cohorts of up-regulated and down-regulated genes that were co-bound by both GATA6 and SOX17 (See Figure 2.16A, group 1 and group 4) were affected by FOXA2 depletion. FOXA2 deletion prevented the activation of genes affected by GATA6/SOX17 co-regulated (Figure 2.16A, group 1), SOX17 only (group 2) and GATA6 only (group 3), and SOX17 only (group 2) genes (Figure 2.20A and 2.20B). On the contrary, a cohort of GATA6/SOX17 down-regulated genes (group 4) were upregulated in the absence of FOXA2 (Figure 2.20A and 2.20B). The data support the view that FOXA2a is a dual-function TF. Interestingly, transcriptomics of cXEN

cells lacking both GATA6/FOXA2 and SOX17/FOXA2 were very similar to that of FOXA2-KO

(Figure 2.20C). These data underscore the notion that FOXA2 maintains the activation potential

of the VE gene program in (PE-like) cXEN cells upon eviction of GATA6 or SOX17 on VE

enhancers.
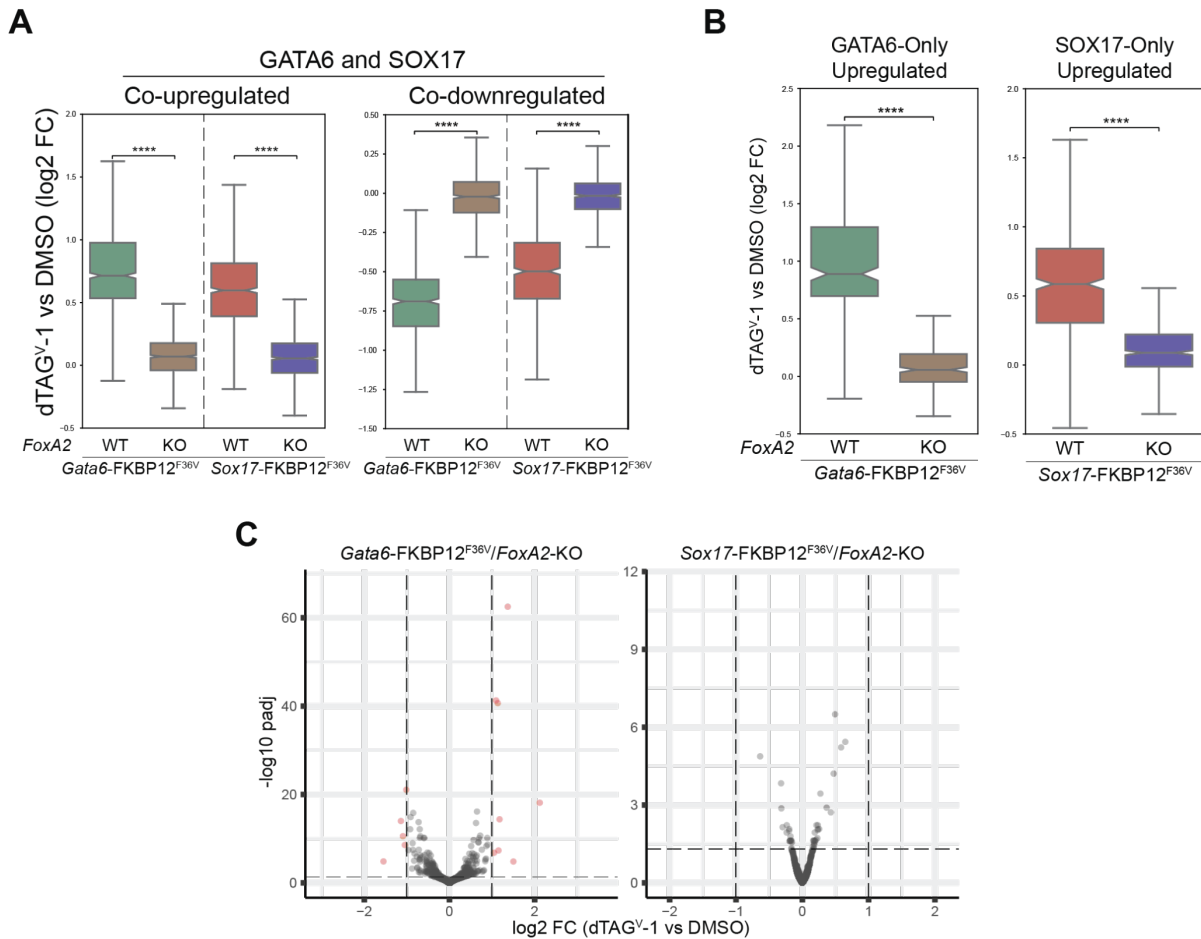
Figure 2.20. FOXA2 counteracts GATA6 and SOX17 in regulating visceral endoderm gene program. (A) Boxplot displaying log2 fold changes of GATA6 or SOX17 depletion (dTAG^{V}-1 versus DMSO) in *FoxA2*-KO or *FoxA2*-WT background for GATA6 and SOX17 Co-upregulated and Co-downregulated modules identified in Figure 2.16A. **** denotes p<0.0001 (Wilcoxon Test) (B) Boxplots displaying log2 fold changes of genes affected after GATA6 or SOX17

depletion in *FoxA2*-KO or WT cell lines. Genes identified in GATA6-Only or SOX17-Only Upregulated modules in Figure 2.16A were used. **** denotes p<0.0001 (Wilcoxon Test). (C) Volcano plots comparing the transcriptomes of GATA6 or SOX17 depleted versus the control on *Gata6*-FKBP12^{F36V}/*FoxA2*-KO or *Sox17*-FKBP12^{F36V}/*FoxA2*-KO cell lines. The number of differentially expressed genes in dTAG^V-1 *Gata6*-FKBP12^{F36V}/*FoxA2*-KO vs DMSO *Gata6*-FKBP12^{F36V}/*FoxA2*-KO are zeros. 0 genes > (1 log2 fold change) and 0 genes < -1 (log2 fold change). The number of differentially expressed genes in dTAG^V-1 *Sox17*-FKBP12^{F36V}/*FoxA2*-KO vs DMSO *Sox17*-FKBP12^{F36V}/*FoxA2*-KO are 8 and 5, respectively. 8 genes > (1 log2 fold change) and 5 genes < -1 (log2 fold change).

We also investigated how the loss of *FoxA2* affects cohorts of *Myc/Mycn* targets as *Myc/Myc* appeared to play a central role in PE cell fate specification. *Mycn* expression was significantly increased in FoxA2-KO cXEN cells relative to WT cXEN cells (Figure 2.21A). Consequently, GSEA analyses of MYC target genes were induced as well (Figure 2.21B). Depletion of GATA6 or SOX17 alone, in FoxA2-KO background, can no longer downregulate *Mycn* expression (Figure 2.21A) and (MYC/MYCN-mediated) GS co-downregulated module (Figure 2.20A), indicating that FOXA2 is a potent repressor for *Mycn* in PE cells. Thus, MYCN activity is finely controlled by GATA6, SOX17, and FOXA2, perhaps preventing the abnormal growth of PE cells. Taken together, FOXA2 counteracts GATA6 and SOX17 in regulating both the VE gene program and MYC/MYCN pathways.
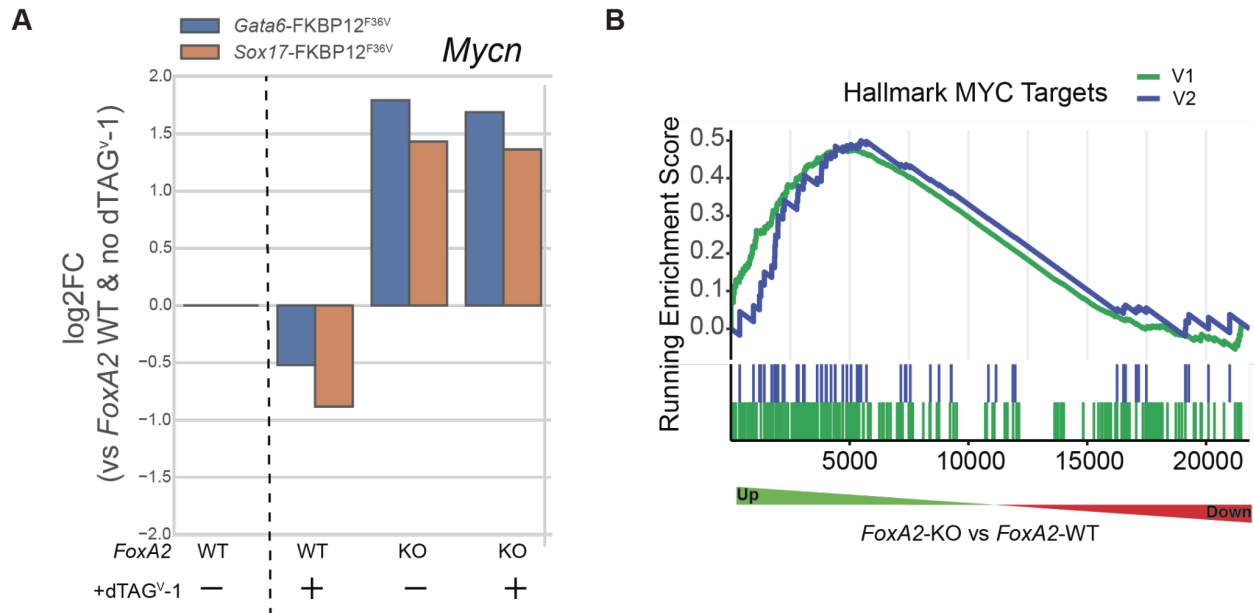
Figure 2.21. FOXA2 represses *Mycn* expression. (A) Barplot displaying fold changes in *Mycn* expression comparing to *Gata6*-FKBP12^F36V (blue) or *Sox17*-FKBP12^F36V (orange) parental cell lines in the absence of dTAG^V-1 (first group in the x- axis) for *FoxA2*-KO or *FoxA2*-WT backgrounds with and without 48h dTAG^V-1 treatment. (B) Barcode plot showing gene set enrichment analyses (GSEA) on MYC target genes against ranked genes according to expression changes in *FoxA2*-KO versus *FoxA2*-WT cXEN cells.

2.2g VE specification requires BMP signaling and integration of Gata6, Sox17, Mycn, and FoxA2 network

GO term analysis of GATA6 and SOX17 targets (Figure 2.4C) identified BMP signaling pathways as critical components affected after GATA6/SOX17 depletion in cXEN cells (Figure 2.22A). GSEA analysis demonstrated that both canonical (mediated by SMAD proteins) and non-canonical (transduced by ERK cascade) BMP signaling pathways were utilized (Figure

68

2.22B). The expression of individual BMP signaling components such as BMP ligands, a BMP receptor, Smad1, and multiple MAPK kinases were identified (Figure 2.22A). This suggests that GATA6 and SOX17 attenuate BMP signaling in cXEN cells.
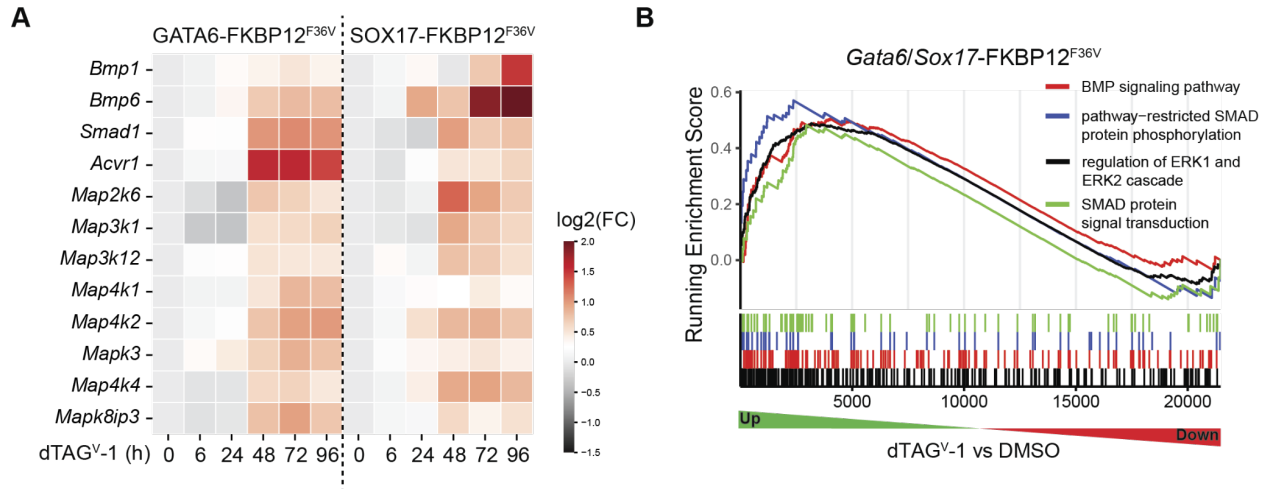
Figure 2.22. GATA6 and SOX17 represses BMP signaling pathways. (A) Heatmap displaying log2 fold changes in normalized TPMs of GATA6- or SOX17-depleted cXEN cells relative to the control for selected genes involved in BMP signaling and ERK pathways as annotated in gene ontology databases. (B) Barcode plot showing gene set enrichment analyses (GSEA) on BMP signaling pathway components against ranked genes according to the expression changes in 96h dTAG^V-1 versus DMSO treated *Gata6*-FKBP12^{F36V} or *Sox17*-FKBP12^{F36V} cXEN cells.

To understand the impact of BMP signaling on cXEN cells, cXEN cells were treated with BMP4 and the gene expression changes were measured using RNA-seq. BMP4 treatment downregulated the expression of PE-associated TFs, such as *Gata6, Sox17*, and *Mycn*, while upregulated the expression of VE-associated TFs (Figure 2.23A). We also performed an enrichment analysis of hallmark gene sets (Figure 2.23B) and detected induction of gene terms

related to VE gene regulation (Artus et al., 2012; Paca et al., 2012) along with suppression of MYC/MYCN pathways (Figure 2.23B and 2.23C). This recapitulated the transcriptomic phenotype of GATA6/SOX17-depleted cXEN cells (Figure 2.23D). In summary, we proposed that PE and VE are antagonistic cell fates whose divergence is determined by the integration of BMP signaling and the transcription factor network wired by GATA6, SOX17, MYCN, and FOXA2 (Figure 2.24).
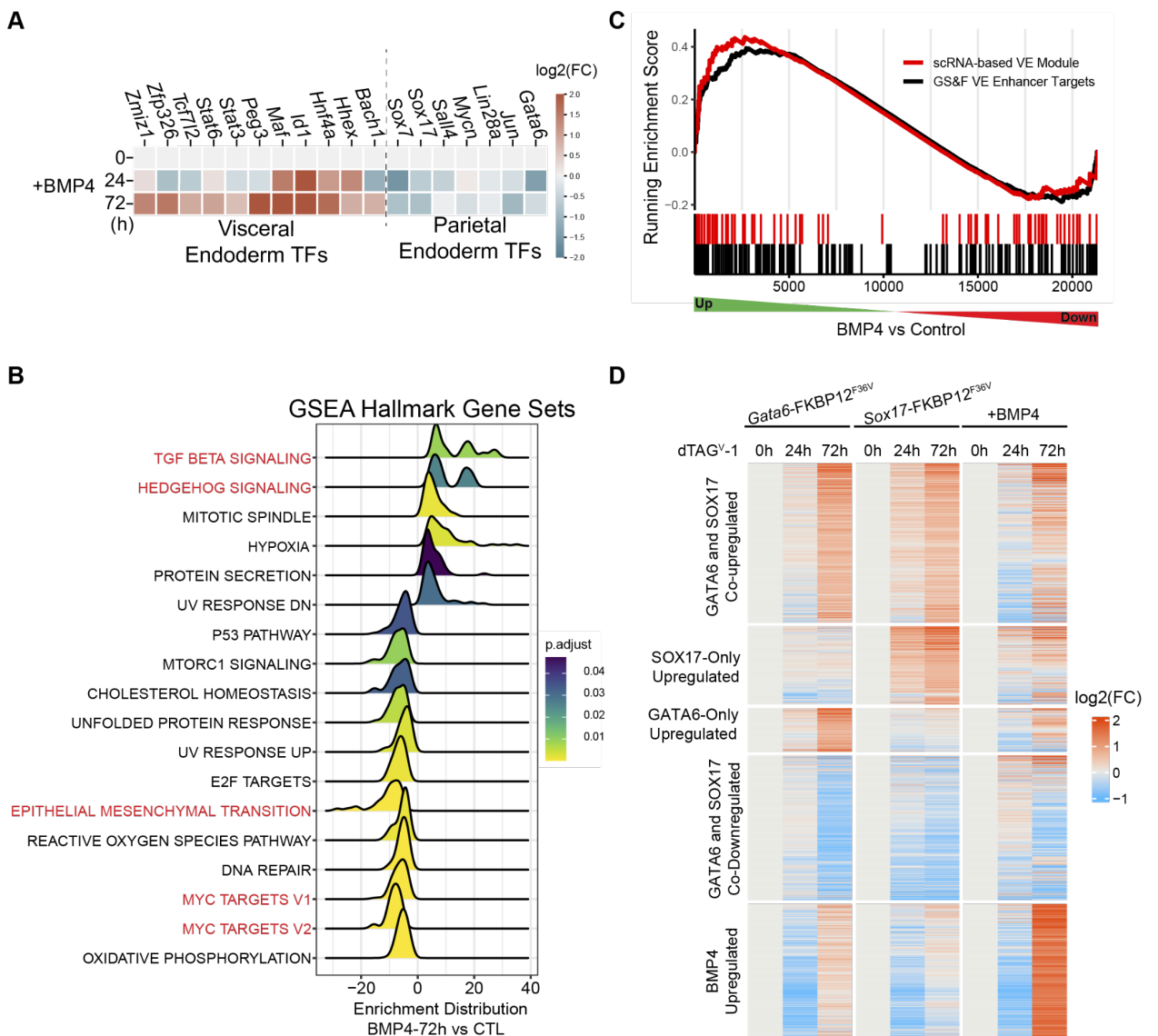
Figure 2.23. BMP signaling is a signaling cue for VE cell differentiation. (A) Heatmap displaying log2 fold changes in normalized TPMs of BMP4 treated cXEN cells relative to the control for differentially expressed (q < 0.05, 72 h vs. 0h DESeq2) TFs in parietal endoderm transcription factor (TF) set (see Figure 2C) or visceral endoderm TF set (see Figure 2.5 C and D). (B) Ridgeplot displaying the distributions of expression changes comparing 72h BMP4 treated cXEN cells versus the control for core genes of the enriched pathways from the MSigDb collection. The significance of pathway enrichments was color coded. (C) Barcode plot showing GSEA on scRNA-based or GS&F-cobound VE enhancer-based gene sets against ranked genes according to the expression changes in 72h BMP4 treated cXEN cells versus the control. (D) Heatmap displaying fold changes (log2 normalized TPM versus 0h) of genes from gene modules identified in Figure 2.16A as well as 72h BMP4 uniquely induced genes during the course of GATA6 depletion, SOX17 depletion, or BMP4 treatment.
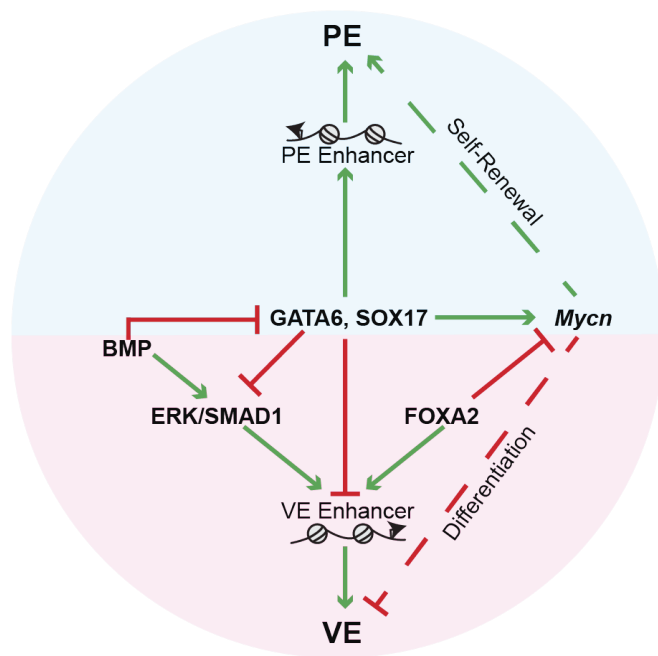
Figure 2.24. Summary of proposed model. GATA6 and SOX17 repress the VE gene program and activate *Mycn*, which is opposed by FOXA2. Reciprocal repression between GATA6 and SOX17 with BMP signaling modulates the relative expression between PE and VE gene programs.

2.3 Discussion

Here we attempted to bridge the knowledge gap about the lineage segregation mechanism of primitive endoderm (PrE) into PE versus VE cells. Through careful mining of the scRNA-seq datasets centered around E4.5 embryos, we have defined transcriptional activities closely associated with PrE, PE, and VE cell states at the onset of PE/VE cell lineage segregation. This systematic analysis allowed us to identify previously identified TFs involved in PrE development as well as other TFs that have not been linked with ExEn development. We discovered that PrE is a bipotential progenitor cell, and that the decision to become PE or VE is dictated by an interplay between GATA6, SOX17 and FOXA2, as well as the contribution of BMP signaling.

2.3a TF network underlying PE and VE specification

The genetic roles of GATA6 and SOX17 in PrE have been studied extensively. GATA6-KO

mice are embryonically lethal and are required for PrE specification *in vivo*, where in the absence

of GATA6, ICM cells adopt an epiblast cell fate (Cai et al., 2008; Koutsourakis et al., 1999;

Schrode et al., 2014). These mouse phenotypes were further corroborated where GATA6 was

found to be a potent reprogramming of mouse embryonic stem cells (mESCs) to extraembryonic

endoderm (XEN) cell lines (Wamaitha et al., 2015). Due to the early GATA6 phenotypes, the

role of GATA6 in PE-VE lineage divergence is not well-understood, though, GATA6 protein

expression is lost in future VE cells and maintained in PE cells (Cai et al., 2008; Wallingford et

al., 2013). SOX17 is detected at the early blastocyst stage and later strongly expressed in PE-

destined cells relative to VE cells (Artus et al., 2011). SOX17-KO mice can survive until

midgestation (Kanai-Azuma et al., 2002). Though, extraembryonic endoderm (XEN) cell lines

can be established by SOX17 overexpression (McDonald et al., 2014). Additionally,

overexpression of SOX17 in 8-cell stage embryos influences Epi-PrE cell fate allocation (Morris

et al., 2010). To clarify SOX17's role in ExEn development, SOX17-KO embryos were induced

in a diapause state where SOX17 was found to play a role in PrE lineage maintenance (Artus et

al., 2011). While these analyses establish the essential functions of GATA6 and SOX17 in

extraembryonic structure development, the regulatory mechanism underlying the process has

remained elusive.

Here, we have built a core TF network that controls PE and VE cell fates, largely based

on a bioinformatics approach (Figure 2.24). Specifically, we compared the enhancer landscapes

between PE (XEN cells) and VE cells to identify TF binding motifs that are associated with PE-

specific or VE-specific enhancers. Both PE and VE enhancers were enriched with GATA and

SOX motifs, suggesting the importance of these TFs for PE and VE cell differentiation. A notable feature of TF binding motifs in VE cells is the strong co-enrichment of a FOXA2 (VE TF) binding motif with GATA6 and SOX17 (PE TFs) binding motifs suggesting a functional interaction in VE specification. To directly address this question, we have used a dTAG-inducible FKBP12^{F36V} degron system (FKBP12^{F36V}-based PROTAC system) to deplete GATA6 and SOX17 in PE-like cXEN cells and interrogated the transcriptomic alterations in the presence and absence of FOXA2. Specifically, we found a TF regulatory circuit where GATA6 and SOX17 suppress the VE gene program, and the activation potential is maintained by FOXA2. We propose a model during PrE lineage divergence, where the absence of FOXA2, GATA6 and SOX17 promote PE cell formation, while suppressing the VE gene program in future PE cells. On the contrary, in the presence of FOXA2, PrE cells differentiate toward the VE cell lineage. It is currently unknown how the occupancies of GATA6 and SOX17 impart repression on VE enhancers, while the same TFs impart activation on PE enhancers. We postulate one potential mechanism is that the motif combination (enhancer grammar) (Jindal and Farley, 2021) influences GATA6 and SOX17 with FOXA2 (compared to that without FOXA2) physical occupation on enhancers, thereby dictating distinct enhancer activities. Alternatively, enhancer activities may be influenced by alterations in epigenetic states of PE versus VE enhancers by differential TF recruitment of epigenetic cofactors and/or remodelers.

In addition to GATA6, SOX17 and FOXA2 TF motifs, our analysis suggests the involvement of HNF1B and HNF4A in VE differentiation, which agrees with the previous findings of HNF1B and HNF4A requirement in VE development (Barbacci et al., 1999; Chen et al., 1994; Coffinier et al., 1999; Duncan et al., 1997). We had only investigated the function of

FOXA2, the highest enriched VE-associated motifs in the VE enhancer. But the motifs of HNF1B and HNF4A were also considerably enriched with strong footprinting even when VE enhancers assume a latent state. How HNF1B and HNF4A cooperate with FOXA2 to potentiate VE gene program is unclear. We also found other key VE TF motifs such as LHX which play an important role during gastrulation given its expression in anterior VE (associated with emVE tissues) (Costello et al., 2015). Additionally, we discovered several PE TF motifs such as JUN/FOS and TEAD which have yet to be characterized in both PE-VE lineage divergence and PE development. In the future, it will be useful to understand the roles of these TFs in context with enhancer activities and the DNA sequence features of the VE and PE enhancers as they may manifest the evolutionary strategy that led to the extraembryonic endoderm when the first amniotes evolved from amphibian ancestors approximately 340 million years ago (Ferner and Mess, 2011).

Our gene interference analysis revealed that *Mycn*, which is preferentially expressed in PE cells (Figure 2.5C and Table 2.1), is activated by GATA6 and SOX17. *Myc* and *Mycn* are concordantly upregulated in PE cells (Figure 2.5C and Table 2.1) suggesting that PE cells undergo an extensive self-renewable process. These findings explain why previously isolated ExEn stem cell lines (e.g. cXEN) are predominantly resembling the lineage of rapidly growing PrE or PE cell types. Interestingly, FOXA2 is a dominant repressor of *Mycn* (Fig 2.21A and 2.21B), and modestly expressed in the PE cells (despite being a VE-associated TF) (Figure 2.5D and Table 2.1) (Dufort et al., 1998). This suggests that FOXA2 may safeguard PE cells from uncontrolled cell growth by regulating the expression of *Mycn*, reflecting its tumor suppressor role in cancer cells (Sahoo et al., 2022). In line with this notion, FOXA2-KO embryos

abnormally accumulate PrE/PE-like cells (Dufort et al., 1998), which may allude to dysregulated *Mycn* expression.

2.3b BMP signaling input is required for PE and VE segregation

We observed a profound upregulation of genes encoding BMP signaling components in either GATA6 or SOX17 deficient cXEN cells (Fig 2.22A and 2.22B). Motivated by this observation, we investigated the genome-wide transcriptional changes in cXEN cells after BMP4 treatment. BMP4 downregulated the expression of PE-associated genes encoding *Gata6, Sox17*, and *Mycn* (Fig 2.23A), while upregulating VE-associated genes such as *Hnf4a* and *Hex* (Fig 2.23A). This finding is consistent with the previous observation that BMP4 treatment directs XEN cells to a VE phenotype (Artus et al., 2012; Paca et al., 2012). We propose that BMP signaling is a potent inducer for VE cell fate in part through antagonizing the activity of GATA6 and SOX17. It also fits with the known expression of *Bmp4* in E4.5 blastocysts, which show high *Bmp4* expression levels in the proximal epiblast where the future VE cells adjacently reside (Coucouvanis and Martin, 1999; Graham et al., 2014; Mochel et al., 2015). The future PE cells are localized further away from the BMP source in epiblast and migrate along mural TE (mTE). Collectively, our data provide compelling evidence that BMP signaling acts as the developmental cue for VE versus PE cell fate decision.

We propose the following mechanism may underplay during PE and VE cell specification. Initially, PrE cells express *Gata6* and *Sox17*, where PrE cells are bipotential, and co-express other genes involved in PE and VE gene regulatory programs (Fig 2.4A and 2.4B). The PrE cells receive BMP signals from the adjacent Epi cells, work with FOXA2, to promote

the VE gene regulatory program. The exact mechanism of how BMP signaling impacts the VE cell decision is currently unclear. On the other hand, the PrE cells that detach from the EPI cells, acquire PE cell fate via upregulation of *Gata6* and *Sox17*, possibly due to being away from BMP signals, but closer to paracrine parathyroid hormone-related protein (PHTrP) signals from adjacent mTE cells. Recently developed *in vitro* PrESC cell lines (Ohinata et al., 2022) that model the PrE progenitor cells will be valuable to test the current proposed model. It will be also important to delineate the respective contributions of canonical and noncanonical BMP signaling pathways to the VE gene program. The use of a BRE-gal reporter system that specifically responds to canonical BMP signaling during preimplantation development will be useful in determining when and where canonical BMP signaling affects the lineage segregation (Javier et al., 2012; Mochel et al., 2015).

Additionally, the precise role of BMP signaling in PE-VE cell lineage specification *in vivo* will need to be determined in developing blastocysts. Effects of BMP signals in PE/VE lineage divergence have yet to be finely delineated due to the requirement of BMP signals in both embryonic and extraembryonic trophectoderm and endoderm derived tissues (Beppu et al., 2000; Chu et al., 2004; Coucouvanis and Martin, 1999; Sirard et al., 1998; Sozen et al., 2021; Yang et al., 1998). Thus, a particularly important question is to address how modulation of BMP signaling and the interplay among GATA6, SOX17, and FOXA2 shifts the balance of PE and VE cells in developing embryos. Recent advances in the synthetic embryo model, a mixture of self-assembled ESC, TSC, and XEN stem cells (Amadei et al., 2021; Amadei et al., 2022; Sozen et al., 2018; Tarazi et al., 2022; Zhang et al., 2019), could provide a potent reductionist approach to conduct functional dissection in embryo development. Besides demonstrating here that cXEN

cells are genetically editable cells, we proved the principle that the FKBP12^{F36V}-based PROTAC system is compatible with XEN cells by using dTAG^{V}-1 for FKBP12^{F36V}-mediated protein degradation (Nabet et al., 2018; Nabet et al., 2020). The application of PROTAC in XEN or associated endodermal stem cell systems (Amadei et al., 2021; Amadei et al., 2022; Ohinata et al., 2022; Tarazi et al., 2022) will considerably expand the targetable genes for functional studies. Remarkably, the acute and inducible nature of the PROTAC system is particularly tailored for interrogating temporal patterns during synthetic embryo development. By interfering with distinct genes for ESCs, TSCs, and XEN when assembling synthetic embryos, we envision that it will be a powerful tool for revealing the complexities in embryo development.

2.4 Acknowledgements

AI09599, AI102853 and AI102853.

2.5 Author Contributions

2.6 METHODS

cXEN cell line derivation

cXEN cell line were derived from mESCs cells (E14TG2a) using a previously established chemical induction protocol (Niakan et al., 2013) in a cXEN derivation medium consisting of 0.010 M Retinoic Acid (Sigma, 50-185-8562), 10 ng/mL Activin A 24 ng/mL (R&D Systems, 338AC010), Fgf2 (R&D Systems, 23-3FB0-10), and 1 ug/mL Heparin (Sigma, H3393). mESCs were maintained in KnockOut DMEM (Thermo Fisher,10829018) supplied with 15% FBS (R&D Systems, S10250H), 1mM MEM Non-Essential Amino Acids (Thermo Fisher, 11140-050), 2mM Glutamax (Thermo Fisher, 35050061), 100 U/mL Pen/Strep (Thermo Fisher, 15140122), 0.1 mM 2(β)-ME (Sigma, M-3148), and 1000 U/mL LIF (Cell Guidance Systems, GFM200), before change to cXEN derivation medium. The established cXEN cell lines were maintained, and passaged in Advanced RPMI (Thermo Fisher,12633012) supplemented with 15% FBS (R&D Systems, S10250H), 2 mM Glutamax (Thermo Fisher, 35050061), 100 U/mL Pen/Strep (Thermo Fisher, 15140122), 0.1 mM 2(β)-ME (Sigma, M-3148).

CRISPR/Cas9 cell line generation

We cloned the FKBP12^{F36V}-mNeonGreen-(3X)HA-tag-P2A-Blasticidin cassette flanked by the homology arms of target proteins into pUC19 backbone. The homology arms, amplified from mouse genomic DNA or synthesized using gBlock, were designed as inframe insertion at the C-terminus of the targeted proteins. SgRNA were designed to cut around the stop codons of the endogenous loci of the targeted genes. The repair template and sgRNAs were validated by sanger sequencing before use for further experiments. For generation of FKBP12^{F36V} cell lines, cXEN cells were seeded the day before transfection on a 10 cm^2 tissue culture plate. 12 ug of plasmid

DNA by mixing equimolar amounts of sgRNA and homology repair template were transfected

using 40 ul of Lipofectamine 3000 with 24 ul of P3000 reagent (Thermo Fisher, L3000008).

Two days after transfection, the transfected cells were selected in 10 ug/mL of Blasticidin for a

week (Cayman Chemical, NC1445974). Single cells from the mNeonGreen positive population

were then sorted into 96-well round bottom plates pre-coated with irradiated CF-1 MEFs

(Thermo Fisher, A34180) using a BDFACSAria Fusion Sorter. The homozygous clones were

screened by PCR genotyping using primers that can detect successful insertions. The clones were

verified by flow cytometry and western blots. The genotyping primers used for detecting

successful knock-in in the *Gata6* loci: FWD: "GGTCAAGACGGCCTCTACATAGGTG",

REV: "CTGGTGAAATAGGTTTCCATATGATATCCC". The genotyping primers used for

detecting successful knock-in in the *Sox17* loci: FWD:

"AGCTCAGCGGTCTACTATTGCAACTACC", REV:

"AGCACTCAGCACAGCATTTGCAGGG". The sgRNA used for targeting *Gata6* loci are as

follows: sense: "CACCGCCAAGAATCCTGTCGCACGG", antisense:

"AAACCCGTGCGACAGGATTCTTGGC". The sgRNA used for targeting *Sox17* loci are as

follows: sense: "CACCGCGGTTGCCGACCCGACCTGA", antisense:

"AAACTCAGGTCGGGTCGGCAACCGC".

To generate FOXA2 knockout lines on the GATA6 or SOX17 FKBP12^{F36V} knockin

lines, two sgRNAs were designed to delete the sequence encoding for the DNA binding domain.

cXEN knockin cell lines were seeded the day before transfection on a 6-well tissue culture plate.

2.5 ug of total plasmid DNA, using equimolar amounts of sgRNA, were transfected with 5 ul of

lipofectamine 2000 (Thermo Fisher, 11668019) for 24h. The BFP positive cells were then sorted

81

using a BDFACSAria into 96-well round bottom plates pre-coated with irradiated CF-1 MEFs

(Thermo Fisher, A34180). The homozygous clones were screened by genotyping. The

genotyping primers used for detecting *FoxA2* forkhead deletion: FWD:

"AACATGTCATCCTATGTGGGCGC" and REV: "GGTCTTCTTGCCTCCGCTACTG". The

sgRNAs used for deletion of *FoxA2*'s forkhead domain: set_1: sense:

"CACCGCACACTTGAAGCGCTTCTGG", antisense:

"AAACCCAGAAGCGCTTCAAGTGTGC", set_2: sense:

"CACCGTGGCGTGTGTGTAGCTGCGT", antisense:

"AAACACGCAGCTACACACACGCCAC"

BMP4 Treatment

cXEN cells were plated on gelatin N2B27 + 1% FBS overnight. N2B27 media is a 1:1 mix of

DMEM/F12 (Thermo Fisher, 11320033), and Neurobasal media (Thermo Fisher, 21103049)

supplemented with B-27 Plus (Thermo Fisher, A3582801) and N-2 (Thermo Fisher, 17502048).

The media was then changed to N2B27 without serum supplemented with 25 ng/ml BMP4

(R&D Systems, 314-BP-010/CF). Day 1 (24h) and Day 3 (72h) cultures were harvested for

RNA-seq.

RNA-seq

Total RNA was extracted using Trizol (Thermo Fisher, 15596026). Strand-specific RNA-seq

libraries were generated with 1000 ng of total RNA using NEBNext Ultra II Directional RNA

Library Prep Kit for Illumina (E7760S). Libraries were paired-end sequenced on the NovaSeq

6000 with 100 cycles.

ATAC-seq

50,000 cXEN cells were dissociated using Accutase (Thermo Fisher, A1110501). ATAC-seq was performed using the ATAC-seq kit (Active Motif, 53150) which included components for tagmentation, tagmented DNA clean-up, and library PCR amplification per manufacturer's instructions. Libraries were paired-end sequenced on the NovaSeq 6000 with 100 cycles.

CUT&RUN

500,000 cXEN and its derivative cell lines were dissociated using Accutase (Thermo Fisher, A1110501). Then, CUT&RUN was performed using pAG-MNase (Epicypher, 15-1016) using a cell preparation from (Fujiwara et al., 2021), and then following the standard protocol published by the Henikoff lab (Meers et al., 2019; Skene and Henikoff, 2017). Half of the DNA eluted from the CUT&RUN reaction was used for library prep using the NEBNext Ultra II system based on Nan Liu's CUT&RUN library prep protocol but without size selection (Liu et al., 2018). Libraries were paired-end sequenced on the NovaSeq 6000 with 100 cycles. Antibodies used in these experiments were as follows: anti-GATA6 (5851S, Cell Signaling Technology), anti-SOX17 (AF1924, R&D Systems), anti-FOXA2 (07-633, EMD Millipore Sigma), anti-H3K27ac (MABE647, EMD Millipore Sigma), anti-H3K4me3 (39159, Active Motif), and anti-H3K4me1 (ab8895, Abcam).

Flow Cytometry

Cells were dissociated and resuspended in media for analyzing $FKBP12^{F36V}$ NeonG fluorescence protein knock-in lines using the ACEA Novocyte Quanteon (Agilent). For surface staining, cells were dissociated, washed in 1X PBS, stained by anti-CD140A(PDGFRA) (17-1401-81, Thermo Fisher) for 15 min. Then, stained cells were washed 2X in PBS, and resuspended in FACS buffer (PBS+1% FBS) before FACS analysis.

Western Blotting

Antibodies used for western blot analysis were as follows: anti-GATA6 (5851S, Cell Signaling Technology), anti-SOX17 (AF1924, R&D Systems), anti-FOXA2 (07-633, EMD Millipore Sigma), anti-HA (901503, Biolegend), or anti-TUBULIN (T6199, Sigma). The cells were harvested in a 1X SDS loading buffer, boiled at 95°C before being subjected to western blotting.

Data Quantification and Statistical Analyses

Cut&Run and ATAC-seq Data Analysis

Paired-end fragments were trimmed using trimmomatic (Bolger et al., 2014). Paired-end reads were aligned using Bowtie2 (Langmead and Salzberg, 2012) with flags: --local --very-sensitive --dovetail --phred33 -I 10 -X 700. The fragments were filtered for a MAPQ score of 30 prior to calling peaks. MACS2 (Zhang et al., 2008) was used for narrowPeak calling for Cut&Run transcription factors and ATAC-seq. HOMER (Heinz et al., 2010) was used for calling H3K27ac peaks centered on nucleosome free regions (NFR).

Motif enrichment analysis was performed using HOMER (Heinz et al., 2010) on cXEN and E6.5 VE H3K27ac peaks. H3K27ac-enriched enhancers were defined by filtering out H3K27ac peaks with high H3K4me3 signals through k-means clustering by Deeptools (Ramírez et al., 2014). Tornado plots and meta profiles of Cut&Run and ATAC-seq datasets were generated by Deeptools' plotHeatmap and plotProfile subcommands (Ramírez et al., 2014). Enhancer sets for nearby genes were annotated by HOMER (Heinz et al., 2010). ATAC-seq footprinting analyses were performed using the TOBIAS pipeline with motifs supplied from the JASPAR CORE 2022 vertebrates (Castro-Mondragon et al., 2021) and Hocomoco V11 (Mouse)

(Kulakovskiy et al., 2017). To plot the Vplots, VplotR (Serizay and Ahringer, 2021; Serizay et al., 2020) was used to analyze fragment distributions subsequently visualized by complexHeatmap (Gu et al., 2016).

RNA-seq Data Analysis

All RNA-seq samples were aligned using HISAT2 (Kim et al., 2019) to the mm10 genome. Raw counts were obtained using HT-seq (Putri et al., 2022). TPM values were obtained using Stringtie (Pertea et al., 2015). To compare gene expressions between E6.5 VE RNA-seq (Zhang et al., 2018) and cXEN cells (Figure 3E), DESeq2's median of ratios (Love et al., 2014) were used for transcript level comparison.

Weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008) was performed on Gata6-FKBP12^{F36V} and Sox17-FKBP12^{F36V} time-course RNA-seq (0, 6, 24, 48, 72, 96h) series using the blockwiseModules function with a "signed" network, TOM type, a soft power threshold of 14, minModuleSize of 150, pamStage set to True, and bicor set to True. Differentially expressed genes for FoxA2 Mutant cell lines and BMP4 treated samples were analyzed by DESeq2 (Love et al., 2014). RNA-seq heatmaps were visualized using the complexHeatmap package or using the Seaborn package.

Gene ontology overrepresentation analyses and GSEA plots were performed using ClusterProfiler (Wu et al., 2021; Yu et al., 2012). GSEA analyses used a rank ordered list from DESeq2 test statistic ("stat" column) output (Love et al., 2014). The GS&F-cobound VE enhancer targets were generated by extracting the differentially expressed genes by DESeq2 by comparing the E6.5 VE versus E6.5 Epiblast RNA-seq (Zhang et al., 2018), which were then overlapped with the annotated nearby genes of the GS&F-cobound VE enhancers (Table S7).

The scRNA-seq VE module set was reduced for GSEA ranking by filtering for kME >0.175, and using the top 15% of genes. Statistical analyses were done using the SciPy statistical package using the Wilcox test (Virtanen et al., 2020).

scRNA-seq Data Analysis

scRNA-seq data of E4.5 embryos were obtained and integrated from (Cheng et al., 2019; Mohammed et al., 2017; Nowotschin et al., 2019; Qiu et al., 2022) using seurat V4 (Hao et al., 2021). Cell clustering, UMAP, and violinplots for transcriptional signatures were generated by seurat V4 (Hao et al., 2021). The transcriptional signatures were calculated by the UCell algorithm (Andreatta and Carmona, 2021; Hao et al., 2021). Weighted gene correlation network analysis of single cell transcriptomes were performed by hdWGCNA (Morabito et al., 2021). Statistical analyses were done using the Wilcox test through R. To generate cell trajectory (Figure 1B), we first integrated E4.5 scRNA-seq data (described above) and E5.5 scRNA-seq data from (Nowotschin et al., 2019) using seurat V4. Then, we converted the integrated seurat object to anndata in order to perform analyses of paga, force-directed graph, and pseudotime using scanpy toolkit (Wolf et al., 2018).

Data and code availability

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE213661. The following publicly available datasets were used in this study: E6.5 visceral endoderm RNA-seq, E6.5 epiblast RNA-seq, E6.5 visceral endoderm ATAC-seq, E6.5 visceral endoderm H3K27ac, E6.5 visceral endoderm H3K4me3 can be found at GSE76505 and GSE125318.

For their contributions in Chapter 2, thanks to: Hanbin Lu, Han Han, Jeff Jiajing Zhou, Aarushi Madan, Wenqi Wang, Ken Cho and Cornelis Murre.

CHAPTER 3:

Transcription factors redundancy in extraembryonic endoderm lineages

3.1 Introduction

Initiation of PrE progenitor specification requires GATA6, a key transcription factor (TF), to establish this lineage during the blastocyst stages of preimplantation development by inducing other extraembryonic endoderm (ExEn) TFs (Koutsourakis et al., 1999; Wamaitha et al., 2015). Loss of GATA6 leads to a lack of an identifiable PrE structure, and subsequently its sub-lineages PE and VE fail to differentiate (Koutsourakis et al., 1999; Wamaitha et al., 2015). Interestingly, GATA4 is also co-expressed following GATA6 activation (Cai et al., 2008). Characterization of GATA4 mice knockouts have shown that although GATA4-KO embryos can specify VE, likely due to GATA6 upregulation to compensate(Kuo et al., 1997; Molkentin et al., 1997). However, GATA4-KO embryos have a problem in VE yolk sac closure(Kuo et al., 1997; Molkentin et al., 1997). Further, protein localization studies have also shown that GATA4 and GATA6 are both expressed in the PrE to PE, but as the VE develops, there is a downregulation and then absence of GATA6 expression (Cai et al., 2008). These differential patterning between the two GATA proteins found in the extraembryonic endoderm (ExEn) lineages suggest there is a dosage requirement of GATA(s) for proper differentiation in PE versus VE as discussed in Chapter 2. Taken together, the co-expression GATA4 and GATA6 in the PE lineage, and the rescue of GATA4-KO embryos by ectopic GATA6 upregulation raises the question of the functional redundancy of transcription factors within the same TF family in the ExEn lineages. We further explore this by comparing GATA4 and GATA6 activities.

3.2 Results

3.2a *Gata* sites recruit GATA4 and GATA6

Gata4 and Gata6 have been shown to cooperatively interact with each other through protein-protein interaction (Charron et al., 1999). To investigate GATA4 and GATA6 association on cXEN cis-regulatory elements (CREs), we performed CUT&RUN profiling using GATA4 and GATA6 specific antibodies. PCA comparison of global binding affinity of GATA4 and GATA6, alongside CUT&RUN profiling of FOXA2 and SOX17, shows clustered GATA4 and GATA6 binding relative to FOXA2 and SOX17 (Figure 3.1A and 3.1B). After peak calling from GATA4 and GATA6 binding profiles, we identified 47550 GATA4 peaks and 41258 GATA6 peaks (Figure 3.1C). We find a large overlap of peaks between GATA4 and GATA6, 30390 peaks respectively, with strong binding colocalization illustrated through the tornado plot (Figure 3.1C and 3.1D). However, GATA4 or GATA6 non-overlapped peaks still retain binding of GATA4 or GATA6, but binding affinity is weaker than the former indicated by weak, normalized signal detection (Figure 3.1D). Correspondingly, motif scores of GATA4/GATA6 cobound sites rank higher than GATA4 or GATA6 non-overlapped peaks, suggesting that motif strength may explain why GATA4/GATA6 overlapped sites are bound more stably (Figure 3.1D).

90
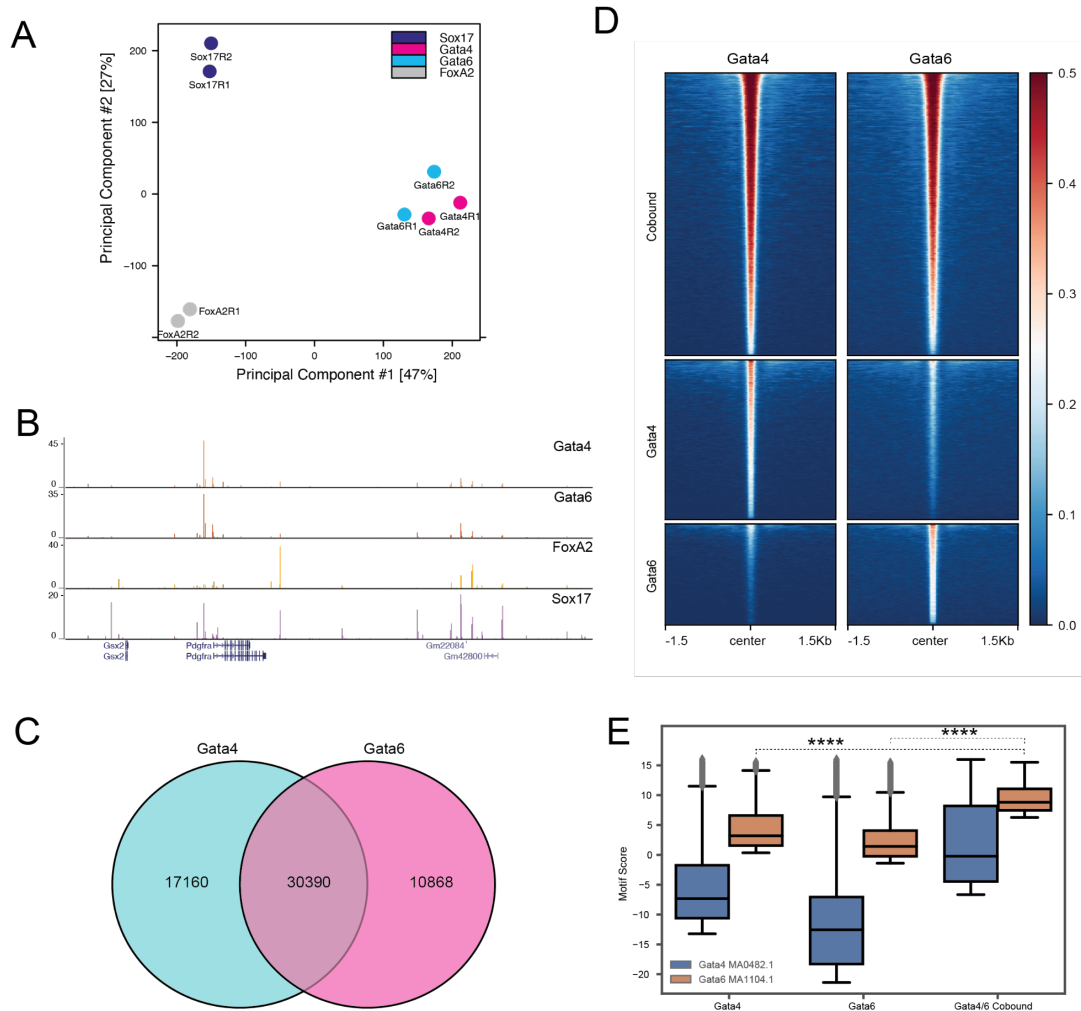
Figure 3.1. GATA4 and GATA6 are cobound. (A) PCA plot of affinity data generated from CUT&RUN assays profiling GATA4, GATA6, FOXA2, and SOX17 occupancy (B) UCSC genome browser visualization of GATA4, GATA6, FOXA2, and SOX17 CUT&RUN at known extraembryonic endoderm gene (PDGFRA). Y-axis represents normalized scaled track height to CPM. (C) Venn diagram overlap of GATA4 and GATA6 replicate merged peaks (D) Tornado plot displaying normalized signals of GATA4 and GATA6.

3.2b Global gain of GATA4 occupancy enrichment upon GATA6 acute depletion

We determined that GATA4 and GATA6 are closely associated with each other at their CREs, which prompted us to examine their binding dependency by utilizing the dTAG-inducible FKBP12^{F36V} degron to the endogenous Gata6 *locus* as described earlier (Figure 2.S4). We confirmed that GATA6 is efficiently degraded by probing GATA6 binding upon application of dTAG^{V1} at various time points (0h, 2h, and 24h) (Figure 3.2A). As expected GATA6 is efficiently degraded by 2h and 24h off the chromatin (Figure 3.2A). Next, to determine whether loss of GATA6 affects binding of GATA4 and other transcription factors (e.g. SOX17 and FOXA2), we profiled their binding through time course analyses in the absence of GATA6 protein (Figure 3.2B). At GATA6 centered peaks, we find a gain in GATA4 protein enrichment suggesting that GATA4 can independently bind at GATA6 sites and possibly compensate for GATA6 binding (Figure 3.2B). In support of this, an increase in GATA4 called peaks by 24h (2016 peaks, respectively) upon GATA6 protein depletion were detected (Figure 3.2C). At these gained GATA4 peaks, we observed a gradual enrichment of GATA4 protein at these gained Gata4 peaks from 2h to 24h that were previously bound by GATA6 (Figure 3.2C). Strikingly, we also observed a gain in FOXA2 protein enrichment which references our previous observations of FOXA2 dependent gene activation upon GATA6 protein depletion (Figure 3.2B, Figure 2.S4). In contrast, we observed minimal changes in SOX17 binding upon GATA6 protein depletion (Figure 3.2B).
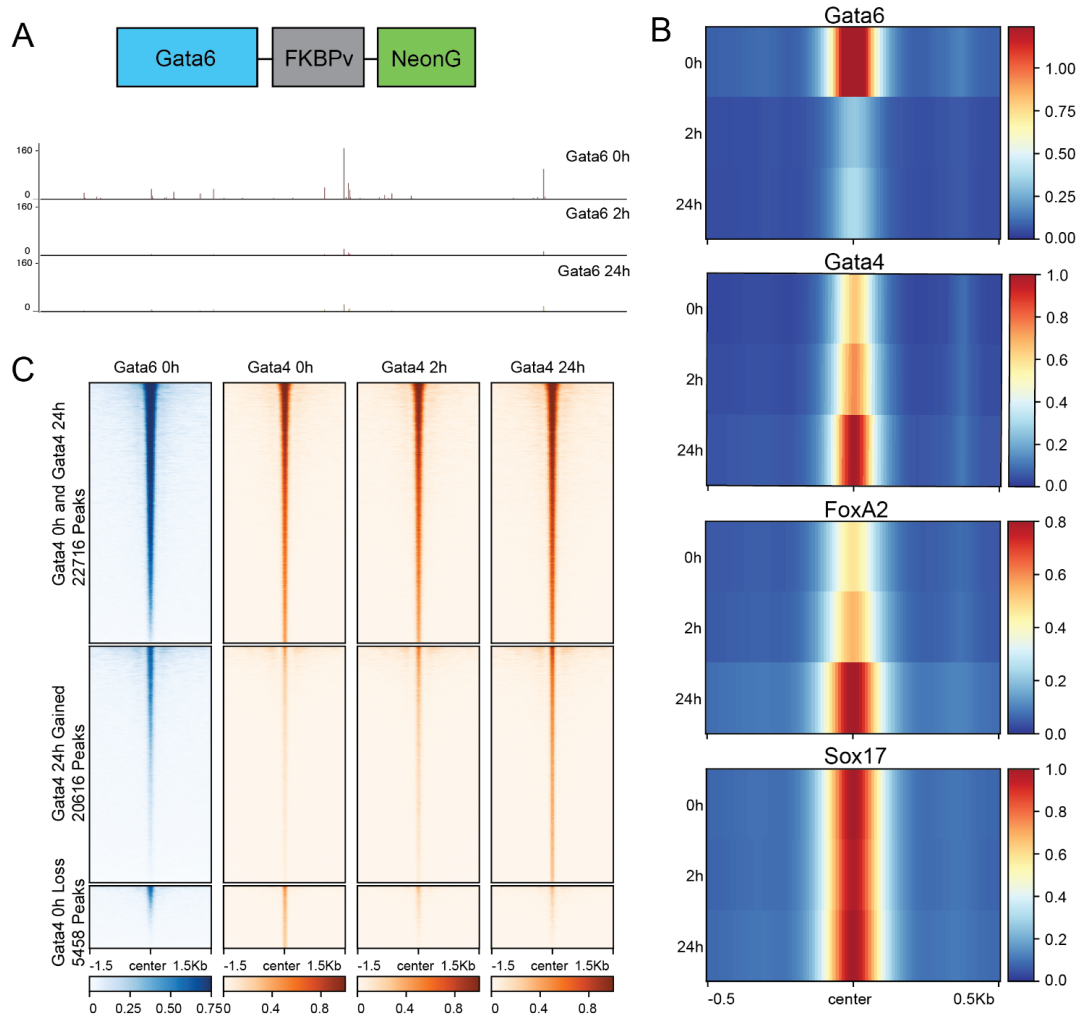
Figure 3.2. Acute depletion of GATA6 leads to an increase in GATA4 and FOXA2 occupancy. (A) Schematic illustrating endogenous GATA6-FKBP12^{F36V} and NeonG knock-in fusion in cXEN cells. UCSC genome browser visualization of GATA6-FKBP12^{F36V} time course protein depletion using dTAG^{V1} of GATA6 occupancy profiling by CUT&RUN for 0h, 2h, and 24h. (B) Averaged normalized signal enrichment (CPM) at GATA6 0h called peaks of Gata6-FKBP12^{F36V} time course protein depletion using dTAG^{V}-1 of GATA6, GATA4, FOXA2, or SOX17 occupancy profiling by CUT&RUN for 0h, 2h, and 24h. (C) Tornado plot displaying normalized signals of GATA6 CUT&RUN without dTAG^{V1} in the GATA6-FKBP12^{F36V} cell

93

line and GATA4 CUT&RUN of GATA6-FKBP12^{F36V} time course protein depletion using

dTAG^V-1 or 0h, 2h, and 24h.

3.3c GATA4 is functionally redundant with GATA6 in XEN cells

To examine GATA4 transcriptomic changes in cXEN cells, we generated a dTAG-inducible

FKBP12^{F36V} degron to the C-terminus of endogenous *Gata4* locus (GATA4-FKBP12^{F36V} line)

by using CRISPR (Figure 3.3A), allowing us to acutely degrade GATA4 proteins upon treating

dTAG^{V1}. These cells were sensitive to degradation using dTAG^{V1} as early as 30 min with

complete degradation by 2h as measured by FACS measured by the fluorescent protein (NeonG)

fused to GATA4 protein (Figure 3.3C). Western blot on HA-tag and native protein further

validated that these TFs were depleted by 24h (Figure 3.3B).
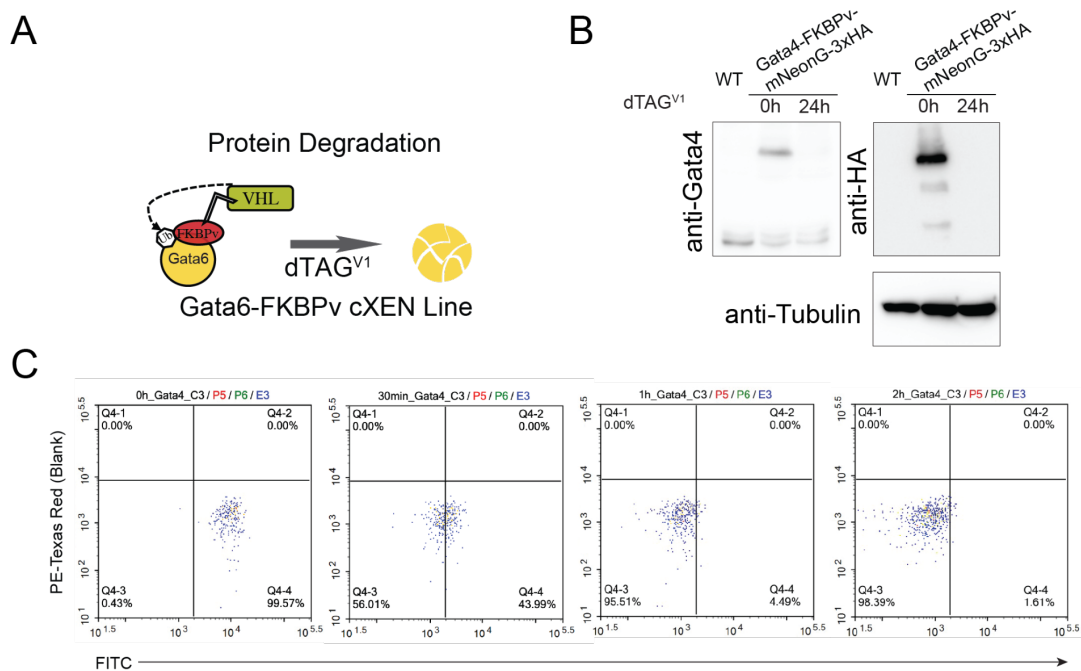
Figure 3.3. GATA4-FKBP12^{F36V} Line Generation. (A) Schematic illustrating the strategy for

acute degradation of GATA4 (B and C) Western blots (B) and flow cytometric analysis (C) on

acute degradation of GATA4-FKBP12^{F36V} cell line.

Since GATA4 is strongly correlated with GATA6 binding activities and can compensate for loss of GATA6 binding, we further investigated GATA4 transcriptomic effects upon GATA4 protein depletion through a time course analysis. We observe GATA4 protein depletion recapitulates GATA6 protein depletion by correlating with our previously identified gene modules (Figure 3.4A, Figure 2.4B, 2.S6C). Additionally, in GATA4 depleted cells, we do not observe a prominent increase in the SOX17-only upregulated gene module (Figure 3.4A, 3.4B). This suggests that transcriptomic effects of GATA4 protein depletion were convergent on redundant TFs that belong to the same TF family given that they bind to similar motif configuration on functional CREs.
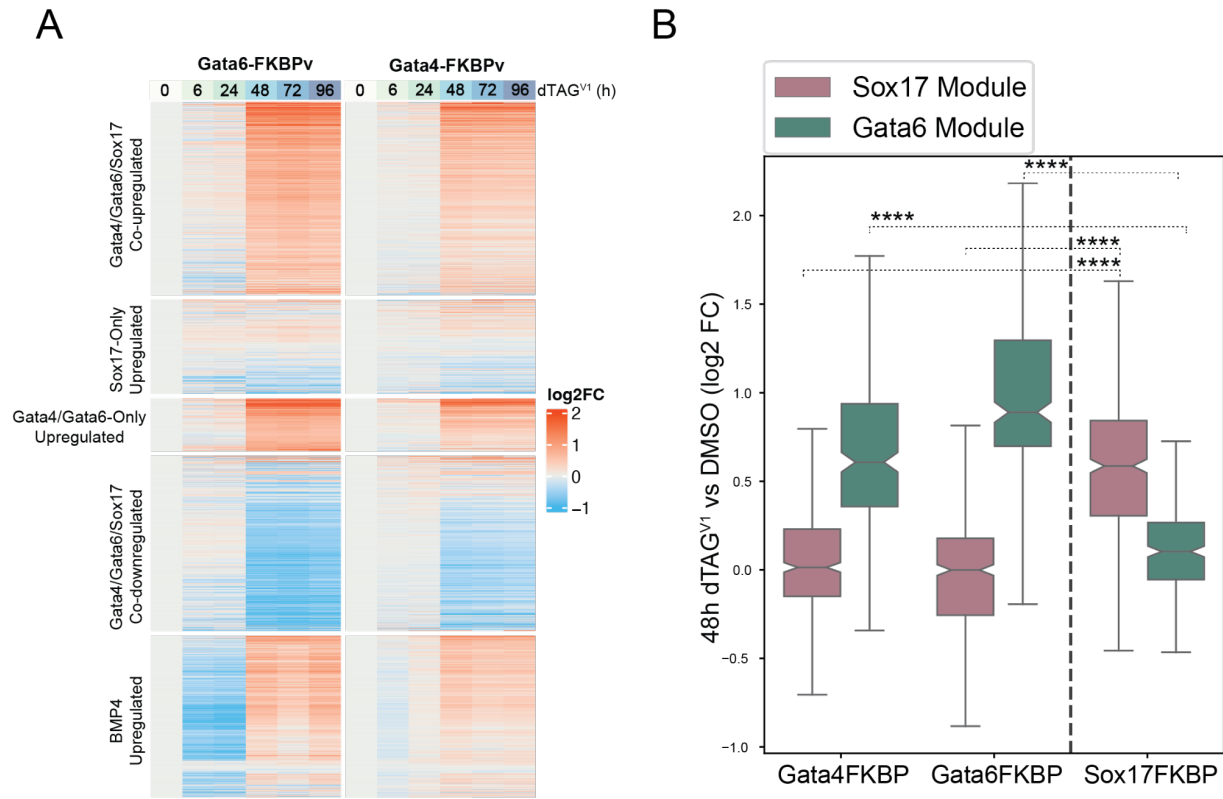
Figure 3.4. GATA4 depletion phenocopies GATA6 depletion. (A) Heatmap displaying fold changes (log2 normalized TPM versus 0h) of genes from gene modules identified in Figure 2.4B and BMP4 upregulated module identified in Figure 2.S6C during the course of GATA6 and GATA4 depletion. (B) Boxplot displaying log2 fold changes of GATA4, GATA6, or SOX17 depletion (48h dTAG^{V1} versus DMSO) GATA6 or SOX17 upregulated gene modules identified in Figure 2.4B. **** denotes $p < 0.0001$ (Wilcoxon Test)

3.3 Method Details

CRISPR/Cas9 cell line generation

We cloned the FKBP12^{F36V}-mNeonGreen-(3X)HA-tag-P2A-Blasticidin cassette and the flanked

homology arms of the target proteins into pUC19 backbone. SgRNA were designed to cut

around the stop codons of the endogenous loci of the targeted genes. The homology arms were

designed to inframe insertion at the C-terminus of the targeted proteins. The homology arm

sequences and sgRNAs were validated by sanger sequencing before use for further experiments.

For generation of FKBP12^{F36V} cell lines, cXEN cells were seeded the day before transfection on

a 10 cm^2 tissue culture plate. 12 ug of plasmid DNA, using equimolar amounts of sgRNA and

homology repair template, were transfected using 40 ul of Lipofectamine 3000 with 24 ul of

P3000 reagent (Thermo Fisher, L3000008). Two days after transfection, the transfected cells

were selected in 10 ug/mL of Blasticidin for a week (Cayman Chemical, NC1445974). Single

cells of mNeonGreen positive population were then sorted using a BDFACSAria Fusion Sorter

in 96-well round bottom plates with irradiated CF-1 MEFs (Thermo Fisher, A34180). The

homozygous clones were genotyped by PCR with primers amplifying the knock-in insertion site

for FKBP12^{F36V} and verified using flow cytometry.

RNA-seq

Total RNA was extracted with Trizol (Thermo Fisher, 15596026). Strand-specific RNA-seq

libraries were generated with 1000 ng of total RNA by using NEBNext Ultra II Directional RNA

Library Prep Kit for Illumina (E7760S). Libraries were paired-end sequenced on the NovaSeq

6000, 100 cycles.

CUT&RUN

500,000 cXEN and its derivative cell lines were dissociated using Accutase (Thermo Fisher, A1110501). Then, CUT&RUN was performed by using pAG-MNase (Epicypher, 15-1016) and following the standard protocol published by the Henikoff lab [citation]. Half of the DNA eluted from the CUT&RUN reaction was used for library prep by using the NEBNext Ultra II system based on Nan Liu's CUT&RUN library prep protocol but without size selection [citation]. Libraries were paired-end sequenced on the NovaSeq 6000, 100 cycles. Antibodies used in these experiments were as follows: anti-Gata6 (5851S, Cell Signaling Technology), anti-Sox17 (AF1924, R&D Systems), anti-FoxA2 (07-633, EMD Millipore Sigma).

Flow Cytometry

Cells were dissociated and resuspended in media for analyzing FKBP12^{F36V} NeonG fluorescence protein knock-in lines where the flow cytometric data was collected using the ACEA Novocyte Quanteon (Agilent).

Western Blotting

Antibodies used for western blot analysis were as follows: anti-Gata4 (sc-25310, Santa), anti-Gata6 (5851S, Cell Signaling Technology), anti-Sox17 (AF1924, R&D Systems), anti-FoxA2 (07-633, EMD Millipore Sigma), anti-HA (#3724, Cell Signaling Technology). The cells were harvested in a 1X SDS loading buffer, boiled at 95°C before being subjected to western blotting.

3.4 Discussion

PE Transcription Factor Redundancy

In chapter 2, we find that perturbing PE-associated TFs promote the VE-cell gene program. However, we did not observe significant effects on biological pathways related to PE identity, with the exception of pathways related to *Myc/Mycn* (Figure 2.1D) upon protein depletion of PE

TFs (GATA6 or SOX17) in chapter 2. This suggests that some of the core PE pathways may perhaps be maintained by redundant TFs within the same family such as GATA4/GATA6 or SOX17/SOX17. Indeed, such compensation of TFs within the same family were proposed through examining single KO phenotypes of GATA4, GATA6, SOX17, or SOX7 (Artus et al., 2011; Bowles et al., 2000; Cai et al., 2008; Kanai-Azuma et al., 2002; Koutsourakis et al., 1999; Kuo et al., 1997; Molkentin et al., 1997; Schrode et al., 2014; Wat et al., 2012). However, such proposals have not been rigorously tested since double heterozygous mice knockouts were unable to be produced. Here, we used the GATA family of TFs as a proof of principle to demonstrate the redundancy of these TFs in cXEN cells. We first show that their functional redundancy stems from their near synonymous occupancies' genome wide. Next, we further correlated single protein depletion changes of GATA4 versus GATA6 transcriptomes to demonstrate their specificity in regulating a "GATA"-specific gene expression. We speculate that SOX7 degron lines may reflect similar phenotypes to SOX17 since SOX7 and SOX17 are highly correlated with each other in the *in vivo* localisation of ExEn cell lineages. To further study the requirement of endodermal linked SOX or GATA factors in the ExEn lineages, then double degron knock-in lines can be generated to examine the requirement of these factors, particularly for the PE GRNs.

CHAPTER 4:

Concluding Discussion

Outlook for Expanding Signaling and Gene Regulatory Models for Extraembryonic

Endoderm Development

Molecular regulation of the primitive endoderm and its divergence to PE and VE relies

on extrinsic signaling to transmit downstream molecular cues which are governed by complex

TF networks. Various TFs have been implicated in the development of the PE and VE by

examining mutant phenotypes in loss of function TFs during early embryogenesis. However,

given the rapid and dynamic period of lineage establishment of these tissues during implantation,

understanding how these endodermal TFs exert their transcriptional control and their activities at

cis-regulatory modules (CRMs) remains challenging. In this dissertation, I performed a

characterization on transcription factors that modulate PE versus VE cell fate through scRNA-

seq analyses combined with single and combinatorial gene editing approaches in cXEN cells to

generate a first molecular model for PE vs VE cell fate choice.

In chapter 2, I identified sets of TFs that were associated with PE versus VE

transcriptomes. I established a PE-like cell line, called cXEN cells, and compared lineage

specific enhancers found in cXEN cells and in VE cells of E6.5 embryos to narrow down the

core PE or VE TFs that regulate PE-VE cell fate choice. Of interest, I find motifs of PE TFs

(GATA and SOX) and VE TFs (FOXA2, HNF1B, HNF4A) on VE enhancers in PE cells and VE

cells. While in PE cells, I find primarily the enrichment of PE TFs (e.g. GATA, SOX, JUN,

TEAD) rather than VE TFs. I further interrogated the functions of *Gata6*, *Sox17*, and *FoxA2* in

our studies.

In this study, I showed that the genome of cXEN cells can be edited successfully to

obtain homozygous knock-in clones by a CRISPR/Cas9-mediated gene editing method. Such

efforts would further expand genetic perturbation studies in cXEN cells and other recently

developed ExEn stem cell lines, such as PrESCs (Ohinata et al., 2022). Single protein depletion of PE TFs demonstrated their regulation of *Mycn*-mediated self-renewal pathways. However, other biological pathways related to PE function were not observed to be affected as discussed in chapter 3. Gene editing approaches to individually or simultaneously target co-expressed genes such as *Gata4* and *Gata6* or *Sox17 and Sox7* will resolve the functional importance of redundant TFs during PE development. Additionally, the advent of synthetic embryo protocols combining mESCs, mTSCs, and XEN cells allows for tissue-specific genetic perturbation platforms. For example, *Eomes* is an important TF for both trophoblast cells but also emVE cells. By targeting *Eomes* specifically in the mTSCs, then one can examine cell autonomous roles of *Eomes* in trophoblast differentiation while keeping intact cell-to-cell interaction with both embryonic and ExEn cell types. Further improvements of novel genetic perturbation platforms may also be available *in vivo* with a recent study demonstrating the use of the dTAG system targeting *Nelfb* locus using a mouse model (Abuhashem et al., 2022). However, generating a tissue-specific knock-in of PROTAC gene systems in an animal model organism may be a limiting factor for some tissues due to a lack of robust lineage promoters that have yet to be tested. Nevertheless, we envision that a combined use of synthetic embryos and the PROTAC system will elucidate the function of many genes involved in the extraembryonic structure development and resolve the molecular mechanism underlying the ExEn GRN program.

Additionally, we identified biological pathways of PE or VE along with key gene modules which may serve as a resource for interrogating PE or VE functional pathways (Figure 2.3A and 2.3B. To examine how these biological pathways may be regulated, we focused on TF regulators that have not been well-understood in the ExEn development. Some of the TFs identified were signal-induced TFs that were associated with either the PE or VE cell state. Our

scRNA-seq analysis is likely to serve as a key resource for future studies interrogating TFs regulating PE or VE cell states. For example, our single cell analyses revealed that *Gata6* and *Sox17* ranks as top core PE TFs, which are also TFs essential for the initiation and lineage segregation of the progenitor PrE. I developed a method to systematically destroy GATA6 or SOX17 factors, and examined their functions as discussed earlier. Similar gene editing strategies can be employed for other candidate TFs identified in our PE or VE TF modules.

I note that signal-induced TFs such as *Jun* and *Tead1/4* TFs are also implicated in PE development (Figure 2.5C and Table 2.1). In this regard, PE cells appear to rely on mechanical cues mediated by Rho/Rock signaling, which is required for migration along the mural TE/TGCs (LaMonica et al., 2009; Mills et al., 2009). Since *Tead1/4* is highly correlated with the PE cell state, Hippo signaling may be a potential signaling pathway regulating mechanical cues required for PE identity. Upregulation of *Jun* indicates the involvement of the JNK pathway, which promotes cell proliferation and growth, consistent with the PE self-renewal property. Other PE cell enriched TFs were *Creb3l2* and *Atf6* (Figure 2.5C and Table 2.1), which encode proteins involved in endoplasmic reticulum biogenesis pathways (Li et al., 2000; Panagopoulos et al., 2006). I also independently found the same pathway enrichment in our single cell module (Figure 2.3A). These TFs may function to regulate processing of proteins for EMT and ECM secretion in endoplasmic reticulum organelles, which are abundant in the PE cells (Adamson and Ayers, 1979; Clark et al., 1975; COOPER et al., 1981; Hogan et al., 1980; Hogan et al., 1982; Salamat et al., 1995; Smith and Strickland, 1981; Wallingford et al., 2013). Next, we note the enrichment of genes encoding factors such as *Sall4*, *Klf5*, *Tbx3*, and *Lin28* (Figure 2.2C). These proteins were previously shown to function in PE cell self-renewal through examining

XEN cells and have dual functions during epiblast and PrE lineage segregation from E3.5 to E4.5 (Azami et al., 2017; Lim et al., 2008; Lu et al., 2011).

In the VE TF module, we note prominent, well-known VE TFs such as *Hnf4a* and *Hnf1b* (Figure 2.5D and Table 2.1) as discussed in chapter 1. Other TFs that likely contribute to VE nutrient delivery functions are signal-induced TFs such as *Stat3* as previously described with *Stat6* being a related family TF (Takeda et al., 1997) (Figure 2.5D and Table 2.1). I also find well-characterized VE TFs related to the emVE subtype, such as *Hhex*, *Eomes*, and *Lhx1*, which act within emVE to form the the DVE or AVE for anterior-posterior patterning as discussed in chapter 1 (Figure 2.5D and Table 2.1). In our study, I targeted FOXA2, a well-known VE regulator as referenced in chapter 1. Additional VE TFs were also identified, but lacked the functional characterization in the VE cells (Figure 2.5D and Table 2.1). However, some of these TFs may be involved in the VE nutritional support functions. For instance, iron supply pathways may potentially be regulated by *Bach1/Maf/Mafk* TFs (Ogawa et al., 2001; Sun et al., 2002), and TFs related to lysosome biogenesis and autophagy may be in part to Peg3 and Ddit3 TF candidates which has been shown in other cellular systems (Neill et al., 2017; Yang et al., 2021) (Figure 2.5D and Table 2.1). Thus, these TF candidates have yet to be characterized in VE development. Taken together, we identified novel and known regulators for PE or VE development. Overall, unique TF candidates may contribute to the diversification of key physiological and developmental pathways, along with signal-induced TFs to allow for the bifurcation of the PE and VE. Notably, key lineage progenitor PrE TFs associated with self-renewal govern the PE TF network which suggests these TFs may serve as a gatekeeper between PE versus VE differentiation.

Lastly, we found a distinct enhancer repertoire for both PE and VE enhancers suggesting these cis-regulatory elements have evolved during the emergence of extraembryonic endoderm structures. PE enhancers are highly enriched in GATA, SOX, KLF, TEAD and JUN (Figure 2.8A). In contrast, VE enhancers are enriched in GATA, FOX, SOX, HNF1B, and HNF4A (Figure 2.8B). All of these TFs are well-known to be involved in endodermal identity but the combinatorial action of these TFs yield widely distinctive endodermal identities which is evident in the formation of both PE and VE tissues as demonstrated in chapter 2. To further support this notion, these endodermal TFs are also important in the specification of the definitive endoderm. However, the top TFs regulating the definitive endoderm enhancers appear to have a unique combination of endodermal TFs relative to what I had found in the PE and VE enhancers. For instance, definitive endoderm enhancers appear to be highly associated with EOMES in combination with GATA6, SOX17, and FOXA2 (Heslop et al., 2021). In contrast, EOMES was not highly enriched at PE nor VE enhancers suggesting it is not strongly associated with activation of PE- nor VE- specific biological pathways (Figure 2.3). Though, EOMES plays a role in activation of other TFs in the anterior visceral endoderm as discussed in chapter 1. Similar inference of endodermal TF enhancer repertoire can be said of that in differentiated pancreatic lineages where enhancers were enriched in FOX, GABPA, TEAD, SPDEF, RXRA, GATA motifs (Geusz et al., 2021). Likewise, specification of endodermal tissues in amphibian counterparts appear to be highly dependent on combinatorial functions of VEGT, OTX2, and FOXH1, a TF combination different from its mammalian counterparts (Paraiso et al., 2019). To conclude, further interrogation and comparison of combinatorial action of TFs and its cis-regulatory sequences will be needed to understand the evolutionary role of how these diverse endodermal lineages arised within and across different species.

REFERENCES

Abuhashem, A., Lee, A. S., Joyner, A. L. and Hadjantonakis, A.-K. (2022). Rapid and efficient degradation of endogenous proteins in vivo identifies stage-specific roles of RNA Pol II pausing in mammalian development. *Dev Cell* **57**, 1068-1080.e6.

Adamson, E. D. and Ayers, S. E. (1979). The localization and synthesis of some collagen types in developing mouse embryos. *Cell* **16**, 953–965.

Adler, R. R., Brenner, C. A. and Werb, Z. (1990). Expression of extracellular matrix-degrading metalloproteinases and metalloproteinase inhibitors is developmentally regulated during endoderm differentiation of embryonal carcinoma cells. *Development* **110**, 211–220.

Alarcon, V. B. (2010). Cell Polarity Regulator PARD6B Is Essential for Trophectoderm Formation in the Preimplantation Mouse Embryo. *Biol Reprod* **83**, 347–358.

Amadei, G., Lau, K. Y. C., Jonghe, J. D., Gantner, C. W., Sozen, B., Chan, C., Zhu, M., Kyprianou, C., Hollfelder, F. and Zernicka-Goetz, M. (2021). Inducible Stem-Cell-Derived Embryos Capture Mouse Morphogenetic Events In Vitro. *Dev Cell* **56**, 366-382.e9.

Amadei, G., Handford, C. E., Qiu, C., Jonghe, J. D., Greenfeld, H., Tran, M., Martin, B. K., Chen, D.-Y., Aguilera-Castrejon, A., Hanna, J. H., et al. (2022). Synthetic embryos complete gastrulation to neurulation and organogenesis. *Nature* 1–3.

Andreatta, M. and Carmona, S. J. (2021). UCell: Robust and scalable single-cell gene signature scoring. *Comput Struct Biotechnology J* **19**, 3796–3798.

Ang, S.-L. and Rossant, J. (1994). HNF-3β is essential for node and notochord formation in mouse development. *Cell* **78**, 561–574.

Aplin, J. D. and Kimber, S. J. (2004). Trophoblast-uterine interactions at implantation. *Reprod Biol Endocrin* **2**, 48.

Artus, J., Panthier, J.-J. and Hadjantonakis, A.-K. (2010). A role for PDGF signaling in expansion of the extra-embryonic endoderm lineage of the mouse blastocyst. *Development* **137**, 3361–3372.

Artus, J., Piliszek, A. and Hadjantonakis, A.-K. (2011). The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17. *Dev Biol* **350**, 393–404.

Artus, J., Douvaras, P., Piliszek, A., Isern, J., Baron, M. H. and Hadjantonakis, A.-K. (2012). BMP4 signaling directs primitive endoderm-derived XEN cells to an extraembryonic visceral endoderm identity. *Dev Biol* **361**, 245–262.

Assmat, E., Vinot, S., Gofflot, F., Linsel-Nitschke, P., Illien, F., Châtelet, F., Verroust, P., Louvet-Valle, S., Rinninger, F. and Kozyraki, R. (2005). Expression and Role of Cubilin in the Internalization of Nutrients During the Peri-Implantation Development of the Rodent Embryo1. *Biol Reprod* **72**, 1079–1086.

Azami, T., Waku, T., Matsumoto, K., Jeon, H., Muratani, M., Kawashima, A., Yanagisawa, J., Manabe, I., Nagai, R., Kunath, T., et al. (2017). Klf5 maintains the balance of primitive endoderm versus epiblast specification during mouse embryonic development by suppression of Fgf4. *Development* **144**, 3706–3718.

Barbacci, E., Reber, M., Ott, M. O., Breillat, C., Huetz, F. and Cereghini, S. (1999). Variant hepatocyte nuclear factor 1 is required for visceral endoderm specification. *Development* **126**, 4795–4805.

Barbera, J. P. M., Clements, M., Thomas, P., Rodriguez, T., Meloy, D., Kioussis, D. and Beddington, R. S. (2000). The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. *Development* **127**, 2433–2445.

Bedzhov, I., Graham, S. J. L., Leung, C. Y. and Zernicka-Goetz, M. (2014a). Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. *Philosophical Transactions Royal Soc B Biological Sci* **369**, 20130538.

Bedzhov, I., Leung, C. Y., Bialecka, M. and Zernicka-Goetz, M. (2014b). In vitro culture of mouse blastocysts beyond the implantation stages. *Nat Protoc* **9**, 2732–2739.

Behrendtsen, O. and Werb, Z. (1997). Metalloproteinases regulate parietal endoderm differentiating and migrating in cultured mouse embryos. *Dev Dynam* **208**, 255–265.

Belaoussoff, M., Farrington, S. M. and Baron, M. H. (1998). Hematopoietic induction and respecification of A-P identity by visceral endoderm signaling in the mouse embryo. *Development* **125**, 5009–5018.

Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E. J. and Constam, D. B. (2006). The Nodal Precursor Acting via Activin Receptors Induces Mesoderm by Maintaining a Source of Its Convertases and BMP4. *Dev Cell* **11**, 313–323.

Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T., et al. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* **11**, 4267.

Beppu, H., Kawabata, M., Hamamoto, T., Chytil, A., Minowa, O., Noda, T. and Miyazono, K. (2000). BMP Type II Receptor Is Required for Gastrulation and Early Development of Mouse Embryos. *Dev Biol* **221**, 249–258.

Bessonnard, S., Mot, L. D., 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.

Blij, S., Parenti, A., Tabatabai-Yazdi, N. and Ralston, A. (2015). Cdx2 Efficiently Induces Trophoblast Stem-Like Cells in Naïve, but Not Primed, Pluripotent Stem Cells. *Stem Cells Dev* **24**, 1352–1365.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.

Bowles, J., Schepers, G. and Koopman, P. (2000). Phylogeny of the SOX Family of Developmental Transcription Factors Based on Sequence and Structural Indicators. *Dev Biol* **227**, 239–255.

Brennan, J., Lu, C. C., Norris, D. P., Rodriguez, T. A., Beddington, R. S. P. and Robertson, E. J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**, 965.

Brickman, J. M., Jones, C. M., Clements, M., Smith, J. C. and Beddington, R. S. (2000). Hex is a transcriptional repressor that contributes to anterior identity and suppresses Spemann organiser function. *Development* **127**, 2303–2315.

Burtscher, I. and Lickert, H. (2009). Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. *Development* **136**, 1029–1038.

Cai, K. Q., Capo-Chichi, C. D., Rula, M. E., Yang, D. and Xu, X. (2008). Dynamic GATA6 expression in primitive endoderm formation and maturation in early mouse embryogenesis. *Dev Dynam* **237**, 2820–2829.

Candia, A. F., Watabe, T., Hawley, S. H., Onichtchouk, D., Zhang, Y., Derynck, R., Niehrs, C. and Cho, K. W. (1997). Cellular interpretation of multiple TGF-beta signals: intracellular antagonism between activin/BVg1 and BMP-2/4 signaling mediated by Smads. *Development (Cambridge, England)* **124**, 4467–80.

Casanova, J. E. and Grabel, L. B. (1988). The role of cell interactions in the differentiation of teratocarcinoma-derived parietal and visceral endoderm. *Dev Biol* **129**, 124–139.

Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., et al. (2021). JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, gkab1113-.

Charron, F., Paradis, P., Bronchain, O., Nemer, G. and Nemer, M. (1999). Cooperative Interaction between GATA-4 and GATA-6 Regulates Myocardial Gene Expression. *Mol Cell Biol* **19**, 4355–4365.

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.

Cheloha, R. W., Gellman, S. H., Vilardaga, J.-P. and Gardella, T. J. (2015). PTH receptor-1 signalling—mechanistic insights and therapeutic prospects. *Nat Rev Endocrinol* **11**, 712–724.

Chen, W. S., Manova, K., Weinstein, D. C., Duncan, S. A., Plump, A. S., Prezioso, V. R., Bachvarova, R. F. and Darnell, J. E. (1994). Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. *Gene Dev* **8**, 2466–2477.

Chen, L., Yabuuchi, A., Eminli, S., Takeuchi, A., Lu, C.-W., Hochedlinger, K. and Daley, G. Q. (2009). Cross-regulation of the Nanog and Cdx2 promoters. *Cell Res* **19**, cr200979.

Cheng, L. and Grabel, L. B. (1997). The Involvement of Tissue-Type Plasminogen Activator in Parietal Endoderm Outgrowth. *Exp Cell Res* **230**, 187–196.

Cheng, S., Pei, Y., He, L., Peng, G., Reinius, B., Tam, P. P. L., Jing, N. and Deng, Q. (2019). Single-Cell RNA-Seq Reveals Cellular Heterogeneity of Pluripotency Transition and X Chromosome Dynamics during Early Mouse Development. *Cell Reports* **26**, 2593-2607.e3.

Chickarmane, V. and Peterson, C. (2008). A Computational Model for Understanding Stem Cell, Trophectoderm and Endoderm Lineage Determination. *Plos One* **3**, e3478.

Cho, K. W. and Blitz, I. L. (1998). BMPs, Smads and metalloproteases: extracellular and intracellular modes of negative regulation. *Current Opinion in Genetics & Development* **8**, 443–449.

Cho, L. T. Y., Wamaitha, S. E., Tsai, I. J., Artus, J., Sherwood, R. I., Pedersen, R. A., Hadjantonakis, A.-K. and Niakan, K. K. (2012). Conversion from mouse embryonic to extra-embryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. *Development* **139**, 2866–2877.

Chu, G. C., Dunn, N. R., Anderson, D. C., Oxburgh, L. and Robertson, E. J. (2004). Differential requirements for Smad4 in TGFβ-dependent patterning of the early mouse embryo. *Development* **131**, 3501–3512.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M. and Zaret, K. S. (2002). Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**, 279–289.

Clark, C. C., Tomichek, E. A., Koszalka, T. R., Minor, R. R. and Kefalides, N. A. (1975). The embryonic rat parietal yolk sac. The role of the parietal endoderm in the biosynthesis of basement membrane collagen and glycoprotein in vitro. *J Biol Chem* **250**, 5259–5267.

Cockburn, K., Biechele, S., Garner, J. and Rossant, J. (2013). The Hippo Pathway Member Nf2 Is Required for Inner Cell Mass Specification. *Curr Biol* **23**, 1195–1201.

Cockroft, D. L. and Gardner, R. L. (1987). Clonal analysis of the developmental potential of 6th and 7th day visceral endoderm cells in the mouse. *Development* **101**, 143–155.

Coffinier, C., Thepot, D., Babinet, C., Yaniv, M. and Barra, J. (1999). Essential role for the homeoprotein vHNF1/HNF1beta in visceral endoderm differentiation. *Development* **126**, 4785–4794.

COOPER, A. R., KURKINEN, M., TAYLOR, A. and HOGAN, B. L. M. (1981). Studies on the Biosynthesis of Laminin by Murine Parietal Endoderm Cells. *Eur J Biochem* **119**, 189–197.

Costello, I., Nowotschin, S., Sun, X., Mould, A. W., Hadjantonakis, A.-K., Bikoff, E. K. and Robertson, E. J. (2015). Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. *Gene Dev* **29**, 2108–2122.

Coucouvanis, E. and Martin, G. R. (1999). BMP signaling plays a role in visceral endoderm differentiation and cavitation in the early mouse embryo. *Development* **126**, 535–546.

Daane, J. M. and Downs, K. M. (2011). Hedgehog signaling in the posterior region of the mouse gastrula suggests manifold roles in the fetal-umbilical connection and posterior morphogenesis. *Dev Dynam* **240**, 2175–2193.

Damert, A., Miquerol, L., Gertsenstein, M., Risau, W. and Nagy, A. (2002). Insufficient VEGFA activity in yolk sac endoderm compromises haematopoietic and endothelial differentiation. *Development* **129**, 1881–1892.

Dietrich, J.-E. and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. *Development* **134**, 4219–31.

Donovan, A., Lima, C. A., Pinkus, J. L., Pinkus, G. S., Zon, L. I., Robine, S. and Andrews, N. C. (2005). The iron exporter ferroportin/Slc40a1 is essential for iron homeostasis. *Cell Metab* **1**, 191–200.

Downs, K. M. (2020). Is extra-embryonic endoderm a source of placental blood cells? *Exp Hematol* **89**, 37–42.

Dufort, D., Schwartz, L., Harpal, K. and Rossant, J. (1998). The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis. *Development* **125**, 3015–3025.

Duncan, S. A., Nagy, A. and Chan, W. (1997). Murine gastrulation requires HNF-4 regulated gene expression in the visceral endoderm: tetraploid rescue of Hnf-4(−/−) embryos. *Development* **124**, 279–287.

Dyer, M. A., Farrington, S. M., Mohn, D., Munday, J. R. and Baron, M. H. (2001). Indian hedgehog activates hematopoiesis and vasculogenesis and can respecify prospective neurectodermal cell fate in the mouse embryo. *Development* **128**, 1717–1730.

Ferkowicz, M. J. and Yoder, M. C. (2005). Blood island formation: longstanding observations and modern interpretations. *Exp Hematol* **33**, 1041–1047.

Ferner, K. and Mess, A. (2011). Evolution and development of fetal membranes and placentation in amniote vertebrates. *Resp Physiol Neurobi* **178**, 39–50.

Filimonow, K. and Fuente, R. de la (2021). Specification and role of extraembryonic endoderm lineages in the periimplantation mouse embryo. *Theriogenology* **180**, 189–206.

Fowler, K. J., Mitrangas, K. and Dziadek, M. (1990). In vitro production of Reichert's membrane by mouse embryo-derived parietal endoderm cell lines. *Exp Cell Res* **191**, 194–203.

Fujiwara, Y., Tanno, Y., Sugishita, H., Kishi, Y., Makino, Y. and Okada, Y. (2021). Preparation of optimized concanavalin A-conjugated Dynabeads® magnetic beads for CUT&Tag. *Plos One* **16**, e0259846.

Gadue, P., Huber, T. L., Paddison, P. J. and Keller, G. M. (2006). Wnt and TGF-β signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proc National Acad Sci* **103**, 16806–16811.

Geusz, R. J., Wang, A., Lam, D. K., Vinckier, N. K., Alysandratos, K.-D., Roberts, D. A., Wang, J., Kefalopoulou, S., Ramirez, A., Qiu, Y., et al. (2021). Sequence logic at enhancers governs a dual mechanism of endodermal organ fate induction by FOXA pioneer factors. *Nat Commun* **12**, 6636.

Goldstein, B., Rogelj, S., Siegel, S., Farmer, S. R. and Niles, R. M. (1990). Cyclic adenosine monophosphate-mediated induction of F9 teratocarcinoma differentiation in the absence of retinoic acid. *J Cell Physiol* **143**, 205–212.

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., Eils, R. and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849.

Hamilton, W. B., Mosesson, Y., Monteiro, R. S., Emdal, K. B., Knudsen, T. E., Francavilla, C., Barkai, N., Olsen, J. V. and Brickman, J. M. (2019). Dynamic lineage priming is driven via direct enhancer regulation by ERK. *Nature* **575**, 355–360.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J.,

Wilk, A. J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29.

Harris, T. M. and Childs, G. (2002). Global gene expression patterns during differentiation of F9 embryonal carcinoma cells into parietal endoderm. *Funct Integr Genomic* **2**, 105–119.

Hayamizu, T. F., Wicks, M. N., Davidson, D. R., Burger, A., Ringwald, M. and Baldock, R. A. (2013). EMAP/EMAPA ontology of mouse developmental anatomy: 2013 update. *J Biomed Semant* **4**, 15.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**, 576–589.

Heslop, J. A., Pournasr, B., Liu, J.-T. and Duncan, S. A. (2021). GATA6 defines endoderm fate by controlling chromatin accessibility during differentiation of human-induced pluripotent stem cells. *Cell Reports* **35**, 109145.

Hirate, Y., Hirahara, S., Inoue, K., Suzuki, A., Alarcon, V. B., Akimoto, K., Hirai, T., Hara, T., Adachi, M., Chida, K., et al. (2013). Polarity-Dependent Distribution of Angiomotin Localizes Hippo Signaling in Preimplantation Embryos. *Curr Biol* **23**, 1181–1194.

Hirate, Y., Hirahara, S., Inoue, K., Kiyonari, H., Niwa, H. and Sasaki, H. (2015). Par-aPKC-dependent and -independent mechanisms cooperatively control cell polarity, Hippo signaling, and cell positioning in 16-cell stage mouse embryos. *Dev Growth Differ* **57**, 544–556.

Hogan, B. L. M., Cooper, A. R. and Kurkinen, M. (1980). Incorporation into Reichert's membrane of laminin-like extracellular proteins synthesized by parietal endoderm cells of the mouse embryo. *Dev Biol* **80**, 289–300.

Hogan, B. L. M., Taylor, A. and Adamson, E. (1981). Cell interactions modulate embryonal carcinoma cell differentiation into parietal or visceral endoderm. *Nature* **291**, 235–237.

Hogan, B. L. M., Taylor, A. and Cooper, A. R. (1982). Murine parietal endoderm cells synthesise heparan sulphate and 170K and 145K sulphated glycoproteins as components of Reichert's membrane. *Dev Biol* **90**, 210–214.

Holmes, W. R., Mochel, N. S. R. de, Wang, Q., Du, H., Peng, T., Chiang, M., Cinquin, O., Cho, K. and Nie, Q. (2017). Gene Expression Noise Enhances Robust Organization of the Early Mammalian Blastocyst. *Plos Comput Biol* **13**, e1005320.

Home, P., Saha, B., Ray, S., Dutta, D., Gunewardena, S., Yoo, B., Pal, A., Vivian, J. L., Larson, M., Petroff, M., et al. (2012). Altered subcellular localization of transcription factor TEAD4 regulates first mammalian cell lineage commitment. *Proceedings of the National*

Academy of Sciences **109**, 7362–7367.

Hoshino, H., Shioi, G. and Aizawa, S. (2015). AVE protein expression and visceral endoderm cell behavior during anterior–posterior axis formation in mouse embryos: Asymmetry in OTX2 and DKK1 expression. *Dev Biol* **402**, 175–191.

Howe, C. C. and Solter, D. (1980). Identification of noncollagenous basement membrane glycopolypeptides synthesized by mouse parietal entoderm and an entodermal cell line. *Dev Biol* **77**, 480–487.

Huang, D., Guo, G., Yuan, P., Ralston, A., Sun, L., Huss, M., Mistri, T., Pinello, L., Ng, H. H., Yuan, G., et al. (2017). The role of Cdx2 as a lineage specific transcriptional repressor for pluripotent network during the first developmental cell lineage segregation. *Sci Rep-uk* **7**, 17156.

Ichimura, T., Hatae, T., Sakurai, T. and Ishida, T. (1994). Three-dimensional architecture of the tubular endocytic apparatus and paramembranous networks of the endoplasmic reticulum in the rat visceral yolk-sac endoderm. *Cell Tissue Res* **278**, 353–361.

Javier, A. L., Doan, L. T., Luong, M., Mochel, N. S. R. de, Sun, A., Monuki, E. S. and Cho, K. W. Y. (2012). Bmp Indicator Mice Reveal Dynamic Regulation of Transcriptional Response. *Plos One* **7**, e42566.

Jedrusik, A., Cox, A., Wicher, K. B., Glover, D. M. and Zernicka-Goetz, M. (2015). Maternal-zygotic knockout reveals a critical role of Cdx2 in the morula to blastocyst transition. *Dev Biol* **398**, 147–152.

Jindal, G. A. and Farley, E. K. (2021). Enhancer grammar in development, evolution, and disease: dependencies and interplay. *Dev Cell* **56**, 575–587.

Julio, M. K., Alvarez, M. J., Galli, A., Chu, J., Price, S. M., Califano, A. and Shen, M. M. (2011). Regulation of extra-embryonic endoderm stem cell differentiation by Nodal and Cripto signaling. *Development* **138**, 3885–3895.

Kanai-Azuma, M., Kanai, Y., Gad, J. M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y., Yonekawa, H., Yazaki, K., Tam, P. P. L., et al. (2002). Depletion of definitive gut endoderm in Sox17 -null mutant mice. *Development* **129**, 2367–2379.

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**, 27–30.

Kaneko, K. J. and DePamphilis, M. L. (2013). TEAD4 establishes the energy homeostasis essential for blastocoel formation. *Development* **140**, 3680–3690.

Kang, M., Piliszek, A., Artus, J. and Hadjantonakis, A.-K. (2013). FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their

initial expression in the mouse. *Development* **140**, 267–279.

Kang, M., Garg, V. and Hadjantonakis, A.-K. (2017). Lineage Establishment and Progression within the Inner Cell Mass of the Mouse Blastocyst Requires FGFR1 and FGFR2. *Dev Cell* **41**, 496-510.e5.

Kawamura, N., Sun-Wada, G.-H., Aoyama, M., Harada, A., Takasuga, S., Sasaki, T. and Wada, Y. (2012). Delivery of endosomes to lysosomes via microautophagy in the visceral endoderm of mouse embryos. *Nat Commun* **3**, 1071.

Kawamura, N., Takaoka, K., Hamada, H., Hadjantonakis, A.-K., Sun-Wada, G.-H. and Wada, Y. (2020). Rab7-Mediated Endocytosis Establishes Patterning of Wnt Activity through Inactivation of Dkk Antagonism. *Cell Reports* **31**, 107733.
Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Gene Dev* **19**, 1129–1155.

Kelly, G. M. and Drysdale, T. A. (2015). Retinoic Acid and the Development of the Endoderm. *J Dev Biology* **3**, 25–56.

Kim, D., Paggi, J. M., Park, C., Bennett, C. and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915.

Kimura-Yoshida, C., Nakano, H., Okamura, D., Nakao, K., Yonemura, S., Belo, J. A., Aizawa, S., Matsui, Y. and Matsuo, I. (2005). Canonical Wnt Signaling and Its Antagonist Regulate Anterior-Posterior Axis Polarization by Guiding Cell Migration in Mouse Visceral Endoderm. *Dev Cell* **9**, 639–650.

Kimura-Yoshida, C., Tian, E., Nakano, H., Amazaki, S., Shimokawa, K., Rossant, J., Aizawa, S. and Matsuo, I. (2007). Crucial roles of Foxa2 in mouse anterior–posterior axis polarization via regulation of anterior visceral endoderm-specific genes. *Proc National Acad Sci* **104**, 5919–5924.

Kono, K., Tamashiro, D. A. A. and Alarcon, V. B. (2014). Inhibition of RHO–ROCK signaling enhances ICM and suppresses TE characteristics through activation of Hippo signaling in the mouse blastocyst. *Dev Biol* **394**, 142–155.

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

Kulakovskiy, I. V., Vorontsov, I. E., Yevshin, I. S., Sharipov, R. N., Fedorova, A. D., Rumynskiy, E. I., Medvedeva, Y. A., Magana-Mora, A., Bajic, V. B., Papatsenko, D. A., et al. (2017). HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* **46**, gkx1106-.

Kumar, A., Lualdi, M., Lyozin, G. T., Sharma, P., Loncarek, J., Fu, X.-Y. and Kuehn, M. R. (2015). Nodal signaling from the visceral endoderm is required to maintain Nodal gene expression in the epiblast and drive DVE/AVE migration. *Dev Biol* **400**, 1–9.

Kuo, C. T., Morrisey, E. E., Anandappa, R., Sigrist, K., Lu, M. M., Parmacek, M. S., Soudais, C. and Leiden, J. M. (1997). GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Gene Dev* **11**, 1048–1060.

Kwon, G. S. and Hadjantonakis, A. (2009). Transthyretin mouse transgenes direct RFP expression or Cre-mediated recombination throughout the visceral endoderm. *Genesis* **47**, 447–455.

Kwon, G. S., Viotti, M. and Hadjantonakis, A.-K. (2008). The Endoderm of the Mouse Embryo Arises by Dynamic Widespread Intercalation of Embryonic and Extraembryonic Lineages. *Dev Cell* **15**, 509–520.

LaMonica, K., Bass, M. and Grabel, L. (2009). The planar cell polarity pathway directs parietal endoderm migration. *Dev Biol* **330**, 44–53.

Lane, E. B., Hogan, B. L. M., Kurkinen, M. and Garrels, J. I. (1983). Co-expression of vimentin and cytokeratins in parietal endoderm cells of early mouse embryo. *Nature* **303**, 701–704.

Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics* **9**, 559.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359.

Lawson, K. A., Dunn, N. R., Roelen, B. A. J., Zeinstra, L. M., Davis, A. M., Wright, C. V. E., Korving, J. P. W. F. M. and Hogan, B. L. M. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Gene Dev* **13**, 424–436.
Leung, C. Y. and Zernicka-Goetz, M. (2013). Angiomotin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and -independent mechanisms. Nat Commun 4, 2251.

Li, M., Baumeister, P., Roy, B., Phan, T., Foti, D., Luo, S. and Lee, A. S. (2000). ATF6 as a transcription activator of the endoplasmic reticulum stress element: thapsigargin stress-induced changes and synergistic interactions with NF-Y and YY1. *Mol Cell Biol* **20**, 5096–106.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P. and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* **1**, 417–425.

Lim, C. Y., Tam, W.-L., Zhang, J., Ang, H. S., Jia, H., Lipovich, L., Ng, H.-H., Wei, C.-L., Sung, W. K., Robson, P., et al. (2008). Sall4 Regulates Distinct Transcription Circuitries in

Different Blastocyst-Derived Stem Cell Lineages. *Cell Stem Cell* **3**, 543–554.

Liu, P., Wakamiya, M., Shea, M. J., Albrecht, U., Behringer, R. R. and Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. *Nat Genet* **22**, 361–365.

Liu, N., Hargreaves, V. V., Zhu, Q., Kurland, J. V., Hong, J., Kim, W., Sher, F., Macias-Trevino, C., Rogers, J. M., Kurita, R., et al. (2018). Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. *Cell* **173**, 430-442.e17.

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.

Lu, R., Yang, A. and Jin, Y. (2011). Dual Functions of T-Box 3 (Tbx3) in the Control of Self-renewal and Extraembryonic Endoderm Differentiation in Mouse Embryonic Stem Cells*. *J Biol Chem* **286**, 8425–8436.

Manejwala, F. M., Cragoe, E. J. and Schultz, R. M. (1989). Blastocoel expansion in the preimplantation mouse embryo: Role of extracellular sodium and chloride and possible apical routes of their entry. *Dev Biol* **133**, 210–220.

Masson, N., Hurst, H. C. and Lee, K. A. W. (1993). Identification of proteins that interact with CREB during differentiation of F9 embryonal carcinoma cells. *Nucleic Acids Res* **21**, 1163–1169.

McCauley, L. K. and Martin, T. J. (2012). Twenty-five years of PTHrP progress: From cancer hormone to multifunctional cytokine. *J Bone Miner Res* **27**, 1231–1239.

McDonald, A. C. H., Biechele, S., Rossant, J. and Stanford, W. L. (2014). Sox17-Mediated XEN Cell Conversion Identifies Dynamic Networks Controlling Cell-Fate Decisions in Embryo-Derived Stem Cells. *Cell Reports* **9**, 780–793.

Meehan, R. R., Barlow, D. P., Hill, R. E., Hogan, B. L. and Hastie, N. D. (1984). Pattern of serum protein gene expression in mouse visceral yolk sac and foetal liver. *Embo J* **3**, 1881–1885.

Meers, M. P., Bryson, T. D., Henikoff, J. G. and Henikoff, S. (2019). Improved CUT&RUN chromatin profiling tools. *Elife* **8**, e46314.

Mesnard, D., Guzman-Ayala, M. and Constam, D. B. (2006). Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. *Development* **133**, 2497–2505.

Messerschmidt, D. M. and Kemler, R. (2010). Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Developmental biology* **344**, 129–37.

Mihajlović, A. I. and Bruce, A. W. (2017). The first cell-fate decision of mouse

preimplantation embryo development: integrating cell position and polarity. *Open Biol* **7**, 170210.

Mills, E., LaMonica, K., Hong, T., Pagliaruli, T., Mulrooney, J. and Grabel, L. (2009). Roles for Rho/ROCK and Vinculin in Parietal Endoderm Migration. *Cell Commun Adhesion* **12**, 9–22.

Miner, J. H., Li, C., Mudd, J. L., Go, G. and Sutherland, A. E. (2004). Compositional and structural requirements for laminin and basement membranes during mouse embryo implantation and gastrulation. *Development* **131**, 2247–2256.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S. (2003). The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES *Cell*s. Cell **113**, 631–642.

Mochel, N. S. R. de, Luong, M., Chiang, M., Javier, A. L., Luu, E., Toshihiko, F., MacGregor, G. R., Cinquin, O. and Cho, K. W. Y. (2015). BMP signaling is required for cell cleavage in preimplantation-mouse embryos. *Dev Biol* **397**, 45–55.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *Cell Reports* **20**, 1215–1228.

Molkentin, J. D., Lin, Q., Duncan, S. A. and Olson, E. N. (1997). Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Gene Dev* **11**, 1061–1072.

Molotkov, A., Mazot, P., Brewer, J. R., Cinalli, R. M. and Soriano, P. (2017). Distinct Requirements for FGFR1 and FGFR2 in Primitive Endoderm Development and Exit from Pluripotency. *Dev Cell* **41**, 511-526.e4.

Morabito, S., Miyoshi, E., Michael, N., Shahin, S., Martini, A. C., Head, E., Silva, J., Leavy, K., Perez-Rosendahl, M. and Swarup, V. (2021). Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat Genet* **53**, 1143–1155.

Morris, S. A., Teo, R. T. Y., Li, H., Robson, P., Glover, D. M. and Zernicka-Goetz, M. (2010). Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc National Acad Sci* **107**, 6364–6369.

Nabet, B., Roberts, J. M., Buckley, D. L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A. L., Erb, M. A., Lawlor, M. A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. *Nat Chem Biol* **14**, 431–441.

Nabet, B., Ferguson, F. M., Seong, B. K. A., Kuljanin, M., Leggett, A. L., Mohardt, M. L.,

Robichaud, A., Conway, A. S., Buckley, D. L., Mancias, J. D., et al. (2020). Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. *Nat Commun* **11**, 4687.

Neill, T., Sharpe, C., Owens, R. T. and Iozzo, R. V. (2017). Decorin-evoked paternally expressed gene 3 (PEG3) is an upstream regulator of the transcription factor EB (TFEB) in endothelial cell autophagy. *J Biological Chem* **292**, 16211–16220.

Niakan, K. K., Ji, H., Maehr, R., Vokes, S. A., Rodolfa, K. T., Sherwood, R. I., Yamaki, M., Dimos, J. T., Chen, A. E., Melton, D. A., et al. (2010). Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Gene Dev* **24**, 312–326.

Niakan, K. K., Schrode, N., Cho, L. T. Y. and Hadjantonakis, A.-K. (2013). Derivation of extraembryonic endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. *Nat Protoc* **8**, 1028–1041.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H. and Smith, A. (1998). Formation of Pluripotent Stem *Cell*s in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. Cell **95**, 379–391.

Niimi, T., Hayashi, Y., Futaki, S. and Sekiguchi, K. (2004). SOX7 and SOX17 Regulate the Parietal Endoderm-specific Enhancer Activity of Mouse Laminin α1 Gene*. *J Biol Chem* **279**, 38055–38061.

Ninomiya, Y., Davies, T. J. and Gardner, R. L. (2005). Experimental analysis of the transdifferentiation of visceral to parietal endoderm in the mouse. *Dev Dynam* **233**, 837–846.

Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R. O., Ogonuki, N., et al. (2009). The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. *Dev Cell* **16**, 398–410.

Nishiyama, A., Xin, L., Sharov, A. A., Thomas, M., Mowrer, G., Meyers, E., Piao, Y., Mehta, S., Yee, S., Nakatake, Y., et al. (2009). Uncovering Early Response of Gene Regulatory Networks in ESCs by Systematic Induction of Transcription Factors. *Cell Stem Cell* **5**, 420–433.

Niwa, H., Miyazaki, J. and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372–376.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation. *Cell* **123**, 917–929.

Notarianni, E. and Fléchon, J.-E. (2001). Parietal Endoderm Cell Line from a Rat Blastocyst. *Placenta* **22**, 111–123.

Nowotschin, S., Setty, M., Kuo, Y.-Y., Liu, V., Garg, V., Sharma, R., Simon, C. S., Saiz, N., Gardner, R., Boutet, S. C., et al. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367.

Ogawa, K., Sun, J., Taketani, S., Nakajima, O., Nishitani, C., Sassa, S., Hayashi, N., Yamamoto, M., Shibahara, S., Fujita, H., et al. (2001). Heme mediates derepression of Maf recognition element through direct binding to transcription repressor Bach1. *Embo J* **20**, 2835–2843.

Ohinata, Y., Endo, T. A., Sugishita, H., Watanabe, T., Iizuka, Y., Kawamoto, Y., Saraya, A., Kumon, M., Koseki, Y., Kondo, T., et al. (2022). Establishment of mouse stem cells that can recapitulate the developmental potential of primitive endoderm. *Science* **375**, 574–578.

Oshima, R. G. (1981). Identification and immunoprecipitation of cytoskeletal proteins from murine extra-embryonic endodermal cells. *J Biol Chem* **256**, 8124–8133.

Paca, A., Séguin, C. A., Clements, M., Ryczko, M., Rossant, J., Rodriguez, T. A. and Kunath, T. (2012). BMP signaling induces visceral endoderm differentiation of XEN cells and parietal endoderm. *Dev Biol* **361**, 90–102.

Panagopoulos, I., Möller, E., Dahlén, A., Isaksson, M., Mandahl, N., Vlamis-Gardikas, A. and Mertens, F. (2006). Characterization of the native CREB3L2 transcription factor and the FUS/CREB3L2 chimera. *Genes Chromosomes Cancer* **46**, 181–191.

Papanayotou, C. and Collignon, J. (2014). Activin/Nodal signalling before implantation: setting the stage for embryo patterning. *Philosophical Transactions Royal Soc B Biological Sci* **369**, 20130539–20130539.

Paraiso, K. D., Blitz, I. L., Coley, M., Cheung, J., Sudou, N., Taira, M. and Cho, K. W. Y. (2019). Endodermal Maternal Transcription Factors Establish Super-Enhancers during Zygotic Genome Activation. *Cell Reports* **27**, 2962-2977.e5.

Perea-Gomez, A., Camus, A., Moreau, A., Grieve, K., Moneron, G., Dubois, A., Cibert, C. and Collignon, J. (2004). Initiation of Gastrulation in the Mouse Embryo Is Preceded by an Apparent Shift in the Orientation of the Anterior-Posterior Axis. *Curr Biol* **14**, 197–207.

Perea-Gomez, A., Cases, O., Lelièvre, V., Pulina, M. V., Collignon, J., Hadjantonakis, A.-K. and Kozyraki, R. (2019). Loss of Cubilin, the intrinsic factor-vitamin B12 receptor, impairs visceral endoderm endocytosis and endodermal patterning in the mouse. *Sci Rep-uk* **9**, 10168.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T. and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295.

Pierre, M., Yoshimoto, M., Huang, L., Richardson, M. and Yoder, M. C. (2009). VEGF and

IHH rescue definitive hematopoiesis in Gata-4 and Gata-6–deficient murine embryoid bodies. *Exp Hematol* **37**, 1038–1053.

Plusa, B., Frankenberg, S., Chalmers, A., Hadjantonakis, A.-K., Moore, C. A., Papalopulu, N., Papaioannou, V. E., Glover, D. M. and Zernicka-Goetz, M. (2005). Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *J Cell Sci* **118**, 505–515.

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.

Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. and Zanini, F. (2022). Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945.

Qiu, C., Cao, J., Martin, B. K., Li, T., Welsh, I. C., Srivatsan, S., Huang, X., Calderon, D., Noble, W. S., Disteche, C. M., et al. (2022). Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat Genet* **54**, 328–341.

Qu, X., Pan, J., Zhang, C. and Huang, S. (2008). Sox17 facilitates the differentiation of mouse embryonic stem cells into primitive and definitive endoderm in vitro. *Dev Growth Differ* **50**, 585–593.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187–W191.

Rayon, T., Menchero, S., Nieto, A., Xenopoulos, P., Crespo, M., Cockburn, K., Cañon, S., Sasaki, H., Hadjantonakis, A.-K., de la Pompa, J. L., et al. (2014). Notch and Hippo Converge on Cdx2 to Specify the Trophectoderm Lineage in the Mouse Blastocyst. *Dev Cell* **30**, 410–422.

Red-Horse, K., Zhou, Y., Genbacev, O., Prakobphol, A., Foulk, R., McMaster, M. and Fisher, S. J. (2004). Trophoblast differentiation during embryo implantation and formation of the maternal-fetal interface. *J Clin Invest* **114**, 744–754.
Robertson, E. J. (2014). Dose-dependent Nodal/Smad signals pattern the early mouse embryo. *Semin Cell Dev Biol* **32**, 73–79.

Rodriguez, A. M. and Downs, K. M. (2017). Visceral endoderm and the primitive streak interact to build the fetal-placental interface of the mouse gastrula. *Dev Biol* **432**, 98–**124**.

Ross, C. and Boroviak, T. E. (2020). Origin and function of the yolk sac in primate embryogenesis. *Nat Commun* **11**, 3760.
Rossant, J. (1995). Development of the extraembryonic lineages. *Seminars Dev Biology* **6**, 237–247.

Rossant, J. and Tam, P. P. L. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136**, 701–713.

Sahoo, S. S., Ramanand, S. G., Gao, Y., Abbas, A., Kumar, A., Cuevas, I. C., Li, H.-D., Aguilar, M., Xing, C., Mani, R. S., et al. (2022). FOXA2 suppresses endometrial carcinogenesis and epithelial-mesenchymal transition by regulating enhancer activity. *J Clin Investigation* **132**, e157574.

Saiz, N., Williams, K. M., Seshan, V. E. and Hadjantonakis, A.-K. (2016). Asynchronous fate decisions by single cells collectively ensure consistent lineage composition in the mouse blastocyst. *Nat Commun* **7**, 13463.

Salamat, M., Götz, W., Horster, A., Janotte, B. and Herken, R. (1993). Ultrastructural localization of carbohydrates in Reichert's membrane of the mouse. *Cell Tissue Res* **272**, 375–381.

Salamat, M., Miosge, N. and Herken, R. (1995). Development of Reichert's membrane in the early mouse embryo. *Anat Embryol* **192**, 275–281.

Schrode, N., Saiz, N., Talia, S. D. and Hadjantonakis, A.-K. (2014). GATA6 Levels Modulate Primitive Endoderm Cell Fate Choice and Timing in the Mouse Blastocyst. *Dev Cell* **29**, 454–467.

Schröter, C., Rué, P., Mackenzie, J. P. and Arias, A. M. (2015). FGF/MAPK signaling sets the switching threshold of a bistable circuit controlling cell fate decisions in embryonic stem cells. *Development* **142**, 4205–4216.

Semoff, S., Hogan, B. L. and Hopkins, C. R. (1982). Localization of fibronectin, laminin-entactin, and entactin in Reichert's membrane by immunoelectron microscopy. *Embo J* **1**, 1171–1175.

Senft, A. D., Costello, I., King, H. W., Mould, A. W., Bikoff, E. K. and Robertson, E. J. (2018). Combinatorial Smad2/3 Activities Downstream of Nodal Signaling Maintain Embryonic/Extra-Embryonic Cell Identities during Lineage Priming. *Cell Reports* **24**,.

Serizay, J. and Ahringer, J. (2021). Generating fragment density plots in R/Bioconductor with VplotR. *J Open Source Softw* **6**, 3009.

Serizay, J., Dong, Y., Janes, J., Chesney, M., Cerrato, C. and Ahringer, J. (2020). Distinctive regulatory architectures of germline-active and somatic genes in C. elegans. *Genome Res* **30**, gr.265934.120.

Sheng, G. and Foley, A. C. (2012). Diversification and conservation of the extraembryonic tissues in mediating nutrient uptake during amniote development. *Ann Ny Acad Sci* **1271**, 97–103.

Shimoda, M., Kanai-Azuma, M., Hara, K., Miyazaki, S., Kanai, Y., Monden, M. and Miyazaki, J. (2007). Sox17 plays a substantial role in late-stage differentiation of the extraembryonic endoderm in vitro. *J Cell Sci* **120**, 3859–3869.

Sirard, C., Pompa, J. L. de la, Elia, A., Itie, A., Mirtsos, C., Cheung, A., Hahn, S., Wakeham, A., Schwartz, L., Kern, S. E., et al. (1998). The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. *Gene Dev* **12**, 107–119.

Skene, P. J. and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856.

Smith, K. K. and Strickland, S. (1981). Structural components and characteristics of Reichert's membrane, an extra-embryonic basement membrane. *J Biol Chem* **256**, 4654–4661.
Smith, K. K. and Strickland, S. (1982). Plasminogen activator production by parietal endoderm cells: Stimulation by a protein from pituitary. *Dev Biol* **90**, 227–237.

Snyder, R. W., Lenburg, M. E., Seebaum, A. T. and Grabel, L. B. (1992). Disruption of the cytoskeleton-extracellular matrix linkage promotes the accumulation of plasminogen activators in F9 derived parietal endoderm. *Differentiation* **50**, 153–162.

Soncin, F., Natale, D. and Parast, M. M. (2015). Signaling pathways in mouse and human trophoblast differentiation: a comparative review. *Cell Mol Life Sci* **72**, 1291–1302.

Sorkin, A. and Zastrow, M. von (2009). Endocytosis and signalling: intertwining molecular networks. *Nat Rev Mol Cell Bio* **10**, 609–622.

Sozen, B., Amadei, G., Cox, A., Wang, R., Na, E., Czukiewska, S., Chappell, L., Voet, T., Michel, G., Jing, N., et al. (2018). Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nat Cell Biol* **20**, 979–989.

Sozen, B., Demir, N. and Zernicka-Goetz, M. (2021). BMP signalling is required for extra-embryonic ectoderm development during pre-to-post-implantation transition of the mouse embryo. *Dev Biol* **470**, 84–94.

Stary, M., Pasteiner, W., Summer, A., Hrdina, A., Eger, A. and Weitzer, G. (2005). Parietal endoderm secreted SPARC promotes early cardiomyogenesis in vitro. *Exp Cell Res* **310**, 331–343.

Stary, M., Schneider, M., Sheikh, S. P. and Weitzer, G. (2006). Parietal endoderm secreted S100A4 promotes early cardiomyogenesis in embryoid bodies. *Biochem Bioph Res Co* **343**, 555–563.

Stephenson, R. O., Yamanaka, Y. and Rossant, J. (2010). Disorganized epithelial polarity and excess trophectoderm cell fate in preimplantation embryos lacking E-cadherin. *Development*

137, 3383–3391.

Stolpe, A. van de, Karperien, M., Löwik, C. W., Jüppner, H., Segre, G. V., Abou-Samra, A. B., Laat, S. W. de and Defize, L. H. (1993). Parathyroid hormone-related peptide as an endogenous inducer of parietal endoderm differentiation. *J Cell Biology* **120**, 235–243.

Stower, M. J. and Srinivas, S. (2014). Heading forwards: anterior visceral endoderm migration in patterning the mouse embryo. *Philosophical Transactions Royal Soc B Biological Sci* **369**, 20130546.

Strickland, S. and Mahdavi, V. (1978). The induction of differentiation in teratocarcinoma stem cells by retinoic acid. *Cell* **15**, 393–403.

Strickland, S., Reich, E. and Sherman, M. I. (1976). Plasminogen activator in early embryogenesis: Enzyme production by trophoblast and parietal endoderm. *Cell* **9**, 231–240.

Strickland, S., Smith, K. K. and Marotti, K. R. (1980). Hormonal induction of differentiation in teratocarcinoma stem cells: Generation of parietal endoderm by retinoic acid and dibutyryl cAMP. *Cell* **21**, 347–355.

Strumpf, D., Mao, C.-A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F. and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093–2102.

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

Sun, J., Hoshino, H., Takaku, K., Nakajima, O., Muto, A., Suzuki, H., Tashiro, S., Takahashi, S., Shibahara, S., Alam, J., et al. (2002). Hemoprotein Bach1 regulates enhancer availability of heme oxygenase-1 gene. *Embo J* **21**, 5216–5224.
Sun-Wada, G.-H. and Wada, Y. (2022). The a subunit isoforms of vacuolar-type proton ATPase exhibit differential distribution in mouse perigastrulation embryos. Sci Rep-uk 12, 13590.

Sun-Wada, G.-H., Tabata, H. and Wada, Y. (2021). Vacuolar-type proton ATPase is required for maintenance of apicobasal polarity of embryonic visceral endoderm. *Sci Rep-uk* **11**, 19355.
Takaoka, K. (2014). New Principles in Developmental Processes. 13–26.

Takaoka, K., Yamamoto, M., Shiratori, H., Meno, C., Rossant, J., Saijoh, Y. and Hamada, H. (2006). The Mouse Embryo Autonomously Acquires Anterior-Posterior Polarity at Implantation. *Dev Cell* **10**, 451–459.

Takaoka, K., Yamamoto, M. and Hamada, H. (2011). Origin and role of distal visceral endoderm, a group of cells that determines anterior–posterior polarity of the mouse embryo. *Nat*

Cell Biol **13**, 743–752.

Takaoka, K., Nishimura, H. and Hamada, H. (2017). Both Nodal signalling and stochasticity select for prospective distal visceral endoderm in mouse embryos. *Nat Commun* **8**, 1492.

Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T. and Akira, S. (1997). Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proc National Acad Sci* **94**, 3801–3804.

Tam, P. P. L. and Loebel, D. A. F. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nat Rev Genet* **8**, 368–381.

Tarazi, S., Aguilera-Castrejon, A., Joubran, C., Ghanem, N., Ashouokhi, S., Roncato, F., Wildschutz, E., Haddad, M., Oldak, B., Gomez-Cesar, E., et al. (2022). Post-gastrulation synthetic embryos generated ex utero from mouse naive ESCs. *Cell* **185**, 3290-3306.e25.

Thomas, P. Q., Brown, A. and Beddington, R. S. (1998). Hex: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors. *Development* **125**, 85–94.

Thompson, J. J., Lee, D. J., Mitra, A., Frail, S., Dale, R. K. and Rocha, P. P. (2022). Extensive co-binding and rapid redistribution of NANOG and GATA6 during emergence of divergent lineages. *Nat Commun* **13**, 4257.

Ueda, Y., Kimura-Yoshida, C., Mochida, K., Tsume, M., Kameo, Y., Adachi, T., Lefebvre, O., Hiramatsu, R. and Matsuo, I. (2020). Intrauterine Pressures Adjusted by Reichert's Membrane Are Crucial for Early Mouse Morphogenesis. *Cell Reports* **31**, 107637.

Verheijen, M. H. G., Karperien, M., Chung, U., Wijuen, M. van, Heystek, H., Hendriks, J. A. A., Veltmaat, J. M., Lanske, B., Li, E., Löwik, C. W. G. M., et al. (1999a). Parathyroid hormone-related peptide (PTHrP) induces parietal endoderm formation exclusively via the Type I PTH/PTHrP receptor. *Mech Develop* **81**, 151–161.

Verheijen, M. H. G., Wolthuis, R. M. F., Bos, J. L. and Defize, L. H. K. (1999b). The Ras/Erk Pathway Induces Primitive Endoderm but Prevents Parietal Endoderm Differentiation of F9 Embryonal Carcinoma Cells*. *J Biol Chem* **274**, 1487–1494.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272.
Wada, Y. (2014). Vacuoles in mammals. *Bioarchitecture* **3**, 13–19.

Wada, Y., Sun-Wada, G.-H., Kawamura, N. and Aoyama, M. (2014). Role of autophagy in embryogenesis. *Curr Opin Genet Dev* **27**, 60–66.

Waldrip, W. R., Bikoff, E. K., Hoodless, P. A., Wrana, J. L. and Robertson, E. J. (1998). Smad2 Signaling in Extraembryonic Tissues Determines Anterior-Posterior Polarity of the Early Mouse Embryo. *Cell* **92**, 797–808.

Wallingford, M. C., Angelo, J. R. and Mager, J. (2013). Morphogenetic analysis of peri-implantation development. *Dev Dynam* **242**, 1110–1120.

Wamaitha, S. E., Valle, I. del, Cho, L. T. Y., Wei, Y., Fogarty, N. M. E., Blakeley, P., Sherwood, R. I., Ji, H. and Niakan, K. K. (2015). Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Gene Dev* **29**, 1239–1255.

Wat, M. J., Beck, T. F., Hernández-García, A., Yu, Z., Veenma, D., Garcia, M., Holder, A. M., Wat, J. J., Chen, Y., Mohila, C. A., et al. (2012). Mouse model reveals the role of SOX7 in the development of congenital diaphragmatic hernia associated with recurrent deletions of 8p23.1. *Hum Mol Genet* **21**, 4115–4125.

Weinstein, D. C., Altaba, A. R. i, Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M. and Darnell, J. E. (1994). The winged-helix transcription factor HNF-3β is required for notochord development in the mouse embryo. *Cell* **78**, 575–588.

Wiley, L. M. (1984). Cavitation in the mouse preimplantation embryo: NaK-ATPase and the origin of nascent blastocoele fluid. *Dev Biol* **105**, 330–342.

Winnier, G., Blessing, M., Labosky, P. A. and Hogan, B. L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Gene Dev* **9**, 2105–2116.

Wolf, F. A., Angerer, P. and Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141.

Xenopoulos, P., Kang, M., Puliafito, A., Di Talia, S. and Hadjantonakis, A.-K. (2015). Heterogeneities in Nanog Expression Drive Stable Commitment to Pluripotency in the Mouse Blastocyst. *Cell Reports* **10**, 1508–1520.

Xiang, Y., Zhang, Y., Xu, Q., Zhou, C., Liu, B., Du, Z., Zhang, K., Zhang, B., Wang, X., Gayen, S., et al. (2020). Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. *Nat Genet* **52**, 95–105.

Xiong, L., Tolen, E. A., Choi, J., Velychko, S., Caizzi, L., Velychko, T., Adachi, K., MacCarthy, C. M., Lidschreiber, M., Cramer, P., et al. (2022). Oct4 differentially regulates

chromatin opening and enhancer transcription in pluripotent stem cells. *Elife* **11**, e71533.

Yagi, S., Tagawa, Y. and Shiojiri, N. (2016). Transdifferentiation of mouse visceral yolk sac cells into parietal yolk sac cells in vitro. *Biochem Bioph Res Co* **470**, 917–923.

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.

Yang, X., Li, C., Xu, X. and Deng, C. (1998). The tumor suppressor SMAD4/DPC4 is essential for epiblast proliferation and mesoderm induction in mice. *Proc National Acad Sci* **95**, 3667–3672.

Yang, C., Xu, X., Dong, X., Yang, B., Dong, W., Luo, Y., Liu, X., Wu, Y. and Wang, J. (2021). DDIT3/CHOP promotes autophagy in chondrocytes via SIRT1-AKT pathway. *Biochimica Et Biophysica Acta Bba - Mol Cell Res* **1868**, 119074.

Yeo, C.-Y. and Whitman, M. (2001). Nodal Signals to Smads through Cripto-Dependent and Cripto-Independent Mechanisms. *Mol Cell* **7**, 949–957.

Yoon, Y., Huang, T., Tortelote, G. G., Wakamiya, M., Hadjantonakis, A.-K., Behringer, R. R. and Rivera-Pérez, J. A. (2015). Extra-embryonic Wnt3 regulates the establishment of the primitive streak in mice. *Dev Biol* **403**, 80–88.

Yu, G., Wang, L.-G., Han, Y. and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics J Integr Biology* **16**, 284–287.

Yu, S., Yehia, G., Wang, J., Stypulkowski, E., Sakamori, R., Jiang, P., Hernandez-Enriquez, B., Tran, T. S., Bonder, E. M., Guo, W., et al. (2014). Global Ablation of the Mouse Rab11a Gene Impairs Early Embryogenesis and Matrix Metalloproteinase Secretion*. *J Biol Chem* **289**, 32030–32043.

Zamparini, A. L., Watts, T., Gardner, C. E., Tomlinson, S. R., Johnston, G. I. and Brickman, J. M. (2006). Hex acts with β-catenin to regulate anteroposterior patterning via a Groucho-related co-repressor and Nodal. *Development* **133**, 3709–3722.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137.

Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., Fidalgo, M., Xia, W., Li, Y., Zhao, Z., et al. (2018). Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat Genet* **50**, 96–105.

Zhang, S., Chen, T., Chen, N., Gao, D., Shi, B., Kong, S., West, R. C., Yuan, Y., Zhi, M., Wei, Q., et al. (2019). Implantation initiation of self-assembled embryo-like structures generated

using three types of mouse blastocyst-derived stem cells. *Nat Commun* **10**, 496.

APPENDIX I:

Baf complex instructs chromatin refolding

AI.1 Introduction

In this appendix, we investigated the alterations in chromatin folding in response to the perturbation of BAF complex, the chromatin remodeler that controls chromatin accessibilities at cis-regulatory elements. We found enhancer syntax instructs the dependency of enhancer on the BAF complex as well as the contact configurations of enhancer-promoter (E-P) hubs. Loss of BAF occupancies weakened the strengths of a subset of E-P interactions. Lastly, we examined how global reduction in enhancer accessibilities affect the nuclear compartmentalization.

AI.2 Results

Enhancer Sequence Dictates Selective Recruitment of Chromatin Remodelers
We developed a MID Hi-C method (unpublished data), unlike in situ HiC, permits the robust detection of E-P interactions and enables us to systematically determine how E-P interactions are regulated. Numerous previous studies have revealed that chromatin remodelers, including the BAF complex, promote chromatin accessibility across enhancers (Alver et al., 2017; Nakayama et al., 2017; Park et al., 2021; Wang et al., 2017). To determine how the BAF complex regulates enhancer accessibility we inserted a dTAG-inducible FKBP12^{F36V} degron (Nabet et al., 2018) into the SMARCA4 locus, encoding for the only ATPase BRG1, in mESCs (Figure AI.2A and AI.2B). We found that dTAG-13 treatment swiftly reduced BRG1 abundance to barely detectable levels (Figure AI.2C). We next conducted a time-course MID Hi-C to interrogate acute (3h) and prolonged (6h and 24h) effects of BAF complex perturbation (Figure AI.2B). We first examined for alterations in chromatin accessibility at sites enhancer and promoter

128

repertoires excluding sites associated with CTCF-bound sites. ATAC-seq reads were overlaid and quantified signal fold changes were compared to t=0 control (Figures AI.1A and AI.1B). We found that chromatin accessibility across enhancers were more vulnerable to BAF depletion in comparison to promoters (Figures AI.1A and AI.1B). Depletion of BAF affected chromatin accessibility at active enhancers to a similar degree as weak enhancers (Figures AI.1A and AI.1B). Interestingly, enhancers demonstrated differential responses to BAF complex perturbation, as evidenced by a wide range of changes in chromatin accessibility (Figure AI.1B). To distinguish between the different responses, we segregated enhancers based on accessibility changes of BAF depletion (3h) into three subsets: BAF-dependent (BD) enhancer (5574, more than 8-fold reduction), BAF-partial-dependent (PD) enhancer (25828, more than 1.5-fold but less than 8-fold reduction), and BAF-independent (BI) enhancer (8970, less than 1.5-fold reduction) (Figure AI.1B). We found that the major reduction in chromatin accessibility of BD and PD enhancers occurred by 3h when the decline in BRG1 plateaued whereas accessibility changes of BI enhancers were moderate during the entire course of perturbation (Figure AI.2D). Consistently, BRG1 occupancies as determined by BRG1 ChIP-seq (Chronis et al., 2017) at BD enhancers were higher than BI enhancers (Figure AI.2E). Notably, DNA motif analysis for the BD and BI enhancer repertoires revealed striking differences in motif constitution (Figure AI.1C). Specifically, we found that BD enhancers were enriched for the motifs for OCT4/SOX2/TCF/NANOG, master pluripotent regulators whose functions are strongly dependent on the BAF complex (esBAF). In contrast, BI enhancers were enriched for the motifs of KLF1, SP5, and NFY that were reported to regulate naive pluripotency. In sum, enhancer sequence dictates selective recruitment of chromatin remodelers.

The data described above indicates that enhancer repertoires can be segregated into BI-enhancers and BD-enhancers. These findings raise the question whether BI-enhancers and BD-enhancers instruct distinct patterns of chromatin folding. As a first approach to address this question, we aggregated MID Hi-C contacts at paired promoters and BI- and BD-enhancers. We found that BI-enhancers were associated with a distinctively insulated chromatin domain structure (Figure AI.1F). The BI-enhancer instructed insulation was characterized by elevated contact frequencies within the domain and marked by stripes emanating from enhancers and promoters that intersected at focal points associated with promoter-enhancer interactions (Figure AI.1F). We found that BI enhancers were able to insulate chromatin interactions crossing the enhancer boundary, albeit weaker when compared to promoters (Figures AI.1E and AI.1F). In contrast, we found that chromatin domains associated with BD enhancers did not insulate against genomic interactions emanating from the promoter or genomic regions located between the promoter and enhancers (Figures AI.1E and AI.1F). Likewise, using Micro-C rather than MID Hi-C we found that BD enhancers differ from BI-enhancers in their ability to enforce insulation beyond the enhancer region (Figure AI.2F). Recent studies suggested that RNA polymerase II occupancy at transcription start sites interferes with loop extrusion resulting in unidirectional stripes and insulations of genomic interactions emanating from the promoter. Consistent with this model, we found that RNA polymerase II occupancy was enriched at BI-enhancers but not at BD-enhancers (Figure AI.1D). Taken together, these data indicate that the chromatin remodeler ATPase BRG1 enables genomic interactions emanating from promoters beyond enhancer boundary elements.

To investigate how BAF orchestrates the E-P interaction hub, we performed a PCA analysis for changes in E-P interaction strength during the course of BRG1 depletion. We found that PCA values associated with loss of BRG1 (3h, 6h, and 24h) clustered but segregated from cells not depleted for BRG1 (0h) (Figure AI.3A). Accordingly, E-P strengths segregated for BI, PD, and BD enhancer repertoires at 0h significantly differed from that observed for 3h, 6h and 24h BRG1-depletion (Figure AI.3B). Conversely, BRG1-depletion for 3 hours when compared to 6 and 24 hours did not reveal significant differences in E-P strength (Figure AI.3B). Thus, a new state of the E-P interactome was established at 3h and maintained beyond 3h, indicating that the changes in E-P interactions were tightly associated with the loss of BAF abundance.

To determine how during the course of BRG1 depletion the EP interaction hub is altered, we generated aggregated contact maps for BD, PD, and BI enhancer repertoires interacting with promoters. We found that the degree of reduction in E-P interaction strengths (intensity at the E-P focal point) correlated well with that of acute changes in chromatin accessibility across enhancers at 3h (Figures AI.3C and AI.4A). Specifically, E-P interactions at BD enhancers were nearly abrogated (Figures AI.3D and AI.4B). Conversely, we found that PE interaction strength at BI enhancers was only modestly attenuated (Figure AI.3C). Despite considerable reductions in the strength upon loss of BRG1 abundance, E-P interactions were retained across the PD enhancer repertoire (Figure AI.3C). Thus, abolishment of E-P interaction requires eradication of factors assembled at the enhancer platform that sequester the BAF complex.

We next focused our analysis on the role of BAF in instructing genomic interactions that span the entire E-P domain. To increase the coverage, we combined the MID Hi-C datasets of samples depleted for BAF during the course of 3, 6 and 24 hours. We found that although

131

genomic interactions emanating from BD enhancers were virtually undetectable under BAF complex perturbation condition, genomic interactions involving the promoter regions were largely intact (Figure AI.3E). Additionally, the contact frequencies of genomic interactions among genomic regions that span the promoter stripe remained elevated above the background (Figure AI.3E). In summary, we found that chromatin accessibility across enhancer repertoires, reflecting transcription factor occupancy and recruitment of associated cofactors, instructs enhancer-communication but is dispensable for genomic interaction across the E-P hub.

At the coarse 3D genomic scale, chromatin is segregated into active (A) and inactive (B) compartments (Lieberman-Aiden et al., 2009). Since activate enhancers are primarily located in the A compartment, we examined whether the global reduction of chromatin accessibility caused by BAF depletion modulates compartmental organization. We found that during the course of BAF depletion (3-24 hours) A/B compartmental remained largely intact (pairwise spearman > 0.94, Figures AI.4C and AI.4D). However, upon closer inspection using PCA analysis we found significant changes in A/B compartmentalization during the course of BAF depletion (Figure AI.5A). To determine whether these changes were orchestrated by changes in the compartmental segregation we generated saddle plots (Gibcus et al., 2018). This analysis captured a progressive increase in contact frequencies at the inter-compartmental interactions, indicating weakening in compartmental segregation that became more prominent during the course of BAF depletion (Figure AI.5B). To determine to what degree compartmental segregation changed upon BAF depletion, segregation scores, which quantifies the strength difference between the intra- and inter-compartmental interaction, were computed for genomic bins spanning 25 Kbp. PCA analysis on the segregation score profiles also detected time-dependent compartmental

segregation differences among samples but a higher percentage of sample-wise variation were captured compared to compartment scores (Figure AI.5D). We found upon BAF depletion a significant decline in compartmental segregation across genomic regions that were associated with a significant decline in chromatin accessibilities (Figure AI.5C). To determine how changes during the course of BAF depletion in compartmental segregation relate to the positioning in the A compartment, we computed segregation differences as a function of compartment score. We found that upon BAF depletion the loss of segregation was increased overtime for regions of higher compartmental scores (Figures AI.5E), correlated well with the reduction of chromatin accessibility (Figures AI.4E). Taken together, these data indicate that the BAF complex enforces compartmental segregation.

In this chapter, we identified two subsets of enhancers based on their dependencies on BAF complex as revealed by the acute changes in the chromatin accessibility after BRG1 depletion. They are not only enriched for different transcription factor motifs, but the underlying E-P interacting neighborhoods also assume distinct configurations. Lastly, we observed global reduction in chromatin accessibilities weaken the compartmental segregation in the nucleus.

AI.3 Materials and Methods

Cell Line Generation

The constructions of the guide RNA (sgRNA) and the donor plasmid for inserting FKBP12^{F36V} to Brg1 endogenous locus were described in our previous paper (Zhu et al., 2020). For generation of FKBP12^{F36V} cell lines, mESCs cells were seeded the day before transfection on a 10 cm tissue culture plate. 12 ug of plasmid DNA comprising equimolar amounts of sgRNA and donor plasmid were transfected using 40 ul of Lipofectamine 3000 with 24 ul of P3000 reagent (Thermo Fisher, L3000008). Two days after transfection, the transfected cells were selected in 10 ug/mL of Blasticidin (Cayman Chemical, NC1445974) for a week. Single cells from the EYFP positive population with successful insertions were then sorted using a BDFACSAria Fusion Sorter in 96-well round bottom plates. Single cells were grown for ~7 days to form single clones. Genotyping primers (Zhu et al., 2020) were used to screen for homozygous clones. The established cell lines were treated with dTAG13 (CAS 2064175-41-1). Efficiency of acute degradation of BRG1 proteins was examined by western blots as previously described (Zhu et al., 2020).

ATAC-seq Analysis

Adaptors in pair-end (PE) sequencing reads were trimmed using Trimmomatic (Bolger et al., 2014). Trimmed PE reads were aligned to mm10 genome using bowtie2 (Langmead and Salzberg, 2012) with flags `--very-sensitive -I 10 -X 700 –dovetail`. Only mapped fragments with sizes between 0 to 200 BP were kept for further analysis. Peaks were called by MACS2 with flags `-g mm -f BEDPE -q 0.01 –SPMR –keep-dup auto`. Then, functionalities of deeptools

(Ramírez et al., 2014) were used to analyze enrichment signals. Briefly, density tracks were generated by `bamCoverage` to be visualized in UCSC genome browser or by pyGenomeTracks (Lopez-Delisle et al., 2020). Clustering of peaks, based on H3K4me3, H3K27ac, and CTCF signal, and tornado plots were generated by `computeMatrix` and `plotHeatmap` to identify enhancers and promoters devoid of CTCF occupancies. Summarized signal profiles were plotted by `plotProfile`. In addition, quantification of peaks was performed through counting the number of reads overlapping with peaks using PyRanges (Stovner and Sætrom, 2019). Homer (Heinz et al., 2010) was used to search enriched motifs within 200 BP from the peak centers.

Compartmental Segregation Analysis

We calculated the averaged interaction strengths with A and B compartments separately for a 25 Kbps bin. Then, according to the compartment that the bin belongs to, we computed the ratio of intra- versus inter-compartmental interaction strengths, a.k.a compartmental segregation score. Segregation tracks were plotted by pyGenomeTracks and PCA analysis was performed as described above. Also, we calculated the averaged changes in segregation scores by stratifying bins according to the compartmental scores.

Figure AI.1. Enhancer dependencies on BAF correlate to the insulation potential. (A) Tornado plots showing the signals of the time-course ATAC-seq, H3K4me3, H3K27ac, H3K4me1, and CTCF at the accessible regions detected from the control (0h). (B) Violin plot showing the fold changes in the chromatin accessibilities for the treatment versus the control. (C) Enriched motifs at BI and BD enhancers. (D) Line plots displaying PolII occupancies at enhancers and promoters. (E) Line plots displaying the column-wise means calculated from the upper (blue) and lower (orange) halves of the APA matrices in (F). (F) APA at the E-P interaction subsets (by column) for MID Hi-C dataset.

A

BRG1-Degron mESC

ATAC | H3 K4me3 | H3 K27ac | H3 K4me1 | CTCF

ATAC Peaks (n=51,390)

-1.5kb center 1.5kb

0 1 2 | 0 1 2 | 0 1 2 | 0 1 2 | 0 1.5 3 | 0 1 2 3 | .05 .1 .150 | 0 1 2

dTAG13   0h   3h   6h   24h

B

log2 Fold Change (FC) vs 0h

-6 -4 -2 0 2

3h
6h
24h

Promoter (n=11,251)

Active Enhancer (n=4,256)

Weak Enhancer (n=35,883)

FC: 1/8 1/1.5

BAF Dependency: dependent partial dependent independent

(BD) (PD) (BI)

C

BI Enhancers

KLF1   1e-184
SP5   1e-180
NFY   1e-170

BD Enhancers

OCT4/SOX2/TCF/NANOG   1e-461
SOX3   1e-288

D

RNAP Occupany

1.4
1.2
1.0
0.8
0.6
0.4
0.2

BI Enhancer
PD Enhancer
BD Enhancer
Promoter

-1.5 KB center 1.5 KB

E

Average Signal

.1
-.1
-.3

-50 kb 50 kb

F

-50 kb

Promoter

50 kb

log2 Ob/Exp

.3

-.3

BAF-independent Enhancer

BAF-dependent Enhancer

Figure AI.2. BAF Complex Perturbation Uncovers BAF-Dependent and BAF-Independent Enhancers. (A) Bar plot showing Brg1 is the only expressed BAF ATPase in mESCs quantified by RNA-seq. (B) Schematic showing the design of experiments. (C) Western blots showing BRG1 protein is completely degraded after applying dTAG13. (D) Trend lines showing the fold changes (vs. 0h) in the chromatin accessibilities (ATAC-seq) during BRG1 depletion. (E) Line plots displaying BRG1 occupancies at enhancers and promoters. (F) APA at the E-P interaction subsets (by column) for Micro-C dataset (heatmap at the bottom). Column-wise mean (along the x axis centering on enhancer) calculated from the upper (blue) and lower (orange) halves of the APA matrices were plotted on the top panel.

A

B

MID Hi-C, ATAC
0h --> 3h --> 6h --> 24h

dTAG13

BRG1-FKBPv-NeonG
0h (Control)

BRG1-FKBPv-NeonG
3h, 6h, 24h

C

BRG1-FKBPv-NeonG

WT 0h 3h 6h 24h

250 kDa

BRG1-FKBPv-NeonG
BRG1

Tubulin

D

Enhancers

BI
PD
BD

E

BI Enhancer
PD Enhancer
BD Enhancer
Promoter

F

Processed from Micro-C Data

-50 kb

Promoter

50 kb

log2 Obr/Exp

BAF-independent
Enhancer

BAF-dependent
Enhancer

139
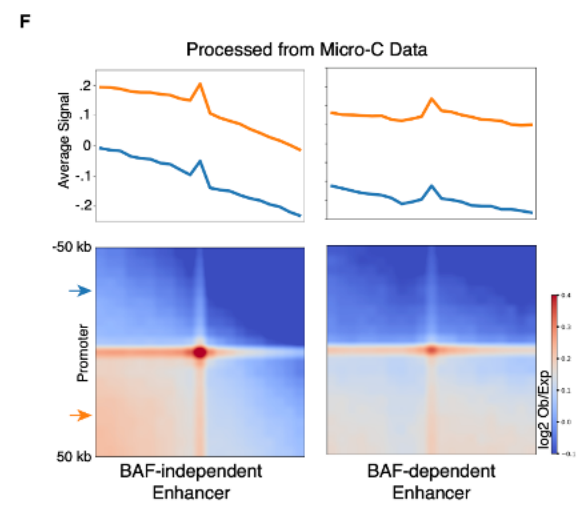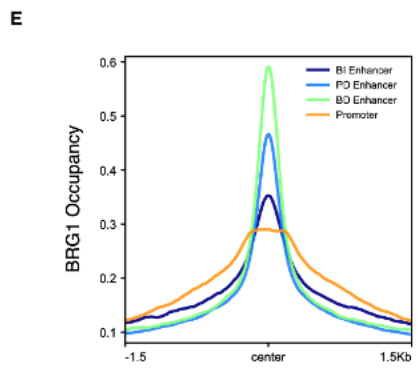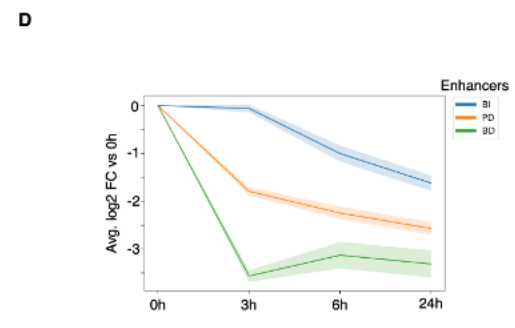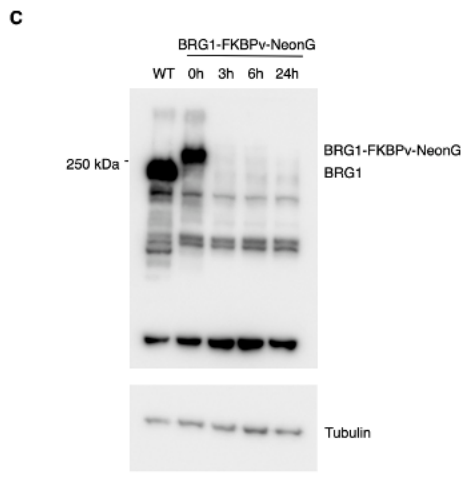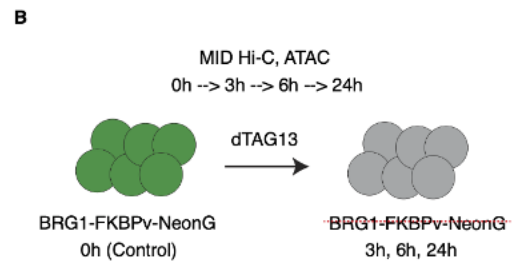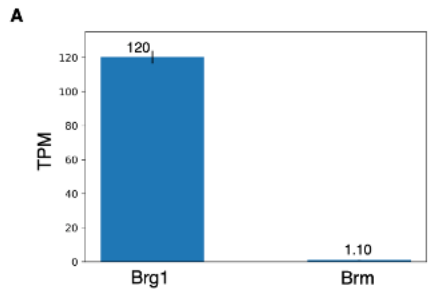
Figure AI.3. BAF complex orchestrates the assembly of E-P interaction neighborhoods. (A) PCA analysis on the profiles of E-P interaction strengths. (B) Pairwise Mann-Whitney U rank tests among time points (by column) for the E-P interaction subsets (by row). (C) APA at the E-P interaction subsets (by column) for time-course MID Hi-C datasets (by row). (D) Virtual 4C displaying the contact frequencies, at 0h and 3h, from the viewpoint (eye symbol at the enhancer). The yellow shade highlights the interacting promoter. (E) APA at the E-P interaction subsets (by column) for combined BRG1-depleted MID Hi-C dataset. Line profiles are generated as described in Figure 3.1.

A

PC2, 17%

- 0h
- 3h
- 6h
- 24h

PC1, 18%

B

0h v.s.

3 h   6 h   24 h   3 h v.s. 6 h   3 h v.s. 24 h   6 h v.s. 24 h

Independent

Partial Dependent

Promoter Interaction with

Dependent

Enhancer BAF Dependency

-log10 P

C

Enhancer BAF Dependency

Dependent   Partial-dependent   Independent

0 h

3 h

6 h

24 h

Promoter

-40 kb   40 kb

log2 Ob/Exp

-0.2   0.5

D

chr1 : 88,528,000 - 88,735,000

H3K4me3
H3K27ac
ATAC 0h
ATAC 3h
CTCF

Virtual 4C

Norm. Contact Freq.

0h

3h

Arl4c

E

Average Signal

.2
.1
0
-.1

Merge dTAG Samples

-40 kb   40 kb

log2 Ob/Exp

-0.2   0.5

Figure AI.4. BAF perturbation affects E-P interactions and compartmental segregations.

(A) Scatter plots showing the acute changes in E-P interaction strengths at 3h versus the acute changes of the anchors (top panel for promoters and bottom panel for enhancers). (B) Virtual 4C displaying the contact frequencies, at 0h and 3h, from the viewpoint (eye symbol at the enhancer). The yellow shade highlights the interacting promoter. (C) Heatmap showing the pairwise spearman correlations among the compartmentalizations for time-course MID Hi-C datasets. (D) Example tracks of the compartment scores. (E) Line plots displaying the acute fold changes (at 3h) in chromatin accessibilities versus the compartmental scores.

A

Avg. Promoter ATAC Diff (3 hr v.s. 0hr)

spearmanr=0.02

Avg. Enhancer ATAC Diff (3 hr v.s. 0hr)

spearmanr=-0.47

Enhancer-Promoter Interaction Diagonally Z-normed Signal Diff (3 hr v.s. 0hr)

B

chr8: 79,390,000 - 79,701,000

H3K4me3

H3K27ac

ATAC 0h

ATAC 3h

CTCF

Virtual 4C

Smad1   Otud4

C

Compartment Score Correlation

Brg1FV_0hr_B1
Brg1FV_0hr_B2
Brg1FV_6hr_B1
Brg1FV_24hr_B2
Brg1FV_6hr_B2
Brg1FV_24hr_B1
Brg1FV_3hr_B2
Brg1FV_3hr_B1

matrix_7

D

chr1: 10 - 100 MB

Brg1FV_0hr_B1
Brg1FV_0hr_B2
Brg1FV_3hr_B1
Brg1FV_3hr_B2
Brg1FV_6hr_B1
Brg1FV_6hr_B2
Brg1FV_24hr_B1
Brg1FV_24hr_B2

E

log2 ATAC FC

Compartment Score

- 3hr
- 3hr SE
- 6hr
- 6hr SE
- 24hr
- 24hr SE

143

Figure AI.5. BAF complex perturbation reduces compartmental segregation. (A) PCA analysis on the profiles of compartmental scores. (B) Saddle plot displaying the segregation pattern for the time-course MID Hi-C datasets. The summarized statistics (bottom left corner) were calculated from the difference of top 20% A-A and top 20% B-B interactions (intra-compartment, diagonal) versus A-B interactions (inter-compartment, off-diagonal) (Gibcus et al., 2018). (C) Example track of the compartmental segregation and epigenetic marks. (D) PCA analysis on the profiles of segregation scores. (E) Line plots displaying the acute fold changes (at 3h) in compartmental segregation versus the compartmental scores.

A Compartment Score PCA

B

C

chr3:32.0-36.5 MB

D Segregation Score PCA

E
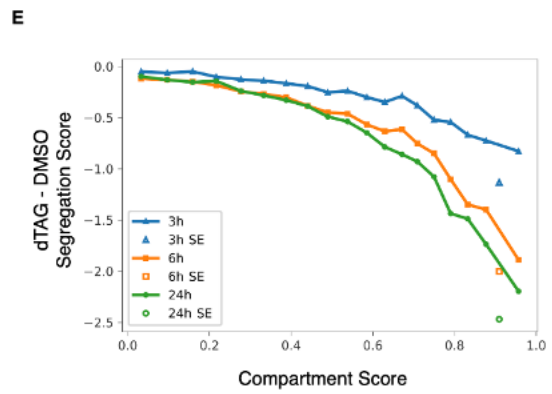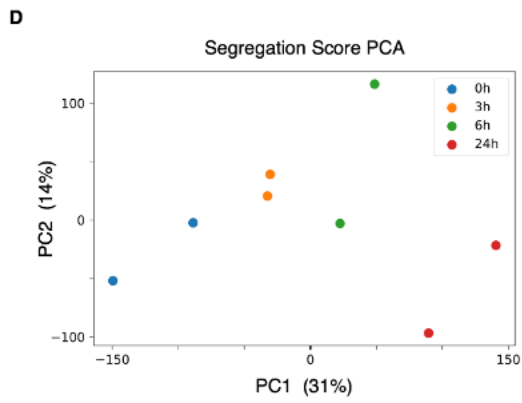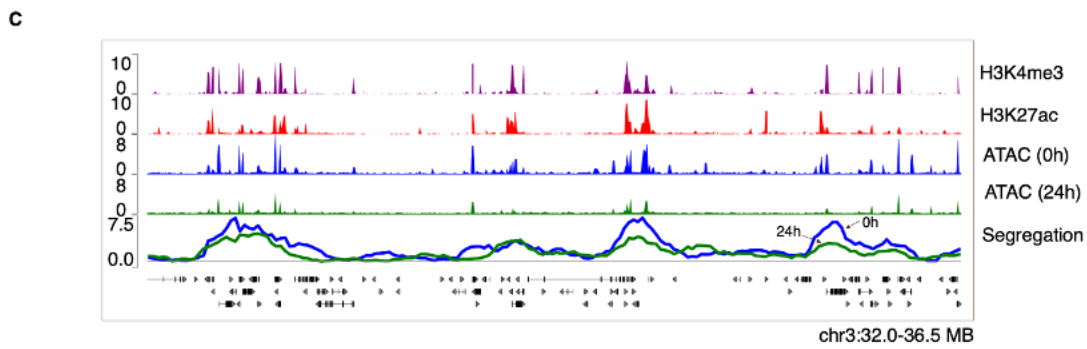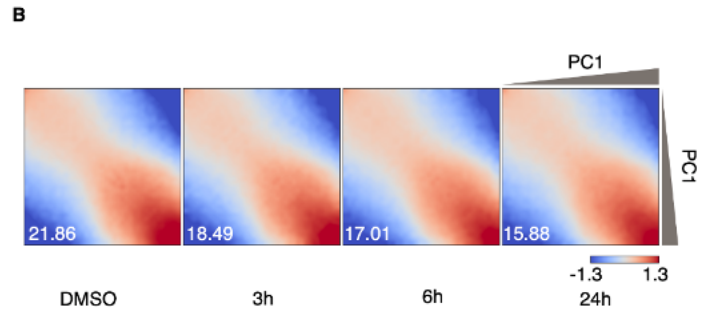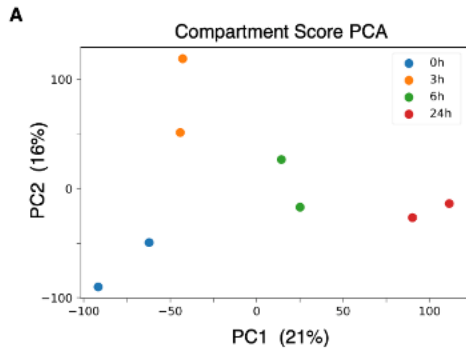
AI.4 References

Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester, H.E., Wang, W., Haswell, J.R., Park, P.J., and Roberts, C.W.M. (2017). The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. Nat Commun 8, 14648. https://doi.org/10.1038/ncomms14648.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell 168, 442-459.e20. https://doi.org/10.1016/j.cell.2016.12.016.

Gibcus, J.H., Samejima, K., Goloborodko, A., Samejima, I., Naumova, N., Nuebler, J., Kanemaki, M.T., Xie, L., Paulson, J.R., Earnshaw, W.C., et al. (2018). A pathway for mitotic chromosome formation. Science 359, eaao6135. https://doi.org/10.1126/science.aao6135.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell 38, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Lieberman-Aiden, E., Berkum, N.L. van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science 326, 289–293. https://doi.org/10.1126/science.1181369.

Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F., and Manke, T. (2020). pyGenomeTracks: reproducible plots for multivariate genomic data sets. Bioinformatics 37, btaa692-. https://doi.org/10.1093/bioinformatics/btaa692.

Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat Chem Biol 14, 431–441. https://doi.org/10.1038/s41589-018-0021-8.

Nakayama, R.T., Pulice, J.L., Valencia, A.M., McBride, M.J., McKenzie, Z.M., Gillespie, M.A., Ku, W.L., Teng, M., Cui, K., Williams, R.T., et al. (2017). SMARCB1 is required for

widespread BAF complex–mediated activation of enhancers and bivalent promoters. Nat Genet 49, 1613–1623. https://doi.org/10.1038/ng.3958.

Park, Y.-K., Lee, J.-E., Yan, Z., McKernan, K., O'Haren, T., Wang, W., Peng, W., and Ge, K. (2021). Interplay of BAF and MLL4 promotes cell type-specific enhancer activation. Nat Commun 12, 1630. https://doi.org/10.1038/s41467-021-21893-y.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 42, W187–W191. https://doi.org/10.1093/nar/gku365.

Stovner, E. B., & Sætrom, P. (2020). PyRanges: efficient comparison of genomic intervals in Python. *Bioinformatics*, *36*(3), 918-919. https://doi.org/10.1093/bioinformatics/btz615

Wang, X., Lee, R.S., Alver, B.H., Haswell, J.R., Wang, S., Mieczkowski, J., Drier, Y., Gillespie, S.M., Archer, T.C., Wu, J.N., et al. (2017). SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. Nat Genet 49, 289–295. https://doi.org/10.1038/ng.3746.

Zhu, Y., Denholtz, M., Lu, H., and Murre, C. (2020). Calcium signaling instructs NIPBL recruitment at active enhancers and promoters via distinct mechanisms to reconstruct genome compartmentalization. Gene Dev 35, 65–81. https://doi.org/10.1101/gad.343475.120.

Label-free assessment of pre-implantation embryo quality by the Fluorescence Lifetime Imaging Microscopy (FLIM)-phasor approach

<u>**Appendix II:**</u>

Label-free assessment of pre-implantation embryo quality by the Fluorescence Lifetime Imaging Microscopy (FLIM)-phasor approach

AII.1 Introduction

Development of quantitative, safe and rapid techniques for assessing embryo quality provides significant advances in Assisted Reproductive Technologies (ART). Instead of assessing the embryo quality by the standard morphologic evaluation, we apply the phasor-FLIM (Fluorescence Lifetime Imaging Microscopy) method to capture endogenous fluorescent biomarkers of pre-implantation embryos as a non-morphological caliber for embryo quality. Here, we identify, under hypoxic and non-hypoxic conditions, the unique spectroscopic trajectories at different stages of mouse pre-implantation development, which is referred to as the developmental, or "D-trajectory", that consists of fluorescence lifetime from different stages of mouse pre-implantation embryos. The D-trajectory correlates with intrinsic fluorescent species from a distinctive energy metabolism and oxidized lipids, as seen with Third Harmonic Generation (THG) that changes over time. In addition, we have defined a non-morphological Embryo Viability Index (EVI) to distinguish pre-implantation embryo quality using the Distance Analysis (DA), a machine learning algorithm to process the fluorescence lifetime distribution patterns. We show, under our experimental conditions, that the phasor-FLIM approach provides a much-needed non-invasive quantitative technology for identifying healthy embryos at the early compaction stage with 86% accuracy. The DA and phasor-FLIM method may provide the opportunity to improve implantation success rates for *in vitro* fertilization clinics.

Determining embryo quality during *in vitro* fertilization (IVF) is one of the most important steps toward successful pregnancy(*1*). The standard non-invasive method to assess embryo quality and viability relies on the visual inspection of embryo morphology according to predefined criteria such as cell division patterns, the number of pronucleoli in cleavage stages(*2, 3*), and the physical characteristics of the blastocyst(*4*). Assisted reproduction through morphological evaluation is labor intensive and highly dependent on the performance of individual physicians trained in these techniques. Development of more quantitative and objective means for assessing embryo quality that are simpler, safer, and faster could provide significant advantages in assisted reproduction by enabling single embryo transfers rather than the implantation of multiple embryos in order to increase the likelihood of a successful pregnancy.

Given the limitations of morphological evaluation, several technologies have been explored for the assessment of embryo viability. These include the measurement of metabolites in embryonic culture media, as well as genomic and proteomic profiling of the embryos themselves(*5*). For example, spectroscopic approaches have been utilized to measure the number of metabolites such as pyruvate, lactate, and glucose in the media during embryo culture (*6, 7*). However, these approaches are time-consuming and require highly-trained personnel to analyze the complex data(*8*). Both genomic and proteomic profiling are equally time consuming and can cause damage to the embryo during the procedure. Here, we apply the phasor-fluorescence lifetime imaging microscopy (FLIM) method and examine the dynamic endogenous biomarker (metabolites as described below) changes during preimplantation embryo development. Based on the quantifiable physiological property changes, we correlate the biomarker changes to the

embryo viability (Fig. 1). This non-invasive phasor-FLIM analysis is sensitive, quick and intuitive.

FLIM produces an image, based on the exponential decay rates at each pixel from a fluorescent sample. The fluorescence lifetime of the fluorophore signal is measured to create the image via FLIM(*9*) (Fig. S1A). When FLIM is coupled with two-photon excitation microscopy, molecules are excited at longer wavelengths (with lower energy photons). This prevents photodamage and allows deeper imaging, resulting in superior image quality(*10*). Since endogenous molecules such as collagen, retinoids, flavins, folate and NADH (nicotinamide adenine dinucleotide) are fluorescent in live cells(*11, 12*), we can collect fluorescence lifetime data to identify these intrinsic fluorescent species. The contributions from these different biochemical species are indicators of an embryo's biochemical property(*13, 14*). In our approach, we measure the fluorescent lifetime signal from integrated images acquired and transform the raw data using the Fourier transformation to the average arrival time of emitted photons in each pixel, represented by polar coordinates "g" and "s" in the transformation function(*12*) (Fig. 1C, Fig. S1A). This allows us to present the data in a two-dimensional graphical representation of the lifetime distributions, known as the phasor plot, for each pixel in the FLIM image (Fig. S1).

Here we have applied the phasor-FLIM approach to pre-implantation mouse embryos and have captured detailed data on their metabolic states at various developmental stages. At each stage, the mouse embryo displays a characteristic phasor-FLIM signature.

For the first time, we defined a unique graphical metabolic trajectory that correlates with energy metabolism and embryo development, which we call the developmental trajectory or "D-trajectory". Initially, embryos uptake pyruvate as their main energy source(*15*). As the embryos develop to later stages, the need for ATP increases in order to activate transcription for proliferation. Then, the embryos switch from glycolysis to oxidative phosphorylation, primarily using glucose as their energy source, which also changes the relative redox potential (NAD+: NADH ratio)(*16*). The spectroscopic signatures from each of these changes are detected and can be used as criteria to identify healthy embryos at each stage in development. We find that the D-trajectory of pre-implantation embryos cultured in nutrient-deficient media deviates significantly from that of the normal media, indicating that lifetime trajectories can be used to detect metabolic alterations in embryos. We have identified a combination of mathematical parameters that are statistically different between healthy and unhealthy pre-implantation embryos based on machine learning information. Therefore, the phasor-FLIM approach provides an objective, non-invasive, and quantitative method to assess the quality of mammalian embryos.

AII.2 Results

The lifetime D-trajectory of pre-implantation embryos

Two different mouse strains (a non-inbred CD1 and an inbred C57BL/6NCrl) were used to acquire a comprehensive representation of the phasor-FLIM distribution patterns of embryos during pre-implantation development (Fig. 2 and Fig. S2a). Fluorescent lifetimes of endogenous fluorescent species, excited at 740nm, were collected at the 2-cell (E1.5), morula (E2.5), compaction (E3.0), early blastocyst (E3.5) and blastocyst stage (E4.5), and pseudo-colored according to the phasor coordinates (Fig. 2a, b). The phasor coordinates, which is the averaged

fluorescent lifetime, of the 2-cell and morula stage embryos have a unique lifetime distribution

pattern distinct from all other cell and tissue types measured (blue arrow, Fig. 2b)(*11*). This

unique phasor lifetime position may reflect special characteristics of totipotent cells, which

mirror low oxygen consumption and preferential utilization of pyruvate oxidation(*17*). On the

other hand, compaction to blastocyst stages display average phasor coordinates typically

observed in pluripotent cells (red arrow, Fig. 2b)(*18, 19*). We refer to this characteristic

developmental time course lifetime distribution pattern as the developmental trajectory or

"D-trajectory". Phasor-FLIM lifetime distributions of individual embryos from both outbred and

inbred mouse strains (Fig. 2c, d) follow the similar developmental trend D-trajectory. In order to

examine whether genetic background of mice influences the D-trajectory, we compared the

trajectories of both CD1 and C57BL/6NCrl strains (Fig. 2c, d). While the average lifetimes (g

and s values) at specific embryonic stages are somewhat variable, the overall D-trajectory

distribution (blue and red arrows) of C57BL/6NCrl is similar to that of CD1 mice. We conclude

that the D-trajectory is a characteristic distribution behavior observed among pre-implantation

mouse embryos. In addition, we have applied time-lapse FLIM imaging to individual embryos

(n=16), and continuously followed at 3-hour time intervals from 2-cell (E1.5) to blastocyst stage

(E4.5) for approximately 60 hours. The *in vitro* developmental trajectory (Fig. S2b) of each

embryo mirrors the D-trajectory (Fig. 2b, with blue and red arrows). Lastly, we have compared

the phasor-FLIM developmental patterns between the pre-implantation embryos cultured under

ambient (20.9% oxygen) and hypoxia condition (12.8% oxygen, trigas of 5% O_2, 5% CO_2, 90%

N_2 mixed with atmosphere) (Fig. S2c-d). After 4 hours of incubation, the 2-cell (E1.5), morula

(E2.5), compaction (E3.0), early blastocyst (E3.5) and blastocyst stage (E4.5) embryos were

subjected to FLIM collection of endogenous fluorescent species, excited at 740nm. The D-

154

trajectories of embryos were similar between embryos grown under the ambient and hypoxic condition (Fig. S2d). Although we noticed slightly shifts towards the right for the hypoxic condition, presumably due to the higher glycolysis rate, the shifts for the s and g coordinates are not statistically significant. In sum, two combined lifetime trajectories (blue and red arrows) encompass the overall D-trajectory for normal pre-implantation embryo development.

Reactive oxygen species (ROS) plays a key role in cellular metabolism and homeostasis[20, 21] and ROS production has been linked to an increase in oxidized lipids[22]. The red arrow (the right to left-downward shift, Fig. 2b) in the D-trajectory is presumably due to an increasing fractional contribution of ROS as well as the oxidized lipids which have a fluorescence lifetime distribution of 7.89ns and fall on the same published location (coordinates) of the semi-circle in the phasor plot (Fig. S1b) (12). This behavior is consistent with the model that an increase in aerobic respiration and metabolism as well as b-oxidation during pre-implantation mouse development (12) requires more efficient energy production from oxidative phosphorylation (23, 24). We have confirmed the presence of active ROS production with fluorogenic marker 2', 7'-dichlorofluorescin diacetate (DCF-DA, also known as H_2DCFDA) staining (Fig. S3).

In order to better characterize the lipid droplets distribution during embryonic development, we have employed third-harmonic generation (THG) microscopy imaging (Fig. 3) with Deep Imaging Via Emission Recovery (DIVER) microscope (Fig. 3). The interfaces heterogeneity can be detected with the third order nonlinearity c^3. Given that the process is ultra-fast for structures with THG signals, the lifetime is approximately zero. Figure 3a shows the representative THG intensity images acquired in the same field of view as that of the FLIM images of Figure 3b. The

phasor plot of the THG images appears at the coordinate of $s = 0$ and $g = 1$. Movie S1 shows the 3D structure of the lipid droplets of embryos from different stages(*25*). Furthermore, we quantify the co-localization correlation of the long lifetime specie in the FLIM images (red) with the lipid droplets (green) in THG images (Fig. 3c, d). During embryonic development, the oxidized lipid signature, color-coded in red for the long lifetime species, (same direction as red arrow in Figure 2b) accumulated. The Mander's split co-localization correlation coefficients increase from 0.0099 to 0.3907 (where a coefficient of 1 is perfect correlation and 0 is complete lack of correlation) with embryonic development, suggesting that the phasor-FLIM distribution changes during these stages are due to increased lipid accumulation. We also characterized the lipid droplets distribution during embryonic development using the 3D THG image (Fig. 3a, e). Cleavage stage embryos have a large amount of small, densely packed lipid droplets, whereas post-cleavage stage embryos have large lipid droplets of the low density. The dramatic changes for both the lipid oxidation and lipid volume size start after compaction stages. These findings demonstrate that the dynamic difference in lipid oxidation can be detected by phasor-FLIM.

Fluorescence lifetime trajectories reveal metabolic states of pre-implantation mouse embryos

The D-trajectory is complex because it is composed of lifetimes from various endogenous fluorescent biochemical species. We first hypothesized that the major component responsible for the shifts in the D-trajectory is intracellular NADH changes based on its fundamental role in energy production during embryogenesis. To test this hypothesis, we first measured the metabolic activity of intracellular NADH(*26*). The bound form of NADH is linked to energy production through oxidative phosphorylation, whereas the free form of NADH is associated

with glycolysis(*27*). The phasor coordinates of free NADH maps on the right side of the plot

with a lifetime of 0.38 ns and the protein bound form of NADH (bound with lactate

dehydrogenase) maps on the left at 3.4ns (Fig. S1b). This lifetime distribution of the free and

bound forms of NADH in the phasor plot was previously described as the metabolic or M-

trajectory(*11*).

Next, embryos were treated with known biochemical inhibitors of oxidative phosphorylation and

glycolysis(*27*). Oxidative phosphorylation was inhibited at the early compaction stage with a

cocktail of rotenone and antimycin A (R&A) (500nM) by inhibiting complex I and complex III

of the electron transport chain. Embryos were imaged after a 4-hour culture period (Fig. 4a). The

FLIM images showed increased fractional contributions of free NADH (shorter lifetimes) when

compared to controls (Fig. 4a). This shift towards glycolytic metabolism is seen in a dose-

dependent manner (Fig. S4), indicating that embryos cultured in R&A have decreased oxidative

phosphorylation activities (Fig. 4a, b). We also cultured the early blastocyst stage embryos in

1mM 2-Deoxy-D-Glucose (2DeoxyG), an analog of glucose, to inhibit glycolysis (Fig. 4c). The

glucose analog treatment shifted the phasor-FLIM distribution to longer lifetime (an increase of

bound NADH) (Fig. 4c), which correlates with a decrease in glycolysis (Fig. 4c, d). These

findings suggest that the source of the changes seen in the phasor coordinates throughout the pre-

implantation stages in the D-trajectory is in part due to the contribution from metabolic shifts of

NADH.

FLIM does not disrupt embryonic development under 10mW laser power

157

In order to ensure the safety of the FLIM imaged embryos, we determined the optimum laser power to avoid DNA damages (*28*), while allowing the rapid and robust acquisition of the FLIM signal on mouse pre-implantation embryos. We exposed 2-cell (E1.5) and morula (E2.5) stage CD1 and C57/BL6NCrl embryos to varying laser powers (1.5, 3.5, 10, and 15mW) and examined the effect on the developmental progression of embryos until the blastocyst stage (Fig. S5, 6). In order to capture FLIM-signals of embryos taken with 1.5mW laser power, 4 times longer exposure time was required than the embryos collected at 3.5mW, 10mW, 15mW laser powers due to their low signal to noise ratio. The majority of embryos exposed to 1.5 mW and 3.5 mW laser power developed to the blastocyst and there were no significant differences between the control (non-imaged) and embryos imaged at the 2-4cell stage or morula-compaction stage, irrespective of strain differences (CD1 or C57BL/6NCrl) (Fig. S5, 6). However, at 10mW, approximately 20% and 35% of CD1 embryos imaged at the 2-cell and compaction stages, respectively, fail to progress to the blastocysts. At 15mW, nearly 50% of CD1 and C57BL/6NCrl embryos imaged at the 2-cell stage were arrested before the compaction stage, while approximately 30% of CD1 and 12% of C57/BL6NCrl embryos imaged at the compaction stage failed to develop to blastocysts. We conclude that CD1 embryos are more sensitive to the laser damage, and 3.5mW is the ideal laser power for our FLIM analysis.

Next, we examined the activation of the DNA repair pathway in the embryo by conducting immunofluorescence staining for anti-phosphorylated Histone 2AX (H2AXs139), a novel marker for DNA-double strand breaks (*29, 30*). Both the non-imaged control and FLIM-imaged embryos were indistinguishable and did not show any signs of DNA repair pathway activation at 3.5mW

(Fig. S7a). However, embryos exposed to 1.5mW laser power, which required longer laser

exposure time (12 minutes, instead of ~3 minutes) showed the sign of DNA damage (Fig. S7a).

Next, we examined the effects of FLIM on the rate of pregnancy. Specifically, FLIM-imaged and

control BRE-gal embryos (reporter mouse line responding to the endogenous levels of BMP

signaling during development) *(41)* at E2.5 were allowed to early blastocyst stage (E3.5) and

control and FLIM-imaged embryos were implanted into females. E18.5 embryos were collected

through caesarean section (C-section) and counted for the implantation efficiency (Fig. S7b, and

Supplementary Table 1). The average live birth rates were 49% for FLIM-imaged group and

43% for the non-imaged BRE-gal control group based on three independent experiments

(Supplementary Table 1). Student t-test reveals that there is no significant statistical difference

between FLIM-imaged and non-imaged group. We conclude that FLIM imaging of the morula

stage embryo at 3.5mW excitation is safe to use and employed in the subsequent experiments.

FLIM distinguishes pre-implantation embryos under stress conditions

Given that early cleave stage embryos utilize aspartate, pyruvate, and lactate for energy

metabolism (*31*) we tested whether the unique lifetime distribution patterns of an embryo

cultured under altered physiological states can be detected by the changes in spectroscopic

distributions of phasor-FLIM.

We cultured 2-cell and morula stage embryos in standard mouse embryo culture

media (KSOMaa), flushing and holding media (FHM: DMEM-pyruvate free with

HEPES), and saline solution (PBS). Brightfield images and FLIM data were

159

collected at 4 hours and 24 hours after the treatment (Fig. 5). The FLIM data were collected once at the first time-point (4 hours). The 2-cell stage embryos cultured under KSOMaa, FHM and PBS were morphologically normal (Fig. 5a, top). However, the embryos in high-stress conditions (FHM and PBS) show distinct lifetime distribution patterns on the phasor-plot when compared to that of KSOMaa cultured embryos (Fig. 5b, c, Fig S8a). Subsequently, we find that the embryos under high-stress conditions fail to cleave normally and remain at the 2-cell stage, unlike KSOMaa controls (Fig. 5b, Fig. S8a). We performed the similar analysis using compaction stage embryos and found that within a few hours under high-stress culture conditions, the phasor-FLIM lifetime trajectories of embryos deviate from those cultured in KSOMaa even before the embryos show any signs of abnormal morphology (Fig. 5d-f, Fig S8b). The cell division in FHM and PBS cultured embryos also slowed down significantly (Fig. S8b). We conclude that phasor-FLIM is a sensitive method to detect the changes in embryo metabolism upon cellular stress.

Derivation of the embryo viability index (EVI) for assessing the developmental potential of the pre-implantation embryo

The phasor distribution analysis of pre-implantation mouse embryos allows us to distinguish between normal and highly stressed embryos (Fig. 5). Therefore, we determined whether the developmental potential of pre-implantation embryos is predictable through phasor-FLIM analysis. We first performed time-lapse phasor-FLIM imaging of embryos from the 2-cell stage for ~60 hours to identify the most desirable stage to predict the developmental potential of embryos (Fig. 6a, Movie S2). At the end of the 60-hour culture period, we classified embryos as healthy (H) if they reached the normal full expanded blastocyst stage showing a tightly packed ICM and cohesive epithelium shaped TE cells, or not healthy (NH) if embryos were arrested

before reaching the blastocyst stage or displaying abnormal blastocyst morphology (Fig. 6a). We then applied the distance analysis (DA) algorithm (*32*) to identify key spectroscopic parameters that could differentiate healthy (H) from unhealthy (UH) embryos by machine learning. Using the DA algorithm, the 3D phasor histogram was separated into 4 sections based on the phasor coordinates (g, s) intensity, from which, 6 parameters were extracted from each section, generating a total of 24 parameters (see Methods). The healthy embryos (H group) were used as the control set and the unhealthy embryos (UH group) were used as the sample set. Each of these sets included images from multiple embryos from each stage in development. Next, we calculated the average and variance of the training set, which includes two groups (H and UH), and weighted 20 parameters (g, s, the secondary moment a, b and angle from 4 sub-layers, intensity excluded) in each set from 3D phasor plot. After optimizing the weights to maximize the difference between unhealthy and healthy group embryos, we applied these weights to index a new score called the EVI or Embryo Viability Index (Methods). This partition metric defines the degree of separation of the test embryos from the average of the training set where -1 to -10 are unhealthy embryos, and +1 to +10 are healthy embryos.

Next, we examined the DA data from 2-cell, 4-cell, and the early compaction stage to determine the best binary classification model using receiver operating characteristic (ROC) curves (Fig. 6b, c, Fig. S9a, b). We have classified the embryos predicted to be healthy in positive values (EVI<0, in blue), and embryos predicted to be unhealthy in negative values (EVI>0, in red). The plot of true positive rates against false positive rates gives an area under the ROC curve (AUC) for 2-cell, 4-cell, and the early compaction stage embryos, which were 0.739, 0.728, and 0.916, respectively. We conclude that the spectroscopic characteristics of the early compaction stage

embryos (prediction accuracy with the highest AUC) possess the best parameters for separating embryos that will develop into normal blastocysts (Fig. 6b, c, Fig. S9a, b).

An embryo viability prediction pipeline was developed based on the DA of phasor-FLIM images of the early compaction stage embryos (Fig. 6d). We have FLIM imaged embryos at the early compaction stage and all of these embryos were allowed to develop to the blastocyst equivalent stage. The resulting embryos were classified as H or UH. We then selected a small number of healthy (H) and unhealthy (UH) embryos and obtained an EVI training data set. The remaining unselected embryos were also subjected to the DA program as "unknowns" (test set) to test the predictability of EVI. In experiment 1, we followed the development of 35 morphologically healthy looking early compaction stage embryos (pooled from 4 mating pairs), until the blastocyst stage (Fig. 6e, f). Of the 34 embryos, 18 developed to normal blastocysts and thus assigned as healthy (H), and 16 embryos that failed to reach the blastocyst were assigned as unhealthy (UH). When we applied EVIs that were determined by the training set, 83.3% of healthy embryos (15 out of 18 embryos) and 75.0% of unhealthy embryos (12 out of 16 embryos) were correctly predicted by EVI (Fig. 6e, f). Subsequently, we performed another 4 biologically independent experiments using a total of 134 embryos and the results are shown in Supplementary Table 2 and Supplementary Figure 6c, d. We achieved 85.9% accuracy (n=134) where a total of 88.5% healthy embryos (n=96) and 73.7% unhealthy embryos (n=38) were identified. Based on the results, we conclude that the DA program is able to predict the development potential of pre-implantation embryos at the early compaction stage.

AII.3 Discussion

Here we report that the phasor-FLIM represents a promising new approach for assessing the quality of pre-implantation mouse embryos. First, we have applied the phasor-FLIM analysis to capture developmental states during pre-implantation development. The spectroscopic trajectory, which we are calling the "D-trajectory" (D for development), is attributed to a combination of metabolic fluorescent species and production of ROS in conjunction with oxidized lipid metabolism within the embryo (Fig. 2c, d), and this trajectory correlates well with other measurements of embryonic development. Second, the intrinsic lifetime trajectory of pre-implantation embryos cultured in nutrient-deficient media deviates from the normal lifetime distribution, indicating that the lifetime trajectory can be used to detect metabolic changes in embryos. Third, we have applied the DA program that uses spectroscopic parameters from 3D phasor histograms of embryos and shown that EVI is a non-morphological, quantitative index that can provide useful information on the quality of pre-implantation embryos.

Other spectroscopic technologies have emerged as a non-invasive means of revealing embryo viability via detection of various metabolic states of common molecules associated with embryo development. Raman, near-infrared, Nuclear Magnetic Resonance (NMR), and Fourier-transform infrared spectroscopy can also detect the metabolic states of pyruvate, lactate, glucose, and oxygen during pre-implantation mammalian development (*33-35*). However, at present time these technologies suffer from a number of shortcomings. It is challenging for these approaches to analyze the data in the short time window needed for the host transfer of embryos. The data analyses are technically demanding and may not be intuitively obvious for the general clinical use. The technologies require fluid samples collected from the embryo culture media and the data are inherently noisier. Nonetheless, in the future, with improvements, these spectroscopic

approaches are likely to provide physicochemical parameters that will be useful in quantitating the influence of ovulation induction, oocyte retrieval, and *in vitro* culture procedures.

Development of qualitative and objective means for assessing embryo quality and viability that are safer and faster will provide significant advances in IVF and animal breeding facilities. If phasor-FLIM is to be applied for diagnostic purposes, it will be crucial to establish that the procedure does not perturb gene expression after the procedure. To date, embryos subjected to phasor-FLIM analysis appeared to be morphologically normal, and we did not detect signs of apoptosis or aberrations in nuclear morphology. However, it is possible that the phasor-FLIM procedure causes other alterations that cannot be easily detected by these morphological criteria. In the future, it will be important to perform additional molecular characterizations (i.e., DNA sequencing) to eliminate the possibility. In addition, our future experiments will include the assessment of implantation efficacy of these indexed embryos. Overall, this work has the potential to improve our understanding of energy metabolism in developing mammalian embryos and advance the ART field directly.

AII.4 Methods:

Animals: Animals were treated according to standards set by UC Irvine's University Laboratory Animal Resources (ULAR). CD1 and C57BL/6NCrl females were purchased from Charles River Laboratories. All animal procedures were performed with strict adherence to National Institutes of Health office of laboratory animal welfare (NIH OLAW) and Institutional Animal Care and Use Committee (IACUC) guidelines.

Ethics Statement: Mice used for these experiments were used in accordance with the regulations overseen by the University of California Irvine Institutional Animal Care and Use Committee (IACUC) who assures that the use of live, vertebrate animals in research, testing, teaching or related activities is scientifically justified in accordance to Federal regulations and accreditation standards. All the techniques and procedures in this project have been refined to provide for maximum comfort and minimal stress to the animals. We confirm all experiments were performed in accordance to the guidelines and regulations by the protocol animal welfare assurance number approved by University of California Irvine Institutional Animal Care and Use Committee: A3416-01.UCI has been accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care, International (AAALAC) since 1971.

Pre-implantation mouse embryo collection: Females at 21-24 days old were superovulated with pregnant mare serum gonadotropin (PMSG, Sigma) and 48 hours later with human chorionic gonadotropin (hCG, Sigma). Matings were set each evening after hCG injections. The following morning a vaginal plug was considered 0.5 days post fertilization and embryos were collected at desired stages by flushing oviducts or uterine horns. For our time course collection (intrinsic fluorescence FLIM and THG measurements) superovulation and matings were staggered and all the embryos were collected the same day except late blastocysts (E4.5) were generated by dissecting at E3.5 (one day before imaging) and cultured till next day.

Embryo culture: Embryos were cultured at 12.8% for hypoxia condition or 20.9% O2 (measured using Neofox oxygen sensor), with 5% CO2 in nitrogen balance at 37°C. The drop size used was on average ~10 embryos/20 µl drop (1 drop per dish) except for the prediction test, where the

drop size was on average 1 embryo/3µl drop (~10 drops/dish). Embryos were cultured and

imaged on Matek Glassbottom Dishes (P35G-1.5-14-C). Single embryo cultures were used for

embryo viability prediction (of 1 embryo per 3 µl of KSOMaa) to prevent the mobility of

embryos and provide stable environment, to avoid ROS accumulation or influence of

neighboring embryos, and to create a library for prediction and embryonic developmental

potential. All other experiments were performed on group cultures ~10 embryos/20 µl drop (1

drop per dish).

Fluorescence lifetime imaging microscopy (FLIM): Fluorescence lifetime images of the pre-

implantation embryos were acquired on Zeiss LSM710 (Carl Zeiss, Jena, Germany), a multi-

photon microscope coupled with a Ti: Sapphire laser (Spectra-Physics Mai Tai, Mountain View,

CA) with 80 MHz repetition rate. The FLIM data detection was performed by the

photomultiplier tube (H7422p-40, Hamamatsu, Japan) and a320 FastFLIM FLIMbox (ISS,

Champaign, IL). The pre-implantation mouse embryos were excited at 740nm; an average power

of ~3.5 mW was used as previously in live cells and tissue(*36*). A Zeiss EC Plan-Neofluar

20x/0.5 NA objective (Cart Zeiss, Jena, Germany) was used. The following settings were used

for the FLIM data collection: image size of 256x256 pixels, scan speed of 25.21µs/pixel. A

dichroic filter at 690nm was used to separate the fluorescence signal from the laser light. And the

emission signal is split with 496nm LP filter and detected in two channels using a band pass

filter 460/80 and a 540/50 filter. Every FLIM image was acquired for 50 frames of the same field

of view with 256X256 per frame. Only the blue channel (460/80) data was used for this study.

FLIM calibration of the system was performed by measuring the known lifetime of a fluorophore

coumarin 6 (dissolved in ethanol), which has a known fluorescence lifetime of τ=2.5 ns(*37, 38*).

Embryos were kept in standard culture conditions, 37 °C and at 5% CO_2. FLIM data were acquired and processed by the SimFCS software developed at the Laboratory of Fluorescence Dynamics (LFD).

Converting FLIM data onto phasor coordinates: All FLIM images are transformed onto the phasor plot by the equations below. The g and s coordinates are generated from the fluorescence intensity decay of each pixel in the FLIM image using the following Fourier transformation equations (Fig. S1A):

$$g_i(\omega) = \int_0^\infty I(t)\cos(\omega t)dt \Big/ \int_0^\infty I(t)dt$$
$$s_i(\omega) = \int_0^\infty I(t)\sin(\omega t)dt \Big/ \int_0^\infty I(t)dt$$

Thus, the phasor approach is a fit-free analysis of FLIM imaging, and the g and s coordinates represent the decay curve at each pixel of the image. Therefore, a phasor analysis transforms complicated spectrum and decay of every single pixel into a unique position on the phasor plot.

Third Harmonic Imaging: The third harmonic generation images and the associated FLIM images of the same field of view were collected using the homebuilt Deep Imaging via Enhanced-Photon Recovery (DIVER) microscope. DIVER microscope is an upright laser scanning microscope, the unique feature is the application of wide photocathode area detector which allows collection of photons from a wide area and angle for high efficiency. The third harmonic generation images and intrinsic fluorescence FLIM images were collected using 40x water immersion objective (Olympus Plan Apo) with 1040nm and 740nm excitation respectively. And UG11 and Blue5543 filters were used for THG and endogenous fluorescence

FLIM images collection. An a320 FastFLIM FLIMbox (ISS, Champaign, IL) was used to transfer the data to the phasor plot. Rho110 was used for calibration with known lifetime τ=4 ns(*38*).

Inhibition of oxidative phosphorylation and glycolysis: Embryos were placed in 25µl microdroplets of KSOMaa (Invitrogen) with the appropriate inhibitors covered in mineral oil (Sigma). Both of the two chemical inhibitors, rotenone and antimycin A cocktail (R&A) and 2-Deoxyglucose (2DeoxyG) were dissolved in KSOMaa. For R&A, we prepare the inhibitor to perform dose dependence measurements for a final concentration of 100nM and 500nM. For 2DeoxyG the inhibitor has a final concentration of 1mM. KSOMaa was used as a solvent and culture media for the control group and treatment group embryos.

H2AXs139 staining: CD1 and C57BL/6NCrl post-imaged embryos are rinsed with Tyrode's acid (Sigma) 3 times and placed in holding and flushing media for 5 minutes to allow embryos to acclimate before 30-minute fixation in 4% paraformaldehyde on ice. Embryos were permeabilized using 0.2% Triton X-100 (Fisher). And then embryos were incubated with H2AXs139 (Genetex) at 1:1000 for 1 hour at room temperature. Embryos were rinsed in 1X PBT three times and then stained with AlexaFluor555 at 1:200. Embryos were rinsed in 1X PBS three times before processing for the Hoechst (Sigma) staining for 10minutes to stain the DNA. Finally, embryos were rinsed and imaged in 1X PBS using 780 Zeiss microscope and Zen 2012 software.

Embryo implantation and C-section at E18.5: CD-1 female mice were mated with vasectomized males to generate pseudopregnant females timed to E3.5 for implantation. E2.5 embryos were collected and imaged, and implanted at E3.5. In each experiment, embryos were randomized before imaging to non-imaged and FLIM-imaged groups. After imaging, non-imaged embryos and FLIM-imaged embryos were randomized. The technician transferring the embryos was blinded to which embryos were imaged or non-imaged. Twelve to sixteen embryos were implanted into the left and right uterine horn of pseudopregnant females. E18.5 embryos were collected through C-section and counted for the implantation efficiency. Genotyping was done for experiments that used BRE-gal$^{+/-}$ embryos to differentiate between WT embryos to BRE-gal$^{+/-}$ embryos. Embryos were genotyped with Tissue Direct Phire PCR Kit with the following primers: (LacZ band) Fwd: 5' ATG AGC GTG GTG GTT ATG C 3' Rev: 5' GAT GGT TCG GAT AAT GCG 3' (Hprt band) Fwd: 5' AAG CCT AAG ATG AGC GCA AG 3' Rev: 5' AAG CGA CAA TCT ACC AGA GG 3'

DCF-DA Staining: Embryos are rinsed in Acid Tyrode 3x, washed in KSOMaa 3x, transferred to 5 uM DCF-DA in 1X PBS. Embryos were incubated in DCF-DA solution for 25 min at 5% CO_2 and 37 °C. Embryos were then transferred to Hoechst stain solution for 8 min. Then embryos were placed in KSOMaa and imaged with LSM780 at 5% CO_2 at 37 °C.

Antibody Staining Image analysis for cell number calculation: We used a 3D segmentation pipeline (as previously described)(*39*) to do a 3D reconstruction of embryos and conduct cell number analysis.

Time-lapse FLIM imaging: Bright-field time-lapse images and FLIM data of *in vitro* cultured embryos were collected every 4 hours for a period of approximately 60 hours starting from the 2-cell stage (E1.5) until the blastocyst stage (E4.5). All FLIM images were collected using the Zeiss LSM710 confocal microscope within stage incubator to obtain the normal *in vitro* culture conditions (37 °C, 5% CO_2) (Movie S1).

Distance Analysis Program: The FLIM data collected from individual embryos are placed in either of two categories the H (control group has FLIM signature from the embryos developed to the blastocyst stage) and UH (sample group has FLIM signature from the embryos arrested at compaction stage or even earlier). The distance algorithm can generate a "spectra" from the given (up to 24 parameters) of phasor FLIM distributions corresponding to individual embryos. The 24 parameters are the 2 coordinates for the center of mass g and s, 2 second axial moments a and b after diagonalization, the angle of the distribution from the diagonalization and the total number of pixels in the phasor plot from the 4 slices of the 3D phasor histogram. For each parameter set, we calculate the average of the parameters and the standard deviation. Then we construct a function that we call "distance" in which we calculate the difference of the average of the two sets weighted by the variance of the parameter in each set for the group H and UH respectively.

Using distance analysis, a training set can be generated based on the best weight set that has been chosen to separate the H and UH set embryos according to the distance from the average of each set. Finally, after the training set has been generated, the rest of the embryos were tested, and an embryo viability index (EVI) is calculated for each embryo. Using the EVI index for the spectra

of the training set, we can build the histogram and determine if a member is a true positive (below 0) or a false positive (above 0). Statistical methods such as the area under the curve (AUC) are then used to determine the quality of the training set. If the AUC is close to one, the two groups are more separable since there are less false positives. The Distance approach has been used previously to determine the separation of spectra in human cancer tissues(*32*). More details of the distance analysis calculation can be found in Dr. Ranjit's recent publication (LFD)(*32*).

Statistical analysis

Data are presented as mean ± standard deviation. For the FLIM data, the statistical analyses were performed using student t-test for the g value only, p<0.05 was considered as statistically significant.

The box-whisker plot showing the prediction ability represents the median ± min/max from each indicated group (Training set H and UH group, Tested H and UH group).

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Contributions: M.A.D and S.N. initiated the phasor FLIM on embryo experiment. N.M. and S.N., M.A.D., K.C. designed experiments for tracking embryonic development, metabolic pathway manipulation test, photo-toxicity test, and embryonic viability diagnosis experiments. N.M. and M.A.D. designed the THG experiment. S.N., P.P., and T.Y. handled the mice and retrieved the pre-implantation embryos. N.M., S.N., P.P., and T.Y. did image analysis. N.M. did FLIM data acquisition. N.M. and P.P designed and implemented the implantation assay and hypoxia measurements. S.N. and N.M. wrote the first draft of the manuscript, and all authors contributed to subsequent revisions.

Competing Interests: The authors declare no competing interests.

Figure 1: Schematic of the workflow of the experimental design. (a) We collected FLIM images of embryos from superovulated female mice at the following developmental stages: 2-cell, morula, compaction, early blastocyst, and blastocyst. **(b)** Intrinsic fluorescence lifetimes for each embryo are collected using a Zeiss 710 microscope coupled with a FLIM-box. **(c)** The

FLIM data analysis of the pre-implantation mouse embryo development was performed using the

phasor approach. **(d)** Distance Analysis (DA) program was applied to predict embryo viability.
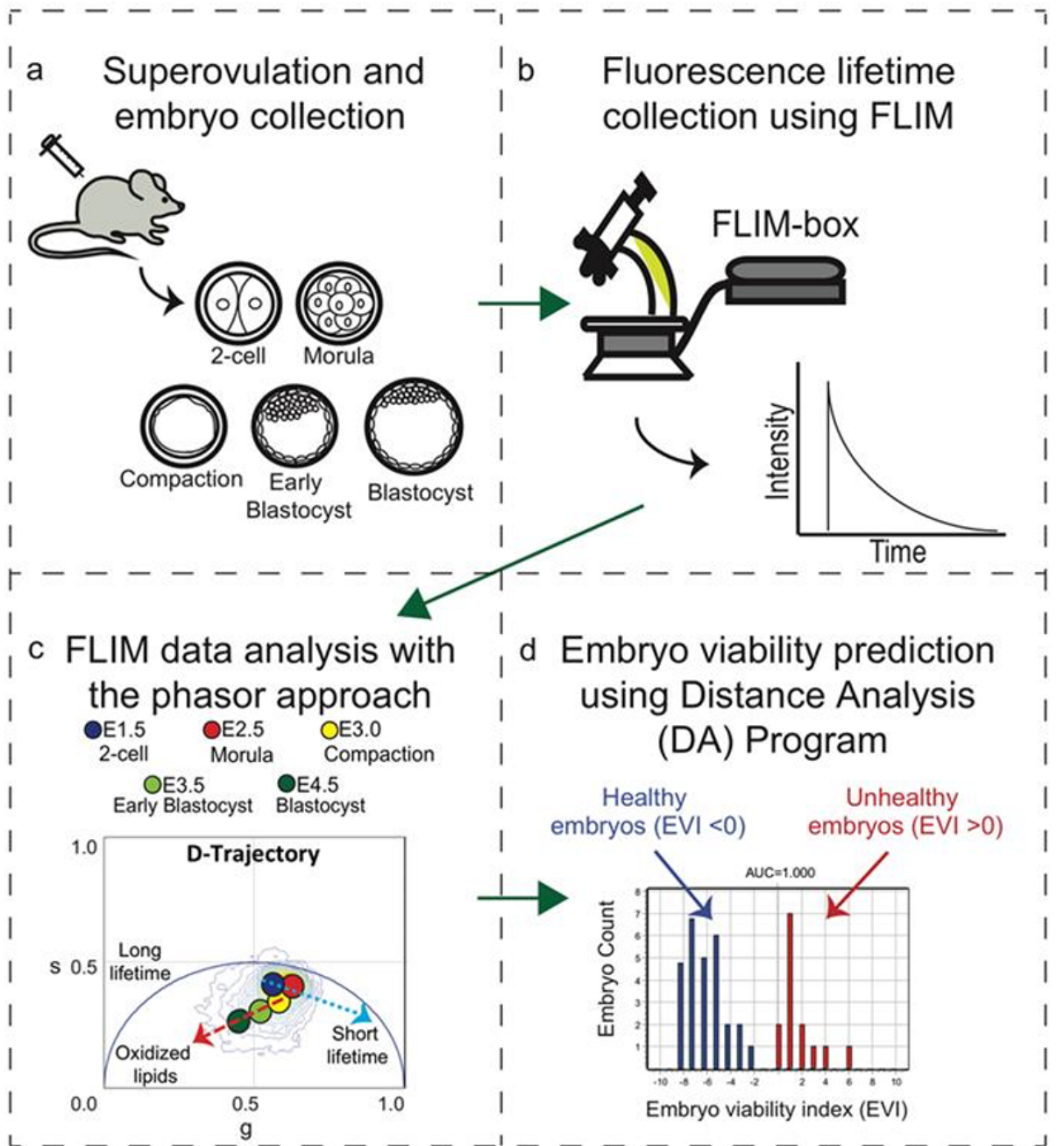
Figure 2: The lifetime trajectory of pre-implantation embryos correlates with embryonic development. (a) Transmission (top row), fluorescence intensity (middle row, 740nm excitation) and FLIM (bottom row) images of representative pre-implantation CD1 mouse embryos at 2-cell (E1.5), morula (E2.5), compaction (E3.0), early blastocyst (E3.5), and blastocyst stage (E4.5). In the FLIM images, the pseudo color displays the fluorescence lifetime. **(b)** Phasor-plot of average fluorescence lifetime of CD1 embryos at the indicated developmental stages demonstrating the D-trajectory (D for development). A blue arrow indicates the fluorescence lifetime change from E1.5 to E2.5 and a red arrow shows the change from E3.0 to E4.5. **(c-d)** Scatter plots show the D-trajectory for CD1 and C57BL/6NCrl embryos. The small window shows the average and standard deviation of each stage. CD1: 2-cell (n=29), morula (n=11), compaction (n=33), early blastocyst (n=50) and blastocyst stage (n=35); C57BL/6NCrl: 2-cell (n=25), morula (n=22), compaction (n=21), early blastocyst (n=38) and blastocyst stage (n=42). c) D-trajectory of CD1 embryos (2-cell, n=8; morula, n=8; compaction, n=12; early blastocyst, n=5; blastocyst, n=8. and d) D-trajectory of C57BL/6NCrl embryos (2-cell, n=7; morula, n=3; compaction, n=17, early blastocyst, n=8; blastocyst, n=21). N= number of embryos analyzed.
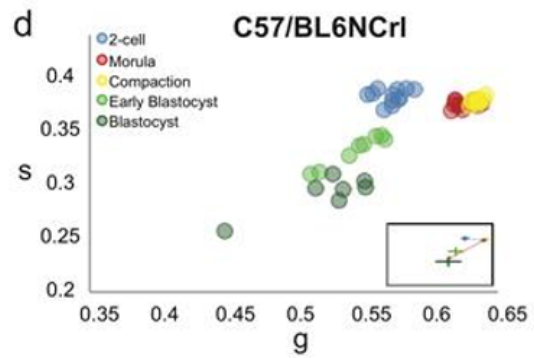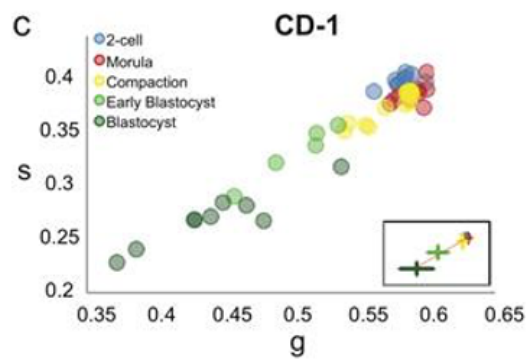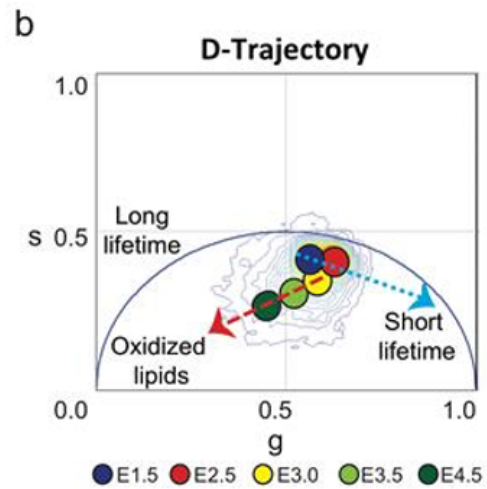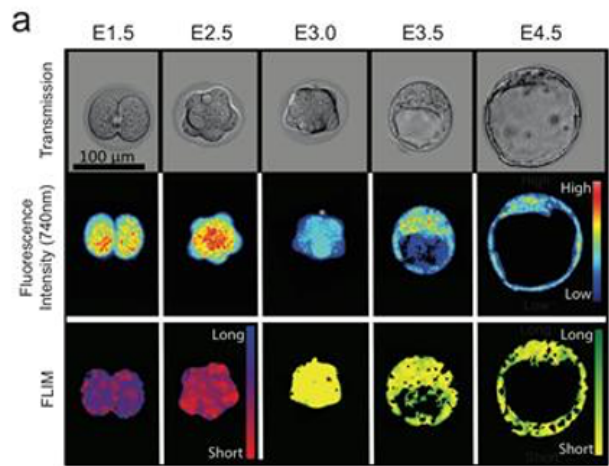
Figure 3: THG and intrinsic fluoresce signal show increasing oxidized lipids during embryonic development.
(a) Representative third harmonic generation images and **(b)** FLIM images during pre-implantation embryonic development, 2-cell, morula, compaction, early blastocyst and blastocyst stage for the same field of view. From blue to red shows the intensity increase. **(d)** Representative optical sections show co-localization (yellow) of the lipid droplet signal (green) in THG images with long lifetime species (red) in FLIM images. Scale bar sets at 100μm. **(d)** Mander's coefficient of the co-localization results during embryo development which shows the proportion co-localization region of the THG channel and FLIM channel correspondence with long lifetime species-oxidized lipids. 2-cell (n=5), morula (n=3), compaction (n=3), early blastocyst (n=4) and blastocyst stage (n=3). Student t-test results (p-value) for morula to 2-cell, compaction, early blastocyst and blastocyst are 0.1923, 0.0823, 0.0091, and 0.0174 respectively. e) Lipid droplets volume characterization during pre-implantation embryo development. 2-cell (n=5), morula (n=5), compaction (n=4), early blastocyst (n=4) and blastocyst stage (n=6). Student t-test results (p-value) for morula to 2-cell, compaction, early blastocyst and blastocyst are 0.5066, 0.6367, 0.1416, and 0.0072 respectively.
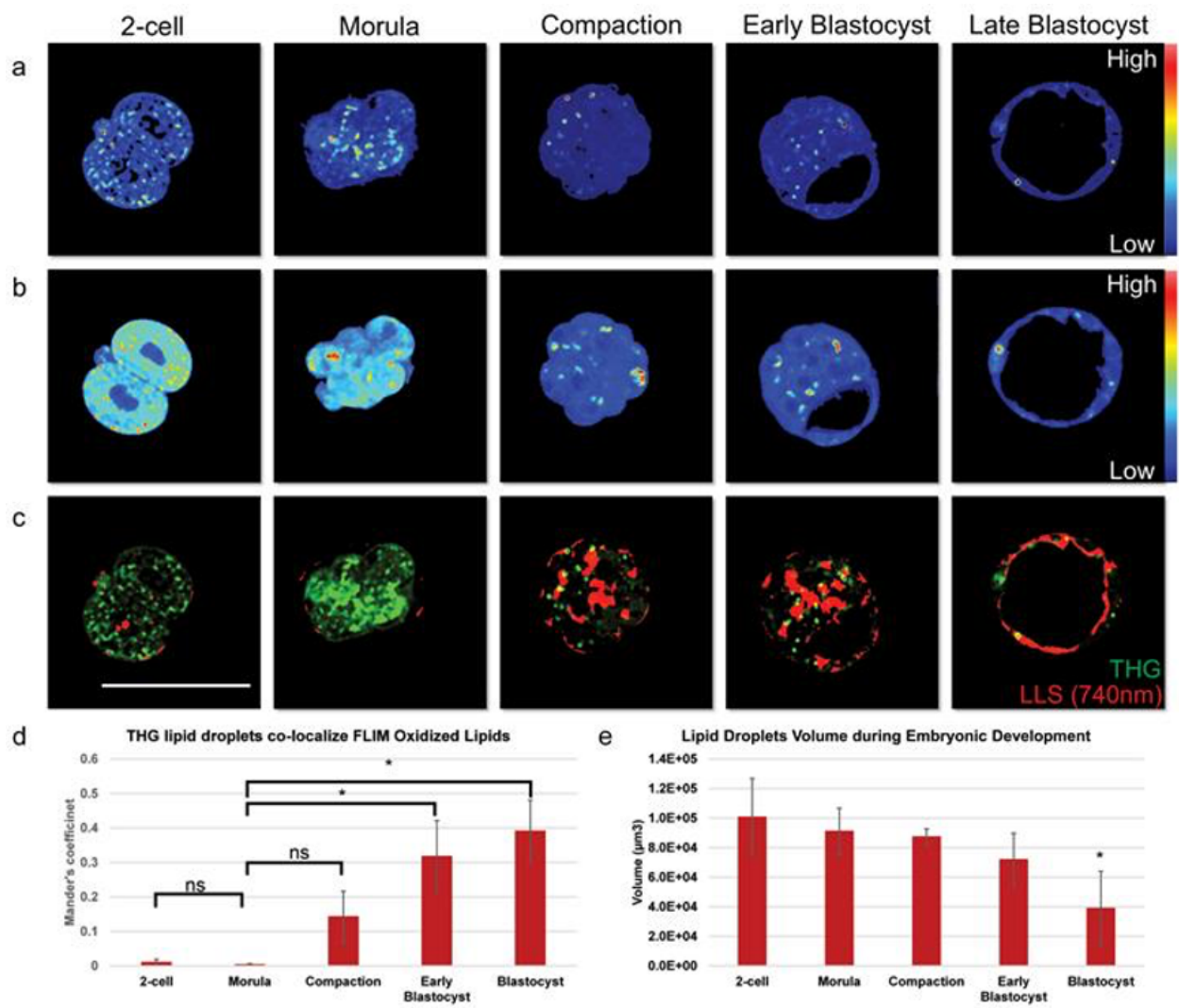
| 2-cell | Morula | Compaction | Early Blastocyst | Late Blastocyst |

a — High / Low

b — High / Low

c — THG / LLS (740nm)

d THG lipid droplets co-localize FLIM Oxidized Lipids

Mander's coefficinet

ns · ns · * · *

2-cell · Morula · Compaction · Early Blastocyst · Blastocyst

e Lipid Droplets Volume during Embryonic Development

Volume (µm3)

2-cell · Morula · Compaction · Early Blastocyst · Blastocyst

177

Figure 4: Fluorescence lifetime trajectories reveal metabolic states of pre-implantation mouse embryos. (a) Transmission (top), fluorescence intensity (middle) and FLIM (bottom) images for control and 4-hour rotenone and antimycin A (R&A) treated embryos. Note a shift from long to short lifetimes (blue to red in FLIM image). **(b)** g and s values of control and R&A-treated embryos for individual embryos. Blue circles are controls (n= 38), red circles are R&A-treated embryos (n= 31), and solid squares and the error bars in the figures means the average and variation of each group (student t-test for g value: p-value= 2.86E-16). FLIM images indicate a rightward shift from long to short lifetimes. **(c)** Transmission (top), fluorescence intensity (middle) and FLIM (bottom) images for control and 2DeoxyG-treated embryos. Note a shift from long to short lifetimes (red to white in FLIM image). **(d)** g and s values of control and 2DeoxyG-treated embryos. Blue squares are controls (n= 12), red circles are 2DeoxyG -treated embryos (n= 13), and the average of each group can be found in the solid colored squares (student t-test for g value: p-value= 3.88E-09). Fluorescence and FLIM images indicate a leftward shift from long to short lifetimes.

a **Oxidative Phosphorylation Inhibition**

Control | 500 nM R&A (4 h)

Transmission

Fluorescence Intensity (740 nm) — High / Low

FLIM — Long / Short

100µm

b
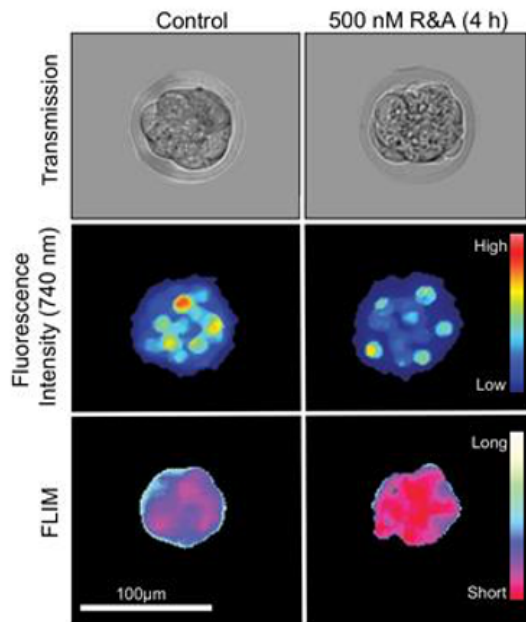
c **Glycolysis Inhibition**

Control | 1 mM 2DeoxyG (8 h)

Transmission

Fluorescence Intensity (740 nm) — High / Low

FLIM — Long / Short

100µm

d
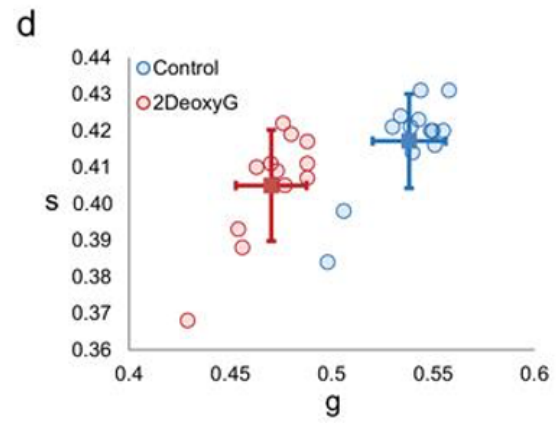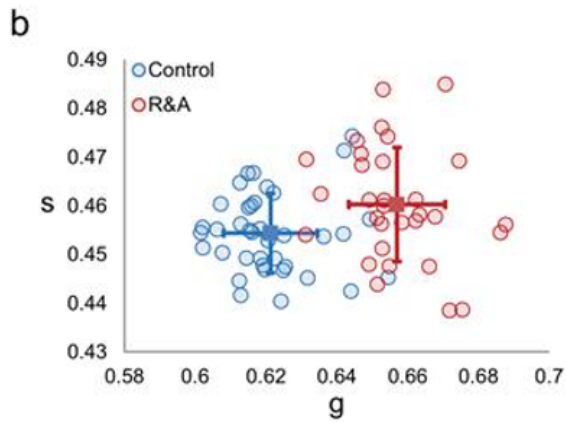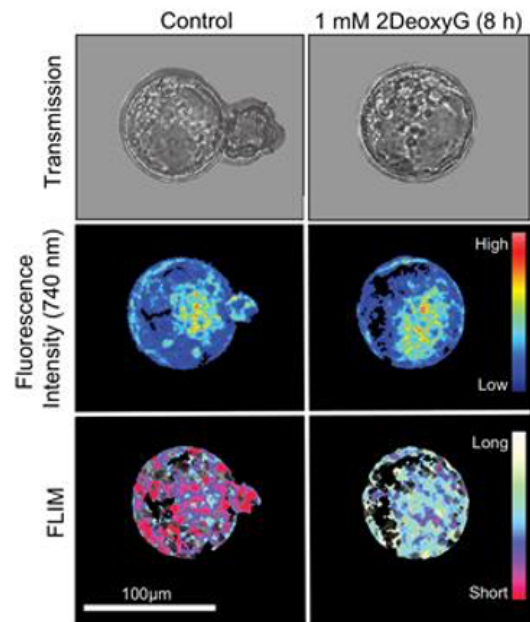
Figure 5: Deviation of intrinsic lifetime trajectory of embryos cultured in nutrient-depleted media. (a) Transmission images of embryos collected at the 2-cell stage and cultured in KSOMaa, FHM, or PBS for 24 hours. **(b)** Representative transmission and FLIM images of embryos in KSOMaa, FHM, or PBS for 4 hours. **(c)** Scatter plot of g and s lifetimes collected from a group of embryos cultured in KSOMaa (n=10), FHM (n=10) and PBS (n=4) for 4 hours. p-value= 0.0002** and 0.01* (student t-test of g value) for the FHM and PBS group compare with KSOMaa group. **(d)** Transmission images of embryos collected at the compaction stage and cultured in KSOMaa, FHM, or PBS for 24 hours. **(e)** Representative transmission and FLIM images of embryos in KSOMaa, FHM, or PBS for 4 hours. **(f)** Scatter plot of g and s of lifetimes collected from a group of embryos cultured in KSOMaa (n= 8), FHM (n=8), and PBS (n=8). p-value = 9.29E-06** and 3.21E-07** (student t-test of g value) for the FHM and PBS group compare with KSOMaa group.
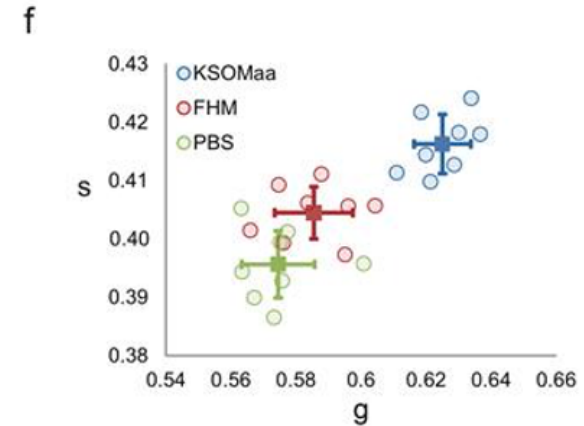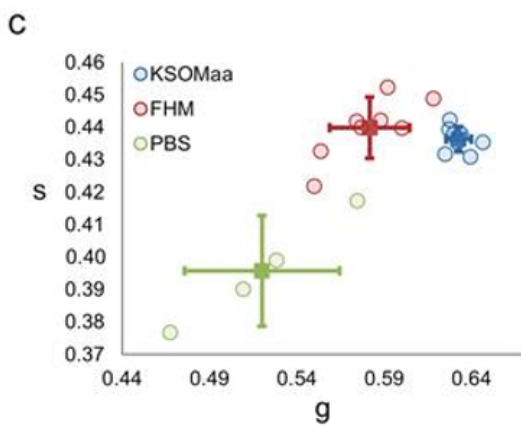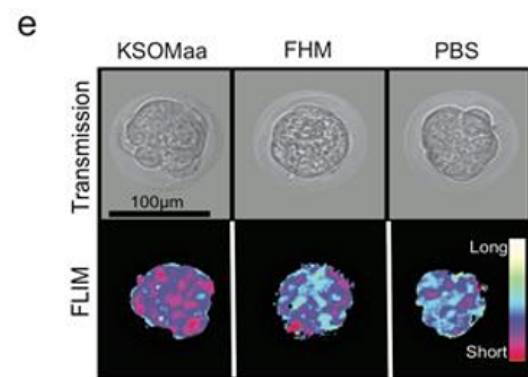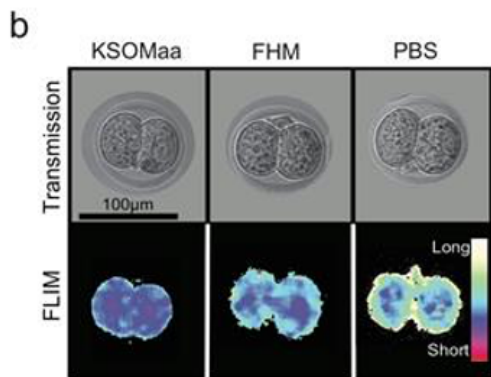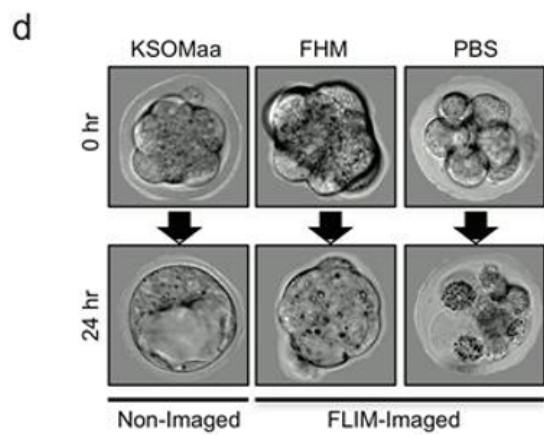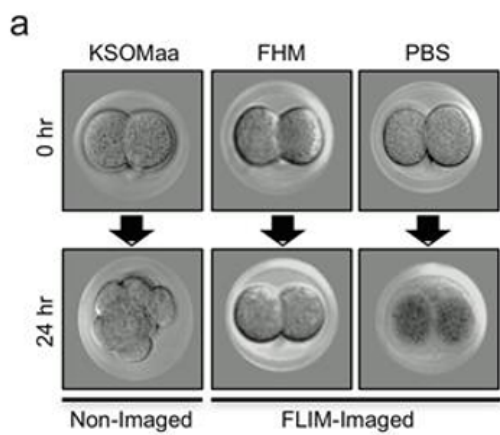
**Figure 6: Derivation of the embryo viability index (EVI) gauging embryo quality.**

(a) Schematic of our experimental setup. Individual embryos (A-F) were followed from the 2-cell to blastocyst stage and classified as healthy (H) and unhealthy (UH) group according to their morphology at E4.5. **(b)** Histogram of embryo viability index (EVI) of early compaction embryos from one representative experiment (H group, n=37; UH group, n=27). The blue and red bars represent the embryo condition determined as healthy and unhealthy at ~60 hours after FLIM imaging at the pre-compaction stage. **(c)** Receiver operating characteristic (ROC) curve shows the performance of the binary classification model developed from lifetime distribution patterns of early developmental stage embryos (2-cell, 4-cell, and early compaction stage). The area under a curve for each stage is 0.739 (2-cell), 0.728 (4-cell) and 0.916 (early compaction). The dashed line in the diagonal is presented as a random bi-classification model. **(d)** Schematic of FLIM-Distance Analysis Pipeline. **(e)** Box-whisker plots of experiment 1 showing a training set of healthy (n=5) and unhealthy (n=7) groups and tested unknowns of healthy (n=18) and unhealthy (n=16) embryos. **(f)** Bar graph of embryo viability index of experiment 1. Training set H is in navy, training set UH is in red. Testing set H is in light blue, and Testing set U is in orange.
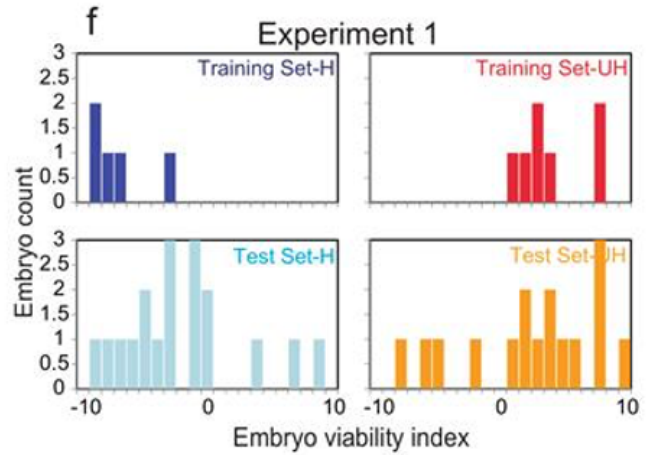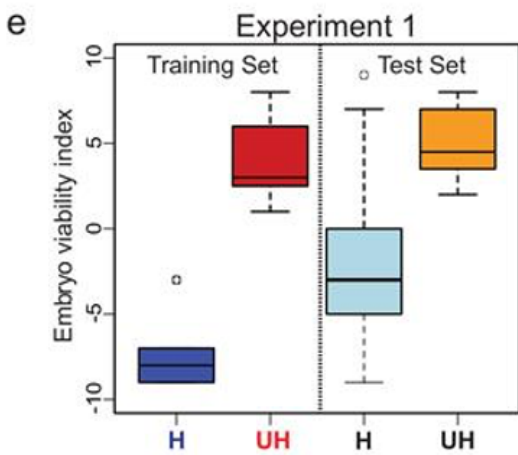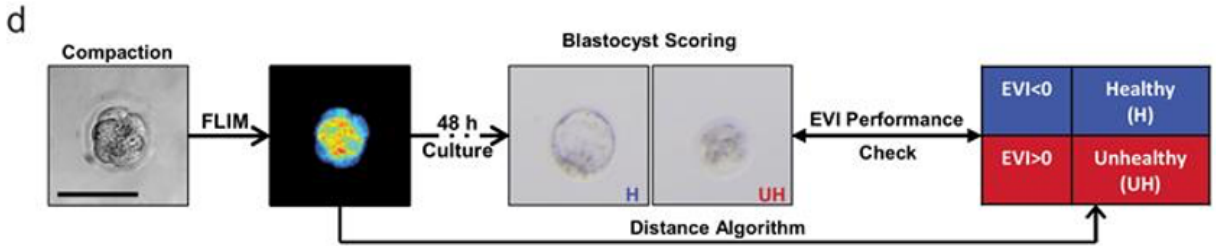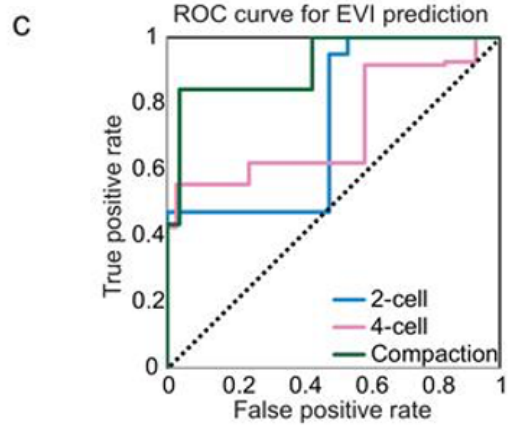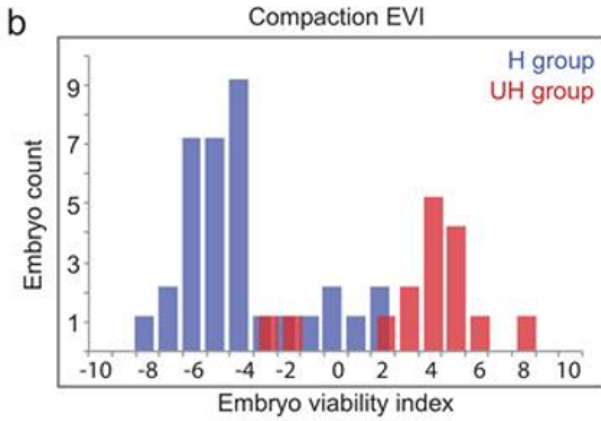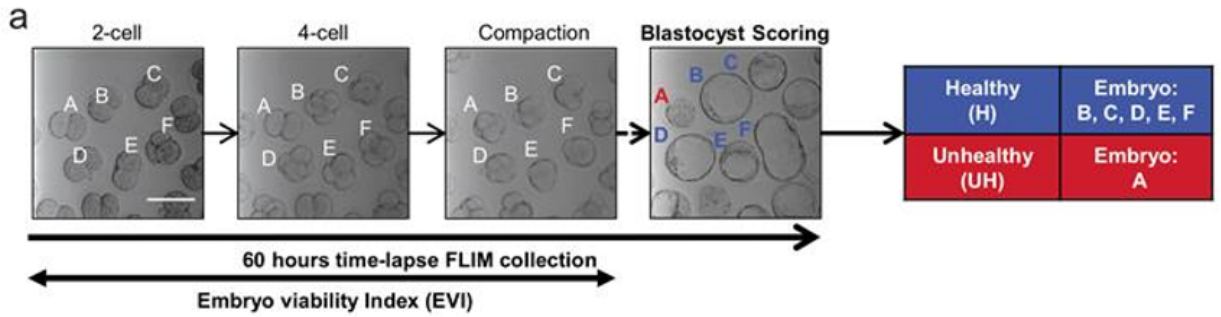
**Supplementary Figure 1: Phasor FLIM analysis.**

(a) Phasor FLIM analysis. During FLIM collection, a pulsed 2-photon laser is used to measure the intensity at short time windows (time arrival of the photons) as a function of time. Instead of fitting the decay curve into an exponential equation (black line), the raw data (intensity at each pixel) is transformed into polar coordinates by plotting the sine (red line) and cosine (blue line) using Fourier transformation, for every pixel in the object, the fluorescence lifetime can be obtained as "phasor lifetime" **(b)** Phasor fingerprint of pure intrinsic biomarkers of free NADH in solution, bound NADH in the presence of lactate dehydrogenase, and a long lifetime species derived from lipid droplets. Given that the free form of NADH exhibits a compact structure with a low fluorescence quantum yield (j=0.019) and a short lifetime of 0.4ns and the extended form of NADH bound to lactate dehydrogenase with a much higher quantum yield (j= 0.099) with a longer fluorescence lifetime up to ~3.4ns, the lifetimes of these two states can be easily distinguished(*40*). Based on the law of phasor addition, any sample containing the combination signature of these three species will fall within the triangle joining the three phasors.
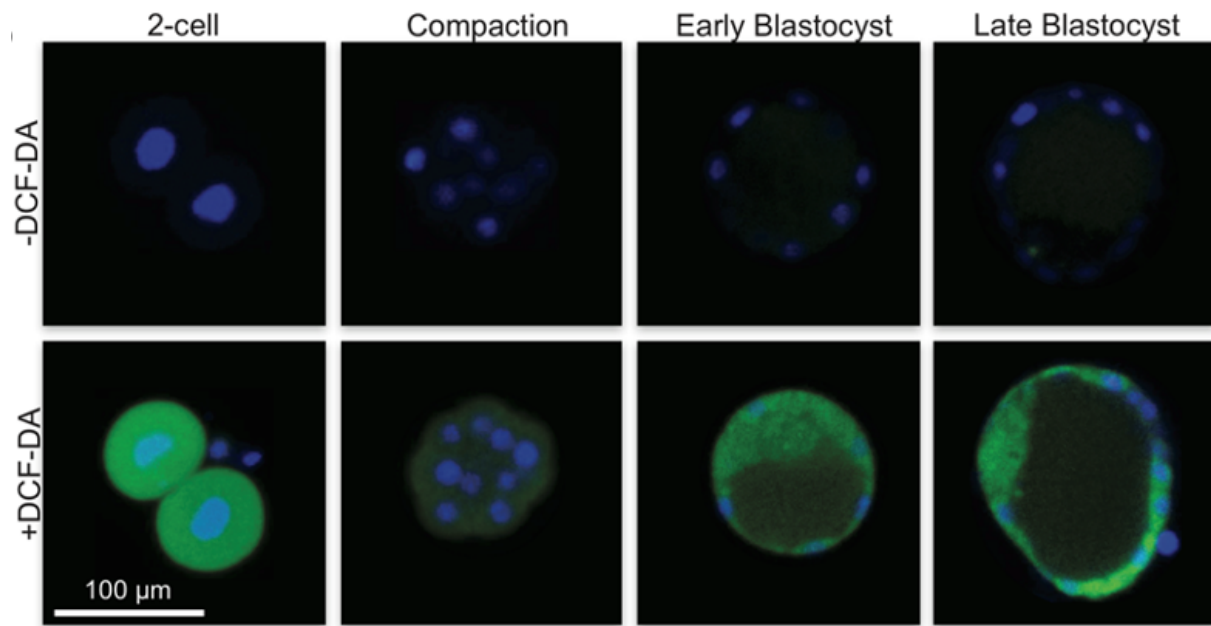
Figure S1
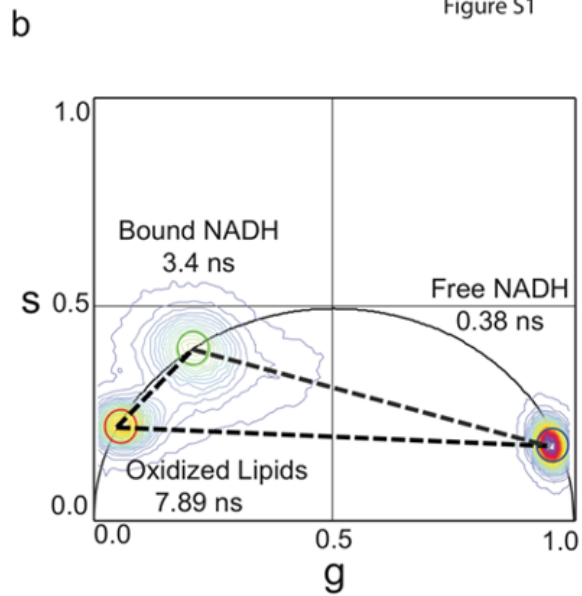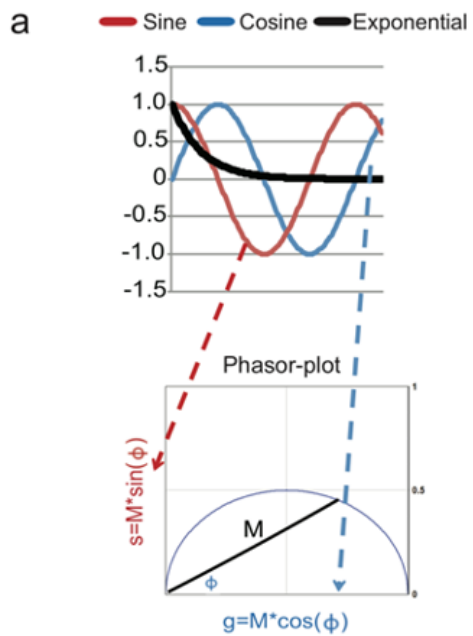
Supplementary Figure 2: Development trajectory composed of two distinct trajectories that correlate with the metabolism of embryonic stages.
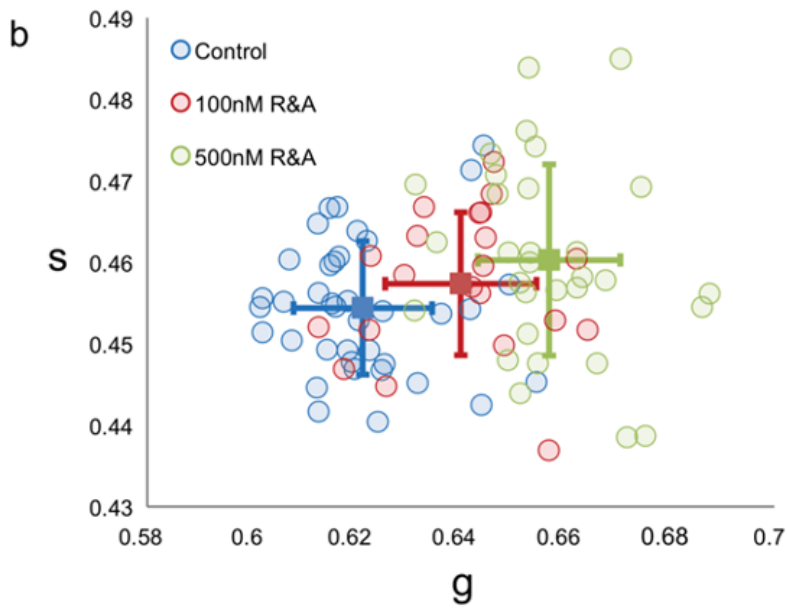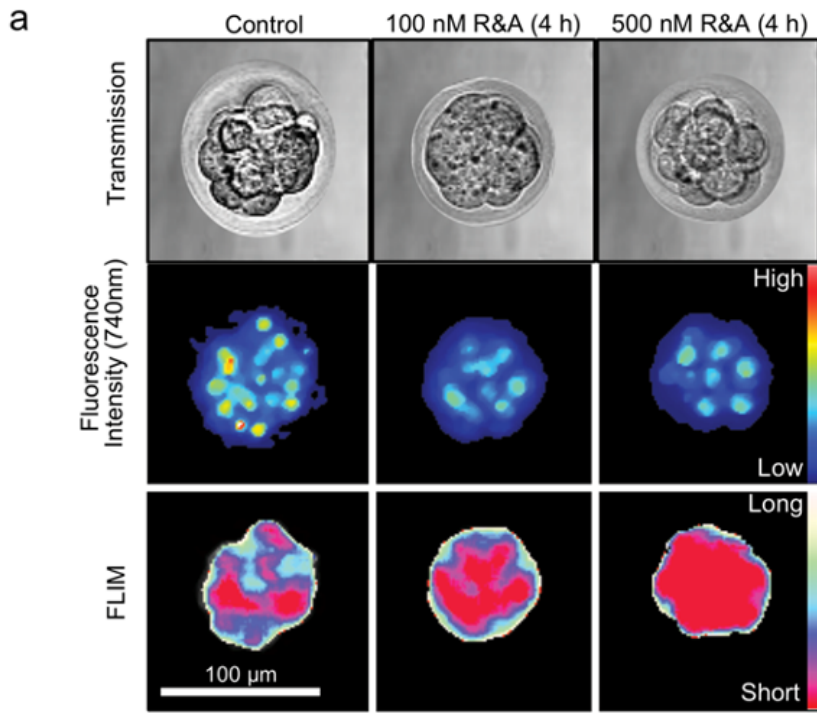
(a) Representative phasor plots for the pre-implantation mouse embryo from early cleavage stage to blastocyst stage (E1.5-E4.5). **(b)** Four examples of the D-Trajectory observed throughout pre-implantation stages (from 2-cell to blastocyst stages). **(c)** Transmission and FLIM image plots for pre-implantation mouse embryos from early cleavage stage to blastocyst stage under 20.9% oxygen and 12.8% oxygen conditions. **(d)** D-Trajectory shows the same trend under an hypoxic condition (12.8% Oxygen, red) (n=19, 16, 11, 15, 13 for E1.5, E2.5, E3, E3.5, E4.5 stage respectively) and regular blood mixture culture condition (20.9% oxygen, blue) (n=29, 8, 11, 14, 11 for E1.5, E2.5, E3, E3.5, E4.5 stage, respectively). The straight lines show the phasor-FLIM patterns of developing embryos. The starting scatter plots (representing E1.5 embryos) is in the top right. The last scatter (E4.5) of the lines is in the bottom left..For g values of hypoxia treatment group, compare with regular blood mixture treatment group, p-value=0.296, 0.018, 0.829, 0.488, 0.584 for E1.5, E2.5, E3, E3.5, E4.5 stage, respectively (student t-test, two-tail test for g). The error bars show the standard deviation for each condition. Scale bar set to 100μm.
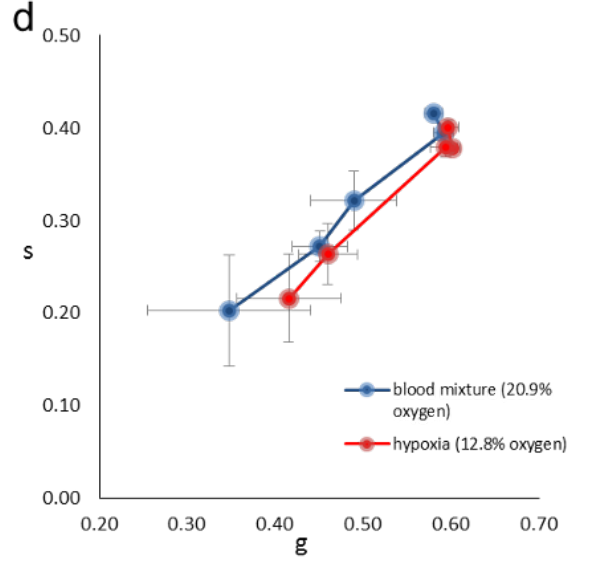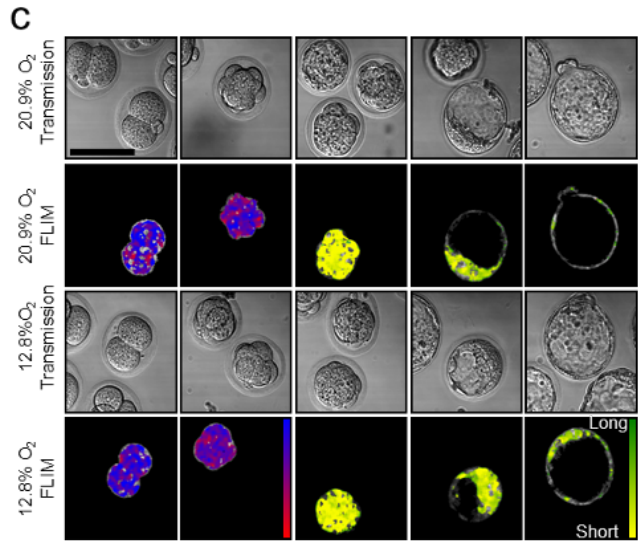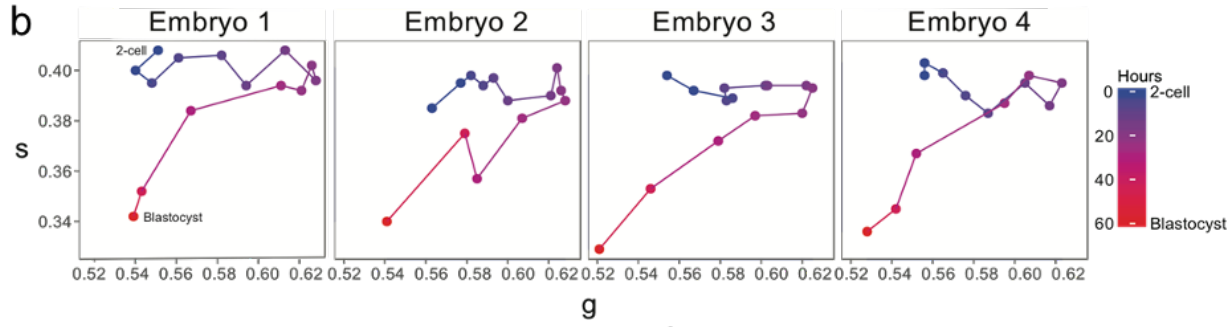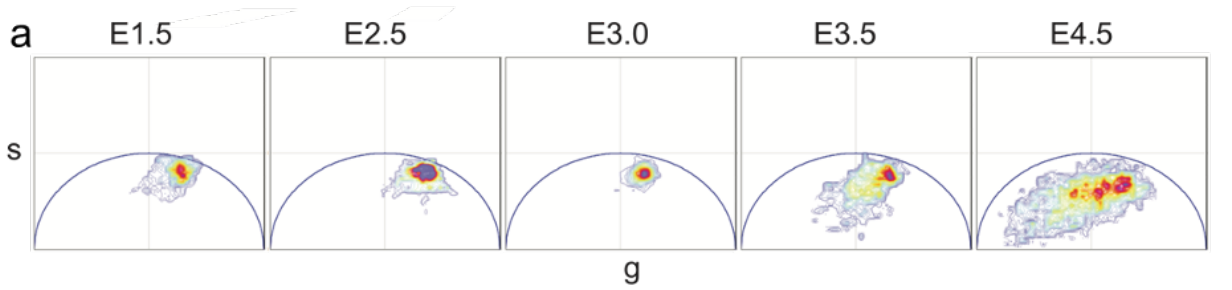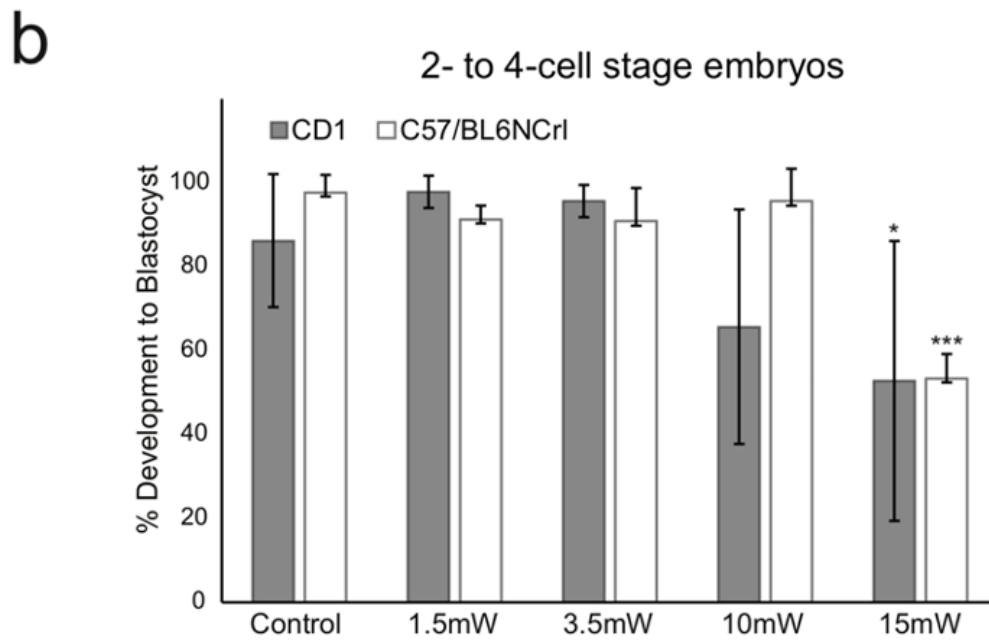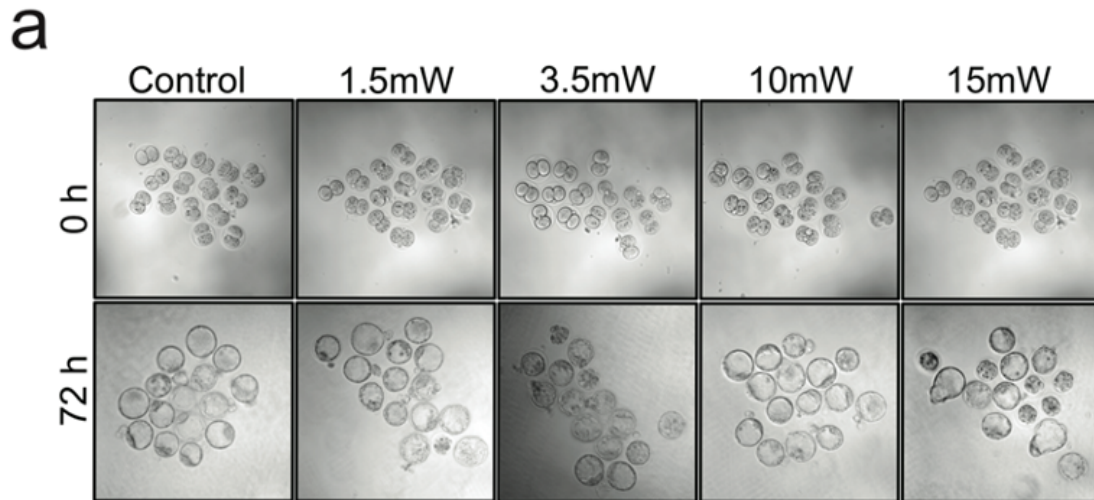
Supplementary Figure 3: Detection of reactive oxygen species in pre-implantation mouse embryos. 2', 7'-dichlorodihydrofluorescein diacetate (DCF-DA) stain across pre-implantation mouse embryos.
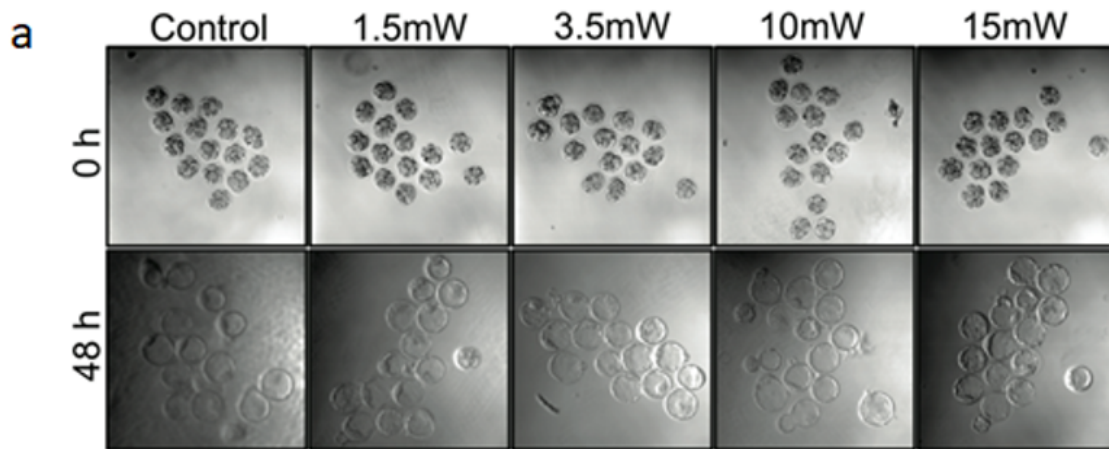
| | 2-cell | Compaction | Early Blastocyst | Late Blastocyst |

100 μm

Supplementary Figure 4: Fluorescence lifetime trajectories reveal dose-dependent metabolic state changes of pre-implantation mouse embryos. **(a)** Transmission (top), fluorescence (middle) and FLIM (bottom) images for control and 4-hour 100nM and 500nM rotenone and antimycin A (R & A) treated embryos, indicating a shift from long to short lifetimes. **(b)** g and s values of control and 4-hour 100nM and 500nM R&A-treated embryos for individual embryos. Blue circles are controls (n= 38), red circles are 4-hour 100nM R&A-treated embryos (n= 21), and green circles are 4-hour 500nM R&A-treated embryos (n=31). The average of each group can be found in the solid squares (for g value of 100nM treatment group and 500nM group compared with control group, p-value=1.76E-5, and 2.86E-16, respectively). FLIM images indicate a rightward shift from long to short lifetimes. Student t-test and two-tail tests were performed.
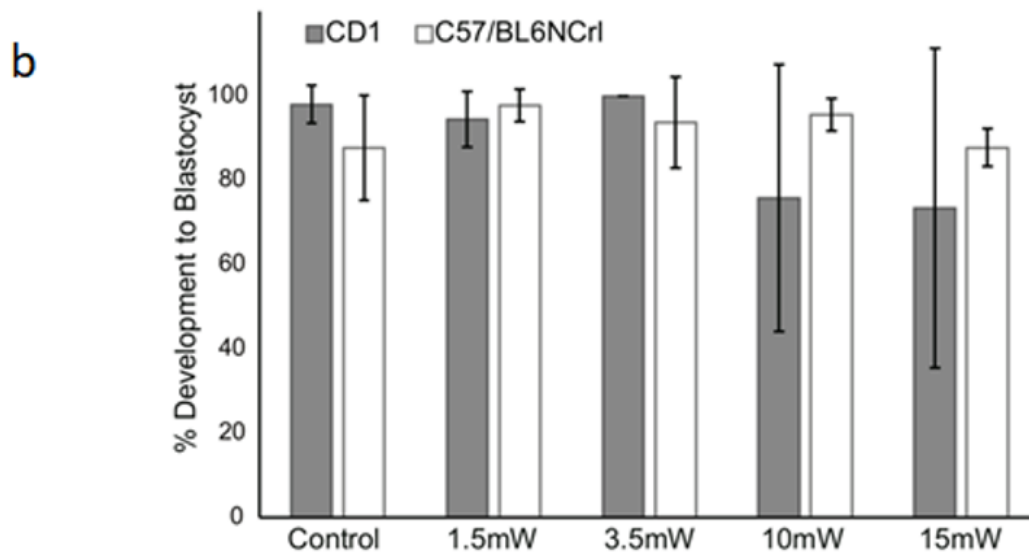
a

| | Control | 100 nM R&A (4 h) | 500 nM R&A (4 h) |

Transmission

Fluorescence Intensity (740nm)

High

Low

FLIM

Long

100 μm

Short

b

Control
100nM R&A
500nM R&A

s

g

Supplementary Figure 5: Validating the safety of FLIM-imaging of E1.5 pre-implantation mouse embryos. (a) Transmission images of embryos before and after imaging and culturing for 72 hours till E4.5. Image size is 708.49*708.49 μm. **(b)** Assessment of embryonic development after imaging and 72-hour *in vitro* culture reported as percent development. CD1: E1.5 non-imaged control (n=72); 1.5mW (n=42); 3.5mW (n=42), 10mW (n=31), 15mW (n=45) imaged embryos. C57BL/6NCrl: E1.5 non-imaged control (n=37); 1.5mW (n=37); 3.5mW (n=36), 10mW (n=38), 15mW (n=37) imaged embryos.

a

	Control	1.5mW	3.5mW	10mW	15mW

0 h

72 h

b

2- to 4-cell stage embryos

% Development to Blastocyst

■CD1 □C57/BL6NCrl

Control   1.5mW   3.5mW   10mW   15mW

Supplementary Figure 6: Validating the safety of FLIM-imaging of E2.5 pre-implantation mouse embryos. (a) Transmission images of embryos before and after imaging and culturing for 48 hours till E4.5. Image size is 708.49*708.49 µm. **(b)** Assessment of embryonic development after imaging and 48-hour *in vitro* culture reported as percent development. For CD1: E2.5 non-image control (n=53); 1.5mW (n=43); 3.5mW (n=43), 10mW (n=28), 15mW (n=28, p<0.05 *) imaged embryos. For C57/BL6NCrl: E2.5 non-image control (n=75); 1.5mW (n=42); 3.5mW (n=47), 10mW (n=44), 15mW (n=41 p<0.001 ***) imaged embryos. Student t-test and one tailed test were performed. Everything is non-significant except the ones annotated.

a

| | Control | 1.5mW | 3.5mW | 10mW | 15mW |

0 h

48 h

b

Morula to Compaction Stage Embryos

■CD1　□C57/BL6NCrl

% Development to Blastocyst

100

80

60

40

20

0

Control　1.5mW　3.5mW　10mW　15mW

Supplementary Figure 7: FLIM does not affect live birth rates. (a) H2AX staining of FLIM imaged pre-implantation embryos. Two strains of embryos were FLIM imaged at E.1.5 and subjected to Hoechst (blue) and H2AX (red) staining at E2.5 for DNA-damage assessment. Control: no FLIM-imaged embryos. For CD1: E2.5 control (n=3); 1.5mW (n=3); 3.5mW (n=3), 10mW (n=3) and 15mW (n=3). For C57/BL6NCrl: E2.5 control (n=9); 1.5mW (n=9); 3.5mW (n=8), 10mW (n=6), and 15mW (n=8). Scale bar set to 20μm. **(b)** Live birth rates of FLIM-imaged embryos. Control (solid blue, n=88) and FLIM-imaged CD1 embryos (diagonal stripes, blue, n=94) were allowed to develop to blastocysts and implanted into 13 pseudo pregnant females (3 independent trials). The pups were collected from C-section on E18.5. There is no statistically significant difference for the live birth rate (p-value, 0.662).
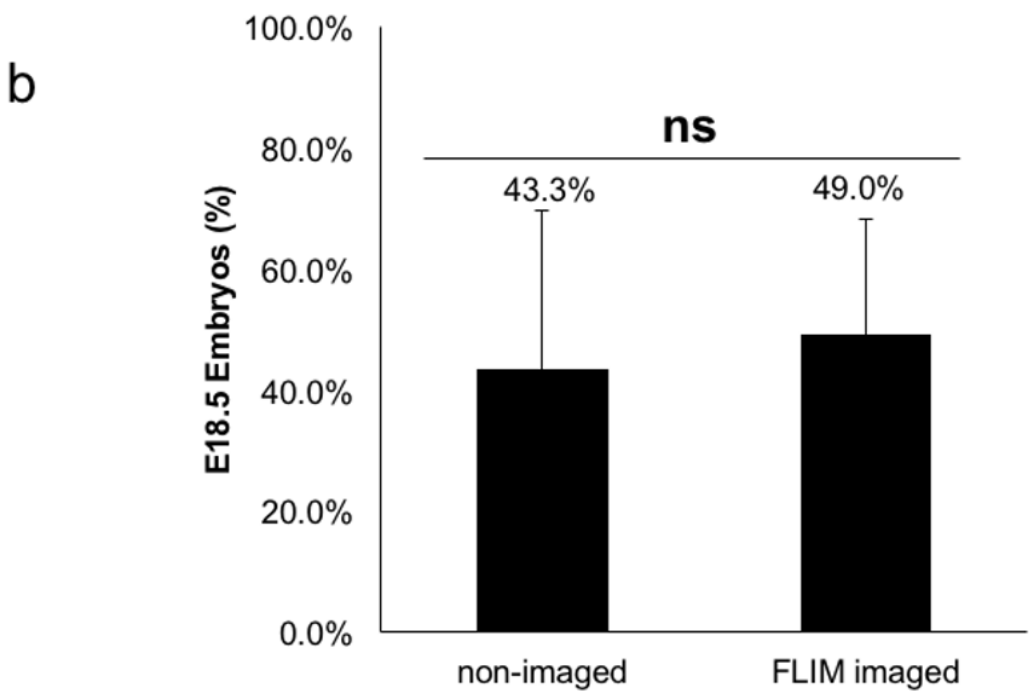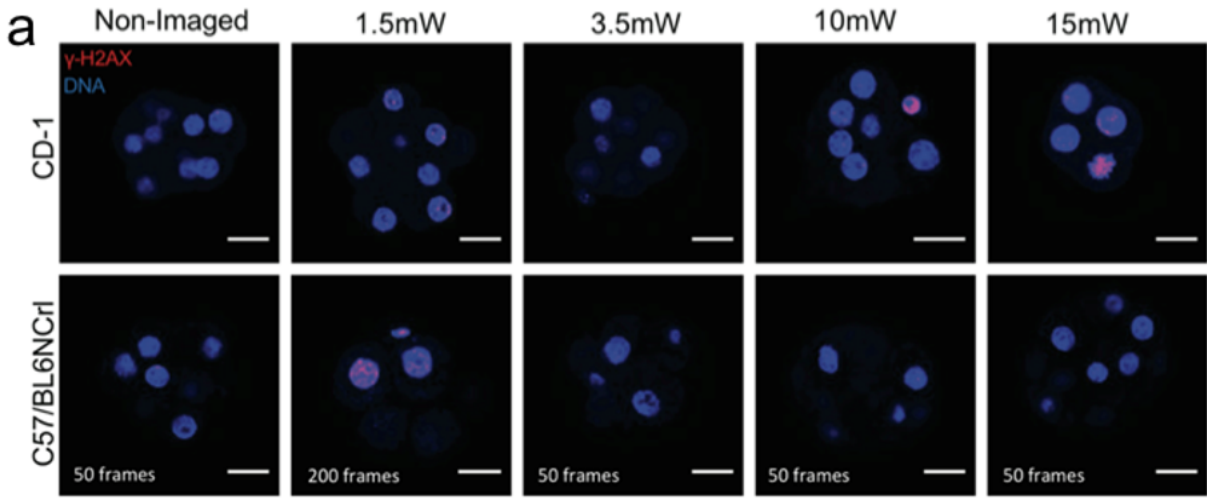
Supplementary Figure 8: Average number of cells per embryo cultured under high stress conditions. (a) Bar graph showing the average number of cells present in an embryo after xx hours of culturing in the indicated media starting at the 2-cell stage. KSOMaa non-image control (n=8), KSOMaa (n=10), FHM (n=11), PBS (n=11), compare to the results from KSOMaa group, p-value = 0.004** and 0.001** for FHM and PBS, respectively. **(b)** Bar graph showing the average number of cells per embryo after continuous culturing of embryos in the indicated media starting at the morula stage. KSOMaa non-image control (n=10), KSOMaa (n=18), FHM (n=11), PBS (n=11), p-value = 0.002** and 0.02* for FHM and PBS, respectively.

a 2-cell stage embryos
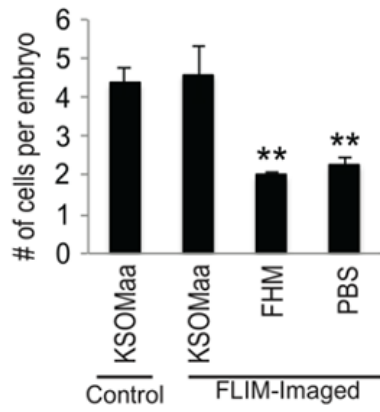
b Compaction stage embryos
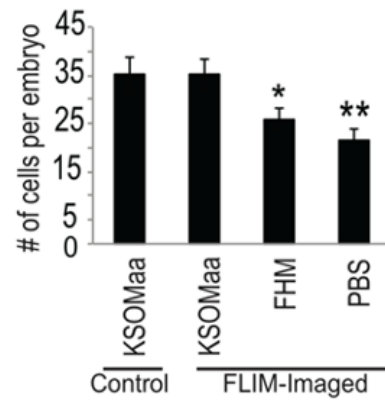
Supplementary Figure 9: Embryo viability index of morula shows the potential to distinguish healthy and unhealthy pre-cleavage stage embryos. (a) Histogram of embryo viability index of 2-cell and morula stage embryos from one representative experiment (2-cell EVI: H group, (n=18), UH group, (n=17). 4-cell EVI, H group, (n=25), UH group, (n=9)). Each blue and red bar represents the morula stage FLIM-fingerprints of healthy (H) and unhealthy (UH) embryos at 60 hours after imaging respectively. **(b)** Receiver operating characteristic (ROC) curve shows the performance of the binary classification model developed from lifetime distribution patterns of pre-compaction stage embryos (2-, 4-, and early compaction) of two time-lapse FLIM tracking experiments. The area under curve for each stage is 0.777 (2-cell, H n=37; UH, n=8), 0.823 (4-cell, H, n=45; UH, n=8) and 1.000 (early compaction, H, n=30; UH, n=2) for experiment 2, and 0.777 (2-cell, H, n=38, UH, n=10), 0.813 (4-cell, H, n=39, UH, n=7) and 0.945 (early compaction, H, n=39, UH, n=6) for experiment 3.**(c)** Box-whisker plots showing training sets of healthy (H) and unhealthy (UH) groups and predication performance on tested embryos for 4 additional experiments Training set H in navy, Training set UH in red, predicated healthy in light blue, and predicated unhealthy in orange. The n number for the training set healthy group was n= 5, 11, 2 and 3. The n number for the training set unhealthy group was n= 4, 3, 2, and 2. The n number for the test unknown healthy
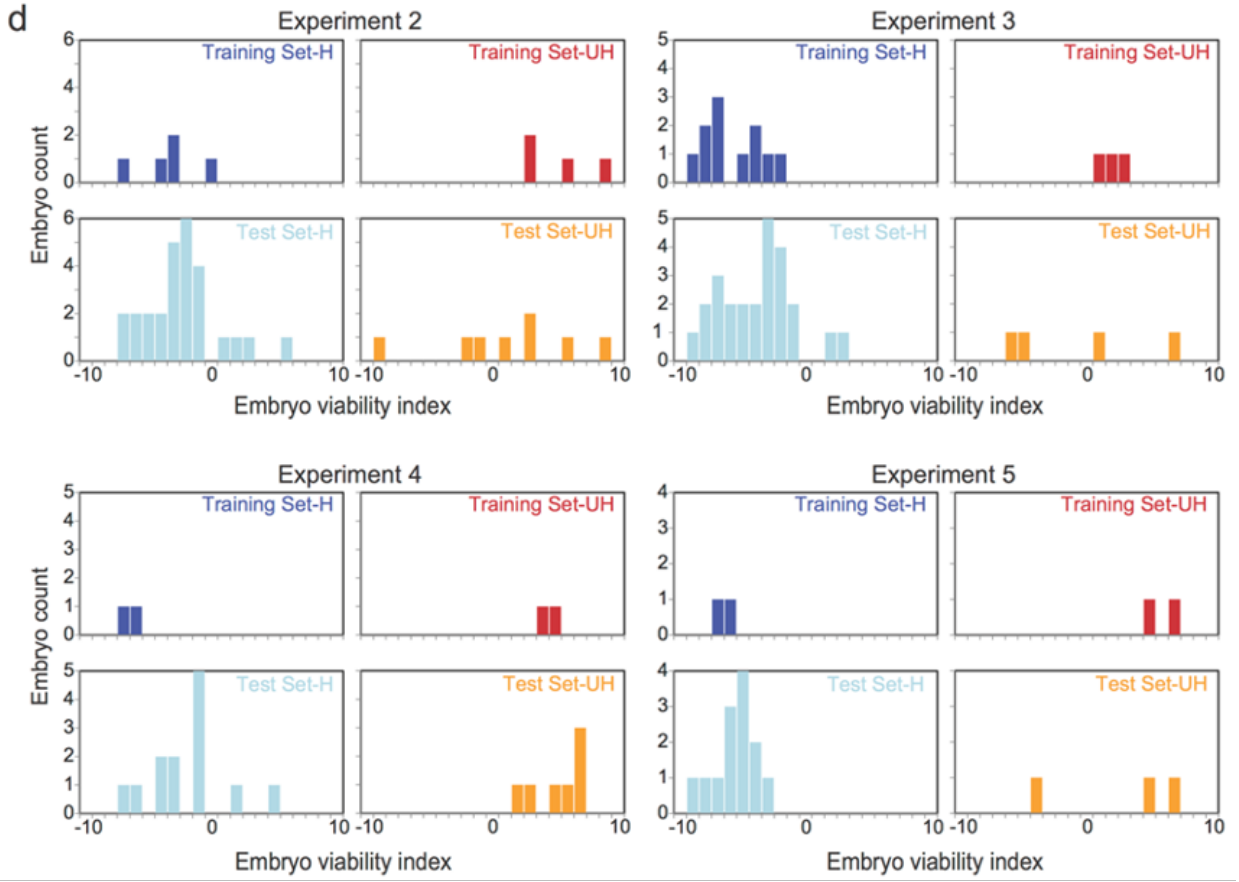
Supplementary Figure 9 (continued): group was n= 27, 25, 13, and 13. The n number for the test unknown unhealthy group was n= 8, 4, 7, and 3. **(d)** Bar graph of embryo viability index of experiment 2, 3, 4 and 5. Training set H in navy, Training set UHs in red, Healthy in light blue, and Unhealthy in orange.
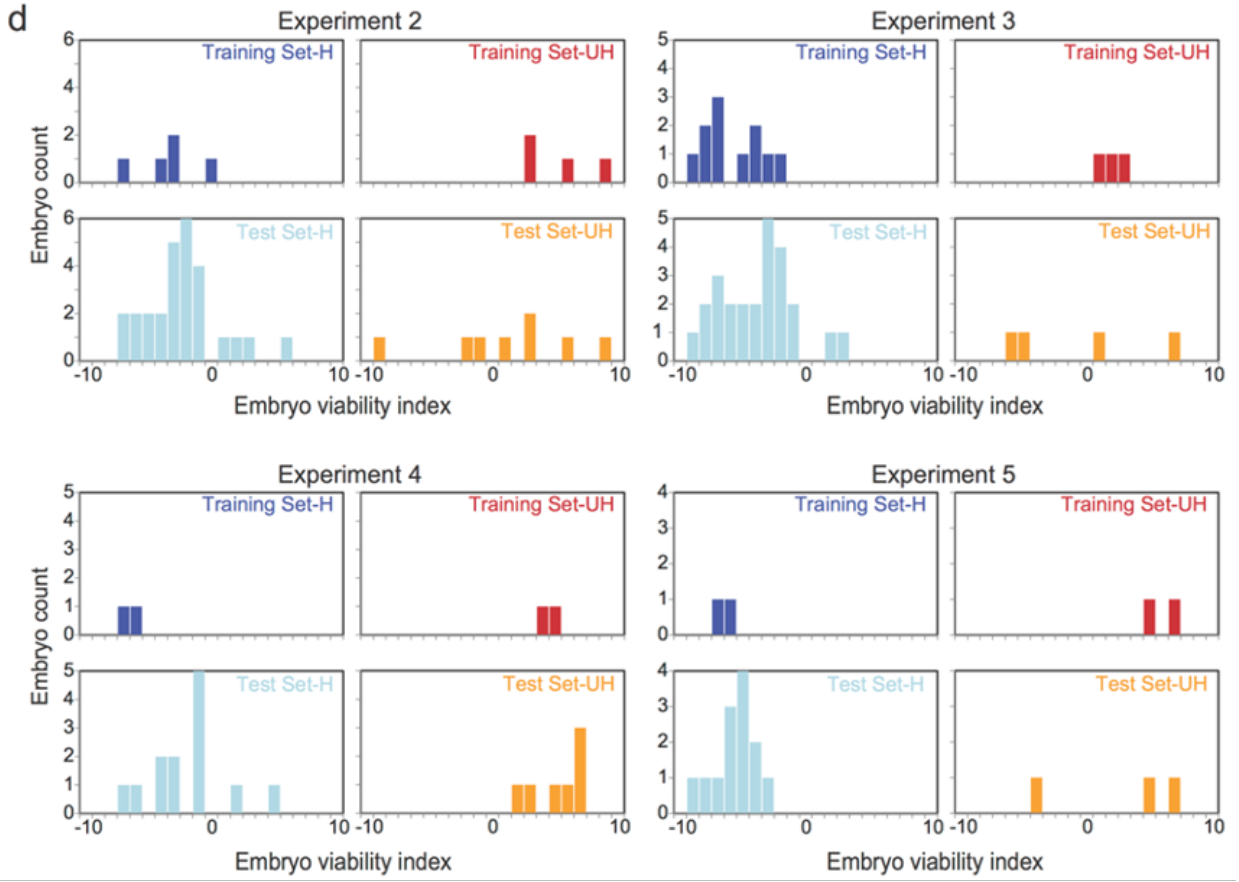
Table 1. Implantation Efficacy at E18.5

	CD-1 Embryos Implanted	Bre-gal^{+-} Embryos Implanted	CD-1 E18.5	Bre-gal^{+-} E18.5
		Trial 0		
Mouse 1	7	7	5/7 (71.4%)	4/7 (57.1%)
Mouse 2	7	7	1/7 (14.3%)	1/7 (14.3%)
Mouse 3	7	7	4/7 (50%)	2/7 (25%)
Mouse 4	7	7	1/7 (12.5%)	4/7 (50%)
Total	**28**	**28**	**11/28 (39.3%)**	**11/28 (39.3%)**
		Trial 1		
	CD-1 Embryos Implanted	Bre-gal^{+-} Embryos Implanted	CD-1 E18.5	Bre-gal^{+-} E18.5
Control 1	8	8	4/8 (50%)	6/8 (75%)
Imaged 1	8	8	6/8 (75%)	2/8 (25%)
Imaged 2	8	8	2/8 (25%)	6/8 (75%)
Imaged 3	8	8	1/8 (12.5%)	2/8 (25%)
Imaged 4	8	8	4/8 (50%)	4/8 (50%)
Total Control	**8**	**8**	**4/8 (50%)**	**6/8 (75%)**
Total Imaged	**32**	**32**	**13/32 (40.6%)**	**14/32 (43.8%)**
		Trial 2		

	CD-1 Embryos Implanted	Bre-gal^{+-} Embryos Implanted	CD-1 E18.5	Bre-gal^{+-} E18.5
Control 1	10	2	1/10 (10%)	1/2 (50%)
Control 2	10	2	1/10 (10%)	0/2 (0%)
Control 3	10	2	5/10 (50%)	1/2 (50%)
Imaged 1	10	2	5/10 (50%)	2/2 (100%)
Imaged 2	10	2	5/10 (50%)	0/2 (0%)
Total Control	**30**	**6**	**7/30 (23.3%)**	**2/6 (33.3%)**
Total Imaged	**20**	**4**	**10/20 (50%)**	**2/4 (50%)**

Trial 3				
	CD-1 Embryos Implanted	Bre-gal^{+-} Embryos Implanted	CD-1 E18.5	Bre-gal^{+-} E18.5
Control 1	14	NA	12/14 (85.7%)	NA
Control 2	14	NA	3/14 (21.4%)	NA
Control 3	14	NA	6/14 (42.9%)	NA
Imaged 1	14	NA	8/14 (57.1%)	NA
Imaged 2	14	NA	9/14 (64.3%)	NA
Imaged 3	14	NA	8/14 (57.1%)	NA
Total Control	**42**	**NA**	**21/42 (50%)**	**NA**
Total Imaged	**42**	**NA**	**25/42 (59.5%)**	**NA**

Supplementary Table 1: Implantation efficacy for the non-image embryos and FLIM-imaged embryos. Trial 0 is aiming for the baseline collection to test the implantation efficacy for the BRE-gal and CD-1 embryos. Trials 1-3 test the implantation efficacy of embryos after FLIM-imaging. The total percentage of E18.5 embryos retrieved from c-section for control (non-image) group and imaged group were 43% and 49%, respectively (p-value, 0.662).

Supplementary Table 2: Statistical analysis of DA program to predict embryo viability

	False positive rate	True positive rate	Accuracy	Precision	Sensitivity	Specificity
Experiment 1	0.250	0.833	0.794	0.789	0.833	0.750
Experiment 2	0.375	0.852	0.800	0.885	0.852	0.625
Experiment 3	0.500	0.920	0.862	0.920	0.920	0.500
Experiment 4	0.000	0.846	0.900	1.000	0.846	1.000
Experiment 5	0.333	1.000	0.938	0.929	1.000	0.667
	0.292	0.890	0.859	0.905	0.890	0.708

Supplementary Table 2: Statistical analysis of Distance Analysis (DA) program as a means to predict embryo viability.

Supplementary Movie 1: 3D Third Harmonic Generation (THG) images of the representative embryos from different pre-implantation embryos. a) 2-cell, b) 8-cell, c) Compaction, d) Early Blastocyst, e) Blastocyst

Supplementary Movie 2: 60-hour time-lapse imaging of pre-implantation embryos from E1.5 to E4.0. Embryos from H group (blue) are the representative healthy embryos that developed to the blastocyst stage after 60 hours of time-lapse imaging. Embryos from UH group (red) are arrested at late compaction stage.

AII.5 References:

1.   T. Baczkowski, R. Kurzawa, W. Głabowski, Methods of embryo scoring in in vitro fertilization. *Reprod Biol* **4**, 5-22 (2004).
2.   P. De Sutter, D. Dozortsev, C. Qian, M. Dhont, Oocyte morphology does not correlate with fertilization rate and embryo quality after intracytoplasmic sperm injection. *Human Reproduction* **11**, 595-597 (1996).
3.   L. A. Scott, S. Smith, The successful use of pronuclear embryo transfers the day following oocyte retrieval. *Human Reproduction* **13**, 1003-1013 (1998).
4.   D. K. Gardner, M. Lane, J. Stevens, T. Schlenker, W. B. Schoolcraft, Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertility and sterility* **73**, 1155-1158 (2000).
5.   R. Scott *et al.*, Noninvasive metabolomic profiling of human embryo culture media using Raman spectroscopy predicts embryonic reproductive potential: a prospective blinded pilot study. *Fertil Steril* **90**, 77-83 (2008).
6.   R. R. González *et al.*, Leptin and leptin receptor are expressed in the human endometrium and endometrial leptin secretion is regulated by the human blastocyst 1. *The Journal of Clinical Endocrinology & Metabolism* **85**, 4883-4888 (2000).
7.   G. Sher, L. Keskintepe, M. Nouriani, R. Roussev, J. Batzofin, Expression of sHLA-G in supernatants of individually cultured 46-h embryos: a potentially valuable indicator of 'embryo competency'and IVF outcome. *Reproductive biomedicine online* **9**, 74-78 (2004).
8.   L. Botros, D. Sakkas, E. Seli, Metabolomics and its application for non-invasive embryo assessment in IVF. *Molecular human reproduction* **14**, 679-690 (2008).
9.   M. A. Digman, V. R. Caiolfa, M. Zamai, E. Gratton, The phasor approach to fluorescence lifetime imaging analysis. *Biophysical journal* **94**, L14-L16 (2008).
10.   J. M. Squirrell, D. L. Wokosin, J. G. White, B. D. Bavister, Long-term two-photon fluorescence imaging of mammalian embryos without compromising viability. *Nat Biotechnol* **17**, 763-767 (1999).
11.   C. Stringari *et al.*, Phasor approach to fluorescence lifetime microscopy distinguishes different metabolic states of germ cells in a live tissue. *Proc Natl Acad Sci U S A* **108**, 13582-13587 (2011).
12.   R. Datta, A. Alfonso-Garcia, R. Cinco, E. Gratton, Fluorescence lifetime imaging of endogenous biomarker of oxidative stress. *Sci Rep* **5**, 9848 (2015).
13.   D. K. Gardner, P. L. Wale, Analysis of metabolism to select viable human embryos for transfer. *Fertil Steril* **99**, 1062-1072 (2013).
14.   P. L. Wale, D. K. Gardner, The effects of chemical and physical factors on mammalian embryo culture and their importance for the practice of assisted human reproduction. *Hum Reprod Update* **22**, 2-22 (2016).
15.   H. J. Leese, A. M. Barton, Pyruvate and glucose uptake by mouse ova and preimplantation embryos. *J Reprod Fertil* **72**, 9-13 (1984).

16.   R. Dumollard, J. Carroll, M. Duchen, K. Campbell, K. Swann, in *Seminars in cell & developmental biology*. (Elsevier, 2009), vol. 20, pp. 346-353.

17.   N. Shyh-Chang, G. Q. Daley, L. C. Cantley, Stem cell metabolism in tissue development and aging. *Development* **140**, 2535-2547 (2013).

18.   C. Stringari, R. Sierra, P. J. Donovan, E. Gratton, Label-free separation of human embryonic stem cells and their differentiating progenies by phasor fluorescence lifetime microscopy. *J Biomed Opt* **17**, 046012 (2012).

19.   C. Stringari, J. L. Nourse, L. A. Flanagan, E. Gratton, Phasor fluorescence lifetime microscopy of free and protein-bound NADH reveals neural stem cell differentiation potential. *PLoS One* **7**, e48014 (2012).

20.   L. Covarrubias, D. Hernández-García, D. Schnabel, E. Salas-Vidal, S. Castro-Obregón, Function of reactive oxygen species during animal development: passive or active? *Developmental biology* **320**, 1-11 (2008).

21.   P. D. Ray, B.-W. Huang, Y. Tsuji, Reactive oxygen species (ROS) homeostasis and redox regulation in cellular signaling. *Cellular signalling* **24**, 981-990 (2012).

22.   H. Esterbauer, Cytotoxicity and genotoxicity of lipid-oxidation products. *Am J Clin Nutr* **57**, 779S-785S; discussion 785S-786S (1993).

23.   D. K. Gardner, M. Lane, I. Calderon, J. Leeton, Environment of the preimplantation human embryo in vivo: metabolite analysis of oviduct and uterine fluids and metabolism of cumulus cells. *Fertil Steril* **65**, 349-353 (1996).

24.   M. Lane, D. K. Gardner, Lactate Regulates Pyruvate Uptake and Metabolism in the PreimplantationMouse Embryo. *Biology of reproduction* **62**, 16-22 (2000).

25.   T. Watanabe *et al.*, Characterisation of the dynamic behaviour of lipid droplets in the early mouse embryo using adaptive harmonic generation microscopy. *BMC Cell Biol* **11**, 38 (2010).

26.   T. Hamaoka, K. K. McCully, V. Quaresima, K. Yamamoto, B. Chance, Near-infrared spectroscopy/imaging for monitoring muscle oxygenation and oxidative metabolism in healthy and diseased humans. *J Biomed Opt* **12**, 062105 (2007).

27.   C. Stringari *et al.*, Metabolic trajectory of cellular differentiation in small intestine by Phasor Fluorescence Lifetime Microscopy of NADH. *Sci Rep* **2**, 568 (2012).

28.   Z. W. Wang *et al.*, Laser microbeam-induced DNA damage inhibits cell division in fertilized eggs and early embryos. *Cell Cycle* **12**, 3336-3344 (2013).

29.   I. Revet *et al.*, Functional relevance of the histone gammaH2Ax in the response to DNA damaging agents. *Proc Natl Acad Sci U S A* **108**, 8663-8667 (2011).

30.   E. Sonoda *et al.*, Collaborative roles of gammaH2AX and the Rad51 paralog Xrcc3 in homologous recombinational repair. *DNA Repair (Amst)* **6**, 280-292 (2007).

31.   M. Lane, D. K. Gardner, Mitochondrial malate-aspartate shuttle regulates mouse embryo nutrient consumption. *Journal of Biological Chemistry* **280**, 18361-18367 (2005).

32.   S. Ranjit, A. Dvornikov, M. Levi, S. Furgeson, E. Gratton, Characterizing fibrosis in UUO mice model using multiparametric analysis of phasor distribution from FLIM images. *Biomedical Optics Express* **7**, 3519-3530 (2016).

33.   E. Seli *et al.*, Noninvasive metabolomic profiling of embryo culture media using Raman and near-infrared spectroscopy correlates with reproductive potential of embryos in women undergoing in vitro fertilization. *Fertil Steril* **88**, 1350-1357 (2007).

34.   C. G. Vergouw *et al.*, Metabolomic profiling by near-infrared spectroscopy as a tool to assess embryo viability: a novel, non-invasive method for embryo selection. *Hum Reprod* **23**, 1499-1504 (2008).

35.   E. Seli *et al.*, Noninvasive metabolomic profiling as an adjunct to morphology for noninvasive embryo assessment in women undergoing single embryo transfer. *Fertil Steril* **94**, 535-542 (2010).

36.   K. König, P. C. So, W. Mantulin, B. Tromberg, E. Gratton, Two-photon excited lifetime imaging of autofluorescence in cells during UV A and NIR photostress. *Journal of microscopy* **183**, 197-204 (1996).

37.   . (2016).

38.   M. Y. Berezin, S. Achilefu, Fluorescence lifetime measurements and biological imaging. *Chem Rev* **110**, 2641-2684 (2010).

39.   M. Chiang *et al.*, Analysis of in vivo single cell behavior by high throughput, human-in-the-loop segmentation of three-dimensional images. *BMC Bioinformatics* **16**, 397 (2015).

40.   R. Datta, A. Alfonso-García, R. Cinco, E. Gratton, Fluorescence lifetime imaging of endogenous biomarker of oxidative stress. *Sci Rep* **5**, 9848 (2015).

41.   A. L., Javier, L. T. Doan, M. Luong, N. S. Reyes de Mochel, A. Sun, E. S. Monuki, K. W. Y. Cho, Bmp Indicator Mice Reveal Dynamic Regulation of Transcriptional Response. PLoS One 7:e42566. (2012).

Material from: "Ning Ma, Nabora Reyes de Mochel, Paula Duyen Pham, Tae Yeon Yoo, Ken W. Y. Cho & Michelle A. Digman, Label-free assessment of pre-implantation embryo quality by the Fluorescence Lifetime Imaging Microscopy (FLIM)-phasor approach., 2019, Scientific Reports"

https://doi.org/10.1038/s41598-019-48107-2