UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Understanding trophoblast development using human embryonic stem cells

Permalink

https://escholarship.org/uc/item/0st352g9

Author

Morey, Robert

Publication Date

2022

Peer reviewed|Thesis/dissertation

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

UNIVERSITY OF CALIFORNIA SAN DIEGO

Understanding trophoblast development using human embryonic stem cells

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

in

Bioinformatics and Systems Biology

by

Robert Eaman Morey

Committee in charge:

Professor Louise Laurent, Chair
Professor Mana Parast, Co-Chair
Professor Melissa Gymrek, Co-Chair
Professor Heidi Cook-Andersen
Professor Alon Goren
Professor Miles Wilkinson

2022

Copyright
Robert Eaman Morey, 2022
All rights reserved.

The dissertation of Robert Eaman Morey is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To Leif and Per

EPIGRAPH

When you get tired...

keep going.

—YNWA

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to apologize to those whom I have forgotten to mention in the upcoming sentences. I am truly sorry for the unintentional oversight. I would like to thank those who have helped me arrive at this milestone. I would first like to thank my family for supporting, encouraging, and believing in me throughout my life. Without their love and support, I would not be authoring this dissertation today. They are the broad shoulders I continue to stand on and any of my accomplishments are theirs as well.

I would like to sincerely thank my advisors Professor Louise Laurent and Professor Mana Parast. I would like to express my gratitude for Dr. Laurent's support, guidance, encouragement, and belief in me throughout my very long, post-graduate career. Her enthusiasm for science and uncovering nature's truths through data is infectious. Her ability to analytically and creatively extract meaning from data has inspired and molded me into the scientist I am today. Likewise, I would like to thank Dr. Parast for her support and guidance throughout my PhD. Her passion for trophoblast development is contagious and her seemingly endless knowledge of placental biology was invaluable to my graduate work. However, most importantly, I would like to thank them both for being kind, empathetic, and thoroughly good people. Although perhaps inadvertently, they have taught me priceless lessons about management, teamwork, and how to set and achieve goals. I can't thank them enough for the mentorship and guidance they have provided me.

Additionally, I would like to express my gratitude to my dissertation committee, Professor Heidi Cook-Andersen, Professor Melissa Gymrek, Professor Alon Goren, and Professor Miles Wilkinson, for their input, advice, and recommendation on the direction of my research. I'm indebted to Professor Goren for his mentorship and guidance, especially while rotating through his lab, and to Professor Gymrek for her kindness and patience despite the havoc I caused on her lab's server. I would also like to thank Professor Cook-Andersen and Professor Wilkinson not only for their kindness and advice, but also for making the southeast corner of the fourth floor an enjoyable, collaborative, and productive environment to conduct my research. Advice and help

from the Cook-Andersen and Wilkinson labs have been instrumental in my success.

Lastly, I would like to thank my many past and present colleagues in the Laurent and Parast labs. Their help has been invaluable throughout my scientific career. Specifically, I would like to thank Dr. Ileana Slavin, for her cell and molecular biology help, Dr. Rathi Thiagarajan, for her bioinformatics help and friendship, Trevor Leonardo for his advice and friendship, Dr. Omar Farah for his trophoblast isolations and placental knowledge, and Dr. Mariko Horii for her kindness, determination, hard work, and trophoblast/preeclampsia expertise.

Chapter 2, in part, contains material by Chen Dekel*, Robert Morey*, Jacob Hanna, Louise C. Laurent, Dalit Ben-Yosef, and Hadar Amir (*these authors contributed equally to this work), and is currently being prepared for submission for publication. The dissertation author was one of the primary investigators and primary authors of this paper.

Chapter 3, in part, has been submitted for publication and contains material that appears in, "Derivation of functional trophoblast stem cells from primed human pluripotent stem cells" by Francesca Soncin*, Robert Morey*, Tony Bui, Daniela F. Requena, Virginia Chu Cheung, Sampada Kallol, Ryan Kittle, Omar Farah, Morgan Meads, Donald Pizzo, Mariko Horii, Kathleen Fisch, and Mana Parast (*these authors contributed equally to this work). The dissertation author was one of the primary investigators and authors on this paper.

Chapter 4, in part, is an adapted reprint of the material, "Transcriptomic Drivers of Differentiation, Maturation, and Polyploidy in Human Extravillous Trophoblast" by Robert Morey, Omar Farah, Sampada Kallol, Daniela F. Requena, Morgan Meads, Matteo Moretto-Zita, Francesca Soncin, Louise C. Laurent, and Mana Parast published in Frontiers in Cell and Developmental Biology, 2021. The dissertation author was the primary investigator and author of this paper.

Chapter 5, the conclusion, contains portions of adapted material currently being prepared for submission for publication by Chen Dekel*, Robert Morey*, Jacob Hanna, Louise Laurent, Dalit Ben-Yosef, and Hadar Amir (*These authors contributed equally to this work). The

dissertation author was one of the primary investigators and authors of this paper. Additionally, chapter five contains portions of adapted material submitted for publication in Stem Cell Reports by Francesca Soncin*, Robert Morey*, Tony Bui, Daniela F. Requena, Virginia Chu Cheung, Sampada Kallol, Ryan Kittle, Omar Farah, Morgan Meads, Donald Pizzo, Mariko Horii, Kathleen M. Fisch, Mana M. Parast (*These authors contributed equally to this work). The dissertation author was one of the primary investigators and authors of this paper. Lastly, chapter 5 contains portions of adapted reprint of material "Transcriptomic Drivers of Differentiation, Maturation, and Polyploidy in Human Extravillous Trophoblast" by Robert Morey, Omar Farah, Sampada Kallol, Daniela F. Requena, Morgan Meads, Matteo Moretto-Zita, Francesca Soncin, Louise C. Laurent, and Mana M. Parast published in *Frontiers in Cell and Developmental Biology* 2021. The dissertation author was one of the primary investigators and authors of this paper.

VITA

1999-2004	Point Loma Nazarene University - B.A., Biology
2009-2011	San Diego State University – M.S., Molecular Biology
2011-2014	San Diego State University – M.B.A., Finance concentration
2017-2022	University of California San Diego – Ph.D., Bioinformatics and Systems Biology

PUBLICATIONS

Morey R, Farah O, Kallol S, Requena D, Meads M, Moretto-Zita M, Soncin F, Laurent LC, Parast MM. Transcriptomic Drivers of Differentiation, Maturation, and Polyploidy in Human Extravillous Trophoblast. Frontiers in Cell and Developmental Biology. 2021 Sep 3; 9(2269).

Horii M, **Morey R**, Bui T, Touma O, Nelson KK, Cho HY, Rishik H, Laurent LC, Parast MM. Modeling preeclampsia using human induced pluripotent stem cells. Sci Rep. 2021 Mar 15;11(1):5877.

Yokoi A, Villar-Prados A, Oliphint PA, Zhang J, Song X, De Hoff P, **Morey R**, Liu J, Roszik J, Clise-Dwyer K, Burks JK, O'Halloran TJ, Laurent LC, Sood AK. "Mechanisms of nuclear content loading to exosomes"; Sci Adv. 2019 Nov 5;(11).

Bauman N, Akella S, Hann E, **Morey R**, Schwartz AS, Brown R, Richardson TH; "Next-Generation Sequencing of Haematococcus lacustris Reveals an Extremely Large 1.35-Megabase Chloroplast Genome"; Genome Announc; 2018 Mar 22;6(12).

Amir H, Touboul T, Sabatini K, Chhabra D, Garitaonandia I, Loring JF, **Morey R**, Laurent LC.; "Spontaneous Single-Copy Loss of TP53 in Human Embryonic Stem Cells Markedly Increases Cell Proliferation and Survival"; Stem Cells; 2017 Apr;35; (4):872-885.

Wolf DP, **Morey R**, Kang E, Ma H, Hayama T, Laurent LC, Mitalipov S.; "Embryonic Stem Cells Derived by Somatic Cell Nuclear Transfer: A Horse in the Race?"; Stem Cells; 2016 Sept 10.

Ma H, Marti Gutierrez N, **Morey R**, Van Dyken C, Kang E, Hayama T, Lee Y, Li Y, Tippner-Hedges R, Wolf DP, Laurent LC, Mitalipov S; "Incompatibility between Nuclear and Mitochondrial Genomes Contributes to Interspecies Reproductive Barrier"; Cell Metabolism; 2016 Aug 9; 24, 1-12.

Sancho-Martinez I, Nivet E, Xia Y, Hishida Y, Aguirre A, Ocampo A, Ma L, **Morey R**, Krause M, Zembrzycki A, Ansorge O, Vazquez-Ferrer E, Dubova I, Reddy P, Lam D, Hishida Y, Wu M, Rodriguez C, O'Leary D, Wahl G, Verma I, Laurent L, and Belmonte JC.; "Establishment of human iPSC-based models for the study and targeting of glioma initiating cells"; Nature Communications; 2016 Feb 22, 22;7:10743.

Mora-Castilla S, Cuong T, Vaezeslami S, **Morey R**, Srinivasan S, Dumdie J, Cook-Andersen H, Jenkins J, Laurent L; "Miniaturization technologies for efficient single-cell library preparations and Next-Generation Sequencing"; Journal of Laboratory Automation; 2016 Feb 18.

Ma H, Folmes CD, Wu J, **Morey R**, Mora-Castilla S, Ocampo A, Ma L, Poulton J, Wang X, Ahmed R, Kang E, Lee Y, Hayama T, Li Y, Van Dyken C, Gutierrez NM, Tippner-Hedges R, Koski A, Mitalipov N, Amato P, Wolf DP, Huang T, Terzic A, Laurent LC, Izpisua Belmonte JC, Mitalipov S; "Metabolic Rescue in Pluripotent Cells from Patients with mtDNA Disease"; Nature; 2015 Aug 13; 524(7564):234-8.

Szabo L, **Morey R**, Palpant NJ, Wang PL, Afari N, Jiang C, Parast MM, Murry CE, Laurent LC, Salzman J; "Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development"; Genome Biology; 2015 June 16.

Gonzalez R, Garitaonandia I, Crain A, Poustovoitov M, Abramihina T, Noskov A, Jiang C, **Morey R**, Laurent LC, Elsworth JD, Snyder EY, Redmond DE Jr, Semechkin R; "Proof of concept studies exploring the safety and functional activity of human parthenogenetic-derived neural stem cells for the treatment of Parkinson's disease"; Cell Transplantation; 2015 Mar 24; 24(4):681-90.

Garitaonandia I, Amir H, Boscolo FS, Wambua GK, Schultheisz HL, Sabatini K, **Morey R**, Waltz S, Wang YC, Tran H, Leonardo TR, Nazor K, Slavin I, Lynch C, Li Y, Coleman R, Gallego Romero I, Altun G, Reynolds D, Dalton S, Parast M, Loring JF, Laurent LC; "Increased Risk of Genetic and Epigenetic Instability in Human Embryonic Stem Cells Associated with Specific Culture Conditions"; PLoS One; 2015 Feb 25; 10(2).

Ma H*, **Morey R***, O'Neil RC, He Y, Daughtry B, Schultz MD, Hariharan M, Nery JR, Castanon R, Sabatini K, Thiagarajan RD, Tachibana M, Kang E, Tippner-Hedges R, Ahmed R, Gutierrez NM, Van Dyken C, Polat A, Sugawara A, Sparman M, Gokhale S, Amato P, Wolf DP, Ecker JR, Laurent LC, Mitalipov S; "Abnormalities in human pluripotent cells due to reprogramming mechanisms."; Nature; 2014 Jul 10; 511(7508):177-83.; ***Co-first authors**.

Morey R, Laurent LC; "Getting off the ground state: X chromosome inactivation knocks down barriers to differentiation"; Cell Stem Cell; 2014 Feb 6; 14(2):131-2.

Thiagarajan R, **Morey R**, Laurent LC; "The Epigenome in Pluripotency and Differentiation" Epigenomics; 2014 February; 6(1):121-37.

Edwards RA, Haggerty JM, Cassman N, Busch JC, Aguinaldo K, Chinta S, Vaughn MH, **Morey R**, Harkins TT, Teiling C, Fredrikson K, Dinsdale EA.; "Microbes, metagenomes and marine mammals: enabling the next generation of scientist to enter the genomic era"; BMC Genomic; 2013 Sept 4.

Jones JC, Sabatini K, Liao X, Tran HT, Lynch CL, **Morey RE**, Glenn-Pratola V, Boscolo FS, Yang Q, Parast MM, Liu Y, Peterson SE, Laurent LC, Loring JF, Wang YC; "Melanocytes derived from transgene- free human induced pluripotent stem cells"; Journal of Investigative Dermatology; 2013 March 20.

Laurent LC, Ulitsky I, Slavin I, Tran H, Schork A, **Morey R**, Lynch C, Harness JV, Lee S, Barrero MJ, Ku S, Martynova M, Semechkin R, Galat V, Gottesfeld J, Izpisua Belmonte JC, Murry C, Keirstead HS, Park HS, Schmidt U, Laslett AL, Muller FJ, Nievergelt CM, Shamir R, Loring JF; "Dynamic changes in the copy number of pluripotency and cell proliferation in human ESCs and iPSCs during reprogramming and time in culture."; Cell Stem Cell; 2011 Jan 7; 8(1):106-18.

Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, Staker B, Pant KP, Baccash J, Borcherding AP, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert JC, Hacker CR, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchyk V, Koenig M, Kong C, Landers T, Le C, Liu J, McBride CE, Morenzoni M, **Morey RE**, Mutch K, Perazich H, Perry K, Peters BA, Peterson J, Pethiyagoda CL, Pothuraju K, Richter C, Rosenbaum AM, Roy S, Shafto J, Sharanhovich U, Shannon KW, Sheppy CG, Sun M, Thakuria JV, Tran A, Vu D, Zaranek AW, Wu X, Drmanac S, Oliphant AR, Banyai WC, Martin B, Ballinger DG, Church GM, Reid CA; "Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays"; Science; 2010 Jan 1; 327(5961):78-81.

ABSTRACT OF THE DISSERTATION

Understanding trophoblast development using human embryonic stem cells

by

Robert Eaman Morey

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California San Diego, 2022

Professor Louise Laurent, Chair
Professor Mana Parast, Co-Chair
Professor Melissa Gymrek, Co-Chair

The human placenta is a transient organ that mediates essential feto-maternal interactions that are critical for a successful pregnancy. Trophoblast cells are specialized placental epithelial cells that arise from the trophectoderm, the first lineage to segregate during embryonic development [139]. Impaired trophoblast development and function is associated with fetal and neonatal, as well as maternal, morbidity and mortality, and has been correlated with diseases later in life [19]. Due to fundamental differences between humans and model organisms, and the ethical and safety concerns of experimental access to early human embryos, human pluripotent stem cells

(hPSCs) and trophoblast stem cells (hTSCs) are valuable tools for investigating early development [168].

This dissertation extends early human placental development research to further characterize the genomic and transcriptomic landscape of hPSCs, hTSCs, and primary trophoblast with the goal of using stem cells to model early human placental development. In the following chapters, I systematically characterize trophoblast precursor cells, hPSCs and hTSCs, and primary trophoblast to increase our understanding of early human placental development and provide a framework for future efforts to model this development. First, I use bioinformatic and experimental methods to characterize two quasi-stable pluripotent substates, termed the "naïve" and "primed" pluripotent states. I highlight the genomic, epigenomic, and transcriptomic differences between the two substates. In the next chapter, I show that hTSCs can be derived from not only naïve pluripotent stem cells but also from the primed substate. Finally, I present the results from comprehensive profiling of first trimester and term extravillous trophoblast (EVT), the invasive trophoblast cells that anchor the placenta to the uterine wall. Overall, this dissertation describes a model of early human placental development and provides a better understanding of normal EVT formation and function, therefore providing future studies an *in vitro* hPSC to EVT framework to build upon.

Chapter 1

Introduction

1.1 Background

Over 70% of human pregnancies result in miscarriage, with a significant proportion occurring in the peri-implantation period [83]. It is during this period that the very first lineage specification occurs, leading to separation of the epiblast, cells which will give rise to the embryo-proper, and trophectoderm (TE), cells which give rise to the epithelial portion of the placenta. However, these events are poorly understood, particularly in human. Overall, the most significant barrier to progress in this field is the lack of experimentally tractable *in vivo* small animal models or *in vitro* cell culture models that accurately represent early human placental development. A model system that recapitulates key human-specific features, and could be manipulated in order to study the effects of genetic and environmental perturbations on placental development, would prove to be invaluable to studying normal and abnormal early placental development and placental-linked pregnancy disease.

Mouse models have been critical in elucidating early developmental events, including specific stem cell states and the role of key genes in the establishment and maintenance of these states. Using mouse models, several groups have shown that there are at least two main

substates of pluripotency, naïve and primed, which represent pre- and post-implantation epiblast, respectively. However, there are numerous shortcomings with using murine systems to study early embryonic development. The dissimilarity between mouse and human pluripotent stem cells (PSCs) has been widely documented [40, 212, 166], including large differences between the species in these two pluripotent substates. In addition, there are numerous known differences in placental morphology, trophoblast cell types and function, and placenta-based pregnancy complications [120, 30, 42]. These differences are particularly stark with respect to the early development of the placenta, including the timing of, and genes involved in, trophoblast lineage specification [21, 168, 202].

Until recently, *in vitro* modeling of placental development was limited to immortalized or transformed versions of already lineage-committed cells of either villous (JEG3, BeWo) or extravillous (HTR8) trophoblast [87, 22, 97]. Recently, conditions were established for the derivation of bona fide human trophoblast stem cells (hTSCs) from blastocyst-stage embryos or early first trimester human placentae [142]. These cells can be propagated indefinitely, and are multipotent, with the ability to differentiate into both syncytiotrophoblast (STB) and extravillous trophoblast (EVT). Although these cells represented an important step toward the establishment of a model of placental development, the fact that these cells were derived from preimplantation embryos or first trimester placentae means that they cannot be obtained from pregnancies for which the clinical outcome is known. However, using the same culture conditions used for the establishment and maintenance of these "primary" hTSCs, we and others have been able to transition naïve hPSCs directly to a hTSC-like state (hPSC-TSCs), and primed hPSCs through a trophoblast progenitor-like intermediate to a similar hPSC-TSC state.

The trophectoderm cell lineage, the lineage that generates trophoblast stem cells and is crucial to the formation of the blastocyst and critical for uterine implantation, undergoes lineage restriction shortly after segregation from the undifferentiated inner cell mass, the cells which give rise to all tissues of the embryo proper [59, 140]. In mouse, naïve stem cell lines derived

directly from the pre-implantation epiblast are unable to make trophoblast derivatives [23, 135], in line with the canonical framework of early trophectoderm lineage restriction. Recent reports indicated that the mouse model of trophectoderm lineage restriction differs from human and that both naïve and primed hPSCs can be converted to trophoblast [49, 39, 63, 31, 116, 210]. Despite this recent progress in modeling development using hPSCs, several outstanding questions remain. First, the question of whether the human primed stem cell substate can be converted to bona vide hTSCs and if so, whether it is therefore better to derive and culture hPSCs in the naïve or primed substate. Second, how best to differentiate hPSC-TSCs so that they accurately represent primary trophoblasts. Therefore, to realize the potential of creating an hPSC *in vitro* model that accurately represents human placentae, it is critical to characterize both hPSCs, the precursors of hTSCs, and primary trophoblasts, the differentiated derivatives of hTSCs.

1.2 Outline

Chapter 2 profiles the genetic, epigenetic, and transcriptomic landscape of isogenic cell lines derived in naïve culture conditions and either kept in naïve conditions or transitioned to the primed substate.

Chapter 3 describes the derivation of functional trophoblast stem cells from primed human pluripotent stem cells.

Chapter 4 characterizes primary trophoblast cells with a focus on extravillous trophoblast and the transcriptomic drivers of trophoblast differentiation.

Chapter 2

Virgin naïve hESCs can be stabilized on the continuum between naïve and primed states as demonstrated by detailed transcriptomic and epigenomic profiling

2.1 Introduction

The appeal of human pluripotent stem cells (hPSCs) arises from their capacity for essentially limitless proliferation and self-renewal, as well as their receptivity for directed differentiation to a broad range of cell types in response to the appropriate lineage-specific cues [46]. Pluripotency, the capacity to differentiate into all three embryonic germ layers, is a transient state in the natural course of embryonic development displayed by inner cell mass (ICM) cells in preimplantation blastocysts. However, it can be captured *in vitro* in the form of embryonic stem cells (ESCs) [141] cultured from ICM cells or by reprogramming somatic cells into induced pluripotent stem cells (iPSCs) [177]. In both cases, the resulting hPSCs must be maintained in culture conditions

that stabilize their pluripotency.

After the first report of *in vitro* conditions for establishment and maintenance of hESCs [184], it was quickly recognized that hESCs differed markedly from naïve mouse embryonic stem cells (mESCs), derived from the ICM of the murine pre-implantation blastocyst, and that in important ways, hESCs are more similar to mouse epiblast stem cells (mEpiSCs), derived from the murine early post-implantation epiblast. Interestingly, hESCS are derived from the ICM of preimplantation embryos, and so it was immediately recognized that the primed state observed for hESCs might be due to dictated by the culture conditions applied to them, rather than their embryonic origin. Several groups have since applied knowledge regarding differences in signaling pathways modulated by mESCs and mEpiSCs culture conditions to promote the naïve or primed phenotypes of hESC cultures, respectively [58, 212, 64, 78, 180, 170, 178, 183, 32, 35, 208, 52, 157]. Like mESCs, naïve human ESCs have been shown to possess characteristics such as increased proliferation and higher clonogenicity [218, 52, 209, 203, 178]. The majority of existing hESC lines were derived in primed conditions, and it has been shown that these primed hESCs can be transitioned to a naïve state by transferring them to naïve conditions (we term these cultures "converted naïve"). A small number of hESC lines were directly derived in naïve conditions. Notably, we were the first to derive naïve hESC lines directly from human blastocysts (we term these lines "virgin naïve") [58].

Recently, there is evidence to suggest that early development exists along a pluripotency continuum and that the various hPSC media stabilize cells at different points along this continuum [166, 40, 224]. Differences in differentiation potential between naïve and primed hPSCs may drive the selection of a particular substate as the optimal starting point for the production of cells for research or therapeutic applications [78, 58, 224]. Recently, Lee et al. reported that lineage-specific differentiation is influenced by the hPSC substate [108].

In this chapter, we compared for the first time the cellular phenotype and transcriptomic, genomic, and epigenomic characteristics over time in culture of matched isogenic pairs of naïve

and primed hESC cultures originating from the same four virgin naïve hESC lines derived in our laboratory. As expected, early passage virgin naïve hESCs displayed shorter doubling times, greater percentages of cells in S-phase, and higher clonogenicity compared to their primed counterparts. The transition from the naïve to the primed state was characterized by changes in the transcriptome and increased DNA methylation. Compared to data from previously reported naïve and primed hESCs, the transcriptomic and epigenomic profiles of our cultures suggested that they occupy moderately naïve and primed states, placing them on either side of the recently described "intermediate" pluripotent state [224].

We observed different phenotypic and genomic changes across time in culture for the naïve and primed hESCs. The naïve cultures were phenotypically stable across time, but displayed increased DNA methylation and decreased expression of previously reported naïve marker genes, while the primed cultures showed increased proliferation at late compared to early passage. With increasing time in culture, both the naïve and primed hESC cultures accumulated genetic aberrations. However, only a small subset of late passage cultures in either condition displayed chromosomal or large subchromosomal aberrations, and the large majority of CNVs and SNVs did not map to known functional areas of the genome. Overall, our results suggest that we have established virgin naïve hESCs that have a moderately naïve phenotype that can be transitioned to the primed state in a facile manner, and that this transition is characterized by expected shifts in mRNA and miRNA expression and DNA methylation. Consistent with many prior studies, extended time in culture for our primed hESCs was associated with phenotypic, genomic, epigenomic, and transcriptomic alterations; our naïve hESCs displayed genomic changes consistent with partial loss of naïve marks, but were phenotypically stable. These findings lead us to suggest that regardless of culture conditions, time in culture should be minimized, that there are advantages and disadvantages for both naïve and primed conditions, and therefore the substate characteristics should drive the choice of condition used.

2.2 Methods

2.2.1 Culture of hESC lines

The use of excess IVF-derived embryos following PGD for the generation of hESCs was approved by the Israeli National Ethics Committee (7/04–043) and is in accordance with the guidelines released by the Bioethics Advisory Committee of the Israel Academy of Sciences and Humanities. Virgin naïve hESC lines were derived in our laboratory at Tel-Aviv Sourasky Medical Center (Lis38_N and Lis39_N published in [58], and Lis 45_N and Lis 46_N). Primed WA09 (H9) and WIBR3 hESC lines were kindly provided by WiCell, University of Wisconsin and Whitehead Institute for Biomedical Research, respectively).

Naïve conditions included culture on MEFs in a modification of the naïve human stem cell medium (NHSM) developed by the Hanna laboratory [58], consisting of: 1:1 Neurobasal medium and DMEM/F-12 (Gibco); 1X Pen/Strep (Biological industries); 2 mM L-Glutamine (Biological industries); 1% non-essential amino acids (Biological industries); 0.7 mM sodium pyruvate (Biological industries); 2% ml B27 supplement (in house produced); 0.2% ml defined lipid concentrate (Gibco); 0.05 mM β-ME (Gibco); 12.5 /mug/ml insulin (Sigma-Aldrich); 100 /mug/ml apo-transferrin (Sigma-Aldrich); 0.02 /mug/ml progesterone (Sigma-Aldrich); 16 /mug/ml putrescine (Sigma-Aldrich); 30 /muM sodium selenite (Sigma-Aldrich); 50 /mug/ml L-ascorbic acid 2-phosphate (Sigma-Aldrich); 0.07% BSA (Gibco); 20 ng/ml human LIF (Peprotech); 5 /muM IWR1 (Biotest); 20 ng/ml human activin A (Peprotech); 1.7 ng/ml FGF2 (Peprotech); 0.2 /muM Chir99021 (Axon Medchem); 1 /muM PD0325901 (Axon Medchem); 0.2 /muM BIRB796 (Axon Medchem); 2 /muM SP600125 (Tocris); 2 /muM PKCi (Tocris); 0.4 /muM Thiazovivin (Peprotech); and 1.5 /muM CGP77675 (Axon Medchem). Primed conditions included culturing on mitotically inactivated mouse embryo fibroblast (MEF) feeder layers in primed hESC culture medium, consisting of: DMEM/F-12 with 20% Knockout Serum Replacement (Gibco); 2 mM L-glutamine (Biological industries); 0.1 mM β-ME (Gibco); 1% non-essential amino acids (Bi-

ological industries); 1X Pen/Strep (Biological industries) and 8 ng/ml FGF2 (Peprotech). For both naïve and primed cultures, the medium was changed daily, passaging was performed every 3-5 days using 0.05% trypsin + EDTA (Biological industries), and 5/muM of Y27632 ROCK inhibitor (Axon Medchem) were used for 24hr before and after passaging.

2.2.2 Quantitative RT-PCR

Total RNA was isolated using the RNeasy mini kit (Qiagen), followed by random hexamer-primed reverse transcription using the Superscript IV RT-PCR kit (Invitrogen). Quantitative Real Time PCR (qRT-PCR) was performed using the FAST SYBR Green Master Mix (Quanta bio). Cycling and analysis were performed using a Rotor Gene 6000 Series instrument (Corbett) and its dedicated data analysis software. Standard curves were performed for each gene in every run, and all PCR reactions were performed for two independent experiments with three technical replicates for each experiment. All qRT-PCR assays included no-template control (NTC) and ACTIN served as the control for normalization of target gene expression. Primer sequences are listed in **Table 2.1**. RNA from a primed WIBR3 culture was used as the reference.

Table 2.1: Primer sequences used for RT-qPCR

Gene	Forward	Reverse
ACTIN	CCACGAAACTACCTTCAACTCC	GTGATCTCCTTCTGCATCCTGT
STELLA	GCATGAAAGAAGACCAACAAACA	TTAGACACGCAGAAACTGCAGGG
TFCP2L1	TCCTTCTTTAGAGGAGAAGC	ACCAACGTTGACTGTAATTC
KLF17	AGCAAGAGATGACGATTTTC	GTGGGACATTATTGGGATTC
KLF4	CGCTCCATTACCAAGAGCTCAT	CACGATCGTCTTCCCCTCTTT
KLF5	CACACTGGTGAAAAGCCATACAA	GCCTGTGTGCTTCCGGTAGT
NANOG	GCAGAAGGCCTCAGCACCTA	AGGTTCCCAGTCGGGTTCA

2.2.3 Immunofluorescence

Cells were grown on Matrigel-coated glass cover slips (13 mm; Marienfeld) in 24-well plates, fixed with 4% paraformaldehyde, and incubated in blocking solution (2.5% BSA) with 0.1% Triton to enable staining of intracellular markers. Cells were incubated with primary antibody for 1 hr at RT, washed and incubated with secondary antibodies for 1 hr at RT, counterstained with DAPI for nuclear staining, and imaged using an Olympus IX51 inverted light microscope. The following antibodies were used at the indicated dilutions: mouse anti-OCT4 (sc-5279, Santa-Cruz, 1:60), mouse anti-SSEA4 (CST-4755S, cell signaling technology, 1:200), mouse anti-TRA-1-60 (ab16288, Abcam, 1:200) and rabbit anti-TFE3 (HPA023881, Sigma-Aldrich, 1:60).

2.2.4 Teratoma formation

hESCs were harvested, resuspended in their respective medium condition with 10% Matrigel and 20 uM ROCKi, and injected subcutaneously into 6-8 week-old NSG mice (Jackson laboratories). Teratomas generally developed within 7-10 weeks and animals were sacrificed before tumor size exceeded 1.5 cm in diameter. Teratomas were dissected and prepared for conventional FFPE and H and E histology. All animal experiments were conducted according to institutional guidelines under approval by the Weizmann Institute IACUC (approval 00960212-3).

2.2.5 Proliferation and clonogenicity assays

Matched sets of naïve and primed hESCs were seeded onto 6-well plates (200,000 cells/well on Matrigel coated wells). Two and four days following plating, cells were collected using trypsin, and the doubling time was calculated as duration (hours)×log(2)/(log2(final concentration)-log(initial concentration)) [209] using number of cells at day 4 normalized by number of cells at day 2.

Cell cycle distribution was determined using the Click-iT EdU Alexa Fluor 647 Flow Cytometry Kit (Life Technologies) according to the manufacturer's protocol. Cells were also stained with FxCycle Violet (Invitrogen) for total DNA content and analyzed using a BD FACSCanto II Flow Cytometer (BD Biosciences) with BD FACSDiva Software (BD Biosciences).

Clonogenicity was analyzed following plating of 100 cells/well on MEF coated 24-well plates. After 7 days, colonies were stained for alkaline phosphatase using the AP staining kit (Stemgent).

2.2.6 Karyotyping

Cells were incubated with 100 ng/ml colcemid (Biological Industries) for 30 min at 37°C, collected by trypsin, incubated with 0.075 mol/L potassium chloride for 10 min at 37°C, fixed with methanol and acetic acid (1:3) and dropped onto glass slides. Karyotype analysis of chromosome spreads was determined by Giemsa staining of at least 20 different metaphase-stage cells for each culture.

2.2.7 DNA isolation and chromosomal microarray analysis (CMA)

Genomic DNA was extracted from samples using the DNeasy Blood and Tissue Kit (Qiagen). The DNA was amplified, labeled, and hybridized to a 24sure V3 microarray (Illumina) according to the manufacturer's protocol. Scanning was performed using Agilent G2565CA scanner and the arrays were analyzed using the BlueFuse Multi software. The detected CNVs were interpreted by referring to key public databases (ISCA, DGV, Ensembl, and Decipher).

2.2.8 Whole genome sequencing (WGS)

WGS libraries were constructed using the Kapa HyperPlus Kit (Roche Holding AG). Briefly, up to one ug of EDTA-free dsDNA was incubated with the fragmentation enzyme for ten

minutes at 37° C. The fragmented samples were then end-repaired and A-Tailed and Illumina indexed adapters were ligated onto the ends of the dsDNA. The libraries were then quanted (Qubit, ThermoFisher Scientific) and run on the BioAnalyzer (Agilent Technologies Inc.) for quality control purposes. The libraries were then size selected using Ampure XP beads (Beckman Coulter) using a 0.65 Ampure XP to DNA ratio followed by a 0.9 ratio to achieve an average library size of about 400 bp. Due to small input amounts for several of the libraries, library amplification (4 cycles) was performed on a selection of samples. Libraries were sequenced using paired-end 100 bp reads on the NovaSeq S4, and six of the 16 samples were re-sequenced on the HiSeq 4000 to achieve greater depth. Samples were sequenced to an average depth of 27x and assessed for quality using FASTQC (v. 0.11.8). Reads were quality trimmed using Trim Galore (v. 0.4.1) with 30 as the quality cutoff. Bowtie 2 (v. 2.3.4.3) [105] was used to map the reads to GRCh38 (v. GCA_000001405.15) and Picard Tools (v. 2.18.15) was used to fix mate information, merge Bam files for samples with two sequencing runs, and mark duplicates.

2.2.9 DNA-seq single nucleotide variant and InDel calling

SNVs and InDels were identified using the best practice instructions of GATK4 (v. 4.0.11.0) HaplotypeCaller [195]. Bam files were recalibrated using GATK BaseRecalibrator with the "known-sites" dbsnp138, 1000G phase1 high confidence SNPs, and Mills 1000G gold standard indels. The data were then run through the HaplotypeCaller and combined using CombineGVCFs for joint genotyping. Following joint genotyping the SNPs and InDels were recalibrated using the following options for the SNPs: –resource hapmap, known=false,training=true, truth=true, prior =15.0:/hapmap_3.3.hg38.vcf.gz –resource omni, known=false, training=true, truth=true, prior= 12.0:/1000G_omni2.5.hg38.vcf.gz –resource 1000G, known=false, training=true, truth=false, prior=10.0:/1000G_phase1.snps.high _confidence.hg38.vcf.gz –resource dbsnp, known=true, training=false, truth=false, prior=2.0:/Homo _sapiens_assembly38.dbsnp138.vcf -an QD -an FS -an SOR -an MQ -an MQRankSum -an ReadPosRankSum -an DP –mode SNP -tranche 100.0 -tranche

99.9 -tranche 99.0 -tranche 90.0 and for the InDels: –resource mills, known=false, training=true, truth=true, prior =12.0:/Mills_and_1000G_gold _standard.indels.hg38.vcf.gz –resource dbsnp, known=true, training=false, truth=false, prior=2.0: /Homo_sapiens_assembly38.dbsnp138.vcf -an QD -an FS -an SOR -an MQRankSum -an ReadPosRankSum -an DP –mode INDEL -tranche 100.0 -tranche 99.9 -tranche 99.0 -tranche 90.0 –max-gaussians 4. Genotype posteriors using the sample trios as priors was calculated for each trio and possible *de novo* mutations were annotated by VariantAnnotator. Finally, low-quality genotypes ($q < 20$) were removed and high-confidence *de novo* variants were selected. Variants were then annotated using ANNOVAR (v. 2019Oct24) [204]. Filtered variant files were compared using vcftools (v. 0.1.16), vcf-isec. Bedtools (v. 2.25.0) was used to find variants overlapping specific genomic regions.

2.2.10 Copy number variation analysis

Regions of the genome with copy number variations (CNV's) were called using three different algorithms, CNVkit (v. 0.9.6) [179], ERDS (v. 1.1) [228], and PURPLE (v. 2.34) [156]. CNVkit: A joint reference using the parental lines was created using the options: –method WGS, –access access-excludes.GRCh38.bed followed by an estimation of the integer copy number of each segment using cnvkit.py call. The input sample was first deduplicated using the samtools (v. 1.9) command rmdup. The integer copy number output was then filtered for only cnv calls and regions of the same copy number were combined. ERDS: erds_pipeline.pl was used with the filtered vcf files for each sample along with the merged and duplicate marked bam files as input. The output from CNVkit and ERDS were split into deletion and duplication bed files. Bedtools (v. 2.25.0) intersect was then used to look for copy number overlaps between the output of the two CNV calling tools, determine if a duplication or deletion was a *de novo* event, and find CNVs overlapping regions of interest. A CNV was called if it was larger than 2000 bp and found by both CNVkit and ERDS. CNV regions of the same copy number were merged if they were within 100 kb. The filtered CNV calls were then annotated using Homer (v. 4.10.4) [71] annotatePeaks.pl

with hg38 (v. 6.0). PURPLE was used to validate differences in calls between the CMA array and WGS CNV calls.

2.2.11 RNA sequencing

RNA was extracted using the mirVana miRNA Isolation Kit (Life Technologies, Inc.), following the manufacturers' instructions. RNA quality was assessed using a BioAnalyzer 2100 (Agilent Technologies, Inc.). All samples had a RIN greater than 8.5. RNA-seq libraries were constructed in triplicate using the KAPA mRNA HyperPrep Kit (Roche) with 500 ng of input RNA. Libraries were sequenced on a HiSeq 4000 (Illumina, Inc.) with paired-end (2x100 bp) reads. Samples were sequenced to an average depth of 20 million uniquely mapped reads per sample and assessed for quality using FASTQC (v.0.11.8). The reads were mapped to GRCh38.p10 (GENCODE release 26) using STAR (v. 2.7.3a) [47] and annotated using featureCounts (subread v.1.6.3, GENCODE release 26 primary assembly annotation) [113]. The STAR parameters used were: –runMode alignReads –outSAMmode Full –outSAMattributes Standard –genomeLoad LoadAndKeep –clip3pAdapterSeq AGATCGGAAGAGC –clip3pAdapterMMp 1. The featureCounts parameters were: -s 2 -p -t exon -T 13 -g gene_id. Ensembl genes without at least three samples with more than 10 reads were removed from the analysis leaving about 20k genes. BiomaRt (v. 2.42.1) was used to convert Ensembl gene ID's to HUGO gene names. The R (v. 3.6.3) package DESeq2 (v.1.26.0) [117], using a multifactor design formula to account for experimental design variables, was used to perform differential expression analysis and normalize the count matrix. Genes with an adjusted p-value of less than 0.05 were considered significant unless otherwise noted. Data from all lines created for this manuscript were compared to previously published naïve and primed lines listed in (**Table 2.3**). Raw fastq files were downloaded and processed using a similar pipeline to that listed above. Samples sequenced using single end reads were mapped separately from paired-end samples and all samples were combined on Ensembl gene IDs. PCA plots were created using DESeq2, VSD transformed values with limma

removeBatchEffect. Gene Set Enrichment Analysis [173] was performed using the R package fgsea (v. 1.12.0) [100] with the hallmark gene sets [114] and the C5 GO gene sets downloaded from the gsea-msigdb (v. 7.1).

2.2.12 Small RNA-seq

Small RNA-seq libraries were constructed using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (New England Biolabs, Inc., Ipswich, MA). The libraries were pooled, size-selected for products that were 120-135 nucleotides in length using a Pippin Prep with a 3% agarose gel cassette (Sage Science, Beverly, Massachusetts), and run on a MiSeq instrument (Illumina, San Diego, California) at the UC San Diego UCSD Institute for Genomic Medicine (IGM) Genomics Core. Samples that produced adequate numbers of miRNA read counts were then rebalanced to produce similar numbers of miRNA reads and sequenced on a HiSeq 4000 instrument using 1x75 bp reads (Illumina, San Diego, California) at the UC San Diego IGM Genomics Core. The data were trimmed and mapped to GRCh38 using the exceRpt Small RNA-seq Pipeline Workflow implemented in the Genboree Workbench [161]. Micro RNAs were filtered such that at least five reads in at least two samples were retained, resulting in 1035 pass-filter miRNAs. Differential expression was carried out in DESeq2 (v.1.26.0) [117] requiring an adjusted p-value < 0.05 and Log_2 FoldChange > 1 considered differentially expressed unless otherwise noted. As with the long RNA-seq, data from all lines created for this manuscript were compared to previously published naïve and primed lines listed in (**Table 2.4**).

2.2.13 Differential mRNA expression analysis

(**Fig. 2.8A**) Naïve transcriptional changes due to long term culturing, comparing post derivation naïve (p20) samples vs early (p30) and late (p50) naïve samples, were determined by performing differential expression analysis using DESeq2 (v.1.26.0) [117] with an adjusted p-

value < 0.05, a Log_2 fold change > 1, and a base mean > 50. Genes considered to be significantly upregulated/downregulated in p20 naïve samples were required to be upregulated/downregulated in both the early (p20 vs p30) and late (p20 vs p50) differential expression analysis comparisons as well as not upregulated/downregulated in a similar early (p30) primed vs late (p50) primed comparison. The differentially upregulated/downregulated gene lists were then separately input into Enrichr [34, 101, 217] to obtain the gene ontology significantly (adj. p-value < 0.05) enriched biological process terms. The terms were then input into Revigo to summarize and find representative subsets of terms using the semantic similarity algorithm SimRel and using the adjusted p-value with every term [174]. The resulting tree map was then input into CirGO, which allows for the visualization of the most enriched terms in an informative 2D circular graph [100].

2.2.14 Integrative analysis of miRNAs and mRNAs differentially expressed between derivation naïve and early/late naïve cultures

mRNAs that were either upregulated or downregulated in the p20 cultures compared to the later cultures (the same differentially expressed mRNAs identified in the section above, "Differential mRNA expression analysis") were used to create protein-protein interaction networks using the stringApp (v. 1.5.1, confidence (score) cutoff = 0.4, max additional interactors = 0, use smart delimiters) application in Cytoscape (v. 3.8.0). The networks were then clustered using MCL clustering with the clusterMaker2 application (v. 1.3.1, inflation value = 2.0, assumption that edges were undirected, and loops were adjusted before clustering). The largest clusters for the mRNAs that were downregulated and upregulated in the p20 naïve cultures are shown in (**Fig. 2.8B**) and (**Fig. 2.8C**), respectively. Functional enrichment analysis was used to identify pathways enriched for the genes in these clusters.

To integrate the miRNA and mRNA data in relation to the transcriptional changes due to long term culturing, naïve p20 samples vs p30 and p50 naïve samples, we performed differential expression of the small RNA-seq data, requiring miRNAs to be differentially expressed between

both the p20 and p30 timepoints and the p20 and p50 timepoints. Additionally, miRNAs that were differentially expressed between the primed p30 and primed p50 timepoints were filtered out of the resulting miRNA lists. This differentially expressed miRNA list and the differentially expressed mRNA list from the section above ("Differential mRNA expression analysis") were each converted to counts per million and combined by pairing the differentially upregulated mRNA genes with the differentially downregulated miRNAs and vice versa. These combined datasets were then used as input into the gene regulatory network inference algorithm GRNBoost2 using Arboreto (v. 0.1.5) [130]. For each target gene, the algorithm uses a tree-based regression model to predict its expression profile using the expression values of the set of miRNAs. The algorithm outputs importance scores that reflect the degree to which each potential mRNA target is regulated by each miRNAs in the dataset.

For the largest clusters from both the downregulated (**Fig. 2.8B**) and upregulated (**Fig. 2.8C**) p20 mRNA networks, heatmaps comprised of the mRNAs in the clusters plotted against the anticorrelated differentially expressed miRNAs were generated, where each cell is color-coded according to the importance score for its respective miRNA/mRNA pair. Hierarchical clustering was then used to visually identify the miRNAs with the highest composite importance scores across their mRNA targets in each cluster. We then used miRPathDB 2.0 [94] to identify pathways that are predicted to be regulated by these miRNAs, and mapped those to the pathways that are enriched for the mRNA targets in each cluster.

2.2.15 Integrative analysis of miRNAs and mRNAs differentially expressed between naïve and primed cultures

Integration of miRNA and mRNA data in relation to the transcriptional changes between naïve p20 cultures vs late p50 primed cultures was carried out using mirTarBase (Release 8.0) [77, 79]. Directionally opposite differentially expressed miRNAs (adj. p-value < 0.05, Log_2 FoldChange > 1; normalized basemean > 10) and corresponding differentially expressed mRNAs

(target genes) (adj. p-value < 0.05, normalized basemean > 50) were used in this comparison. Pathway analysis on the genes within the largest clustered network and miRNA-targets were analyzed using Enrichr [34, 100, 217].

2.2.16 Mapping of RNA-seq data to single-cell expression data from human preimplantation embryos

RNA-seq samples were compared to single cell expression data from human embryonic stages in a similar manner to Theunissen et al., 2016 [182]. Briefly, raw single cell mRNA-seq counts [220] were downloaded from https://www.ebi.ac.uk/gxa/sc/experiments/E-GEOD-36552/downloads, processed in scanpy (v. 1.3.2) [214] and differential expression was performed in R (v.3.6.3) using the package DESeq2 (v. 1.26.0) for each cell state compared against all other cell states. A gene was determined to be expressed in a specific cell state if it had an adjusted p-value less than 0.05 and a Log_2fold change greater than 3. Genes determined to be expressed in a specific cell state were divided between up and down regulated and compared to the genes that were determined to be up or down regulated in our early naïve cell lines compared to our late primed cell lines using an adjusted p-value cutoff of less than 0.05 and a Log_2fold change greater than 2.

2.2.17 Variant analysis using RNA-seq data

Samples were aligned to the GRCh38 reference genome as described above. Variants were identified using the best practice instructions of GATK4 HaplotypeCaller [195]. Briefly, Picard Tools (v. 2.18.15) was used to first merge bam files and mark duplicates. Next, GATK (v. 4.0.11.0) SplitNCigarReads was used to reformat reads that span introns and base quality recalibration was done to detect and correct for patterns of systematic errors in the base quality scores. The following "known-sites" were used in the BaseRecalibrator step:

Homo_sapiens_assembly38.dbsnp138.vcf, 1000G_phase1.snps.high_confidence.hg38.vcf, and Mills_and_1000G_gold_standard.indels.hg38.vcf. HaplotypeCaller with a –stand-call-conf 20 was then used to call variants. To find high certainty *de novo* variants, the HaplotypeCaller results were filtered not dissimilar to [8]. First, variants were required to be called in two of the three replicates for each sample and not in the low passage naïve sample, unless the samples in question were the low passage naïve samples. Next, only positions with over 20 reads, are uncommon in the general population (allele frequency lower than 0.0001 in the Exome Aggregation Consortium (ExAC Database) (v.0.3) [111]), and cause nonsynonymous single nucleotide variations or stop-gain mutations were assessed. InDels were not used in the analysis. Annotations were obtained using ANNOVAR (v. 2019Oct24) [204] and only variants in genes listed as "Tier 1" in Cancer Gene Census of the Catalogue of Somatic Mutations in Cancer (COSMIC) v. 90 database (https://cancer.sanger.ac.uk/census) [55] with FATHMM scores below –1.5 were considered pathogenic. Single nucleotide variants passing these filters were then checked against DNA-seq reads as well as assessed for possible murine contamination.

2.2.18 DNA Methylation Profiling

DNA from every hESC line and time point was sent to UC San Diego UCSD Institute for Genomic Medicine (IGM) Genomics Core for processing and imaging. The samples were processed according to standard protocols and hybridized to the Illumina Infinium MethylationEPIC array, which interrogates over 850,000 methylation sites across the genome, and imaged on the Illumina iScan (Illumina Inc., San Diego, CA). The data was normalized in R (v. 3.6.3) using the minfi package (v. 1.32.0) [6] with the "funnorm" function. Following normalization, probes that were found to cross-reactive on the EPIC array, probes overlapping genetic variatants at targeted CpG sites, and probes overlapping genetic variatants at single base extension sites for Infinium Type I probes were filtered out [149] leaving a total of 811,063 probes. Annotation BED files were download and analysis was performed using python scripts from the CpGtools package

[210]. Imprinted regions (iDMRs) were taken from [15], and X-chromosome inactivation (XCI) and escape regions (non-XCI) were lifted over to GRCh38 from [14].

2.2.19 Statistical Analysis

For each experiment data were obtained from 2-3 independent biological experiments (with each experiment including at least three replicates). P-values were calculated by paired or unpaired two-tailed Student's t-test using SPSS software, are *P<0.05, **P<0.01, and ***P<0.001.

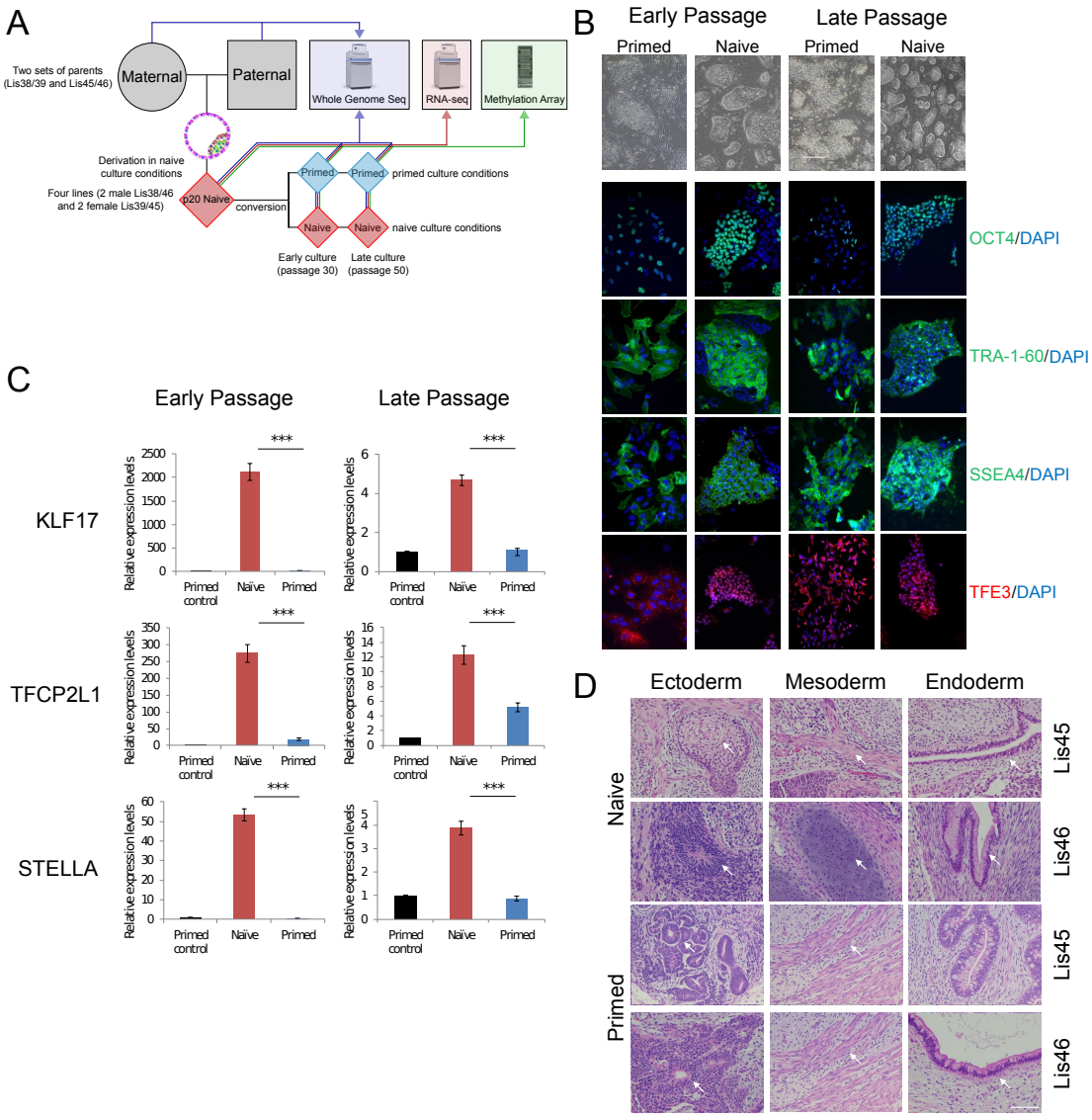
2.3 Results

2.3.1 Confirmation of naïve and primed markers in matched isogenic hESC cultures at early passage

Four hESC lines were derived in our laboratory in naïve conditions, which we term virgin naïve hESCs: Lis38_N, Lis39_N, Lis 45_N and Lis 46_N. The Lis38 and Lis39 lines were derived from blastocysts donated from one set of parents and the Lis45 and Lis46 lines were derived from another set of parents; the Lis38 and Lis46 lines are male and the Lis39 and Lis45 lines are female. After these lines were established, they were expanded, and triplicate biological replicates were collected for genomic analysis at passage 20 (the p20 derivation naïve samples) (**Fig. 2.1A**). Since the genetic background of hPSC lines may result in line-to-line variability, independently of their pluripotent state, we used these original virgin naïve hESC cultures to generate matched isogenic pairs of naïve and primed hESC cultures. This was done at passage 20, by transitioning a portion of each culture to primed conditions, while another portion was kept in naïve conditions. The primed and naïve cultures were grown in parallel for another 10 passages and at p30, phenotyping and collection of samples for genomic analysis was performed (the p30 Early primed and p30 Early naïve samples) (**Fig. 2.1A**). Finally, these cultures were grown for an

additional 20 passages and at p50, phenotyping and collection of samples for genomic analysis was performed (the p50 Late primed and p50 Late naïve samples) (**Fig. 2.1A**).

Figure 2.1: (*next page*): Validation of naïve and primed states at derivation, and at early and late passages.
(A) Study design - four virgin naïve hESC lines, Lis38; Lis39; Lis 45 and Lis 46, were derived and cultured in naïve conditions. At passage 20 (p20, Derivation naïve), half of the cells from each line were transferred to primed conditions (conversion), and half continued to be cultured in naïve conditions. After 10 passages cells in the naïve (p30, Early naïve) and primed (p30, Early primed) cultures were analyzed. Cells were analyzed again at after an additional 20 passages (p50, Late naïve and Late primed). (B) Colony morphology of virgin naïve cells and their isogenic primed counterparts (scale bar-200 μm) and immunofluorescence staining for the pluripotent markers OCT4, TRA-1-60 and SSEA4 and for the naïve marker TFE3 (scale bar-100 μm). (C) Relative mRNA expression levels of the naïve markers KLF17, TFCP2L1 and STELLA as determined by qRT-PCR in naïve (red column) and their primed counterparts (blue column) lines, relative to primed control (black column). All data are expressed as mean $\pm$ SEM of two independent biological experiments, each was conducted in triplicates. ***p$<$.005 by paired t-test. (D) In vivo differentiation is was demonstrated by teratoma formation at p40. Tissues of representing the three germ layers were identified using hemotoxylin and eosin staining and marked by white arrows (scale bar 50 μm).


A
Two sets of parents (Lis38/39 and Lis45/46)
Maternal
Paternal
Whole Genome Seq
RNA-seq
Methylation Array
Derivation in naive culture conditions
Four lines (2 male Lis38/46 and 2 female Lis39/45)
p20 Naive
conversion
Primed
Primed
primed culture conditions
Naive
Naive
naive culture conditions
Early culture (passage 30)
Late culture (passage 50)
B
Early Passage
Late Passage
Primed
Naive
Primed
Naive
OCT4/DAPI
TRA-1-60/DAPI
SSEA4/DAPI
TFE3/DAPI
C
Early Passage
Late Passage
KLF17
TFCP2L1
STELLA
Relative expression levels
***
Primed control
Naïve
Primed
D
Ectoderm
Mesoderm
Endoderm
Naive
Primed
Lis45
Lis46
Lis45
Lis46

The pluripotency of Derivation naïve, Early naïve, Early primed, Late naïve, and Late primed cultures was confirmed by demonstrating expression of pluripotency-associated markers (similar expression of OCT4, TRA-1-60, SSEA4 and NANOG (**Fig. 2.1B, Fig. 2.2A, Fig. 2.3B, Fig. 2.4B**). The naïve status of the Derivation naïve, Early naïve, and Late naïve cultures was confirmed by demonstrating domed colony morphology and nuclear staining of TEF3 (in contrast to flat colony morphology and cytoplasmic staining in the primed cultures) (**Fig. 2.1B, Fig. 2.2B, Fig. 2.3A-B, Fig. 2.4A-B**), and higher expression of KLF17, TFCPL1, STELLA, KLF5 and KLF4 compared to primed cultures (**Fig. 2.1C, Fig. 2.2C, Fig. 2.3C, Fig. 2.4C**). The differentiation capacity of the naïve and primed hESC lines was demonstrated by their ability to generate mature teratomas comprised of cells representing all three germ layers following xenotransplantation into NSG mice (**Fig. 2.1D**).

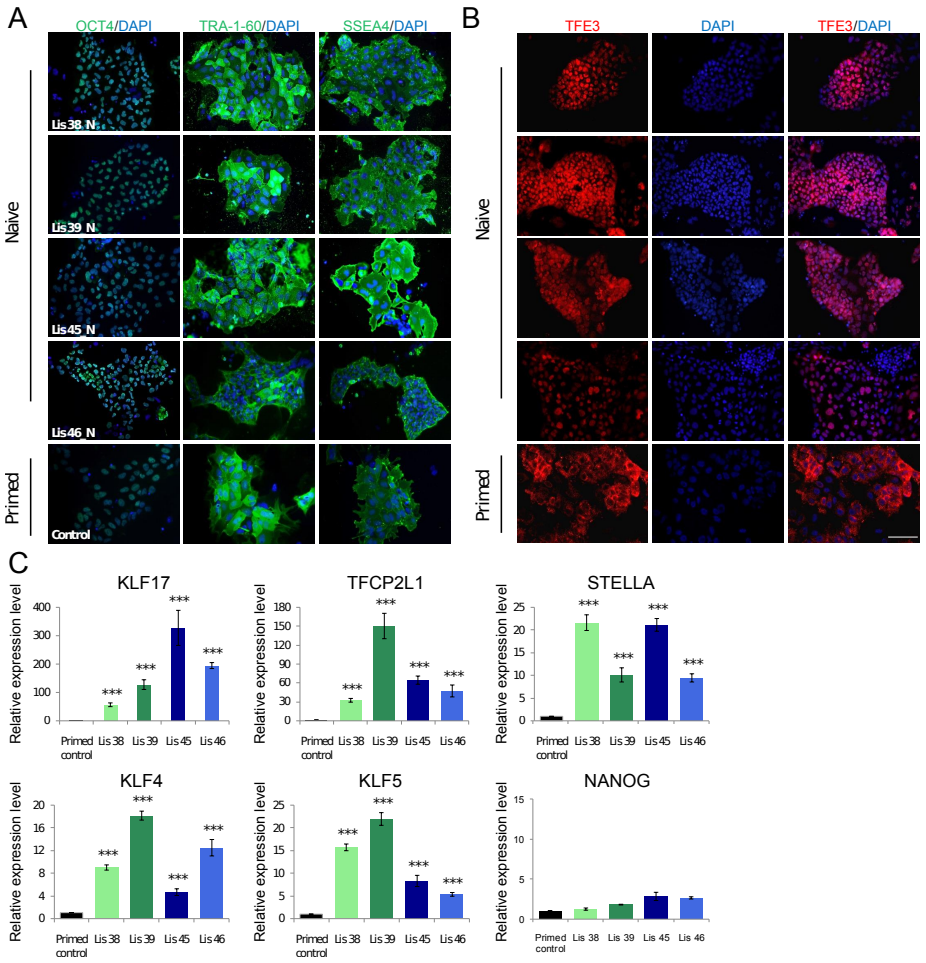


Figure 2.2: Characterization of virgin naïve hESC lines.
(A) Immunofluorescence staining for pluripotency markers, OCT4, TRA-1-60 and SSEA4 (A) and the naïve marker TFE3 (B) in the virgin naïve and primed control hESC lines (scale bar-100 μm). (C) Relative mRNA expression levels of the naïve markers KLF17, TFCPL1, STELLA, KLF5 and KLF4 and a pluripotency marker (NANOG) as determined by qRT-PCR in virgin naïve at p20 (Derivation naïve) and primed control (black column) hESC cultures. All data are expressed as mean ± SEM of two independent biological experiments, each was conducted in triplicates. ***$p<0.005$ by one sample t-test.

Figure 2.3: (*next page*): Validation of naïve and primed states at early passage after conversion. (A) Colony morphology of early passage (p30, Early naïve, top row) and their early passage isogenic primed counterparts (p30, Early primed, bottom row; scale bar-200μm). (B) Immunofluorescence staining for pluripotent markers, OCT4, TRA-1-60 and SSEA4 and for the naïve marker TFE3 in Early naïve (top 4 rows) and isogenic Early primed (bottom 4 rows) hESC cultures (scale bar-100 μm). (C) Relative mRNA expression levels of naïve markers KLF17, TFCP2L1 and STELLA as determined by qRT-PCR in Early naïve (red column) and Early primed (blue column) cultures, relative to primed control (black column). All data are expressed as mean ± SEM of two independent biological experiments, each was conducted in triplicates. *$p<0.05$, **$p<0.01$, ***$p<0.005$ by paired t-test.

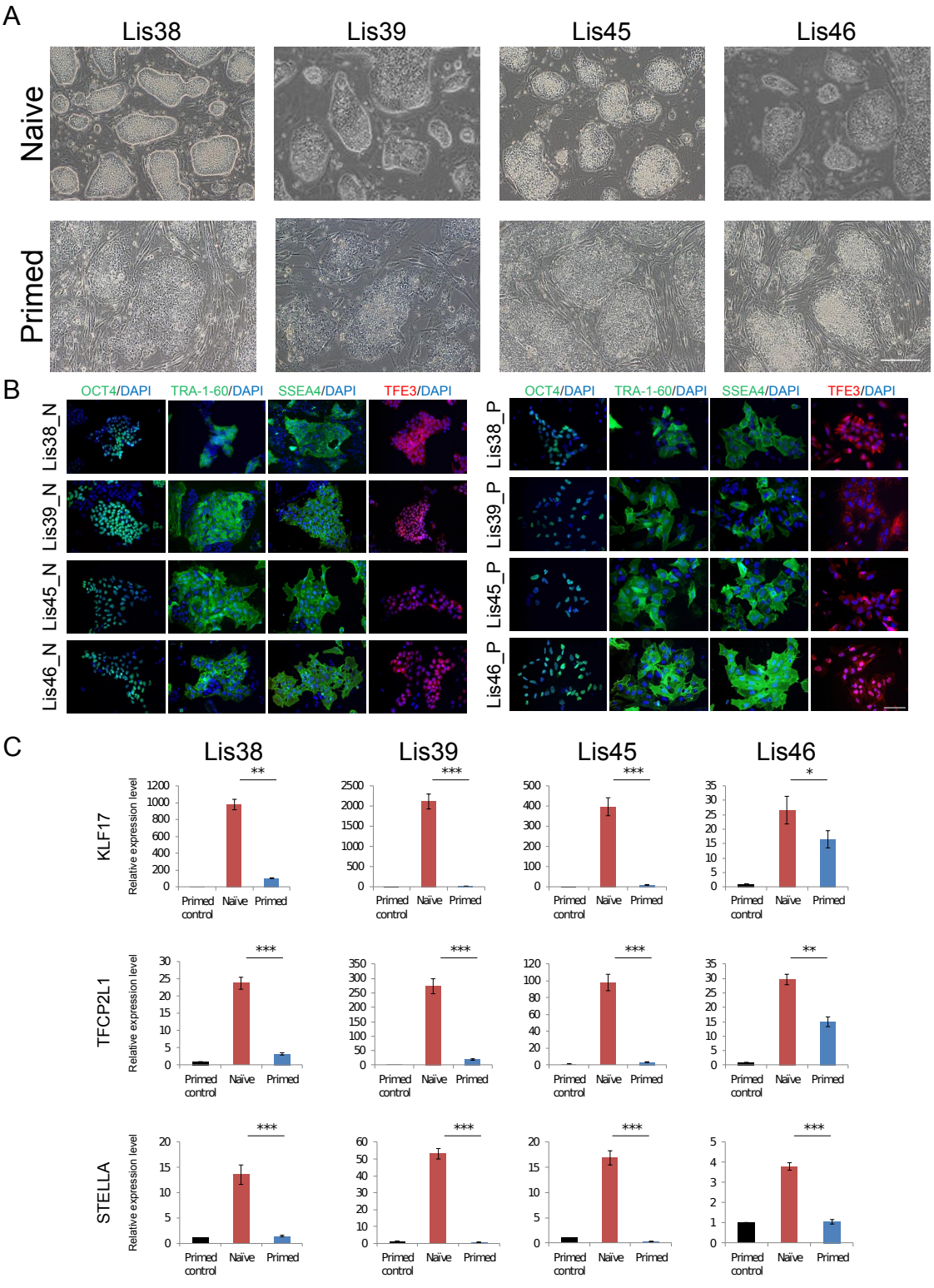
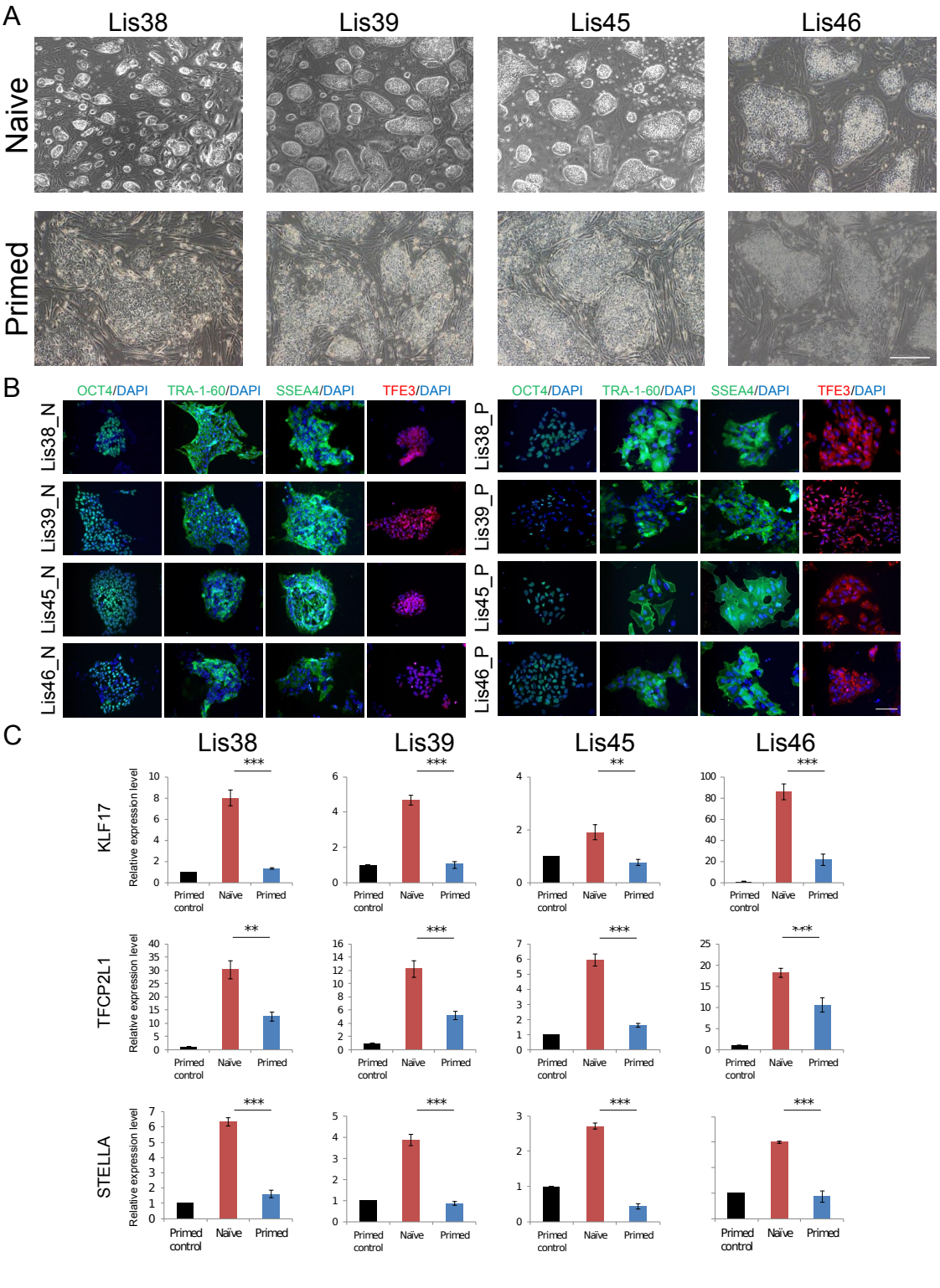

A
Lis38
Lis39
Lis45
Lis46
Naive
Primed
B
OCT4/DAPI
TRA-1-60/DAPI
SSEA4/DAPI
TFE3/DAPI
Lis38_N
Lis39_N
Lis45_N
Lis46_N
Lis38_P
Lis39_P
Lis45_P
Lis46_P
C
KLF17
TFCP2L1
STELLA
Relative expression level
Primed control
Naïve
Primed

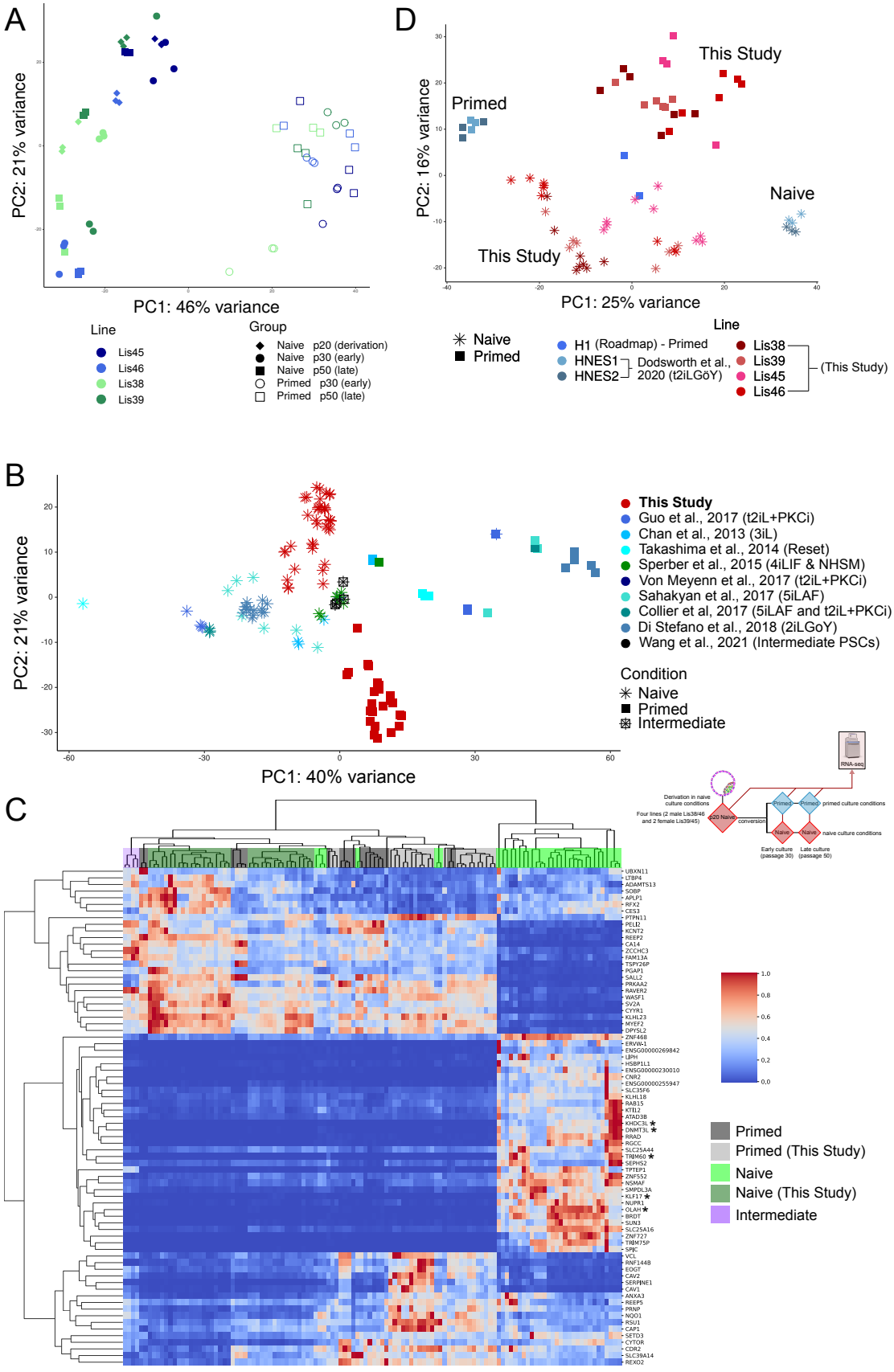
Figure 2.4: (*next page*): Analyzing the naïve and primed states of hESC lines after prolonged culture.
(A) Colony morphology of p50 Late naïve (top row) and isogenic Late primed (bottom row) hESC cultures (scale bar-200 μm). (B) Immunofluorescence staining for pluripotency markers, OCT4, TRA-1-60 and SSEA4 and for the naïve marker TFE3 in naïve and isogenic primed hESCs (scale bar-100 μm). (C) Relative mRNA expression levels of the naïve markers KLF17, TFCP2L1 and STELLA as determined by qRT-PCR in Late naïve (red column) and Late primed (blue column) hESC cultures, relative to primed control cultures (black column). All data are expressed as mean ± SEM of two independent biological experiments, each was conducted in triplicates. *$p<0.05$, **$p<0.01$, ***$p<0.005$ by paired t-test.


A
Lis38
Lis39
Lis45
Lis46
Naive
Primed
B
OCT4/DAPI
TRA-1-60/DAPI
SSEA4/DAPI
TFE3/DAPI
Lis38_N
Lis39_N
Lis45_N
Lis46_N
Lis38_P
Lis39_P
Lis45_P
Lis46_P
C
KLF17
TFCP2L1
STELLA
Relative expression level
Primed control
Naïve
Primed

2.3.2 Transcriptome analysis places the isogenic naïve/primed hESC cultures along the previously reported developmental continuum

Transcriptomic profiling of all four isogenic naïve and primed hESC lines was performed using RNA-seq. Samples were collected at three timepoints: Derivation naïve cultures at p20, Early naïve and primed cultures at p30, and Late naïve and primed cultures at p50. Triplicate cultures in each condition at each timepoint were collected and analyzed, for a total of 60 samples. Principal component analysis (PCA) on normalized data revealed that the cultures separated along the first principal component by pluripotent state (naïve vs primed, (**Fig. 2.5A**)). The second principal component divided our naïve cultures by time in culture, with the p20 (Derivation naïve) cultures clustering above the p50 (Late naïve) cultures for each hESC line, but with the p30 (Early naïve) cultures occupying variable positions (clustering more closely to the p20 cultures for Lis38, more closely to the p50 culture for Lis46, close to both p20 and p50 for Lis45, and far from both p20 and p50 for Lis39) (**Fig. 2.5A**). Differential expression analysis of our cultures resulted in 278 genes upregulated and 668 genes downregulated (adj p-value < 0.05, basemean > 50, paired analysis by line and timepoint) in the naïve cultures compared to their isogenic primed cultures at similar time points (**Fig. 2.6A**). The genes upregulated in our naïve cultures were enriched for the ectoderm differentiation pathway (adj p-value < 0.005) whereas the top enriched pathways for genes downregulated in our naïve cultures were Focal Adhesion and the VEGFA-VEGFR2 Signaling Pathway (**Table 2.2**).

Figure 2.5: (*next page*): Transcriptomic profiling of early and late naïve and primed cultures in the context of previously published reports.
(A) Principal component analysis of all 60 (triplicates of each condition and timepoint) samples using all genes passing filter (see methods). (B) RNA-seq principal component analysis of all 60 (triplicates of each condition and timepoint) samples in this study along with previously published naïve, intermediate, and primed lines (**Table 2.3**) using all genes passing filter (see methods). (C) Heatmap displaying the top 100 genes, ranked by their loadings on the first principal component (see **Fig. 2.5B**). Samples and genes are hierarchically clustered using Euclidean distance. A "*" has been placed next to the five genes previously reported to be naïve markers shared by three different naïve conversion methods. (D) Small RNA-seq principal component analysis of all samples in this study along with previously published naïve and primed lines (**Table 2.4**) using all miRNAs passing filter (see methods).


A
PC2: 21% variance
PC1: 46% variance
Line
Lis45
Lis46
Lis38
Lis39
Group
Naive p20 (derivation)
Naive p30 (early)
Naive p50 (late)
Primed p30 (early)
Primed p50 (late)
D
PC2: 16% variance
PC1: 25% variance
This Study
Primed
Naive
This Study
Naive
Primed
Line
H1 (Roadmap) - Primed
HNES1
HNES2
Dodsworth et al., 2020 (t2iLGöY)
Lis38
Lis39
Lis45
Lis46
(This Study)
B
PC2: 21% variance
PC1: 40% variance
This Study
Guo et al., 2017 (t2iL+PKCi)
Chan et al., 2013 (3iL)
Takashima et al., 2014 (Reset)
Sperber et al., 2015 (4iLIF & NHSM)
Von Meyenn et al., 2017 (t2iL+PKCi)
Sahakyan et al., 2017 (5iLAF)
Collier et al, 2017 (5iLAF and t2iL+PKCi)
Di Stefano et al., 2018 (2iLGoY)
Wang et al., 2021 (Intermediate PSCs)
Condition
Naive
Primed
Intermediate
RNA-seq
Derivation in naive culture conditions
Four lines (2 male Lis38/46 and 2 female Lis39/45)
p20 Naive
conversion
Primed
Primed
primed culture conditions
Naive
Naive
naive culture conditions
Early culture (passage 30)
Late culture (passage 50)
C
Primed
Primed (This Study)
Naive
Naive (This Study)
Intermediate

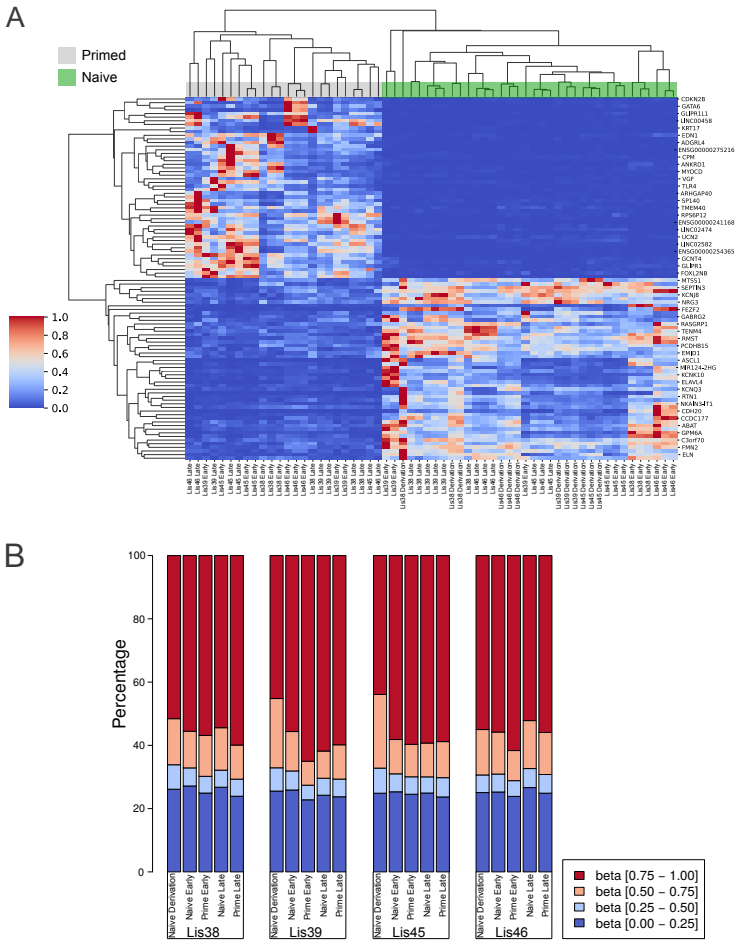


Figure 2.6: Transcriptomic and epigenomic characterization of our naïve and primed cultures. (A) Heatmap displaying the top 100 differentially expressed genes by fold change (50 upregulated in naïve vs 50 upregulated in primed). (B) Global methylation levels from EPIC DNA methylation array. One sample from each line at each condition and timepoint. Methylation levels are displayed as a percentage of four beta value bins (beta values range from 0 to 1 and are used to measure the percentage of methylation. They are the ratio of the methylated probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities))

Previous reports have suggested that instead of two pluripotent states, there exists a developmental continuum between the naïve and primed states [166, 212], including a formative (and intermediate) state [224]. To assess where our naïve and primed cultures lie in relation to previously published human naïve, primed, and intermediate pluripotent stem cell lines, we combined our transcriptomic data with data from several previous studies. Using PCA, the

Table 2.2: Gene set enrichment analysis for upregulated and downregulated genes in the naïve cultures compared to their isogenic primed cultures at similar time points using WikiPathways 2019 in Enrichr.

WikiPathways 2019 – Pathways enriched in genes upregulated in Naive cultures

Term	Adj. p-value	Score
Ectoderm Differentiation WP2858	0.003274123	60.825

WikiPathways 2019 – Pathways enriched in genes downregulated in Naive cultures

Term	Adj. p-value	Score
Focal Adhesion WP306	2.474994E-10	156.06
VEGFA-VEGFR2 Signaling Pathway WP3888	1.314557E-08	105.69
Primary Focal Segmental Glomerulosclerosis FSGS WP2572	1.367432E-06	141.58
Nuclear Receptors Meta-Pathway WP2882	2.848667E-05	45.891
NRF2 pathway WP2884	2.848667E-05	64.988
Lung fibrosis WP3624	5.342485E-05	95.692
Integrin-mediated Cell Adhesion WP185	5.342485E-05	69.769
Ebola Virus Pathway on Host WP4217	5.342485E-05	60.332
Senescence and Autophagy in Cancer WP615	7.130343E-05	64.184
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway WP3932	0.000120781	36.198
Myometrial Relaxation and Contraction Pathways WP289	0.000137519	47.105
Pathogenic Escherichia coli infection WP2272	0.000315784	73.725
Prostaglandin Synthesis and Regulation WP98	0.000357064	81.412
Differentiation Pathway WP2848	0.000540748	71.432
Hepatitis C and Hepatocellular Carcinoma WP3646	0.000605166	68.502
Arrhythmogenic Right Ventricular Cardiomyopathy WP2118	0.000615547	52.877
Androgen receptor signaling pathway WP138	0.000785945	45.278
Endochondral Ossification WP474	0.0008113679	53.952
TGF-beta Receptor Signaling WP560	0.001092899	56.178
MAPK Signaling Pathway WP382	0.001304751	25.837
miR-509-3p alteration of YAP1/ECM axis WP3967	0.002554352	104.6
PI3K-Akt Signaling Pathway WP4172	0.002793695	19.915
Apoptosis Modulation and Signaling WP1772	0.002887426	33.89
Photodynamic therapy-induced AP-1 survival signaling. WP3611	0.002887426	46.622
Nanoparticle-mediated activation of receptor signaling WP2643	0.003291976	65.289
Copper homeostasis WP3286	0.003515558	43.016
Ferroptosis WP4313	0.003515558	49.828
Extracellular vesicles in the crosstalk of cardiac cells WP4300	0.003515558	83.662
Interferon type I signaling pathways WP585	0.004199503	39.792
TGF-beta Signaling Pathway WP366	0.00494669	24.468
Regulation of Actin Cytoskeleton WP51	0.0051118098	22.881
Selenium Micronutrient Network WP15	0.005616929	28.686
Mammary gland development pathway - Puberty (Stage 2 of 4) WP2814	0.00618824	94.086
Osteopontin Signaling WP1434	0.00618824	94.086
Oxidative Stress WP408	0.00618824	47.116

previously published naïve and primed lines separated largely on the first principal component, with the naïve cultures to the left and the primed cultures to the right (**Fig. 2.5B**). Our naïve lines were positioned slightly to the left of hESCs cultured in similar naïve culture conditions (NHSM) or Intermediate conditions (**Fig. 2.5B, Table 2.3**). Our primed cultures were positioned slightly to

the right of the NHSM and Intermediate hESCs. Our naïve and primed lines were also separated from each other on the second principal component. Performing hierarchical clustering using the top 100 genes ranked by their loadings on the first principal component, our naïve and primed lines clustered together with all of the previously published primed lines, the intermediate lines (FTW), the four NHSM derived naïve lines (**Fig. 2.5C, Table 2.3**). These 100 genes include five genes previously reported to be markers shared among three disparate naïve conversion methods (DNMT3L, KHDC3L, TRIM60, KLF17, OLAH) [128].

To focus on the transcriptomic changes that occurred in conversion from our naïve to our primed conditions, we identified the genes that were differentially up- and downregulated in our p20 Derivation naïve cultures compared to our p50 Late primed cultures (adj p-value < 0.05, Log_2 fold change > 2) and asked how many of these genes were also differentially expressed (adj p-value < 0.05, Log_2 fold change > 2) in each stage of human embryonic development compared to all the other stages in a previously reported single-cell RNA-seq dataset [220]. Comparing the gene expression changes between these datasets, we found that our naïve lines most closely resembled the late blastocyst stage (data not shown).

We next asked how similar the miRNA profiles of our matched isogenic naïve and primed cultures were to those in the small number of previously published miRNA datasets from naïve and primed hESC cultures. We performed PCA on our cultures in the context of two primed samples from the Encode project (WA01 hESC line) and six primed and six naïve samples (HNES hESC line) (**Fig. 2.5D, Table 2.4**) [48]. The naïve and primed cultures from the NHES line separated predominantly on the first principal component (PC) and to a lesser extent on the second PC, while the naïve and primed cultures from our lines separated mostly on the second PC. For both the NHES and our lines, the naïve cultures clustered lower on PC2, and the primed cultures clustered higher on PC 2. The primed WA01 cultures were positioned near the middle on both PCs. Differential expression analysis (adj p-value < 0.05) revealed that the top five differentially

Table 2.3: Publicly available RNA-seq samples used in Figure 2.5B

Sample Name	Cell Line	Description	Subset	BioProject	Study	Treatment
ERR1924246	S6EOS	S6EOS_R1_s	Primed	PRJEB20388	Guo, et al., 2017	Primed
ERR1924247	S6EOS	S6EOS_R2_s	Primed	PRJEB20388	Guo, et al., 2017	Primed
ERR1924248	S6EOS	S6EOS_R3_s	Primed	PRJEB20388	Guo, et al., 2017	Primed
ERR1924232	S6EOS	cR_S6EOS_p18_R2	Naive	PRJEB20388	Guo, et al., 2017	tt2iLGo
ERR1924233	S6EOS	cR_S6EOS_p18_R3_s	Naive	PRJEB20388	Guo, et al., 2017	tt2iLGo
ERR1924234	S6EOS	cR_S6EOS_p26_R1_s	Naive	PRJEB20388	Guo, et al., 2017	tt2iLGo
ERR1924235	S6EOS	cR_S6EOS_p26_R2_s	Naive	PRJEB20388	Guo, et al., 2017	tt2iLGo
ERR1924236	S6EOS	cR_S6EOS_p26_R3_s	Naive	PRJEB20388	Guo, et al., 2017	tt2iLGo
ERR361241	WA01	hESCs_RNA_Rep1	Primed	PRJEB4879	Chan et al., 2013	Primed
ERR361243	WA01	hESCs_RNA_Rep3	Primed	PRJEB4879	Chan et al., 2013	Primed
ERR361245	WA01	hESCs_RNA_Rep2	Primed	PRJEB4879	Chan et al., 2013	Primed
ERR361240	WA01	3iL_hESCs_RNA_Rep2	Naive	PRJEB4879	Chan et al., 2013	3iL
ERR361242	WA01	3iL_hESCs_RNA_Rep1	Naive	PRJEB4879	Chan et al., 2013	3iL
ERR361244	WA01	3iL_hESCs_RNA_Rep3	Naive	PRJEB4879	Chan et al., 2013	3iL
ERR590401	WA09	H9_R3	Primed	PRJEB7132	Takashima et al., 2014	Primed
ERR590408	WA09	H9-R1	Primed	PRJEB7132	Takashima et al., 2014	Primed
ERR590410	WA09	H9-R2	Primed	PRJEB7132	Takashima et al., 2014	Primed
ERR590400	WA09	H9_reset_R1	Naive	PRJEB7132	Takashima et al., 2014	t2iLGo+NANOG+KLF2
SRR1561941	WA01	H1 mRNA rep 1	Primed	PRJNA259889	Sperber et al., 2015	Primed
SRR2297450	WA01	H1 4iLIF rep1	Naive	PRJNA259889	Sperber et al., 2015	4iLIF
SRR2297451	WA01	H1 4iLIF rep2	Naive	PRJNA259889	Sperber et al., 2015	4iLIF
SRR2297452	Lis1	Lis1 rep1	Naive	PRJNA259889	Sperber et al., 2015	4iLIF
SRR2297453	Lis1	Lis1 rep2	Naive	PRJNA259889	Sperber et al., 2015	4iLIF
SRR1561942	ELF	ELF rep1	Naive	PRJNA259889	Sperber et al., 2015	2iL-I-F
SRR1561943	ELF	ELF rep2	Naive	PRJNA259889	Sperber et al., 2015	2iL-I-F
SRR4241925	WA09	RNA-seq_hESC_2	Naive	PRJNA342888	Von Meyenn et al., 2016	t2iL+PKCi
SRR4289957	UCLA1	UCLA1_Primed_Rep1	Primed	PRJNA343938	Sahakyan et al., 2017	Primed
SRR4289958	UCLA1	UCLA1_Primed_Rep2	Primed	PRJNA343938	Sahakyan et al., 2017	Primed
SRR4289959	UCLA1	UCLA1_Reprimed	Primed	PRJNA343938	Sahakyan et al., 2017	5i/L/FA
SRR4289950	UCLA1	UCLA1_Clone12_XIST_Pos_Rep1	Naive	PRJNA343938	Sahakyan et al., 2017	5i/L/FA
SRR4289951	UCLA1	UCLA1_Clone12_XIST_Pos_Rep2	Naive	PRJNA343938	Sahakyan et al., 2017	5i/L/FA
SRR4289952	UCLA1	UCLA1_Clone4_LatePassage_XIST_Pos	Naive	PRJNA343938	Sahakyan et al., 2017	5i/L/FA
SRR4289953	UCLA1	UCLA1_Clone4_XIST_Neg_Rep1	Naive	PRJNA343938	Sahakyan et al., 2017	5i/L/FA
SRR4289954	UCLA1	UCLA1_Clone4_XIST_Neg_Rep2	Naive	PRJNA343938	Sahakyan et al., 2017	5i/L/FA
SRR4289955	UCLA1	UCLA1_Clone9_XIST_Pos_Rep1	Naive	PRJNA343938	Sahakyan et al., 2017	5i/L/FA
SRR4289956	UCLA1	UCLA1_Clone9_XIST_Pos_Rep2	Naive	PRJNA343938	Sahakyan et al., 2017	5i/L/FA
SRR5151105	WA09	Primed_Rep1	Primed	PRJNA360413	Collier et al., 2017	Primed
SRR5151106	WA09	Primed_Rep2	Primed	PRJNA360413	Collier et al., 2017	Primed
SRR5151102	WA09	Naive_Rep1	Naive	PRJNA360413	Collier et al., 2017	t2iLGo+NANOG+KLF2
SRR5151103	WA09	Naive_Rep2	Naive	PRJNA360413	Collier et al., 2017	t2iLGo+NANOG+KLF2
SRR5151104	WA09	Naive_Rep3	Naive	PRJNA360413	Collier et al., 2017	t2iLGo+NANOG+KLF2
SRR6748810	WIBR3	WIBR3 Primed	Primed	PRJNA434588	Di Stefano et al., 2018	Primed
SRR6748811	UCLA1	UCLA1 Primed	Primed	PRJNA434588	Di Stefano et al., 2018	Primed
SRR6748812	UCLA4	UCLA4 Primed	Primed	PRJNA434588	Di Stefano et al., 2018	Primed
SRR6748813	UCLA5	UCLA5 Primed	Primed	PRJNA434588	Di Stefano et al., 2018	Primed
SRR6748814	UCLA9	UCLA9 Primed	Primed	PRJNA434588	Di Stefano et al., 2018	Primed
SRR6748815	WIBR3	WIBR3 naive PD03 1mM	Naive	PRJNA434588	Di Stefano et al., 2018	5i/L/FA
SRR6748816	WIBR3	WIBR3 naive PD03 0.5mM	Naive	PRJNA434588	Di Stefano et al., 2018	m5i/L/FA
SRR6748817	UCLA1	UCLA1 naive PD03 1mM	Naive	PRJNA434588	Di Stefano et al., 2018	5i/L/FA
SRR6748818	UCLA1	UCLA1 naive PD03 0.5mM	Naive	PRJNA434588	Di Stefano et al., 2018	m5i/L/FA
SRR6748819	UCLA4	UCLA4 naive PD03 1mM	Naive	PRJNA434588	Di Stefano et al., 2018	5i/L/FA
SRR6748820	UCLA4	UCLA4 naive PD03 0.5mM	Naive	PRJNA434588	Di Stefano et al., 2018	m5i/L/FA
SRR6748821	UCLA5	UCLA5 naive PD03 1mM	Naive	PRJNA434588	Di Stefano et al., 2018	5i/L/FA
SRR6748822	UCLA5	UCLA5 naive PD03 0.5mM	Naive	PRJNA434588	Di Stefano et al., 2018	m5i/L/FA
SRR6748823	UCLA9	UCLA9 naive PD03 1mM	Naive	PRJNA434588	Di Stefano et al., 2018	5i/L/FA
SRR6748824	UCLA9	UCLA9 naive PD03 0.5mM	Naive	PRJNA434588	Di Stefano et al., 2018	m5i/L/FA
SRR6748825	UCLA4	UCLA4 naive TAK-733 0.5mM	Naive	PRJNA434588	Di Stefano et al., 2018	4i/L/FA_TAK-733
SRR6748826	UCLA4	UCLA4 naive Refametinib 0.5mM	Naive	PRJNA434588	Di Stefano et al., 2018	4i/L/FA_Refametinib
SRR6748827	UCLA4	UCLA4 naive Cobimetinib 0.5mM	Naive	PRJNA434588	Di Stefano et al., 2018	4i/L/FA_Cobimetinib
SRR6748828	UCLA4	UCLA4 naive t2iLGoY	Naive	PRJNA434588	Di Stefano et al., 2018	t2iLGoY
SRR12430924	FTW-hiPS#1	FTW-hiPSC#1.r1	Intermediate	PRJNA560892	Wang et al., 2021	FTW
SRR12430925	FTW-hiPS#1	FTW-hiPSC#1.r2	Intermediate	PRJNA560892	Wang et al., 2021	FTW
SRR12430926	FTW-hiPS#2	FTW-hiPSC#2.r1	Intermediate	PRJNA560892	Wang et al., 2021	FTW
SRR12430927	FTW-hiPS#2	FTW-hiPSC#2.r2	Intermediate	PRJNA560892	Wang et al., 2021	FTW

expressed miRNA's upregulated in the HNES naïve cultures compared to HNES primed cultures were miR-143-3p, miR-92b-3p, miR-512-3p, miR-199, and miR-363-3p. Of these, miR-92b-3p and miR363-3p were also significantly upregulated in our naïve cultures compared to our primed cultures.

Table 2.4: Publicly available small RNA-seq samples used in Figure 2.5D.

Sample Name	Cell Line	Description	Subset	BioProject	Study
SRR026663	H1	ENCODE_H1_6	Primed	PRJNA259585	Roadmap Epigenomics
SRR020285	H1	ENCODE_H1_8	Primed	PRJNA34535	Roadmap Epigenomics
GSM3669491	HNES1	HNES1_naive_1	Naive	PRJNA526900	Dodsworth et al., 2020
GSM3669492	HNES1	HNES1_naive_2	Naive	PRJNA526900	Dodsworth et al., 2020
GSM3669493	HNES1	HNES1_naive_3	Naive	PRJNA526900	Dodsworth et al., 2020
GSM3669503	HNES1	HNES1_primed_1	Primed	PRJNA526900	Dodsworth et al., 2020
GSM3669504	HNES1	HNES1_primed_2	Primed	PRJNA526900	Dodsworth et al., 2020
GSM3669505	HNES1	HNES1_primed_3	Primed	PRJNA526900	Dodsworth et al., 2020
GSM3669494	HNES2	HNES2_naive_1	Naive	PRJNA526900	Dodsworth et al., 2020
GSM3669495	HNES2	HNES2_naive_2	Naive	PRJNA526900	Dodsworth et al., 2020
GSM3669496	HNES2	HNES2_naive_3	Naive	PRJNA526900	Dodsworth et al., 2020
GSM3669506	HNES2	HNES2_primed_1	Primed	PRJNA526900	Dodsworth et al., 2020
GSM3669507	HNES2	HNES2_primed_2	Primed	PRJNA526900	Dodsworth et al., 2020
GSM3669508	HNES2	HNES2_primed_3	Primed	PRJNA526900	Dodsworth et al., 2020

It is well established from DNA methylation studies that the preimplantation methylome is hypomethylated [167]. Likewise, human naïve hESCs generated under a variety of conversion methods have reported global CpG methylation levels of about 30% whereas primed methylation levels are generally reported to be close to 80% [40]. To interrogate our cultures' global DNA methylation profiles, we performed whole genome DNA methylation array analysis using EPIC DNA methylation arrays (Illumina, Inc.). We found that the average methylation level of our naïve cultures was close to 60%, approximately 2-4% lower than that of our primed cultures (**Fig. 2.6B**).

2.3.3 Long-term culture has differential effects on the phenotypes and transcriptomes of virgin naïve and isogenic primed hESCs

We observed that the doubling times of early virgin naïve cultures were significantly shorter than those of their isogenic primed counterparts for all four hESC lines in this study (**Fig. 2.7A, left**), consistent with higher proliferation rates. We then used EdU labeling to determine whether the increased proliferation of the virgin naïve cells was associated with a

difference in cell cycle regulation, and found that early virgin naïve cultures had significantly higher fractions of cells in S phase compared to their isogenic primed counterparts (**Fig. 2.7B, left**). Clonogenicity, or the ability to expand from single cells to form colonies, is a useful feature that enables experimental manipulations that require cloning of single cells, such as gene targeting. We determined that for each of our four hESC lines, the clonogenicity of the early virgin naïve culture was higher than that of its isogenic primed counterpart (**Fig. 2.7C, left**).

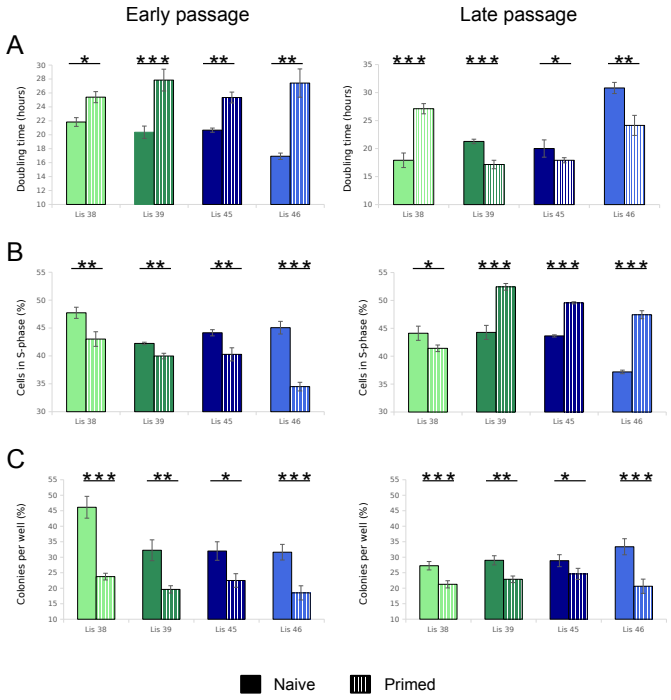


Figure 2.7: Comparing proliferation rate and clonogenicity potential of virgin naïve and isogenic primed hESCs at early and late passage.
(A) Proliferation rate was determined by doubling time assay. (B) Fraction of cells in S-phase was analyzed using EdU incorporation and quantitative DNA staining followed by FACS analysis. (C) Clonogenicity potential determined by alkaline phosphatase staining of colonies 7 days after seeding at low density. All data are expressed as mean ± SEM of two independent biological experiments, each was conducted in triplicates. *p<.05, **p<.01, ***p<.005 by paired t-test.

We then investigated the effect of extended culture on these phenotypic properties for the virgin naïve and primed cultures. The colony formation assay showed that late naïve cells were more clonogenic than their isogenic primed counterparts (**Fig. 2.7C, right**). However, analysis of proliferation rate, as indicated by either doubling time or fraction of cells in S phase,

demonstrated that only one hESC line (Lis 38) maintained a higher proliferation rate in the late naïve culture compared to the primed culture; for the other three lines (Lis 39, Lis 45, and Lis 46), the late naïve cells proliferated more slowly than their isogenic primed counterparts (**Fig. 2.7A and Fig. 2.7B, right**).

We next sought to determine whether these phenotypic changes were associated with transcriptomic alterations that occurred during long-term culture in naïve conditions. We first identified changes in mRNA expression associated with long-term culture in naïve conditions by performing differential expression analysis among cultures collected at the post-derivation (p20), early (p30) and late (p50) timepoints (please see "Differential mRNA expression analysis" in the Methods section). By taking the mRNAs that were upregulated (adj. p-value < 0.05, Log_2 fold change > 1) at the naïve p20 timepoint compared to both the early (p30) and late (p50) naïve timepoints, and removing the mRNAs that were also upregulated in early primed compared to the late primed timepoint, we identified 177 mRNAs that were uniquely upregulated at the earliest naïve state. These mRNAs were significantly enriched (adj. p-value < 0.05) for gene ontology (GO) biological process terms relating to positive regulation of DNA transcription and the regulation of pathway-restricted SMAD protein phosphorylation, including the previously reported naïve markers NODAL and LEFTY2 [52] (**Table 2.5**). This set of early naïve upregulated genes also contained the classical pluripotency markers OCT4 and NANOG, the pre-implantation embryo-associated gene TEAD4, and the naïve marker KLF4. The corresponding analysis to identify mRNAs uniquely downregulated (adj. p-value < 0.05, Log_2 fold change > 1) in the early naïve cultures yielded 157 mRNAs. These mRNAs were significantly enriched (adj. p-value < 0.05) for GO terms relating to development, differentiation, and negative regulation of both the Wnt and insulin-like growth factor receptor signaling pathways, supporting their more naïve state (**Fig. 2.8A, Table 2.6**).

Next, because miRNAs are known to play important roles in regulation of gene expression,

Table 2.5: Gene ontology enrichment analysis (biological processes) for the 177 mRNAs uniquely upregulated at the earliest naïve (p20) timepoint.

Term	Adjusted P-value
regulation of transcription DNA-templated (GO:0006355)	0.00665
cellular modified amino acid catabolic process (GO:0042219)	0.00843
regulation of gene expression (GO:0010468)	0.01381
positive regulation of transcription by RNA polymerase II (GO:0045944)	0.0192
regulation of cell differentiation (GO:0045595)	0.0192
cellular response to cadmium ion (GO:0071276)	0.0192
regulation of transcription by RNA polymerase II (GO:0006357)	0.0192
response to cadmium ion (GO:0046686)	0.01927
positive regulation of transcription DNA-templated (GO:0045893)	0.01927
regulation of pathway-restricted SMAD protein phosphorylation (GO:0060393)	0.02112

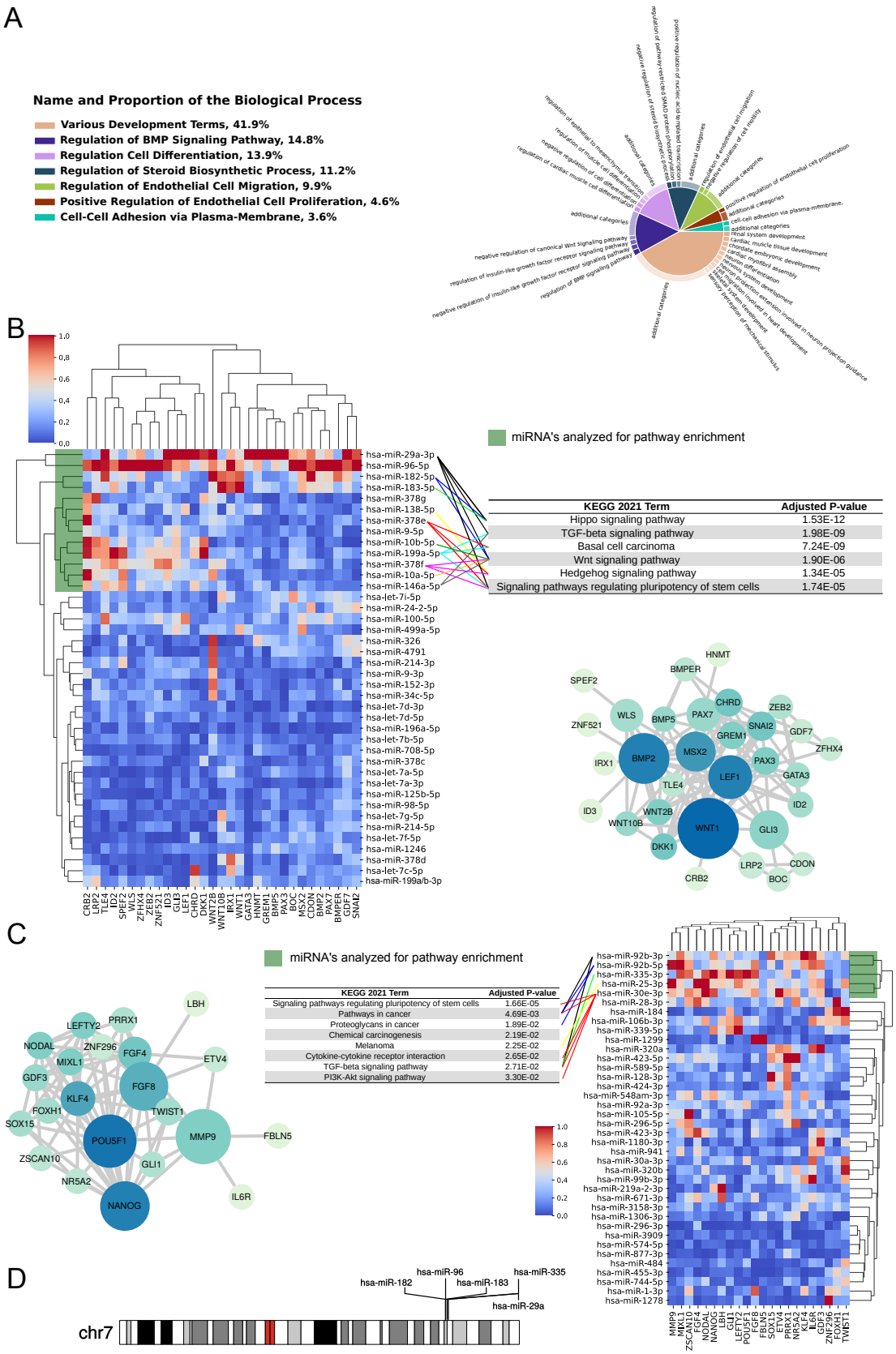
Table 2.6: Gene ontology enrichment analysis (biological processes) for the 157 mRNAs uniquely downregulated at the earliest naïve (p20) timepoint. Top 20 terms shown.

Term	Adjusted P-value
regulation of BMP signaling pathway (GO:0030510)	2.0015E-04
regulation of cardiac muscle cell differentiation (GO:2000725)	5.3299E-04
negative regulation of steroid biosynthetic process (GO:0010894)	5.3299E-04
negative regulation of cell differentiation (GO:0045596)	6.8721E-04
negative regulation of cellular response to growth factor stimulus (GO:0090288)	6.8721E-04
negative regulation of BMP signaling pathway (GO:0030514)	6.8721E-04
cardiac muscle tissue development (GO:0048738)	6.8721E-04
cardiac myofibril assembly (GO:0055003)	6.8721E-04
negative regulation of insulin-like growth factor receptor signaling pathway (GO:0043569)	6.8721E-04
renal system development (GO:0072001)	6.8721E-04
chordate embryonic development (GO:0043009)	6.8721E-04
regulation of pathway-restricted SMAD protein phosphorylation (GO:0060393)	6.8721E-04
regulation of endothelial cell migration (GO:0010594)	6.8721E-04
regulation of muscle cell differentiation (GO:0051147)	7.4195E-04
cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742)	7.4557E-04
skeletal muscle thin filament assembly (GO:0030240)	7.9371E-04
neuron differentiation (GO:0030182)	7.9371E-04
nervous system development (GO:0007399)	9.2864E-04
regulation of insulin-like growth factor receptor signaling pathway (GO:0043567)	9.4121E-04
negative regulation of Wnt signaling pathway (GO:0030178)	1.4210E-03

and have recently been reported to influence human naïve pluripotency [170, 206, 48], we performed a combined mRNA/miRNA analysis to identify regulatory changes that occurred over time in passage in naïve conditions (see "Integrative analysis of miRNAs and mRNAs differentially expressed between derivation naïve and early/late naïve cultures" in the Methods section). We first used STRING to create separate networks of mRNAs that were upregulated and downregulated between the derivation (p20) and early/late naïve cultures (p30 and p50), and performed clustering to highlight closely related genes (networks in (**Fig. 2.8B and Fig. 2.8C**).

The largest subcluster of genes that showed mRNA downregulation in the derivation naïve cultures compared to early/late naïve cultures (**Fig. 2.8B network**) was highly enriched for the Hippo, TGF-beta, Wnt, and Hedgehog signaling pathways (**Fig. 2.8B table**), all of which have been shown to be important in the establishment or regulation of the naïve state [157, 69, 218, 206]. Using hierarchical clustering to identify the miRNAs with the highest composite importance scores across the genes in the subcluster (**Fig. 2.8B heatmap**), we applied miRPathDB 2.0 [94] and found that the majority of these miRNAs displayed significant (p-value < 0.05) enrichments for one or more of the pathways enriched by the set of mRNAs in the cluster (**Fig. 2.8B table**).

Figure 2.8: (*next page*): Effects of long-term culture on the transcriptomics of naïve hESCs. (A) Gene ontology (GO) enrichment analysis of uniquely downregulated genes in the naïve derivation (p20) timepoint compared to the early (p30) and late (p50) timepoints among cultures. Genes that were downregulated in the early primed vs late primed timepoint were removed leaving 157 genes (**Table 2.6**). Revigo and CirGO were used to provide a non-redundant and biologically succinct visualization of GO terms with an adjusted p-value < 0.05. (B) Combined mRNA and miRNA expression analysis to identify regulatory changes occurring over time in naïve conditions. The network (right) was constructed by creating STRING protein to protein interaction networks and then clustering the networks. The network shows the largest cluster of mRNAs downregulated in derivation (p20) naïve cultures compared to early (p30) and late (p50) naïve cultures with the size relative to the Betweenness Centrality score and the color representing the node degree. The heatmap (left) was created using the importance scores generated by integrating the miRNA and mRNA using a gene regulatory inference algorithm. The colors represent the importance scores with reflect the degree to which each potential mRNA target is regulated by each miRNA in the dataset (see methods). The heatmap shows the mRNAs downregulated in the p20 cultures that were also found in the largest clustered network (right). The table shows the significant (adj. p-value < 0.00005) KEGG 2021 pathways enriched in the mRNAs in the largest clustered network. The lines drawn from the miRNAs in the heatmap to the pathways in the table, indicate the pathways predicted to be regulated by the corresponding miRNAs by miRPathDB. Only the miRNAs highlighted in green were analyzed. (C) Combined mRNA and miRNA expression analysis to identify regulatory changes occurring over time in naïve conditions. The network (left) was constructed by creating STRING protein to protein interaction networks and then clustering the networks. The network shows the largest cluster of mRNAs upregulated in derivation (p20) naïve cultures compared to early (p30) and late (p50) naïve cultures with the size relative to the Betweenness Centrality score and the color representing the node degree. The heatmap (left) was created using the importance scores generated by integrating the miRNA and mRNA using a gene regulatory inference algorithm. The colors represent the importance scores with reflect the degree to which each potential mRNA target is regulated by each miRNA in the dataset (see methods). The heatmap shows the mRNAs downregulated in the p20 cultures that were also found in the largest clustered network (right). The table shows the significant (adj. p-value < 0.05) KEGG 2021 pathways enriched in the mRNAs in the largest clustered network. The lines drawn from the miRNAs in the heatmap to the pathways in the table, indicate the pathways predicted to be regulated by the corresponding miRNAs by miRPathDB. Only the miRNAs highlighted in green were analyzed. (D) 5 of 18 miRNAs with the highest importance scores in **Fig. 2.8B** and **Fig. 2.8C** were found in the same miRNA cluster on chromosome 7.


A
Name and Proportion of the Biological Process
Various Development Terms, 41.9%
Regulation of BMP Signaling Pathway, 14.8%
Regulation Cell Differentiation, 13.9%
Regulation of Steroid Biosynthetic Process, 11.2%
Regulation of Endothelial Cell Migration, 9.9%
Positive Regulation of Endothelial Cell Proliferation, 4.6%
Cell-Cell Adhesion via Plasma-Membrane, 3.6%
B
miRNA's analyzed for pathway enrichment
KEGG 2021 Term | Adjusted P-value
Hippo signaling pathway | 1.53E-12
TGF-beta signaling pathway | 1.98E-09
Basal cell carcinoma | 7.24E-09
Wnt signaling pathway | 1.90E-06
Hedgehog signaling pathway | 1.34E-05
Signaling pathways regulating pluripotency of stem cells | 1.74E-05
C
miRNA's analyzed for pathway enrichment
KEGG 2021 Term | Adjusted P-value
Signaling pathways regulating pluripotency of stem cells | 1.66E-05
Pathways in cancer | 4.69E-03
Proteoglycans in cancer | 1.89E-02
Chemical carcinogenesis | 2.19E-02
Melanoma | 2.25E-02
Cytokine-cytokine receptor interaction | 2.65E-02
TGF-beta signaling pathway | 2.71E-02
PI3K-Akt signaling pathway | 3.30E-02
D
chr7

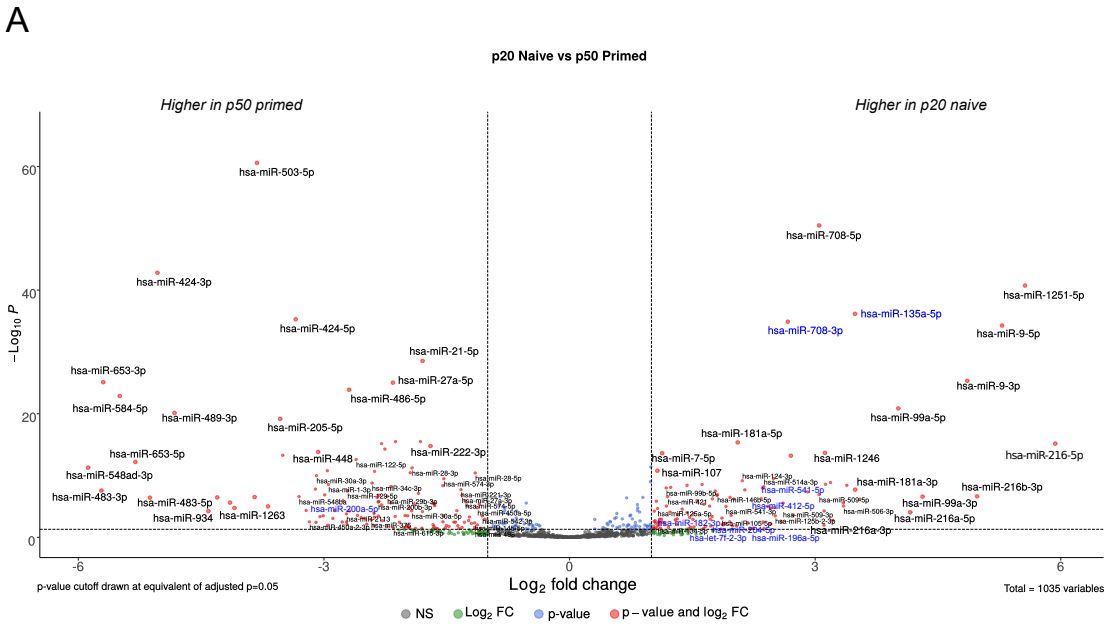
The largest subcluster for the genes that showed mRNA upregulation in the derivation naïve cultures compared to early/late naïve cultures (**Fig. 2.8C network**) was highly enriched for pathways pertaining to pluripotency and the maintenance of pluripotency, namely the TGF-beta and PI3K-Akt signaling pathways [158, 194] as well as pathways associated with cancer (**Fig. 2.8C network**). The miRNA's with the highest composite importance scores across the corresponding mRNAs in this subcluster (**Fig. 2.8C heatmap**) were likewise significantly enriched (p-value > 0.05) for the same pathways (**Fig. 2.8C table**).

Interestingly, 5 of the 18 miRNAs that had the highest importance scores for these two subclusters (4 out of 13 downregulated and 1 out of 5 upregulated miRNAs in the derivation naïve cultures) are encoded by the same miRNA cluster on chromosome 7 (**Fig. 2.8D**). The DNA methylation levels of the CpG islands in this region of chromosome 7 were lower for the naïve p20 samples compared to the naïve and primed p30 and p50 samples, although statistically significantly lower for only two of the four hESC lines (p-value < 0.05, Mann Whitney U test)(data not shown). Moreover, 7 of these highest importance 18 miRNAs are included in the miRPathDB database, and 6 of these 7 were found to have highly similar sets of target genes, as assessed by their Jaccard similarity coefficients; using the "Similar miRNAs" function in miRPathDB, these 6 miRNAs were ranked in the top 20 in the "Jaccard coefficient ('target genes strong)" column out of the over 4000 total miRNAs in the database [94]. Consistent with prior studies, we found that several let-7 family miRNAs were downregulated in derivation naïve cultures compared to later cultures [61, 48, 206]. However, the miR-302/miR-371-373 clusters, previously reported to be associated with the transition from naïve to primed substates [61, 48, 206], were not differentially expressed between our early and late naïve cultures.

We next focused on studying differences in miRNA expression between our naïve and primed cultures in the context of several previous reports [170, 206, 48]. To identify the greatest differences between miRNAs expressed in our naïve and primed cultures, we performed differential expression analysis between our p20 naïve cultures and p50 primed cultures (adj p-value

< 0.05, Log_2 fold change > 1, basemean > 10) and compared the results to recently reported human naïve and primed small RNA-seq data (**Fig. 2.9A**) [206, 48]. Of the 72 miRNAs that were more highly expressed in the p20 naïve cultures, 20 were reported to be consistently upregulated in previously reported naïve lines, and of the 74 miRNAs that were upregulated in the p50 primed cultures, 10 were found to be consistently upregulated in previously reported primed lines (**Fig. 2.9A, miRNAs in blue = previously reported**). Notably, similar to our comparison between the p20 and p30/p50 naïve cultures, the previously reported naïve-associated miR-371-373 cluster [206, 183, 48] was not upregulated in our p20 naïve cultures compared to our p50 primed cultures. To identify miRNAs that regulate the transition from the naïve to primed substates in our study, we combined our list of differentially expressed miRNAs with our list of differentially expressed mRNAs (adj. P-value < 0.05, basemean > 50) from the same comparison and used mirTarBase, a database of experimentally validated miRNA-target gene relations [38], to link changes between our differentially expressed miRNAs and mRNAs (see "Integrative analysis of miRNAs and mRNAs differentially expressed between naïve and primed cultures" in the Methods Section). Only miRNA-target mRNA connections changing in the opposite direction in our p20 naïve vs p50 primed comparison were considered.

Figure 2.9: (*next page*): miRNA differences between our naïve and primed cultures.
(A) Differential expression analysis between our naïve and primed cultures. Volcano plot shows miRNAs significantly different (adj. p-value < 0.05, Log_2 fold change $> |1|$, basemean > 10) between our derivation naïve timepoint samples (p20) and our late (p50) primed cultures. The dotted lines represent a Log_2 fold change of $|1|$ and an adjusted p-value of 0.05. Grey dots are not significant miRNAs, blue dots are miRNAs that have an adj. p-value < 0.05 but not a Log_2 fold change $> |1|$, green dots have a Log_2 fold change $> |1|$ but not an adj. p-value < 0.05, and red dots are those miRNAs considered significantly different by both metrics. miRNAs in blue font are those that were previously reported to be differentially expressed between naïve and primed lines. (B) Gene ontology (GO) enrichment analysis of downregulated genes in the naïve derivation (p20) timepoint that were also targeted by a miRNA that was upregulated in the p20 naïve cultures. Revigo and CirGO were used to provide a non-redundant and biologically succinct visualization of GO terms with an adjusted p-value < 0.05. (C) Genomic locations of miRNA clusters containing multiple miRNAs upregulated in naïve derivation cultures (p20). (D) Boxplots displaying methylation levels of the regions containing the chromosome 14 and chromosome X miRNA clusters in part C above. Boxes span the top 75th percentile to bottom 25th percentile with the line in the middle of the boxes representing the median. Whiskers contain 95% of the data. "*" represents a significant difference between cultures (p-value $<$ 0.05 Mann-Whitney U).


A
p20 Naive vs p50 Primed
Higher in p50 primed
Higher in p20 naive
hsa-miR-503-5p
hsa-miR-424-3p
hsa-miR-424-5p
hsa-miR-21-5p
hsa-miR-653-3p
hsa-miR-27a-5p
hsa-miR-584-5p
hsa-miR-486-5p
hsa-miR-489-3p
hsa-miR-205-5p
hsa-miR-653-5p
hsa-miR-448
hsa-miR-222-3p
hsa-miR-548ad-3p
hsa-miR-483-3p
hsa-miR-483-5p
hsa-miR-934
hsa-miR-1263
hsa-miR-708-5p
hsa-miR-1251-5p
hsa-miR-135a-5p
hsa-miR-708-3p
hsa-miR-9-5p
hsa-miR-9-3p
hsa-miR-99a-5p
hsa-miR-181a-5p
hsa-miR-7-5p
hsa-miR-1246
hsa-miR-216-5p
hsa-miR-107
hsa-miR-181a-3p
hsa-miR-216b-3p
hsa-miR-99a-3p
hsa-miR-216a-5p
hsa-miR-216a-3p
hsa-let-7f-2-3p
hsa-miR-196a-5p
$-Log_{10}$ P
Log_2 fold change
p-value cutoff drawn at equivalent of adjusted p=0.05
Total = 1035 variables
NS
Log_2 FC
p-value
p – value and log_2 FC
B
Name and Proportion of the Biological Process (Inner Ring)
signaling pathways, 24.8%
cell migration, 15.7%
regulation of cell proliferation, 13.8%
extracellular matrix organization, 12.6%
regulated exocytosis, 9.2%
protein phosphorylation, 8.8%
transcription, 5.5%
cell adhesion, 5.2%
immune response, 4.3%
D
Chr.14 region
Chr.X region
Beta Values
Lis38Np20
Lis38Pp50
Lis39Np15
Lis39Pp51
Lis45Np18
Lis45Pp50
Lis46Np19
Lis46Pp51
C
chr14
chr7
chrX
hsa-miR-548h
hsa-miR-342
hsa-miR-345
hsa-miR-433
hsa-miR-432
hsa-miR-136
hsa-miR-541
hsa-miR-412
hsa-miR-134
hsa-miR-889
hsa-miR-382
hsa-miR-487a
hsa-miR-889-3p
hsa-miR-544a
hsa-miR-3774
hsa-miR-1193
hsa-miR-1197
hsa-miR-380
hsa-miR-96
hsa-miR-592
hsa-miR-182
hsa-miR-183
hsa-let-7f
hsa-miR-421
hsa-miR-374c
hsa-miR-891a
hsa-miR-506
hsa-miR-508
hsa-miR-509-2
hsa-miR-509-3
hsa-miR-509-1
hsa-miR-514a
hsa-miR-105

We performed GO enrichment analysis on the 1235 downregulated genes that were also targeted by a miRNA that was upregulated in our p20 naïve cultures, and grouped the resulting enriched (adj p-value < 0.05) GO terms into larger categories using Revigo and CirGO [174, 103]. The largest group of enriched GO terms was signaling pathways, and included numerous pluripotency-associated signaling pathways, including the MAPK cascade, and the Wnt, ERK, and TGF beta pathways (**Fig. 2.9B, Table 2.7**). Notably, the TGF beta pathway was recently reported to be required to maintain the naïve pluripotent state [144]. Consistent with Wang et al., [206], we found that the Hedgehog Signaling pathway was significantly enriched (adj. p-value = 0.02, Wikipathway 2021) in our p20 naïve cultures, with the ciliary G-protein coupled receptor GPR161 targeted by hsa-miR-301a-3p, reported to repress the Hedgehog pathway in the primed substate, being upregulated in our primed cultures. Interestingly, the second and fourth-largest GO term groups were the related cell migration and extracellular matrix organization term groups, respectively (**Fig. 2.9B**). Recent reports have shown that membrane mechanics and cell migration may have important regulatory roles in the exit from naïve pluripotency and help explain morphological differences between the two substates /citeBergert2021. The third-largest group of GO terms related to cell proliferation, suggesting that miRNAs might be involved in regulating the doubling time and cell cycle differences that were noted between our naïve and primed cultures (**Fig. 2.7A and 2.7B; Fig. 2.9B**).

Finally, we noted that there were three genomic locations that contained multiple naïve upregulated miRNAs: one miRNA cluster on chromosome 14 q32 in the pluripotency-associated DLK1-DIO3 locus, which contained 17 of the miRNAs that were upregulated in the p20 naïve cultures, and whose 167 target genes were enriched for the miRNA-regulated DNA damage repair pathway (adj. p-value = 0.002); another miRNA cluster on the X chromosome, which contained six p20 naïve upregulated miRNAs, and whose 71 target genes were enriched for regulation of cell migration (adj. p-value = 0.01); and a third smaller cluster on chr 7 (analysis

Table 2.7: Gene ontology enrichment analysis (biological process 2021) of downregulated genes also targeted by a miRNA upregulated in derivation (p20) naïve cultures. Refers to data in **Figure 2.9B**. Shows top 20 terms.

GO Biological Process 2021 Term Name	Term ID	Adj. P-value
regulation of cell migration	GO:0030334	5.49249E-13
negative regulation of programmed cell death	GO:0043069	4.4803E-09
regulated exocytosis	GO:0045055	5.91794E-09
negative regulation of apoptotic process	GO:0043066	2.36032E-08
cellular protein modification process	GO:0006464	2.36032E-08
extracellular matrix organization	GO:0030198	0.000000027
positive regulation of intracellular signal transduction	GO:1902533	8.43956E-08
platelet degranulation	GO:0002576	2.24847E-07
collagen fibril organization	GO:0030199	2.41274E-07
regulation of apoptotic process	GO:0042981	4.95629E-07
regulation of cell population proliferation	GO:0042127	1.90928E-06
cytokine-mediated signaling pathway	GO:0019221	2.94334E-06
regulation of DNA-templated transcription	GO:0045449	3.1748E-06
regulation of focal adhesion assembly	GO:0051893	4.35571E-06
positive regulation of cell motility	GO:2000147	8.43932E-06
cellular response to growth factor stimulus	GO:0071363	8.94593E-06
supramolecular fiber organization	GO:0097435	1.92518E-05
positive regulation of cell migration	GO:0030335	1.92518E-05
negative regulation of cell migration	GO:0030336	4.70442E-05
protein phosphorylation	GO:0006468	4.73694E-05

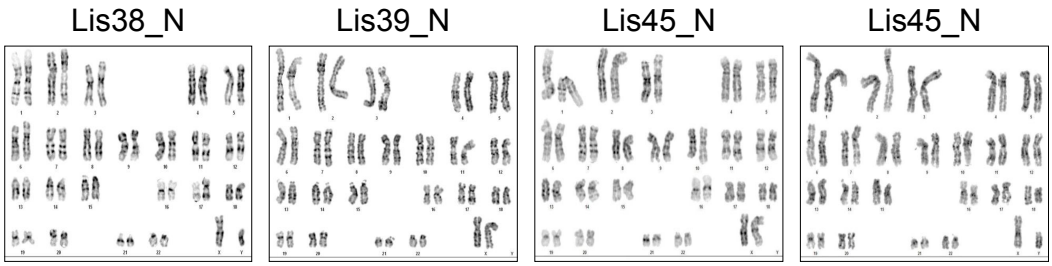
performed using miRTarBase) (**Fig. 2.9C**). Both the Chr.14 and Chr.X loci are known to be epigenetically regulated (imprinted and X-linked, respectively). We therefore analyzed the DNA methylation in the CpG island regions of the miRNAs in these two clusters. Interestingly, we found that the methylation levels of the early naïve cultures for Lis39 and Lis45 (the two female lines) were markedly lower than their late corresponding primed cultures, whereas the p20 naïve cultures for Lis38 and Lis46 (the two male lines) had higher methylation levels (**Fig. 2.9D**). These results suggest that miRNA-mediated mRNA regulation plays a role in the establishment and maintenance of substates of pluripotency and may contribute the phenotypic differences seen in our cultures.

2.3.4 Genetic stability of naïve and isogenic primed hESCs over time in culture

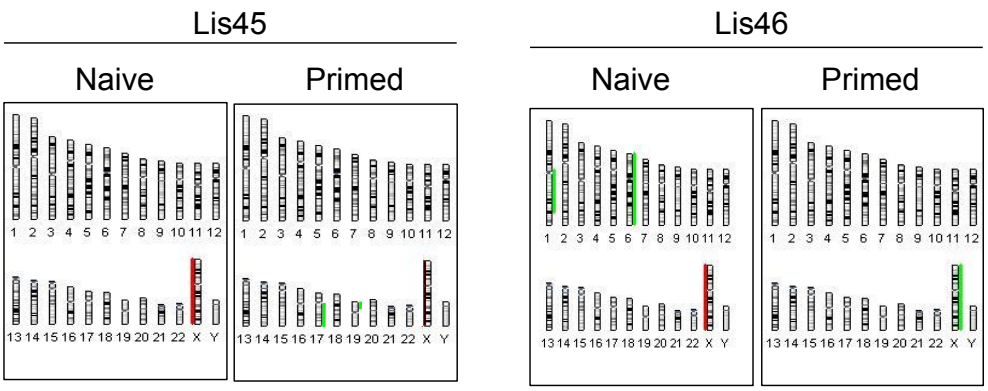
Since high cellular proliferation rate may increase the susceptibility of the cells to genetic instability, which in turn may increase the risk of tumorigenicity, we compared the genetic stability of isogenic naïve and primed hESCs. Karyotype analysis showed that all four early virgin naïve hESC lines retained normal karyotypes as shown by the G-banding (**Fig. 2.10A**). Consistent with the karyotype results, chromosomal microarray analysis (CMA) also showed no genetic aberrations in derivation and early virgin naïve hESCs or isogenic early primed cultures (**Fig. 2.10B**).

Figure 2.10: (*next page*): Cytogenetic analysis of virgin naïve and isogenic primed hESCs over time in culture.
(A) Karyotype analysis of early virgin naïve hESC lines. (B) Chromosomal charts of the CMA aberrations detected in late naïve and isogenic primed cells. Duplications are shown in green and deletions in red. (C) Validation of CMA results by WGS.

A



B



C

Sample	Chr.	Karyotype	CMA Results	WGS CNV Analysis
Lis 45_N early	X	Normal	Normal	4 MB deletion Mosaic Chr. X Deletion
Lis 45_N late	X	-	Entire Chr. X deletion	Entire Chr. X deletion
Lis 45_P late	X	-	Normal	Mosaic Chr. X duplication
Lis 46_N early	X	Normal	Normal	Normal
Lis 46_N late	X	-	Entire Chr. X deletion	Normal
Lis 46_P late	X	-	Entire Chr. X duplication	Chr. X duplication
Lis 45_N early	19	Normal	Normal	Normal
Lis 45_N late	19	-	Normal	Normal
Lis 45_P late	19	-	19 MB duplication	Mosaic Chr. 19 duplication
Lis 45_N early	17	Normal	Normal	Normal
Lis 45_N late	17	-	Normal	Normal
Lis 45_P late	17	-	55 MB duplication	Mosaic Chr. 17 duplication
Lis 46_N early	6	Normal	Normal	61 KB duplication
Lis 46_N late	6	-	Entire Chr. 6 duplication	Mosaic partial Chr. 6 duplication
Lis 46_P late	6	-	Normal	Small mosaic Chr. 6 duplication
Lis 46_N early	1	Normal	Normal	2-400 KB duplication
Lis 46_N late	1	-	99.8 MB duplication	1 MB duplication
Lis 46_P late	1	-	Normal	-

Prior studies have shown that extended culture is the most prominent factor contributing to the accumulation of genetic abnormalities in hPSCs [104, 62, 20, 175, 2, 51, 13, 123, 106, 124, 7, 138, 127, 1]. Indeed, CMA analysis revealed no genetic aberrations in the late cultures for two of the hESC lines (Lis 38 and Lis 39), the late naïve and primed cultures for the two other lines (Lis 45 and Lis 46) displayed large copy number aberrations. Specifically, CMA identified monosomy X in naïve cells for Lis 45 (Lis 45_N); partial duplications of chromosome 17 and 19 in primed cells for Lis 45 (Lis 45_P); a partial duplication of chromosome 1, trisomy 6 and loss of chromosome X in Lis 46_N; and duplication of chromosome X in Lis 46_P (**Fig. 2.10B and 2.10C**). While CMA is widely used method for detecting large chromosomal aberrations in clinical samples, WGS provides higher resolution and the ability to detect mosaic populations of cells, some of which contain a given genetic aberration, and others which do not. WGS confirmed the majority of the CMA results (monosomy X in late Lis 45_N, trisomy X in late Lis 46_P, normal X in early Lis 46_N, normal chr19 in early and late Lis 45_N, and normal chr17 in early and late Lis 45_N; (**Fig. 2.10C**). Moreover, WGS showed that some of the aberrations identify by CMA were in fact mosaic (monosomy X in late Lis 46_N, chr19 duplication in late Lis 46_P, chr17 duplication in late Lis 46_P and trisomy 6 in late Lis 46_N). In other cases, loci that were called normal by CMA were shown to have copy number alterations by WGS (mosaic monosomy X in early Lis 45_N, mosaic trisomy X in late Lis 45_P, a small duplication in chr6 in early Lis 46_N, a mosaic duplication in chr6 in late Lis 46_P and a small duplication in chr1 in early Lis 46_N). In one case, WGS detected a smaller aberration than was found by CMA (1 MB duplication in chr1 in late Lis 46_N) (**Fig. 2.10C**).

2.3.5 Whole Genome Sequencing reveals de-novo CNVs and SNVs that arise in naïve and isogenic primed hESCs over time in culture

We had access to DNA samples for the donor parents, which allowed us to filter out genetic variants that were inherited and focus on those that arose de-novo during derivation and culture.

Genome-wide WGS analysis was performed in order to discover *de novo* aberrations that were acquired during culture and could not be detected by CMA. CNV analysis of sub-chromosomal CNVs (see methods) revealed the presence of de-novo CNVs in both late naïve and isogenic primed hESCs (**Fig. 2.11A and 2.11B**). However, the total number of *de novo* CNVs varied markedly among hESC lines, with the Lis45 and Lis46 lines showing higher numbers of CNVs than Lis38 and Lis39. In addition, the naïve cultures tended to accumulate deletions over time, while the primed cultures tended to accumulate duplications (**Fig. 2.11A and 2.11B**). Since just 15 of the CNVs were found to be located in coding regions, we asked if any of the CNVs had an effect on expression levels of the nearest gene using RNA sequencing (RNA-seq). The RNA-seq analysis showed only two CNVs that are associated with differentially expressed genes (DEG). Both of these CNVs were duplications on chromosome 16 in the same primed line (Lis 46_P), and the corresponding upregulated gene was autocrine motility factor receptor (AMFR). Overexpression of AMFR has been observed in several types of human cancer and its expression has been found to be correlated with more advanced tumor stage and decreased survival rates [36, 136, 189, 90, 80]. Notably, AMFR was significantly upregulated only in Lis 46_P compared to its naïve counterparts, and not compared to the other primed lines. Next, we studied de-novo single nucleotide variations (SNVs). Our results showed that less than 50 SNVs arose during derivation or early culture, and most of these were not seen in late passage culture (**Fig. 2.12A**). During long-term culture, Lis 38 and Lis 39 acquired approximately 60-80 additional SNVs, with no significant differences between the number of SNVs in late naïve and isogenic primed cultures, while in Lis 45 and Lis 46, substantially more SNVs were observed in primed cells compared to late naïve hESCs ($>$100 vs. $<$50 respectively; (**Fig. 2.12A**). Mapping the SNVs, we found that all were located in non-coding regions, mostly in introns and intergenic regions (**Fig. 2.12B**).

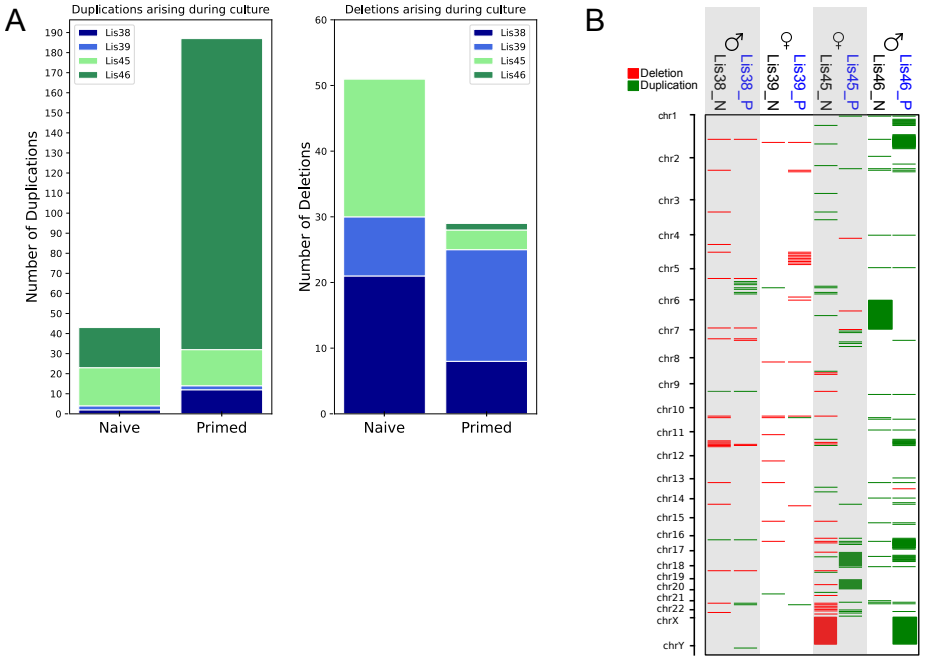
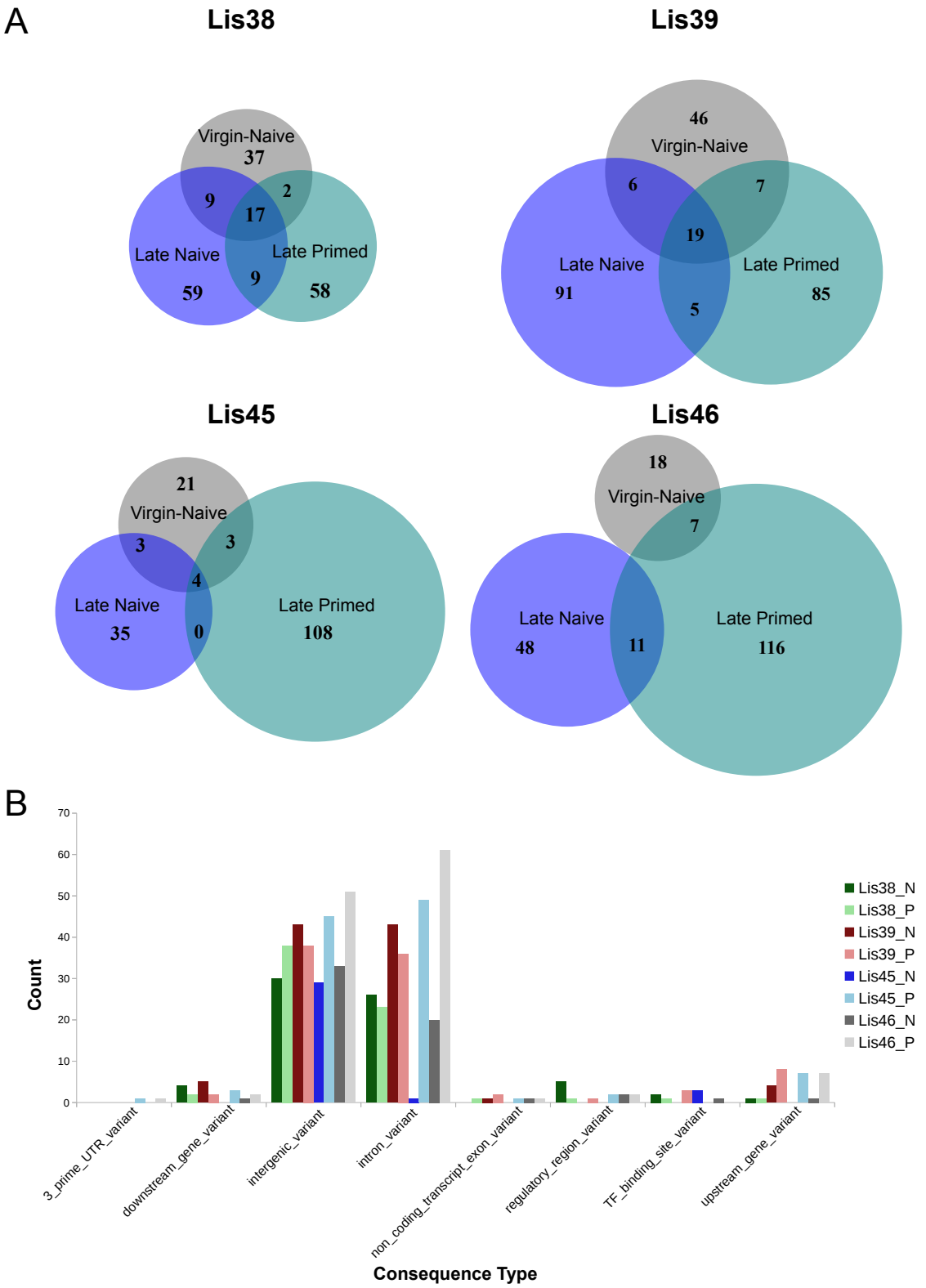


Figure 2.11: Copy number variations arising over time in culture.
(A) Stacked bar chart shows the number of deletions and duplications arising during time in culture by whole genome sequencing.(B) CNV's arising over time in culture. Genomic locations of duplications (green) and deletions (red) in late passage cultures.

Figure 2.12: (*next page*): SNVs in virgin naïve compared to isogenic primed hESCs.
(A) Venn-diagrams presenting the number of SNVs detected in early virgin naïve hESCs, as well as in late naïve and isogenic primed cells, including unique and overlapping SNVs. (B) Bar graph showing the distribution of the detected SNVs throughout the genome for each line in non-coding regions, including: SNVs within introns, up-stream or down-stream to genes, non-coding transcript exons, TF binding sites, regulatory regions, and intergenic regions.


A
Lis38
Virgin-Naive
37
9
2
17
Late Naive
59
Late Primed
9
58
Lis39
46
Virgin-Naive
6
7
19
Late Naive
91
Late Primed
85
5
Lis45
21
Virgin-Naive
3
3
4
Late Naive
35
0
Late Primed
108
Lis46
18
Virgin-Naive
7
Late Naive
48
11
Late Primed
116
B
Count
70
60
50
40
30
20
10
0
3_prime_UTR_variant
downstream_gene_variant
intergenic_variant
intron_variant
non_coding_transcript_exon_variant
regulatory_region_variant
TF_binding_site_variant
upstream_gene_variant
Consequence Type
Lis38_N
Lis38_P
Lis39_N
Lis39_P
Lis45_N
Lis45_P
Lis46_N
Lis46_P

2.3.6 Genetic aberrations within enhancers and super enhancers at late passage

Apart from mutations in the coding sequences that may affect the encoded protein structurally and functionally, genetic variants can also alter gene expression levels. Gene expression levels are modulated by cis-regulatory elements, such as enhancers, which regulate spatiotemporal expression of target gene by transcription factors binding, and super-enhancers, that consist of clusters of enhancers. Since the majority of CNVs and all SNVs observed at late passage were located in non-coding regions, we investigated whether they were associated with such regulatory regions in light of a recent publication that naïve and primed hESCs can be characterized by unique sets of enhancers and super-enhancers [16]. When analyzing CNVs in these regions, we found that the total number of CNVs was larger in the primed cells compared to their naïve counterparts in all four lines, both within enhancers (**Fig. 2.13A**) and super enhancer regions (**Fig. 2.13B**). Moreover, more SNVs were found in these enhancers and super-enhancer regions in all primed hESC lines compared to their naïve counterparts (**Fig. 2.13C**). We then analyzed the effects of these CNVs on the expression of genes that are close to the enhancer using RNA-seq and found that in three out of four lines (Lis 38, Lis 45 and Lis 46) the primed cells had more DEG resulting from duplications or deletions within enhancer or super enhancer regions containing CNVs (**Table 2.8**). A similar analysis was done for SNVs and we found that none of these SNVs was associated with a significant change in gene expression level.

Since the aberrations we detected were in non-coding regions (**Fig. 2.12B**), we investigated whether there are aberrations in regulatory regions of cancer promoting genes (classified as Tier1 in The Cancer Gene Census of COSMIC v86 database). In addition, RNA-seq of the same lines enabled us to evaluate whether these aberrations had transcriptional effects. Our CNV analysis showed six CNVs located within enhancers of Tier1 genes that were also associated with altered gene expression levels in primed hESCs, but not in their naïve counterparts (FBXW7 in Lis39_P;

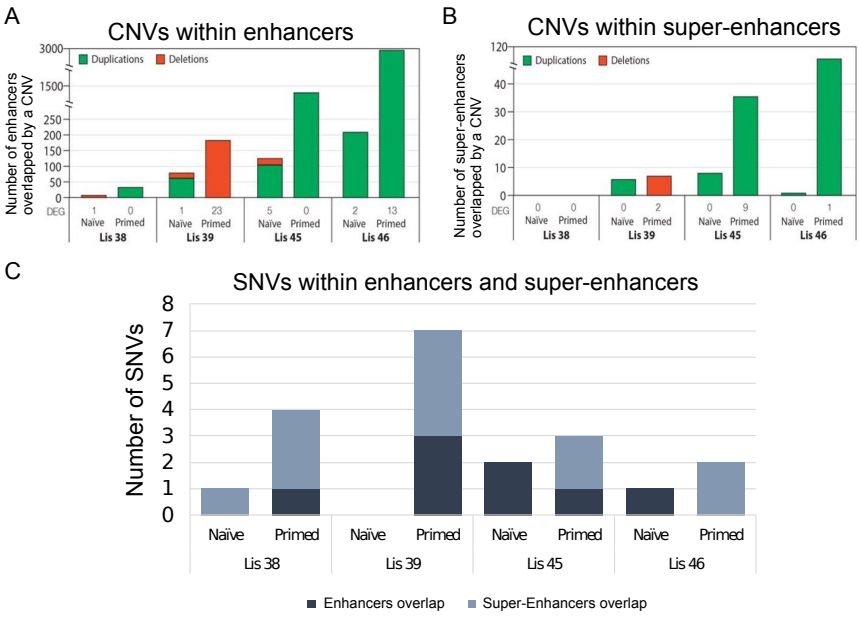


Figure 2.13: Genetic aberrations within regulatory regions.
CNVs located within enhancer (A) or super-enhancer (B) regions in each line, including duplications (green) and deletions (red) with the respective DEG detected by RNA-seq. (C) SNVs located within enhancer (dark gray) or super-enhancer (light gray) regions had no transcriptional effect.

RAC1 and CARD11 in Lis 45_P; and CDH1, SLC45A3 and LMNA in Lis 46_P). However, the only aberration with a statistical and biological significance was a deletion located upstream to the FBXW7 tumor suppressor gene in Lis 39_P that was associated with down-regulation, contrary to the high expression of these gene in all other primed lines. Importantly, no CNVs were found in regulatory regions of Tier1 genes in naïve lines.

In 2019, it was reported that hESCs that were derived in primed conditions and converted to naïve at later passages tend to acquire more SNV mutations in cancer related genes (classified as Tier1 in The Cancer Gene Census of COSMIC v86 database), compared to cells that were continually grown as primed [8]; these results were subsequently found to be likely an artifact of contamination of hPSC lines by mouse feeder cells [171]. In our cultures, we identified only two cancer-associated SNVs, which were found at very low allelic fractions, which appeared to be acquired over time in culture: the first in POU5F1 in late Lis 39_N and the second in COL1A1

Table 2.8: Summary of CNVs in enhancer regions. Table shows the number of duplications and deletions in super-enhancers and non super-enhancers as defined by [16].

Super-Enhancers

Number of Duplicated Super-Enhancer Regions	Sample Name	Enhancer Associated Genes Up-Regulated	Number of Cosmic Cancer Tier 1or 2 Genes	DE Genes (adj pval)
5	Lis39Np51	0 (out of 5 genes)	0	N/A
8	Lis45Np51	0 (out of 8 genes)	0	N/A
36	Lis45Pp50	0 (out of 14 genes)	0	N/A
1	Lis46Np51	0 (out of 1 genes)	0	N/A
108	Lis46Pp51	1 (out of 62 genes)	0	SYT7 (0.005)

Number of Deleted Super-Enhancer Regions	Sample Name	Enhancer Associated Genes Down-Regulated	Number of Cosmic Cancer Tier 1or 2 Genes	DE Genes (adj pval)
6	Lis39Pp51	2 (out of 5 genes)	0	ELF2 (0.00003), TMEM13L (0.0046)

Non Super-Enhancers

Number of Duplicated ACTIVE Regions	Sample Name	Enhancer Associated Genes Up-Regulated	Number of Cosmic Cancer Tier 1or 2 Genes	DE Genes (adj pval)
105	Lis45Np51	3 (out of 38 genes)	0	ATG4C (0.02), BCL2L11 (0.007), HM13 (0.015)
211	Lis46Np51	2 (out of 40 genes)	0	MAEA (0.04), RHOU (0.017)
35	Lis38Pp50	0 (out of 14 genes)	0	N/A
2	Lis39Pp51	0 (out of 2 genes)	0	N/A
1426	Lis45Pp50	0 (out of 400 genes)	0	N/A
2970	Lis46Pp51	13 (out of 981 genes)	0	C16orf8 (0.000003), CCS (0.000005), ANXA (0.0004), PELI3 (0.0006), RNF187 (0.0006), AKTIP (0.002), MYLK3 (0.003), TNN (0.004), ADAP2 (0.004), SYT7 (0.005), DUSP10 (0.02), CFDP1 (0.02), FTH1 (0.04)
66	Lis39Np51	1 (out of 14 genes)	0	TM9SF4 (0.012)

Number of Deleted ACTIVE Regions	Sample Name	Closest Gene Down-Regulated	Number of Cosmic Cancer Tier 1or 2 Genes	DE Genes (adj pval)
8	Lis39Np51	0 (out of 4 genes)	0	N/A
179	Lis39Pp51	23 (out of 51 genes)	1 (FBXW7)	TIGD4 (0.000002), FBXW7 (0.000004), ELF2 (0.00003), GNPNAT1 (0.00004), ARFIP1 (0.00005), DDHD1 (0.00006), GATB (0.0002), GAB1 (0.0002), NUDT9 (0.0003), GPR137C (0.0005), HSD17B11 (0.0008), BMP4 (0.001), CTSO (0.001), FERMT2 (0.003), SMAD1 (0.004), ZNF827 (0.005), TMEM131L (0.005), USP53 (0.007), ERO1A (0.01), NDUFC1 (0.02), SYNPO2 (0.03), SNX25 (0.039), PDLIM3 (0.04)
18	Lis45Np51	2 (out of 2 genes)	0	ZC2HC1A (0.0001), MMP16 (0.006)
9	Lis38Np49	1 (out of 1 genes)	0	GRID2 (0.008)

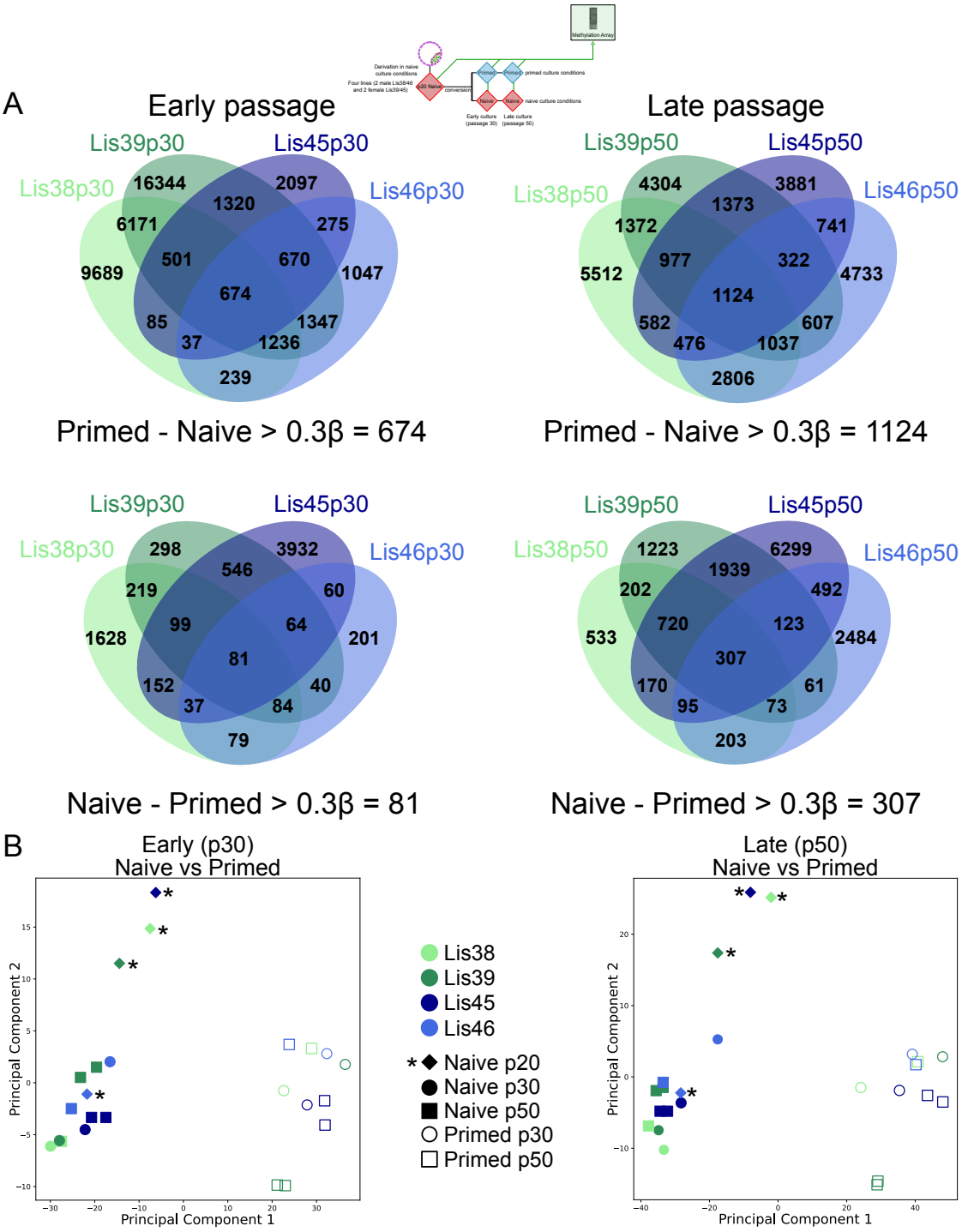
in late Lis 39_P; however, consistent with the Stirparo et al. report, these two variants can be attributed to a low level of contamination by mouse embryonic fibroblast (MEF) feeder cells.

2.3.7 Genome-wide DNA methylation analysis reveals alterations that differ by pluripotency state (naïve vs primed) and time in culture

Recent studies have reported that, similar to the epigenetic profile of human preimplantation embryos [146, 167, 182, 162], the human naïve stem cell state is characterized by

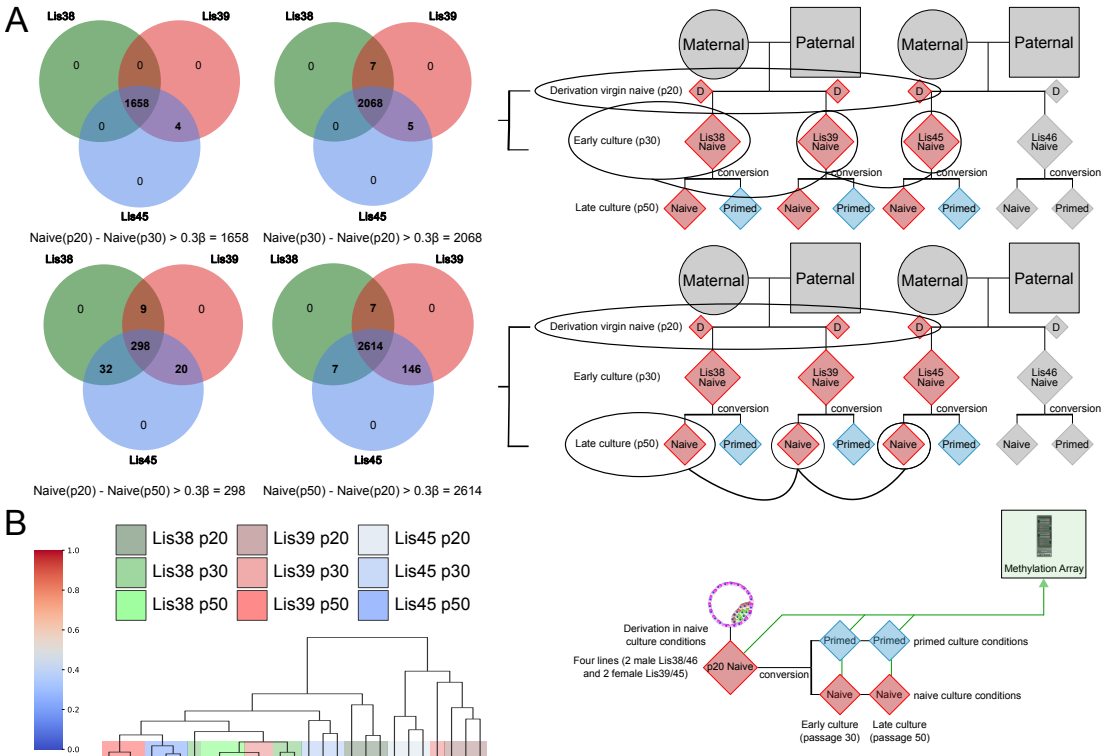
genome-wide DNA hypomethylation [182, 64, 178]. Since our naïve cultures were only slightly hypomethylated compared to their corresponding isogenic primed cultures (**Fig. 2.6B**), we examined which loci were consistently differentially methylated between our naïve and primed cultures. To be considered a differentially methylated site, we required a beta (methylated/(methylated + unmethylated)) difference of at least $|0.3|$ in all four of the paired naïve/primed comparisons. We found there were 674 loci, associated with 307 genes, that were consistently hypomethylated in our early naïve cultures compared to their matched primed cultures and only 81 loci, associated with 47 genes, that were hypomethylated in the early primed cultures compared to their matched naïve cultures (**Fig. 2.14A**). A similar analysis in our late cultures revealed 1124 (462 genes) hypomethylated loci in our naïve cultures compared to just 307 (157 genes) in our primed cultures (**Fig. 2.14A**). PCA using either the early or late passage differentially methylated sites separated the naïve and primed samples along the first principal component (**Fig. 2.14B**). We also noted that for three of the four hESC lines, the derivation (p20) naïve cultures (indicated by "*") separated from their early (p30) and late (p50) passage naïve counterparts along the second principal component (**Fig. 2.14B**). This suggested to us that, in contrast to mRNA expression changes (**Fig. 2.5A**), DNA methylation changes resulting from time in culture occurred predominantly during early culture, and stabilized by p30.

Figure 2.14: (*next page*): Methylation differences between primed and naïve cultures.
(A) Venn diagrams showing differential methylation (beta difference (methylated/(methylated + unmethylated) of at least $|0.3|$ in all four of the paired naïve/primed comparisons) between early and late passage naïve and primed pairs. Venn diagrams show that naïve cultures contain more differentially hypomethylated probes than primed cultures and that there are more differentially methylated probes in the late passage comparisons compared to the early passage comparisons.
(B) Principal component analysis using either the early or late passage differentially methylated probes from part "A". The "*" show the location of the derivation (p20) naïve culture samples.


A
Early passage
Late passage
Lis39p30
Lis45p30
Lis38p30
Lis46p30
16344
2097
1320
6171
275
501
670
9689
1047
674
85
1347
37
1236
239
Primed - Naive > 0.3β = 674
Lis39p50
Lis45p50
Lis38p50
Lis46p50
4304
3881
1373
1372
741
977
322
5512
4733
1124
582
607
476
1037
2806
Primed - Naive > 0.3β = 1124
298
3932
546
219
60
99
64
1628
201
81
152
40
37
84
79
Naive - Primed > 0.3β = 81
1223
6299
1939
202
492
720
123
533
2484
307
170
61
95
73
203
Naive - Primed > 0.3β = 307
B
Early (p30)
Naive vs Primed
Late (p50)
Naive vs Primed
Principal Component 1
Principal Component 2
Lis38
Lis39
Lis45
Lis46
Naive p20
Naive p30
Naive p50
Primed p30
Primed p50

To identify functional DNA methylation changes occurring during the early period after derivation, we focused on the three hESC lines that showed separation along PC2 between the p20 cultures and the p30/p50 cultures (Lis38, Lis39, and Lis45). We identified loci with a $> |0.3\beta|$ difference between both the matched naïve p20 vs naïve p30 and naïve p20 vs naïve p50 cultures for these three lines. We found that there were markedly more sites that were hypomethylated in the p20 cultures than the later cultures, with 4682 sites showing lower DNA methylation levels in the p20 cultures and 1956 sites showing lower DNA methylation levels in the p30/p50 cultures (**Fig. 2.15A**).

Figure 2.15: (*next page*): Methylation differences between naïve cultures due to time in culture. (A) Venn diagrams showing methylation differences between derivation (p20) naïve cultures and both early (p30) and late (p50) naïve cultures. Lis46 was removed from the analysis (see Fig 2.14B). (B) Heatmap showing the gene expression of genes that were both differentially hypomethylated (see part "A") and were differentially upregulated (Log_2 fold change > 1) at the naïve derivation (p20) timepoint when compared to both the early (p30) and the late (p50) naïve timepoints. (C) Scatterplot showing the correlation between methylation levels and expression levels of all genes associated with probes that were hypomethylated (beta difference (methylated/(methylated + unmethylated) of at least $|0.3|$) in virgin cultures (p20) compared to late cultures (p50). The red dots are genes that are both differentially hypomethylated (see part "A") and were differentially upregulated (Log_2 fold change > 1) at the naïve derivation (p20) timepoint when compared to both the early (p30) and the late (p50) naïve timepoints.


A
Lis38
Lis39
Lis45
0
0
0
1658
0
4
0
Naive(p20) - Naive(p30) > 0.3β = 1658
0
7
0
2068
0
5
0
Naive(p30) - Naive(p20) > 0.3β = 2068
0
9
0
298
32
20
0
Naive(p20) - Naive(p50) > 0.3β = 298
0
7
0
2614
7
146
0
Naive(p50) - Naive(p20) > 0.3β = 2614
Maternal
Paternal
Derivation virgin naive (p20)
D
Early culture (p30)
Lis38 Naive
Lis39 Naive
Lis45 Naive
Lis46 Naive
conversion
Late culture (p50)
Naive
Primed
Methylation Array
Derivation in naive culture conditions
Four lines (2 male Lis38/46 and 2 female Lis39/45)
p20 Naive
primed culture conditions
naive culture conditions
Early culture (passage 30)
Late culture (passage 50)
Refers to part A)
B
Lis38 p20
Lis38 p30
Lis38 p50
Lis39 p20
Lis39 p30
Lis39 p50
Lis45 p20
Lis45 p30
Lis45 p50
1.0
0.8
0.6
0.4
0.2
0.0
FAM71E1
GRID1
CSMD1
BRINP2
MYT1L
BCL11B
SLC6A17
HHLA1
IGF1
CR1L
TRIM61
ADAT3
HPDL
HUS1B
IL6
RPL10L
LEFTY2
DND1
MYH7B
NLRP7
TRIML2
SMARCA2
ABHD12B
GDF3
NODAL
SMAD7
UMODL1
MOG
ZYG11A
FXYD3
PLBD1
NKAPL
DPPA3
ERVH48-1
ABCC13
ZNF534
ATP10B
HPGDS
Lis39 LateN3
Lis39 LateN1
Lis39 LateN2
Lis45 LateN3
Lis45 LateN1
Lis45 LateN2
Lis38 EarlyN1
Lis38 LateN3
Lis38 LateN1
Lis38 LateN2
Lis39 EarlyN1
Lis39 EarlyN3
Lis38 EarlyN2
Lis38 EarlyN3
Lis45 EarlyN1
Lis45 EarlyN2
Lis45 EarlyN3
Lis38 DerivationN3
Lis38 DerivationN1
Lis38 DerivationN2
Lis45 DerivationN1
Lis45 DerivationN2
Lis45 DerivationN3
Lis39 EarlyN2
Lis39 DerivationN2
Lis39 DerivationN1
Lis39 DerivationN3
C
Genes associated with hypomethylated sites (p50 - p20)
Gene Expression (Log2 fold change)
DNMT3L
CSMD1
TRIM61
DPPA2
DPPA3
LEFTY2
Genes upregulated in naive p20 vs naive p50
6
4
2
0
−2
−4
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
Correlation between β difference (methylation) and log2 fold change
RNA-seq
Derivation in naive culture conditions
Four lines (2 male Lis38/46 and 2 female Lis39/45)
p20 Naive
primed culture conditions
naive culture conditions
Early culture (passage 30)
Late culture (passage 50)
Refers to parts B) & C)

To find functional differences related to the differentially methylated sites, we identified 1581 mRNAs that were differentially expressed (Log_2 fold Change $> |1|$) between both the matched naïve p20 vs naïve p30 and naïve p20 vs naïve p50 cultures for these three lines, of which 57 were associated with differentially methylated loci: 38 genes showed low DNA methylation and high gene expression at p20 compared to p30/p50; eight showed high DNA methylation and low gene expression at p20; 10 showed low DNA methylation and low gene expression at p20; and 1 gene that showed high DNA methylation and high gene expression at p20. Thus, the largest group of genes that showed both differential DNA methylation and gene expression was hypomethylated and upregulated in the p20 cultures, and this group contained several previously identified naïve marker genes, including LEFTY2 and DPPA3, as well as TRIM61, a gene closely related to the previously reported naïve marker gene TRIM60 [209, 128] (**Fig 2.15B and 2.15C**).

Gene enrichment analysis of the genes that were hypomethylated and showed higher expression in the p20 cultures compared to the p30/50 cultures revealed significant enrichments for pathways relating to NODAL and TGF-beta signaling, which have been associated with naïve pluripotency [52, 209], SMAD protein phosphorylation and BMP signaling pathway, which have been reported to be important in the acquisition of multi-lineage competence in cells departing naïve pluripotency [134], and maintenance of pluripotency (SMAD)[225] (**Fig. 2.16**).

Loss of DNA methylation and imprinting have been linked to genomic instability [146, 65]. Recent reports have shown that while genome-wide naïve DNA hypomethylation is reversible upon conversion to a primed substate, naïve hPSC methylation levels in imprinted regions are dramatically reduced and not reestablished when naïve hPSCs are transitioned into primed conditions [182, 146]. We analyzed DNA methylation levels at regions associated with single-isoform imprinted genes that were differentially methylated among naïve and primed cultures according to a previous report [15]. We found that our naïve cultures were slightly more hypomethylated than our primed cultures but not in a consistent manner (**Fig. 2.17A**). Moreover, unlike previously

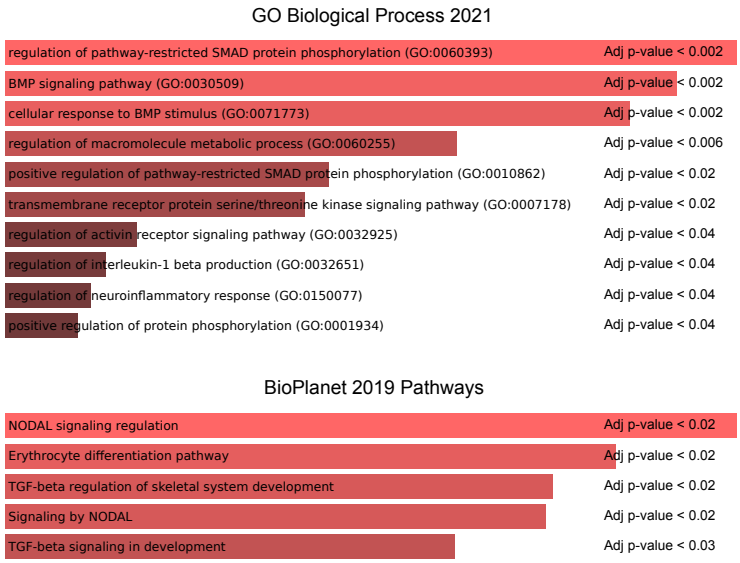
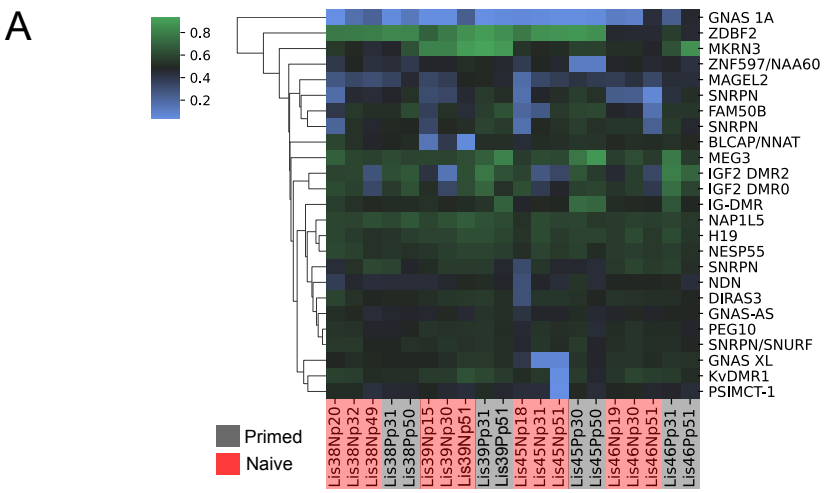
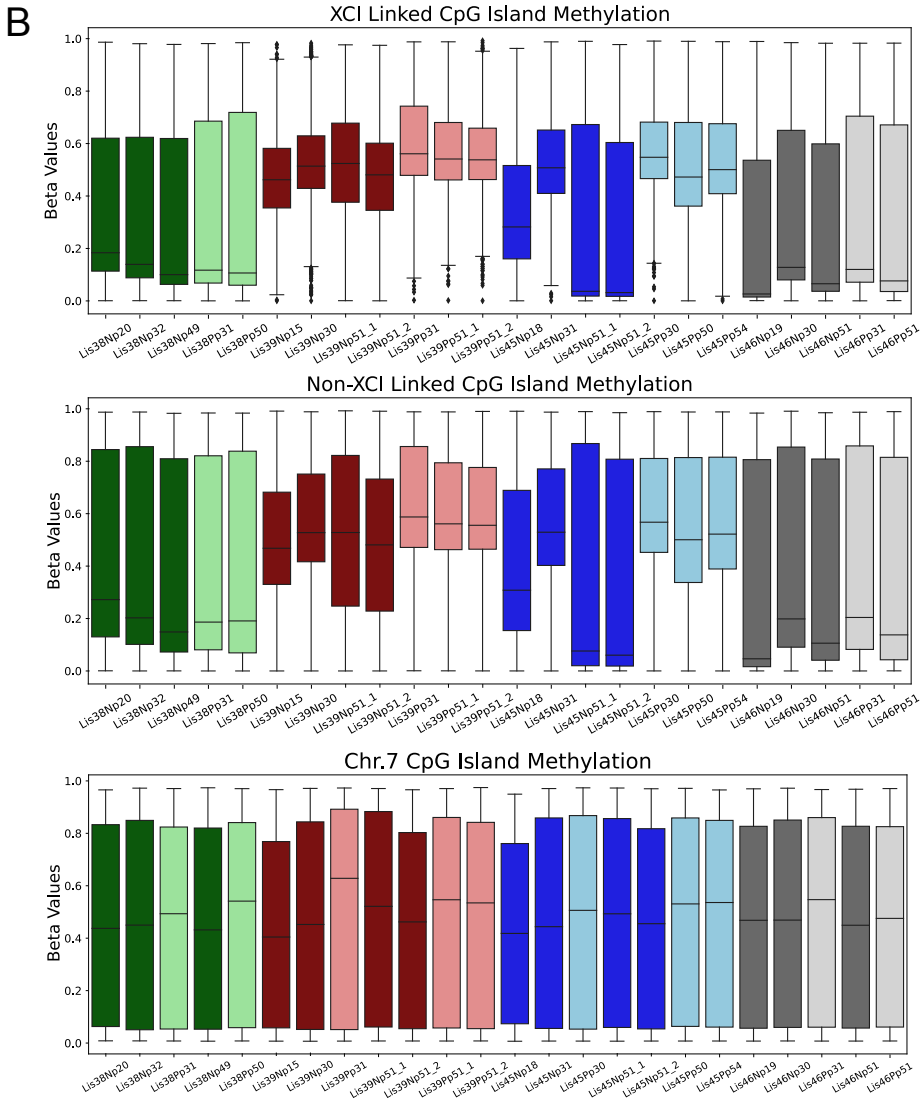


Figure 2.16: Gene ontology enrichment analysis of hypomethylated and high expressed early (p20) naïve cultures.
Gene ontology enrichment analysis of genes both hypomethylated and showing higher expression in the derivation (p20) cultures compared to the early (p30) and late (p50) cultures.

reported 5i naïve lines [182, 146], several imprinted loci that were hypomethylated in our derivation (p20) naïve cultures converted to a hemi-methylated status after transition to a primed state, and in some cases also after extended time in culture in our naïve conditions (e.g., MAGEL2, SNRPN, FAM50B). With the exception of a handful of loci (such as GNAS 1A and ZDBF2, which were hypomethylated and hypermethylated in close to all of our cultures, respectively), our primed cultures were hemimethyated at nearly all of the evaluated imprinted loci. In contrast, the later passage naïve cultures displayed sporadic losses of DNA methylation at imprinted loci, suggesting greater epigenetic stability at imprinted loci in our primed cultures.

Figure 2.17: (*next page*): Methylation at imprinted and X-Chromosome inactivation sites. (A) Heatmap displaying the DNA methylation at regions associated with single-isoform imprinted genes. (B) Methylation levels of the CpG islands associated with X-Chromosome Inactivation (XCI) genes (top boxplot), non-XCI linked genes on the X-chromosome (middle), and with all genes found on chromosome 7 (bottom). Boxes span the top 75th percentile to bottom 25th percentile with the line in the middle of the boxes representing the median. Whiskers contain 95% of the data.


A
0.8
0.6
0.4
0.2
GNAS 1A
ZDBF2
MKRN3
ZNF597/NAA60
MAGEL2
SNRPN
FAM50B
SNRPN
BLCAP/NNAT
MEG3
IGF2 DMR2
IGF2 DMR0
IG-DMR
NAP1L5
H19
NESP55
SNRPN
NDN
DIRAS3
GNAS-AS
PEG10
SNRPN/SNURF
GNAS XL
KvDMR1
PSIMCT-1
Lis38Np20
Lis38Np32
Lis38Np49
Lis38Pp31
Lis38Pp50
Lis39Np15
Lis39Np30
Lis39Np51
Lis39Pp31
Lis39Pp51
Lis45Np18
Lis45Np31
Lis45Np51
Lis45Pp30
Lis45Pp50
Lis46Np19
Lis46Np30
Lis46Pp31
Lis46Pp51
Primed
Naive


B
XCI Linked CpG Island Methylation
Non-XCI Linked CpG Island Methylation
Chr.7 CpG Island Methylation
Beta Values
1.0
0.8
0.6
0.4
0.2
0.0
Lis38Np20
Lis38Np32
Lis38Np49
Lis38Pp31
Lis38Pp50
Lis39Np15
Lis39Np30
Lis39Np51_1
Lis39Np51_2
Lis39Pp31
Lis39Pp51_1
Lis39Pp51_2
Lis45Np18
Lis45Np31
Lis45Np51_1
Lis45Np51_2
Lis45Pp30
Lis45Pp50
Lis45Pp54
Lis46Np19
Lis46Np30
Lis46Np51
Lis46Pp31
Lis46Pp51

X chromosome inactivation (XCI) is characterized by methylation of XCI-linked CpG islands (CGI) on one of the two X chromosomes in female somatic cells [164]. Thus, female somatic cells are hemimethylated at XCI-linked CGIs, while male somatic cells are unmethylated at these loci. Previous reports have shown that both female preimplantation embryos and female human naïve hPSCs contain two actively transcribed X chromosomes [143, 148, 182] and are unmethylated at XCI-linked CGIs. On the other hand, female primed lines typically contain one active X chromosome and have hemimethlated XCI-linked CGIs. We assessed the methylation at the CGIs of XCI-linked, non-XCI-linked X chromosome, and autosomal genes in our cultures (**Fig. 2.17B**). As expected, the male lines were largely unmethylated on the X chromosome. For the two female lines (Lis39 and Lis45), we saw similar DNA methylation levels between the XCI-linked and non-XCI-linked CGIs for both naïve and primed cultures. Also in the female lines, the p20 naïve cultures showed slightly lower levels of DNA methylation on the X chromosome compared to the p30 and p50 naïve and primed cultures, which all had similar X chromosome methylation levels with the exception of the late Lis45 cultures, which were essentially unmethylated on the X chromosome, consistent with the X chromosome deletion seen in the CMA and WGS data (**Fig. 2.10C**). However, the median beta values in all of the naïve female cultures (except for late passage Lis45) were close to 0.3-0.6, consistent with a largely hemimethylated state. Despite this somewhat unexpected hemimethylated pattern, inspection of the RNA-seq data showed that female naïve cultures showed biallelic expression of XCI-linked genes (data not shown), suggesting that they contain two active X chromosomes whereas only Lis45 primed cultures showed biallelic expression.

2.4 Acknowledgements

Chapter 2 is an adapted reprint of material currently being prepared for submission for publication by Chen Dekel*, Robert Morey*, Jacob Hanna, Louise Laurent, Dalit Ben-Yosef, and

Hadar Amir (*These authors contributed equally to this work). The dissertation author was one of the primary investigators and authors of this paper.

Chapter 3

Derivation of functional trophoblast stem cells from primed human pluripotent stem cells

3.1 Introduction

Much of what we know about placental development and trophoblast lineage specification comes from studies in rodents. However, recent reports have highlighted several key differences between these animal models and human placental development [169, 21, 168]. One significant difference between mouse and human involves trophoblast invasion and the vascular remodeling process mediated by these invading cells. It is generally accepted that trophoblast invasion and vascular remodeling is crucial to fetal growth and survival and key to the pathogenesis of preeclampsia [72, 121]. In human, extravillous trophoblast (EVT) invade the maternal endometrium as well as the myometrium but in mice, trophoblast invasion does not extend into the myometrium [121]. Additionally, compared to mice, specification of the trophectoderm (TE), the cells that give rise to the epithelial cells of the placenta, occurs later in human and lacks

expression of key mouse TE genes such as EOMES [21, 168]. Differences such as these lower the utility of mouse models and necessitate human *in vitro* models.

We have reported that hPSCs can be differentiated into trophoblast in culture conditions containing bone morphogenetic protein-4 (BMP4), and that they do so through a process that mimics human placental development. Namely, the differentiating hPSCs first form $KRT7^{+}/TP63^{+}$ cytotrophoblast (CTB), and then further differentiate into either hCG-secreting STB or invasive EVT that express cell-surface HLA-G [112]. We have demonstrated that this method can be used to model both normal development and disease [76]. However, this method is limited by its inability to maintain and expand the CTB-like cells obtained at the end of the first step. Recently, Okae and colleagues described culture conditions for the derivation and maintenance of human trophoblast stem cells (hTSCs) *in vitro* from both preimplantation blastocysts and early first trimester placental tissues [142]. These primary human TSCs can be maintained and expanded indefinitely and differentiated to both EVT- and STB-like cells. Although a vital first step in the quest to model placental development, Okae's protocol was limited to starting material from early first trimester human placental tissue or blastocyst, posing both ethical and practical challenges, namely the inability to model abnormal placental development because tissue would need to be collected prior to knowing pregnancy outcomes.

In this chapter, we applied a culture media recently developed for the maintenance of iTSC, reprogrammed from term CTB [11], to hPSC-derived CTB, and show that this transitions the cells into bona fide TSC. These cells resemble placenta-derived TSC and can be differentiated into functional syncytiotrophoblast and extravillous trophoblast, both *in vitro* and *in vivo*. In addition, we compare these cells to those recently derived from naïve hPSC [84], and show that they resemble naïve hPSC-derived TSC, both during the transition from pluripotency to TE, and in the final CTB-like state. We conclude that, similar to their naïve counterparts, primed hPSC can be converted to self-renewing TSC following a short induction period with BMP4 containing media.

3.2 Methods

3.2.1 Isolation of primary cytotrophoblast and derivation of human trophoblast stem cells and mesenchymal stem cells

Human placental tissues were collected under a UCSD Human Research Protections Program Committee Institutional Review Board-approved protocol; all patients provided informed consent for collection and use of these tissues. Primary cytotrophoblast (CTB) were isolated from a total of 10 early first trimester (gestational age range of 5 to 7 weeks) and 10 late first trimester (gestational age range of 10 to 14 weeks) normal placentae, defined as those coming from patients undergoing elective termination of pregnancy in the absence of known structural fetal abnormalities. CTB isolation was performed as previously described [132]. Briefly, chorionic villi were minced, washed in Hanks' balanced salt solution (Gibco) and digested three times with DNase I (Roche) and trypsin (Gibco). The cells were then pelleted and separated on a Percoll gradient (Sigma-Aldrich) and subjected to sequential magnetic activated cell sorting (MACS) selection (Miltenyi Biotec). Specifically, cells were first subjected to negative selection using a PE-conjugated antibody against HLA-G (EXBIO MEM-G/9) to remove extravillous trophoblast (EVT). The unbound fraction was collected and CTB were selected using APC-conjugated antibody against EGFR (Biolegend 352906) and tested for purity using flow cytometry. CTB preparations yielding greater than 90% EGFR positivity were considered adequate and used in downstream experiments.

Human trophoblast stem cell (hTSC) lines were derived from early first trimester placentae as previously described [132]. Cells were plated on collagen IV coated 12-well plates at a density of 200k cells per well, and cultured in the same media used for culture of pluripotent stem cell-derived TSC (described below). Human mesenchymal stem cells (MSC) were isolated from umbilical cord (UC) tissues as previously described [76].

3.2.2 Human pluripotent stem cell culture and differentiation into TE-like cells

Trophoblast differentiation of human pluripotent stem cells was performed under a protocol approved by the UCSD Institutional Review Board and Embryonic Stem Cell Research Oversight Committee. Two human embryonic stem cell (hESC) lines (WA01/H1 and WA09/H9, obtained from WiCell Institute, Madison, WI, USA), and one iPSC line (reprogrammed from human dermal fibroblast/HDF from ScienCell Research Laboratories, using CytoTune-iPSC 2.0 Sendai Reprogramming Kit from ThermoFisher) [187] were used in this study. Prior to differentiation, hPSCs were converted to feeder-free conditions in StemFlex (ThermoFisher) on Geltrex-(ThermoFisher) coated plates (using 1:200 diluted Geltrex). Differentiation into trophectoderm (TE)-like cells was performed using the first step of the two- step trophoblast differentiation previously established in our lab (detailed in [75]). In brief, hPSCs were dissociated using TrypLE Express (ThermoFisher) and plated onto Geltrex coated plates in StemFlex in the presence of 5μM Y-27632 (Selleck Chemicals). The next day media was changed to first-step differentiation media: DMEM/F12 (ThermoFisher), with 1x ITS (Millipore-Sigma), 64μg/ml L-ascorbic acid (Millipore-Sigma), 543μg/ml NaHCO 3 (Fisher Scientific), 2% BSA (Gemini), 10 ng/mL BMP4 (ThermoFisher), and 2μM IWP2 (Selleck Chemicals). Media was changed every day for 4 days. Conversion into TE-like cells was confirmed based on surface EGFR expression of over 90% by flow cytometry.

3.2.3 Conversion of hPSC-derived TE to TSC-like cells

Following the first-step of trophoblast differentiation above, hPSC-derived TE-like cells were dissociated using TrypLE Express (ThermoFisher) and pelleted by centrifugation at 1000 rpm for 5min. hPSC-derived TE-like cells were resuspended in TSC media and plated with a 1:1 split ratio onto collagen IV (MilliporeSigma, 5μg/mL) coated plates. We used TSC media

that was established in the laboratory of Dr. John Kessler for optimal growth and passage of primary term CTB reprogrammed into self-renewing TSC-like cells [11], which is composed of: Advanced DMEM/F12 (ThermoFisher), 1x N2 (ThermoFisher), 1x B27 (ThermoFisher), 1x Glutmax (ThermoFisher), 150μM 1-thioglycerol (MilliporeSigma), 0.05% BSA (Gemini), 1% Knockout serum replacement (KSR, ThermoFisher), 2μM CHIR99021 (MilliporeSigma), 0.5μM A83-01 (Tocris), 1μM SB431542 (MilliporeSigma), 5μM Y27632 (Selleck Chemicals), 130μg/mL Valproic Acid sodium salt (MilliporeSigma), 100ng/mL recombinant human FGF2 (BioPioneer), 50ng/mL recombinant human EGF (RD Systems), 50ng/mL recombinant human HGF (Stem Cell Technologies), and 20ng/mL Noggin (RD System). Media was replenished every day. Cells were passaged using TrypLE when confluent (approximately every other day) with a split ratio of 1:2 onto collagen IV coated plates. Conversion into TSC was deemed complete after at least 5 passages with surface EGFR expression of over 90%, and HLA-G expression of 20% or less, by flow cytometry.

3.2.4 Statistical analysis

All experiments were performed at least in triplicate. Bar chart data display mean fold change and standard deviation of the mean. Statistical analysis was done using GraphPad Prism 9. Student's t-test was performed to determine significance of differences between groups, and the level of significance is represented with * as indicated in the figures.

3.2.5 Flow cytometric analysis

Flow cytometry was conducted using live cells. Cells were collected and incubated at room temperature for one hour in 200μL FC buffer (1% BSA,10% FBS in PBS) with 1μg APC-conjugated mouse anti-human EGFR antibody (clone AY13, BioLegend) and 2μg PE conjugated mouse anti-HLA-G antibody (MEM-G/9, ExBio) in combination, or 0.125μg APC-conjugated

mouse anti-HLA-A2 (clone BB7.2, BD Pharmigen), 0.22μg PE-conjugated mouse anti-HLA-Bw6 (clone REA143, Miltenyi Biotec), and 0.5μg PE-conjugated rat IgG2b anti- CD49f (ITGA6, clone GoH3, BioLegend) as single antibody stains. APC-conjugated mouse IgG

(cloneMOPC-21, BioLegend), PE-conjugated mouse IgG (cloneMOPC-21, BioLegend), and PE-conjugated rat IgG2b (clone 141945, R&D Systems) were used at the same concentration as isotype IgG controls. Cells were washed 3 times with Wash buffer (0.1% FBS in PBS) and analysis was carried out using a BD FACS-Canto Flow Cytometer.

3.2.6 *In vitro* differentiation of primary and hPSC-derived TSC

Differentiation into extravillous trophoblast (EVT) and syncytiotrophoblast (STB) was performed based on protocols slightly modified from Okae et al. [142]. Briefly, for EVT differentiation, TSCs were dissociated by using TrypLE Express and resuspended in complete EVT media, composed of DMEM/F12 no HEPES (ThermoFisher), 0.3% BSA (Gemini), 1x ITS-X (ThermoFisher), 100μM 2-mercaptoethanol (ThermoFisher), 2.5μM Y-27632 (Selleck Chemicals), 4% KSR (ThermoFisher), 7.5μM A83-01 (Tocris), 100ng/mL NRG1 (Abcam), and supplemented with 2% Matrigel (Corning). 200,000 cells were plated per well of a fibronectin- (MilliporeSigma, 20μg/mL) coated 6-well plate. On the 3rd day of differentiation, media was changed with EVT media without NRG1, and Matrigel was added to a final concentration of 0.5%. Differentiation was deemed complete on day 6 and cells were collected for analysis. For STB differentiation, TSCs were plated at 30% confluence. At this point, media was switched to STB differentiation media, composed of DMEM/F12 no HEPES (ThermoFisher), 0.3% BSA (Gemini), 1x ITS-X (ThermoFisher), 100μM 2 mercaptoethanol (ThermoFisher), 2.5μM Y-27632 (Selleck Chemicals), 4% KSR (ThermoFisher), 2μM Forskolin (Selleck Chemicals). Media was replaced on day 3, and differentiation was deemed complete on day 6 and cells were collected for analysis.

3.2.7 *In vivo* differentiation of primary and hPSC-derived TSC (tumor formation assay)

Use of mice for these assays was approved by the Institutional Animal Care and Use Committee (IACUC) at UC San Diego. Primary and hPSC-derived TSCs were grown to 90% confluence in TSC medium (as described above) and dissociated with TrypLE. 1.5 x 10 7 cells were resuspended in 0.15 mL of a 1:2 mixture of Matrigel and TSC medium, and subcutaneously injected into the flank or hindleg of 8-12-week-old male NOD-SCID mice (JAX Stock No: 005557). Tumor growths were collected 7-10 days after injection. The tumors were fixed in 10% neutral-buffered formalin overnight at 4°C, then processed and embedded in paraffin. HE staining and IHC were performed on 5-μm sections of these tissues. IHC was performed on a Ventana Discovery Ultra automated stainer (Ventana Medical Systems) at the UC San Diego Advanced Tissue Technology Core lab. Following standard antigen retrieval, performed for 40 min at 95°C as per the manufacturer's protocol (Ventana Medical Systems), the sections were stained using mouse anti-HLAG antibody (1:6000; clone 4H84; Abcam), mouse anti-hCG antibody (1:75; clone 5H4-E2; Abcam), or mouse anti-EGFR antibody (1:15; clone 5B7; Ventana Medical Systems). Staining was visualized using 3,3 - diaminobenzidine (DAB, Ventana Medical Systems) and slides were counterstained with Hematoxylin. Slides were analyzed by conventional light microscopy on an Olympus BX43 microscope.

3.2.8 hCG hormone secretion assays

Cell culture supernatants were collected and stored at -80°C until use. Levels of total hCG were quantified using the hCG ELISA Kit (HC251F, Calbiotech Inc.) following manufacturer's protocol. Briefly, 100 uL of each diluted sample was incubated in the provided plate for 60min at room temperature. Samples were removed and wells were washed 3 times with 300uL 1x Wash Buffer. 100uL of TMB substrate was added to each well and incubated at room temperature for

15min. 50uL of stop solution was added to each well. Absorbance was immediately measured at 450nm using Tecan infinite 200Pro. The concentration (mIU/mL) was calculated based off of the standards provided. The concentrations were normalized to total DNA content, extracted by DNeasy (Qiagen), and quantified by NanoDrop spectrophotometer (ThermoFisher).

3.2.9 RNA isolation, cDNA preparation, and quantitative real-time PCR

Total RNA was isolated using NucleoSpin isolation Kit (Macherey-Nagel). RNA concentration was measured using Nanodrop. cDNA was prepared from total RNA using the Primescipt RT-Kit (Takara bio). Quantitative real-time PCR (qPCR) was performed using TB GREEN (Takara bio) on a Quant-it Studio 5 thermocycler (ThermoFisher). Relative expression of each transcript was calculated using $\Delta\Delta$CT method, normalized to L19 rRNA.

3.2.10 RNA sequencing and analysis

For RNA-seq, total RNA was isolated using the mirVana kit (Thermo Fisher). RNA concentration was measured with the QubitTM RNA BR Assay Kit (Thermo Fisher) and RNA integrity with the Agilent RNA 6000 Nano Kit on an Agilent 2100 Bioanalyzer (Agilent). Only samples with RIN above 7.5 were used. RNA-seq libraries were prepared using the TruSeq Stranded Total RNA Sample preparation kit with Ribo-Zero Gold (Illumina) at the IGM Genomics Center, University of California, San Diego, La Jolla, CA. Libraries were pooled and sequenced on NovaSeq 6000 S1Flow Cell (Illumina) to an average depth of 28 million uniquely mapped reads. Quality control was performed using FastQC (v. 0.11.8) and multiQC (v.1.6). Reads were mapped to GRCh38.p10 (GENCODE release 26) using STAR (v. 2.7.3a) [47] and annotated using featureCounts (subread v.1.6.3, GENCODE release 26 primary assembly annotation) [113]. The STAR parameters used were: –runMode alignReads –outSAMmode Full –outSAMattributes Standard –genomeLoad LoadAndKeep –clip3pAdapterSeq

AGATCGGAAGAGC –clip3pAdapterMMp 1. The featureCounts parameters were: -s 2 -p -t exon -T 13 -g gene_id. Ensembl genes without at least three samples with 10 or more reads were removed from analysis. Normalization was performed using the R (v.3.6.3) package DESeq2 (v. 1.28.1) [117]. BiomaRt (v. 2.42.1) was used to convert Ensembl gene ID's to HUGO gene names. Data visualization was done in R (v. 4.0.2) using the packages ggplot2 (v3.3.3) and pvclust (v.2.2-0), in python (v.3.6.11) using the package seaborn (0.11.0), Qlucore Omics Explorer (v3.6) (Qlucore AB, Lund, Sweden), and BioVenn [81]. Differential expression analysis was performed in Qlucore Omics Explorer: first low expression genes were eliminated by variance (0.01-0.02) and then the two-sided two-group comparison (equivalent to a t-test) with q-value <0.05 was used to identify differentially expressed genes. Up and down-regulated genes were manually separated and analyzed in Metascape for gene ontology and PaGenBase enrichment, and for placental cell type using the PlacentaCellEnrich tool [86]. Hierarchical clustering was done using the R (v. 4.0.2) package pvclust (v.2.2-0) and Qlucore.

For single cell sequencing, a total of 11,259 cells across all timepoints were run on the 10X Genomics platform with the Chromium Next GEM Single Cell 3' kit, with libraries prepared according to the manufacturer's protocol (10X Genomics). Libraries were pooled and sequenced on the Illumina NovaSeq 6000 sequencer to an average depth of 20K reads per cell. Raw fastq files were aligned and quantified with the CellRanger (v 6.0.1) count function with default parameters to the human reference genome (GRCh38-2020-A). The counts matrix was filtered to retain genes expressed in greater than 0.1% of the cells, cells that express greater than 200 unique genes, genes expressed in greater than 10 cells, less than 20% mitochondrial reads, greater than 5% ribosomal reads and those with less than 5% hemoglobin reads. The R package Seurat v4 [68] was used with default values for preprocessing, cell cycle scoring, normalization using scTransform regressing out variables for mitochondrial percentages and cell cycle phase, identifying variable features, linear dimensional reduction, clustering and visualization. The FeaturePlot function was used to visualize gene expression of selected genes.

3.2.11 DNA isolation and ELF5 promoter methylation analysis

Genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen) and quantified using the Qubit dsDNA BR assay kit (ThermoFisher), 500 ng of DNA was then bisulfite converted using the EZ DNA Methylation-Lightning Kit (Zymo Research) as per instructions. Following bisulfite conversion, the upstream promoter region of the ELF5 start site was amplified using a nested PCR and primers as described previously [107].

Primer Set 1:

forward: 5'-GGAAATGATGGATATTGAATTTGA-3'

reverse: 5'-CAATAAAAATAAAAACACCTATAACC-3'

Primer Set 2:

forward: 5'-GAGGTTTTAATATTGGGTTTATAATG-3'

reverse: 5'-ATAAATAACACCTACAAACAAATCC-3'

The 20μl PCR reactions were carried out using 10μl KAPA HiFi HotStart Uracil+ ReadyMix (2X), 0.5μl of the forward and reverse primers, 8μl of water, and 1μl of converted DNA. PCR conditions for the first and second PCR were as follows: 95°C 10:00, 35x (95°C 30s, 45°C 30s, 72°C 30s), 72°C 7:00. 1μl of the reaction product from the first PCR was used in the second PCR. Following both the first and second PCR, the PCR products were purified using DNA Clean Concentrator (Zymo Research) and run on the BioAnalyzer (Agilent) for quality control. 2μl of the purified PCR fragment was then A-tailed using 2μL of Taq polymerase 5X reaction buffer, 0.2μL of 50 mM MgCl2, 0.2mM dATP, and 1U of Taq DNA polymerase (Qiagen) in a total reaction volume of 10μL, and incubated at 70°C for 30 min. The products of the A-tailing reaction were then purified using the DNA Clean and Concentrator kit (Zymo Research). The resulting DNA was ligated into the pGEM-T Easy Vector (Promega) according to the manufacturer's instructions using a 1:3 plasmid:insert molar ratio and incubated overnight at 4°C. The ligation reaction was then transformed into DH5α competent bacteria (NEB, Cat No. C2987H) and plated on LB Agar Plates with Ampicillin, IPTG, and X-gal (Teknova, Cat No.

L1949) and incubated overnight at 37°C. The plasmid DNA was then purified using the QIAprep Spin Miniprep Kit (Qiagen) and sent to Eton Bioscience Inc. for sequencing. The sequences were subsequently aligned and analyzed using the BiQAnalyzer software [118].

3.3 Results

3.3.1 Generation of TSC-like cells from primed hPSC

We previously developed a two-step protocol for trophoblast differentiation of primed hPSCs [75]. This protocol involved an initial induction of cells resembling villous cytotrophoblast (CTB), uniformly expressing TP63, using a combination of BMP4 and the WNT inhibitor, IWP2 [75]. This led to formation of a uniform group of EGFR + CTB stem-like cells, which, though not able to self-renew, could be further differentiated, using BMP4 in the presence of feeder-conditioned media, into a mixed set of terminally differentiated trophoblast, including multinucleated hCG-secreting syncytiotrophoblast (STB) and invasive surface HLA-G^+ extravillous trophoblast (EVT) [75]. Nevertheless, this was not only a heterogenous mixture of trophoblast, but also tended to favor STB formation, with only a minority (10-20%) of EVT in the final culture [75]. Here, we applied a new media to CTB stem-like cells at the end of step 1 of our trophoblast differentiation protocol (**Fig. 3.1A**). This media was developed originally by the Kessler lab for reprogramming of term CTB into TSC-like cells (induced TSC/iTSC) [11], and is a modification of the Okae media (see Methods) used for derivation of TSC from human embryos and early gestation placentae [142]. After passage in this media at least 5 times, we were able to derive cells morphologically resembling primary TSC (**Fig. 3.1B**) and uniformly expressing TSC markers, EGFR and ITGA6 (**Fig. 3.1C**). We successfully derived such cells from both human embryonic stem cell (hESC) lines (WA09/H9 and WA01/H1) and induced pluripotent stem cells (iPSCs) (**Fig. 3.2**), having been able to keep these cells in culture for over 20 passages.

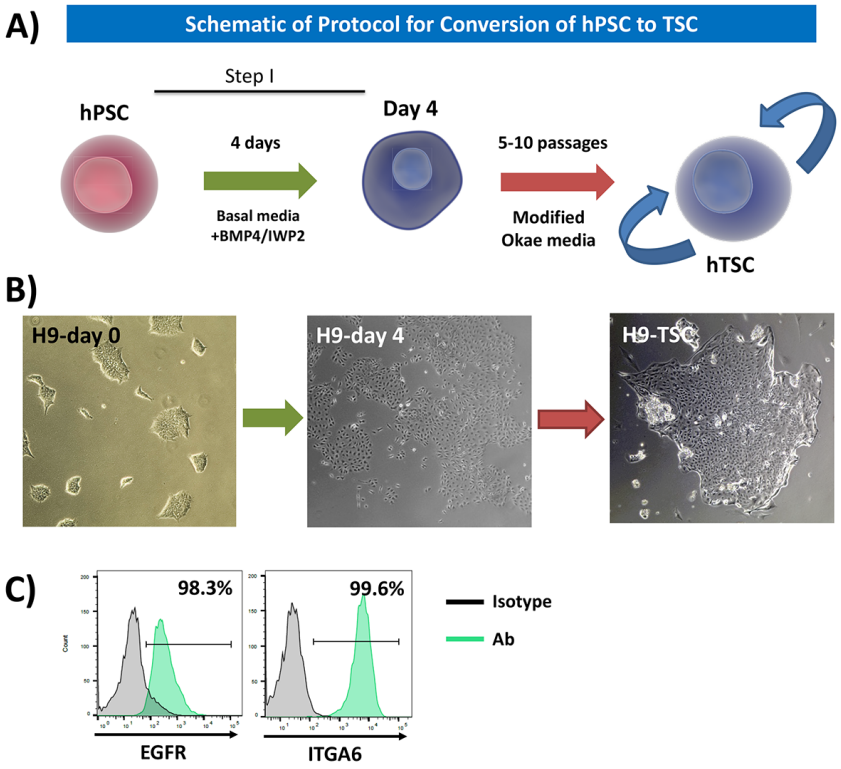


Figure 3.1: Protocol for conversion of primed hPSC into TSC.
(A) Protocol schematic. (B) Morphology of H9 hESC line as undifferentiated (day 0), following 4 days of BMP4/IWP2 treatment (day 4), and after 5 passages in our modified Okae media for TSC (H9-TSC). (C) Flow cytometric analysis of H9-derived TSC for EGFR and ITGA6. Data in B and C are representative of n = 5 independent experiments.

3.3.2 Characterization of cellular identity of hPSC-derived TSC

We next moved to compare the transcriptome of these cells to primary TSC, which we recently derived from early gestation placental tissues in our own laboratory [132], and have confirmed their identity by direct comparison to the Okae cells [142]. We isolated RNA from the hPSCs, in both the undifferentiated state and on days 1-4 post trophoblast induction, as well as from hPSC-derived TSC and the two primary TSC lines (1048 and 1049). Principal component analysis (PCA) and k-means clustering separated undifferentiated hPSC, hPSC treated with BMP4/IWP2 (days 1-4), and both primary and hPSC-derived TSC samples (**Fig. 3.3A**). The transcriptomes of primary TSC and undifferentiated hPSC were then directly compared (variance

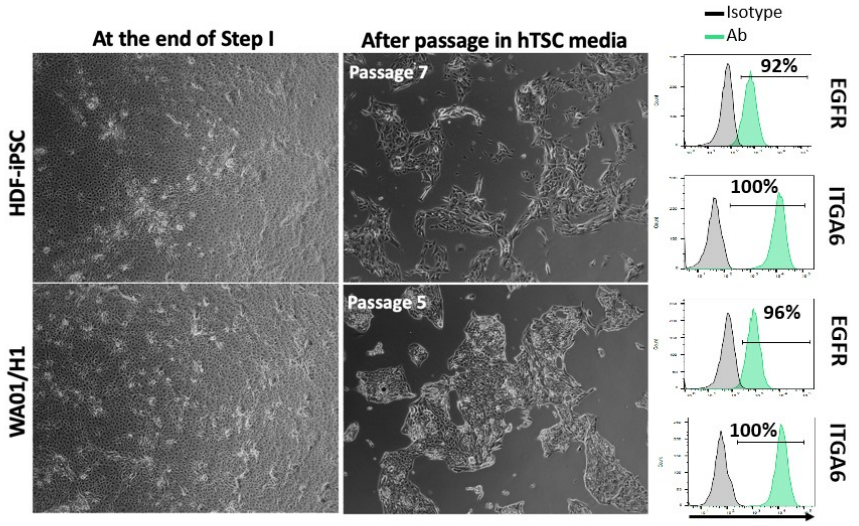
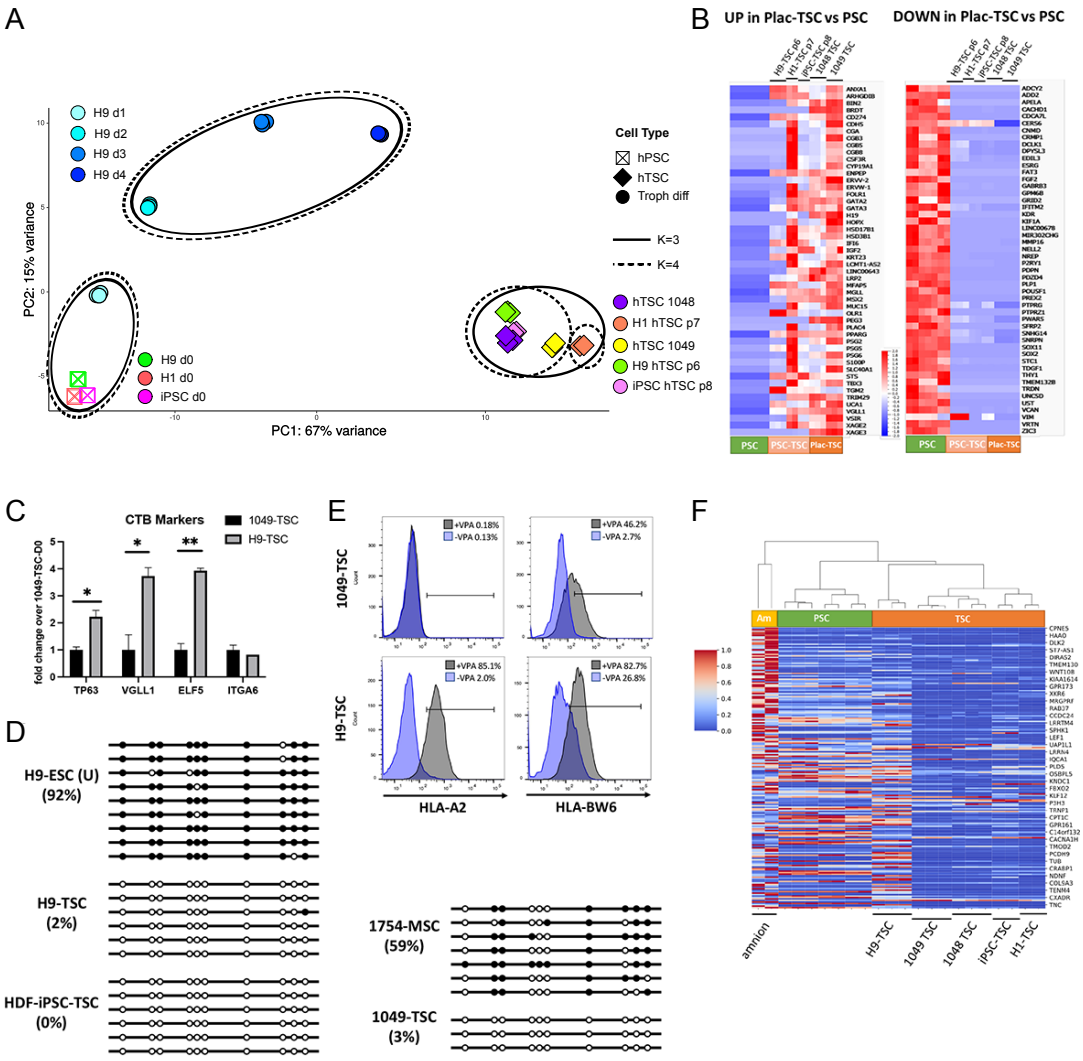


Figure 3.2: Conversion of two additional primed hPSC into TSC.
Conversion of two additional primed hPSC lines (human dermal fibroblast/HDF-derived iPSC and H1-ESC) into TSC, expressing both EGFR and ITGA6 (related to Figure 3.1). Data are representative of 2 independent experiments per line.

= 0.05, two-group analysis $q < 0.05$), leading to identification of 1618 differentially expressed genes, of which 839 genes were upregulated in TSC and 779 in PSC. When the top 50 up and down-regulated genes were evaluated in hPSC-derived TSC, the vast majority showed similar expression levels to primary TSC (**Fig. 3.3B**). The top 50 genes downregulated in primary TSC were all down-regulated in hPSC-derived TSC. Of the top 50 genes upregulated in primary TSC, BRDT was the only gene that failed to be up-regulated in any of the hPSC-TSC lines, while PEG3 and XAGE3 were up-regulated in hPSC-TSC but levels remained lower than in primary TSC (**Fig. 3.3B**). We validated expression levels of a handful of TSC-associated markers (TP63, VGLL1, ELF5 and ITGA6) by qPCR, and found that the hPSC-derived TSC expressed similar or higher levels of these markers (**Fig. 3.3C**).

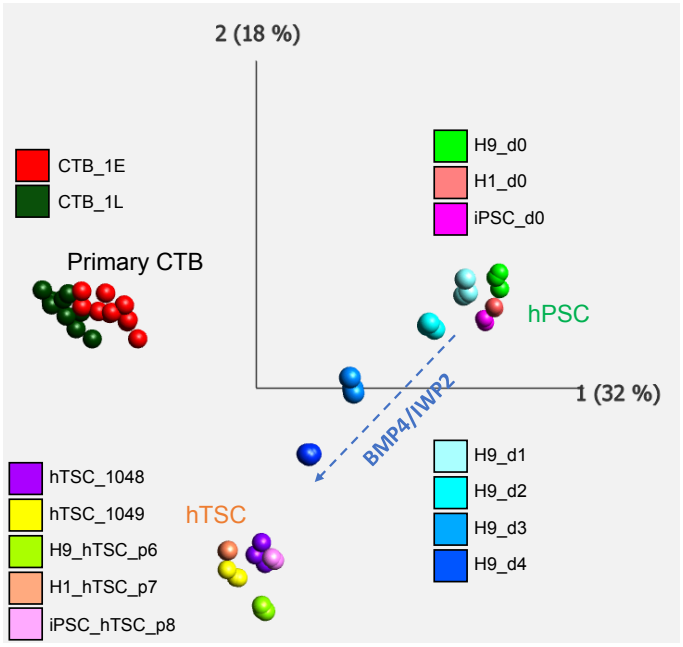
Figure 3.3: (*next page*): Characterization of hPSC-derived TSC.
Principal Component Analysis (A) of hPSC (undifferentiated/day 0), treated with BMP4/IWP2 for 1-4 days, and after 6-8 passages in the modified Okae TSC media (TSC) show that the latter cluster together with primary (placenta-derived) TSC (lines 1048 and 1049). Each dot on the PCA represents a sample from an independent experiment (n = 2 for H1 and iPSC, day 0 and TSC; n = 3 for all other samples). Dotted lines show k-means clustering with k=3 and solid lines show k=4. (B) Heatmap of top 50 genes, either upregulated (left) or downregulated (right), in primary (placenta-derived) TSC (Plac-TSC) vs. undifferentiated hPSC, as expressed in hPSC-derived TSC (PSC-TSC). (C) qPCR of the indicated CTB markers in undifferentiated primary TSC (1049) compared to hPSC (H9)-TSC. Data were normalized to L19 and shown as fold change over undifferentiated 1049 (D0=day 0), and represent mean +/- SD for n = 3 independent experiments. *$p < 0.05$; **$p < 0.01$ by student's t-test. (D) DNA methylation surrounding the ELF5 promoter in undifferentiated (U) H9 ESC, TSC derived from both H9 and a human dermal fibroblast (HDF)-derived iPSC, an umbilical cord derived mesenchymal stem cell line (1754), and a primary (placenta-derived) TSC line (1049). Each line represents a distinct sequenced clone (n= 3 to 9). (E) Flow cytometric analysis of primary (1049) and hPSC (H9)-derived TSC for HLA-A2 and HLA-Bw6 with (grey) and without (purple) valproic acid (VPA) in the culture media. Data are representative of 2 independent experiments. (F) Heatmap of first trimester amnion-specific markers (based on Roost et al. [159]) in undifferentiated hPSC, as well as primary (1048 and 1049 TSC) and hPSC-derived TSC.


A
H9 d1
H9 d2
H9 d3
H9 d4
H9 d0
H1 d0
iPSC d0
PC2: 15% variance
PC1: 67% variance
Cell Type
hPSC
hTSC
Troph diff
K=3
K=4
hTSC 1048
H1 hTSC p7
hTSC 1049
H9 hTSC p6
iPSC hTSC p8
B
UP in Plac-TSC vs PSC
DOWN in Plac-TSC vs PSC
H9-TSC p6
H1-TSC p7
iPSC-TSC p8
1048 TSC
1049 TSC
PSC
PSC-TSC
Plac-TSC
C
CTB Markers
1049-TSC
H9-TSC
fold change over 1049-TSC-D0
TP63
VGLL1
ELF5
ITGA6
D
H9-ESC (U)
(92%)
H9-TSC
(2%)
HDF-iPSC-TSC
(0%)
E
1049-TSC
H9-TSC
+VPA 0.18%
-VPA 0.13%
+VPA 46.2%
-VPA 2.7%
+VPA 85.1%
-VPA 2.0%
+VPA 82.7%
-VPA 26.8%
HLA-A2
HLA-BW6
1754-MSC
(59%)
1049-TSC
(3%)
F
Am
PSC
TSC
amnion
H9-TSC
1049 TSC
1048 TSC
iPSC-TSC
H1-TSC

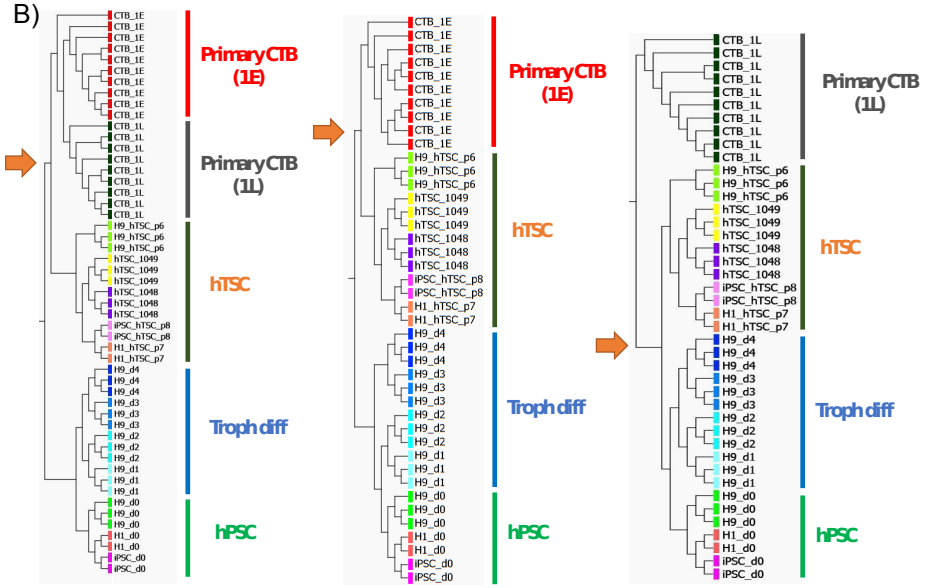
Primary TSC can only be derived from early first trimester (less than 10 weeks gestational age) cytotrophoblast (CTB) [142]. We therefore asked how both our primary and hPSC-derived TSC compare to CTB isolated from early (6-8 weeks, CTB_1E) vs. late (10-14 weeks, CTB_1L) first trimester placentae. Transcriptomic analysis showed that both primary and hPSC-derived TSC were distinct from isolated CTB (**Fig. 3.4A**). Unsupervised hierarchical clustering was performed with primary CTB sub-groups combined and independently. Both primary and hPSC-derived TSC clustered with early first trimester CTB when these cells were included in the analysis, while they clustered with undifferentiated and BMP-treated hPSC cells when compared to late first trimester CTB group only (compare left and center dendrograms to the one on the right, in **Fig. 3.4B**).

Figure 3.4: (*next page*): Comparison of hPSC-derived TSC to primary TSC and CTB (related to Fig. 3.3).
(A) Principal component analysis of undifferentiated hPSC, hPSC after 1-4 days of BMP4/IWP2 treatment, and primary and hPSC-derived TSC, in comparison to CTB isolated from early (5-8 weeks) or late (10-14 weeks) first trimester placentae. For hPSC lines, each dot represents a sample from an independent experiment (n = 2 for H1 and iPSC, day 0 and TSC; n = 3 for all H9 samples); for CTB samples, each dot represents a biological replicate. (B) Dendrograms of the same cells compared to early- and late- first trimester CTB, either together or independently. Arrow points to cluster to which both primary and hPSC-derived TSC belong in each comparison. Primary CTB (1E) are early first trimester primary CTB and Primary CTB (1L) are late first trimester primary CTB.

A)


2 (18 %)
1 (32 %)
CTB_1E
CTB_1L
Primary CTB
H9_d0
H1_d0
iPSC_d0
hPSC
BMP4/IWP2
H9_d1
H9_d2
H9_d3
H9_d4
hTSC_1048
hTSC_1049
H9_hTSC_p6
H1_hTSC_p7
iPSC_hTSC_p8
hTSC

B)


Primary CTB (1E)
Primary CTB (1L)
hTSC
Troph diff
hPSC

We next aimed to go beyond the transcriptome, and characterize two other features of early gestation CTB, in our hPSC-derived TSC, namely methylation status of the ELF5 promoter and expression of class I HLA. The ELF5 promoter is known to be hypermethylated in hPSCs, but hypomethylated in early gestation CTB [73] and primary TSC [142]. Compared to both undifferentiated hPSC and mesenchymal stem cells derived from umbilical cord [76], both primary and hPSC-derived TSCs showed significant hypomethylation of this genomic region (**Fig. 3.3D**). In addition to a hypomethylated ELF5 promoter, villous CTB and TSC, which are thought to reside within the CTB layer, are both known to lack expression of class I HLA (HLA-A and HLA-B) [84, 4]. When we evaluated our cells by flow cytometry, both primary and hPSC-derived TSC expressed class I HLA. Therefore, we re-evaluated our media components and noted that valproic acid (VPA), a histone deacetylase (HDAC) inhibitor, had been previously identified as a compound that promotes class I HLA expression in tumor cells [5, 131, 219, 222]. When we removed this chemical from our media, after only three passages, class I HLA expression significantly decreased in both primary and hPSC-derived TSC (**Fig. 3.3E**), without affecting self-renewal.

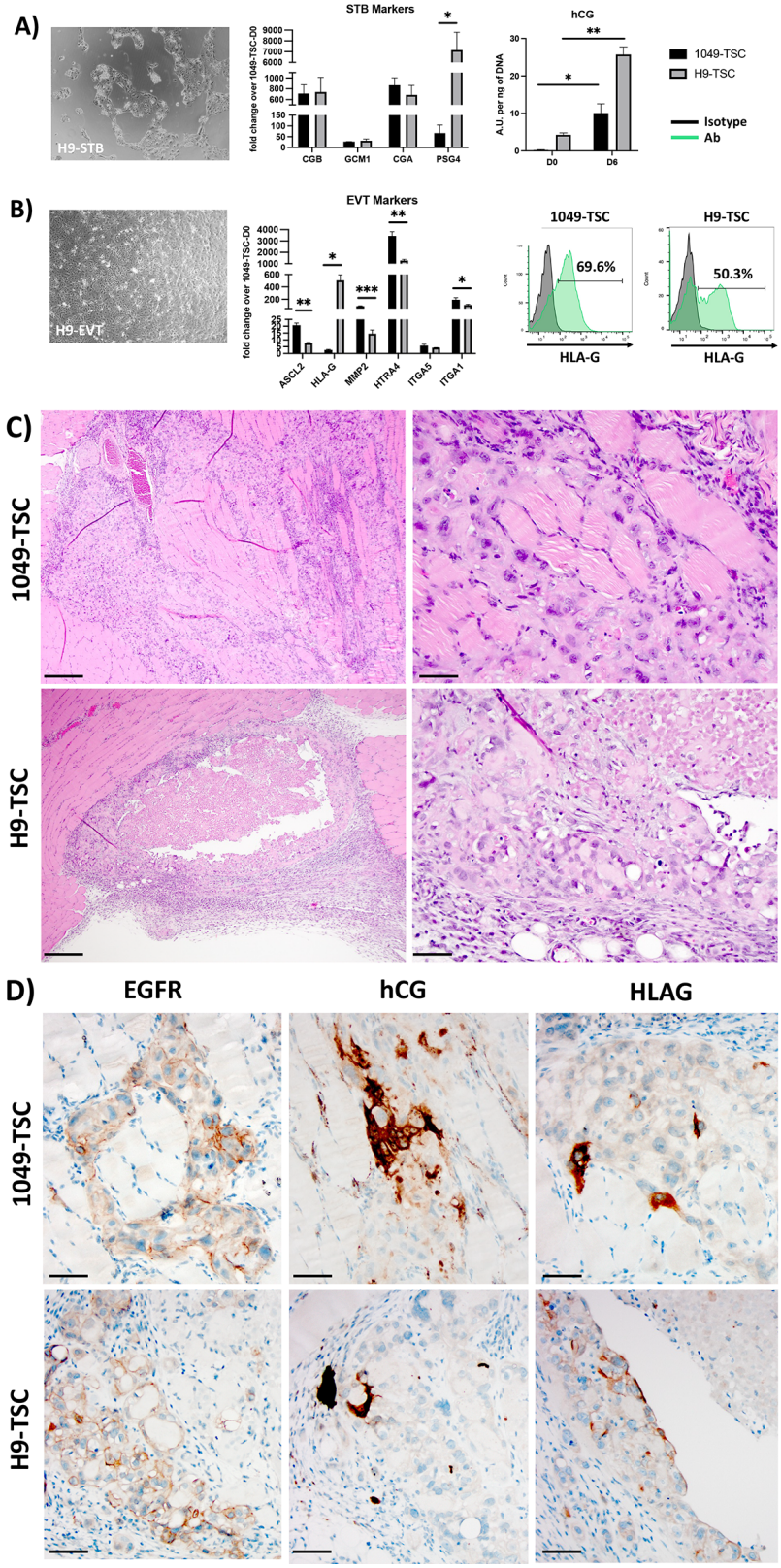
It has been argued that, unlike naïve hPSC, primed hPSC give rise to amnion when treated with media containing BMP4 [84, 63]. To further examine the identity of hPSC-derived TSC, we next evaluated expression of a set of amnion-specific genes [159] in primary TSC, compared to hPSC and their TSC derivatives, and found that, overall, the expression of these genes was lower in hTSC and similar between primary and hPSC-derived TSC (**Fig. 3.3F**). Based on all these shared features with primary TSC, we conclude that we have successfully derived TSC-like cells from primed hPSC.

3.3.3 Functional characterization of hPSC-derived TSC

We next moved to characterize the differentiation ability of primed hPSC-derived TSC. When differentiated into STB, they upregulated appropriate markers, including CGA, CGB,

GCM1, and PSG4 by qPCR, and secreted hCG into the media, at similar levels as primary TSC- derived STB (**Fig. 3.5A**). Similarly, hPSC-derived TSC could be differentiated into EVT, expressing similar levels of EVT markers, including ASCL2, HLAG, MMP2, HTRA4, ITGA5, and ITGA1, and gaining surface HLA-G expression, similar to primary TSC-derived EVT (**Fig. 3.5B**). Finally, when injected into NOD-SCID mice, both primary and hPSC-derived TSC produced mixed trophoblastic tumors, composed of both hCG+ and HLAG+ cells (**Fig. 3.5C and D**). Thus, both *in vitro* and *in vivo*, hPSC-derived TSC showed the capacity to differentiate into both terminally- differentiated trophoblast lineages.

Figure 3.5: (*next page*): *In vitro* and *in vivo* differentiation potential of hPSC-derived TSC. (A) Morphology, lineage-specific gene expression, and hCG secretion of primary (1049) and hPSC (H9)-derived TSC differentiated into syncytiotrophoblast (STB). qPCR data were normalized to L19 and shown as fold change over undifferentiated (day 0/D0) 1049 TSC. hCG secretion was normalized to ng of DNA. Both qPCR and ELISA data represent mean +/- SD for n = 3 independent experiments. *$p < 0.05$; **$p < 0.01$ by student's t-test. (B) Morphology, lineage-specific gene expression, and flow cytometric analysis for surface HLA-G expression of primary (1049) and hPSC (H9)-derived TSC differentiated into extravillous trophoblast (EVT). qPCR data were normalized to L19 and shown as fold change over undifferentiated (day 0/D0) 1049 TSC, and represent mean +/- SD for n = 3 independent experiments. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ by student's t-test. Flow cytometric data are representative of 3 independent experiments. (C) and (D) Tumors generated 10 days following injection of primary (1049) and hPSC (H9)-derived TSC into NOD-SCID mice (representative of n = 2 independent experiments). (C) H&E staining shows the tumor cells invading through muscle (1049-TSC tumor) or forming a tumor with a necrotic center (H9-TSC), both characteristic of human trophoblastic tumors. Scale bars = 250μm for low-power images (left) or 50μm for high-power images (right). (D) Immunohistochemical staining of the same tumors using antibodies against EGFR, hCG, and HLAG shows positively-stained cells (brown) in the TSC-derived lesions. Scale bars = 50μm.


A)
H9-STB
STB Markers
fold change over 1049-TSC-D0
CGB
GCM1
CGA
PSG4
hCG
A.U. per ng of DNA
D0
D6
1049-TSC
H9-TSC
Isotype
Ab
B)
H9-EVT
EVT Markers
ASCL2
HLA-G
MMP2
HTRA4
ITGA5
ITGA1
1049-TSC
69.6%
HLA-G
H9-TSC
50.3%
HLA-G
C)
1049-TSC
H9-TSC
D)
EGFR
hCG
HLAG
1049-TSC
H9-TSC

3.3.4 Characterizing the path from primed pluripotency to TSC

Our trophoblast induction protocol [75] was partially based on a previous study by Kurek et al. [102], showing that BMP4 induces both trophoblast and mesoderm lineages, where only the latter lineage is WNT-dependent. Thus, the addition of IWP2 should allow exclusive differentiation into trophoblast. We evaluated cells at the end of step 1 (BMP4/IWP2 treatment for 4 days), and noted the gradual flattening of the cells from pluripotency to this stage, and gain of EGFR expression (**Fig. 3.6A**). By qPCR, these cells lost expression of pluripotency factors, and gained expression of markers of trophectoderm (TE), including NR2F2 and CDX2, as well as markers of CTB, including VGLL1 and TP63 (**Fig. 3.6B**); with the exception of DNMT3L, markers of naïve pluripotency were either decreased or only slightly increased (**Fig. 3.6B**). The transition from this state to TSC through several rounds of passage in our modified Okae media was characterized by an initial transition of cells into a mesenchymal morphology, with loss of EGFR, and then a gradual re-epithelialization, with re-gaining of EGFR (**Fig. 3.6C**). Transcriptomic analysis of this adaptation phase showed unique changes in the gene expression of cells from passage 0 (p0) to p2, before returning back to a signature that more closely resembled primary TSC (**Fig. 3.6D**). Of the 839 genes differentially up-regulated in primary TSC compared to undifferentiated hPSC, 37% (308 genes) were already up-regulated at the end of step I, while 59% (492 genes) were up-regulated in hPSC-derived TSC (**Fig. 3.7A**). The adaptation did not change the nature of the cells as the most significant tissue-specific signature of all the timepoints was associated with "placenta" (**Fig. 3.7B**). Compared to cells at the end of step 1 (4 days of BMP4/IWP2 treatment), cells at p0 (about 48h after being switched to TSC media) demonstrated down-regulation of genes associated with adherens and tight junctions, and actin cytoskeleton reorganization, which correlates with the observed loss of cell-cell adhesion (**Fig. 3.7C**). Compared to both primary and hPSC-derived TSC, cells at p2 (the transition point furthest away from TSC) had up-regulated genes associated with the epithelial-mesenchymal transition, extracellular matrix reorganization, cell-cell and cell-substrate adhesion (**Fig. 3.7D**).

Figure 3.6: (*next page*): Transition from primed pluripotency to TSC involves a trophectoderm-like intermediate.
(A) Morphology of H9-ESC at days 0-4 following treatment with BMP4/IWP2. At the end of 4 days, the majority of the cells express EGFR by flow cytometric analysis. Data are representative of n = 5 independent experiments. (B) qPCR of H9-ESC at days 0, 2, and 4 following BMP4/IWP2 treatment, for markers of pluripotency (NANOG and POU5F1), naïve pluripotency (DNMT3L, KLF17, DPPA3, and DPPA5), trophectoderm (ENPEP, TACSTD2, NR2F2, and CDX2), and cytotrophoblast (CTB) (TP63, VGLL1, GATA2, GATA3, TFAP2C). Data were normalized to L19 and shown as fold change over undifferentiated H9-ESC (D0=day 0), and represent mean +/- SD for n = 3 independent experiments. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$ by student's t-test. (C) Morphology and EGFR expression of BMP4/IWP2-treated H9-ESC, passaged in modified Okae TSC media. During this time, the cells initially lose EGFR expression and slowly re-gain it after at least 5 passages. Data are representative of n = 3 independent experiments. (D) Principal component analysis of cells transitioning from primed pluripotency (hPSC) through trophoblast induction (using BMP4/IWP2), then undergoing adaptation through 6 passages (p0- p6) in modified Okae TSC media. Arrow points to passage in which the hPSC-derived cells are farthest away from primary TSC. Each dot on the PCA represents a sample from an independent experiment (n = 2 for H1 and iPSC, day 0 and TSC; n = 3 for all H9 samples).

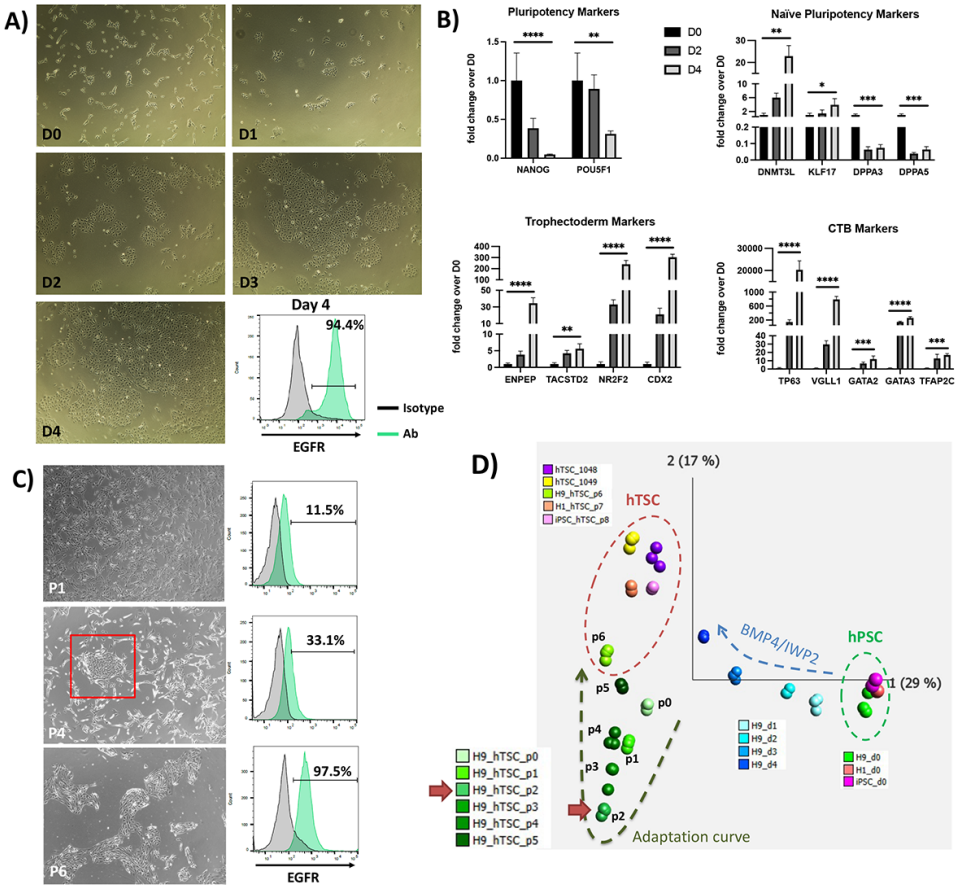
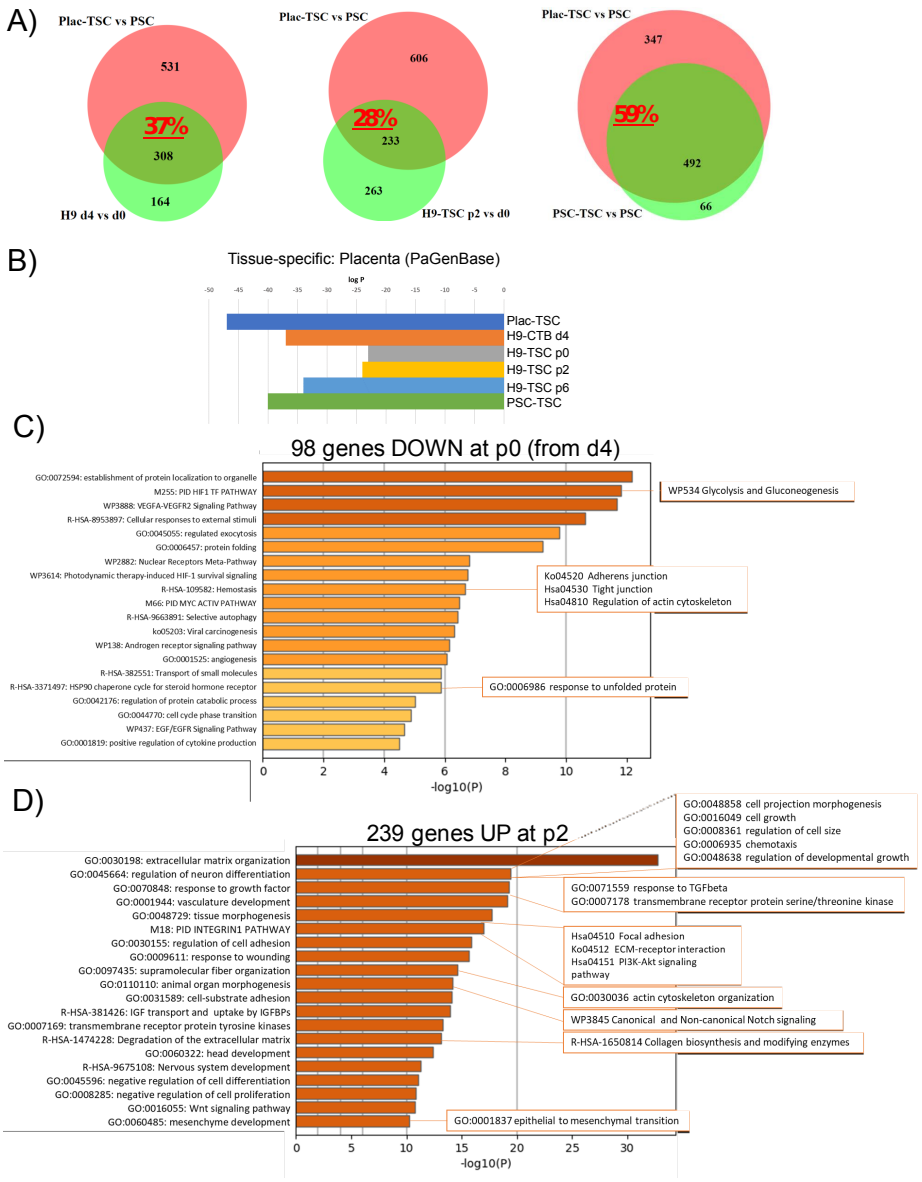

A)
D0
D1
D2
D3
D4
Day 4
94.4%
Isotype
Ab
EGFR
B)
Pluripotency Markers
fold change over D0
NANOG
POU5F1
D0
D2
D4
Naïve Pluripotency Markers
DNMT3L
KLF17
DPPA3
DPPA5
Trophectoderm Markers
ENPEP
TACSTD2
NR2F2
CDX2
CTB Markers
TP63
VGLL1
GATA2
GATA3
TFAP2C
C)
P1
P4
P6
11.5%
33.1%
97.5%
EGFR
D)
hTSC_1048
hTSC_1049
H9_hTSC_p6
H1_hTSC_p7
iPSC_hTSC_p8
hTSC
2 (17 %)
BMP4/IWP2
hPSC
1 (29 %)
p6
p5
p0
p4
p1
p3
p2
H9_hTSC_p0
H9_hTSC_p1
H9_hTSC_p2
H9_hTSC_p3
H9_hTSC_p4
H9_hTSC_p5
H9_d1
H9_d2
H9_d3
H9_d4
H9_d0
H1_d0
iPSC_d0
Adaptation curve

Figure 3.7: (*next page*): Analysis of gene expression changes during adaptation of BMP4/IWP2-treated hPSC in modified Okae TSC media (related to Fig. 3.6).
(A) Proportion of genes (839), differentially up-regulated in primary TSC compared to undifferentiated hPSC, first decreases from 37% at 4 days of BMP4/IWP2 treatment to 28% after two passages in the modified Okae media, but then increases to 59% in hPSC- derived TSC. (B) The most significant tissue-specific signature of hPSC-derived trophoblast at all the timepoints was associated with "placenta." (C) Downregulated genes in passage 0 (p0), 48 hours after switching BMP4/IWP2-treated cells to TSC media, include those associated with adherens and tight junctions and actin cytoskeleton reorganization, consistent with the initial loss of epithelial morphology. (D) Upregulated genes at passage 2 (p2) (the transition point furthest away from TSC), include those associated with the epithelial-mesenchymal transition, extracellular matrix reorganization, and cell-cell and cell-substrate adhesion.


A)
Plac-TSC vs PSC
531
37%
308
164
H9 d4 vs d0
Plac-TSC vs PSC
606
28%
233
263
H9-TSC p2 vs d0
Plac-TSC vs PSC
347
59%
492
66
PSC-TSC vs PSC
B)
Tissue-specific: Placenta (PaGenBase)
log P
-50 -45 -40 -35 -30 -25 -20 -15 -10 -5 0
Plac-TSC
H9-CTB d4
H9-TSC p0
H9-TSC p2
H9-TSC p6
PSC-TSC
C)
98 genes DOWN at p0 (from d4)
GO:0072594: establishment of protein localization to organelle
M255: PID HIF1 TF PATHWAY
WP3888: VEGFA-VEGFR2 Signaling Pathway
R-HSA-8953897: Cellular responses to external stimuli
GO:0045055: regulated exocytosis
GO:0006457: protein folding
WP2882: Nuclear Receptors Meta-Pathway
WP3614: Photodynamic therapy-induced HIF-1 survival signaling
R-HSA-109582: Hemostasis
M66: PID MYC ACTIV PATHWAY
R-HSA-9663891: Selective autophagy
ko05203: Viral carcinogenesis
WP138: Androgen receptor signaling pathway
GO:0001525: angiogenesis
R-HSA-382551: Transport of small molecules
R-HSA-3371497: HSP90 chaperone cycle for steroid hormone receptor
GO:0042176: regulation of protein catabolic process
GO:0044770: cell cycle phase transition
WP437: EGF/EGFR Signaling Pathway
GO:0001819: positive regulation of cytokine production
WP534 Glycolysis and Gluconeogenesis
Ko04520 Adherens junction
Hsa04530 Tight junction
Hsa04810 Regulation of actin cytoskeleton
GO:0006986 response to unfolded protein
0 2 4 6 8 10 12
-log10(P)
D)
239 genes UP at p2
GO:0030198: extracellular matrix organization
GO:0045664: regulation of neuron differentiation
GO:0070848: response to growth factor
GO:0001944: vasculature development
GO:0048729: tissue morphogenesis
M18: PID INTEGRIN1 PATHWAY
GO:0030155: regulation of cell adhesion
GO:0009611: response to wounding
GO:0097435: supramolecular fiber organization
GO:0110110: animal organ morphogenesis
GO:0031589: cell-substrate adhesion
R-HSA-381426: IGF transport and uptake by IGFBPs
GO:0007169: transmembrane receptor protein tyrosine kinases
R-HSA-1474228: Degradation of the extracellular matrix
GO:0060322: head development
R-HSA-9675108: Nervous system development
GO:0045596: negative regulation of cell differentiation
GO:0008285: negative regulation of cell proliferation
GO:0016055: Wnt signaling pathway
GO:0060485: mesenchyme development
GO:0048858 cell projection morphogenesis
GO:0016049 cell growth
GO:0008361 regulation of cell size
GO:0006935 chemotaxis
GO:0048638 regulation of developmental growth
GO:0071559 response to TGFbeta
GO:0007178 transmembrane receptor protein serine/threonine kinase
Hsa04510 Focal adhesion
Ko04512 ECM-receptor interaction
Hsa04151 PI3K-Akt signaling pathway
GO:0030036 actin cytoskeleton organization
WP3845 Canonical and Non-canonical Notch signaling
R-HSA-1650814 Collagen biosynthesis and modifying enzymes
GO:0001837 epithelial to mesenchymal transition
0 5 10 15 20 25 30
-log10(P)

Recently, Io et al. [84] developed a BMP4-based protocol for transitioning naïve hPSCs, first into a TE-like state (after 3 days), characterized by expression of ENPEP, TACSTD2, and NR2F2, and subsequently into a CTB-like state (after 3-15 passages), characterized by SIGLEC6 among other markers (including VGLL1, TP63, and EGFR). We compared our cells to theirs, starting with bulk RNA-seq, from primary CTB isolated by both our groups, as well as naïve and primed hPSC and their trophoblast derivatives. Principal component analysis (PCA) showed separation between the naïve and primed hPSC group and primary CTBs (separated along the PC1 axis, 67% variance), with both the naïve and primed hPSC-derived cells located between these two groups on PC1 (**Fig. 3.8A**). Unsupervised hierarchical clustering showed that, while naïve and primed hPSC and their early derivatives (Io's naïve-TE at day 1, and our BMP4/IWP2-treated cells at days 1-2) clustered together, all the later derivatives (Io's naïve-TE at days 2-3, and our BMP4/IWP2-treated cells at days 3-4, as well as both naïve- and primed-hPSC-derived TSC) clustered together with primary CTB (**Fig. 3.8B**). Within the latter group, our BMP4/IWP2-treated cells (at days 3-4) clustered together with Io's naïve-TE (at days 2-3), suggesting the primed cells go through a TE-like phase, while our primed-TSC (at passage 6-8) clustered together with Io's naïve-CTB (at passage 3-15) (**Fig. 3.8B**).

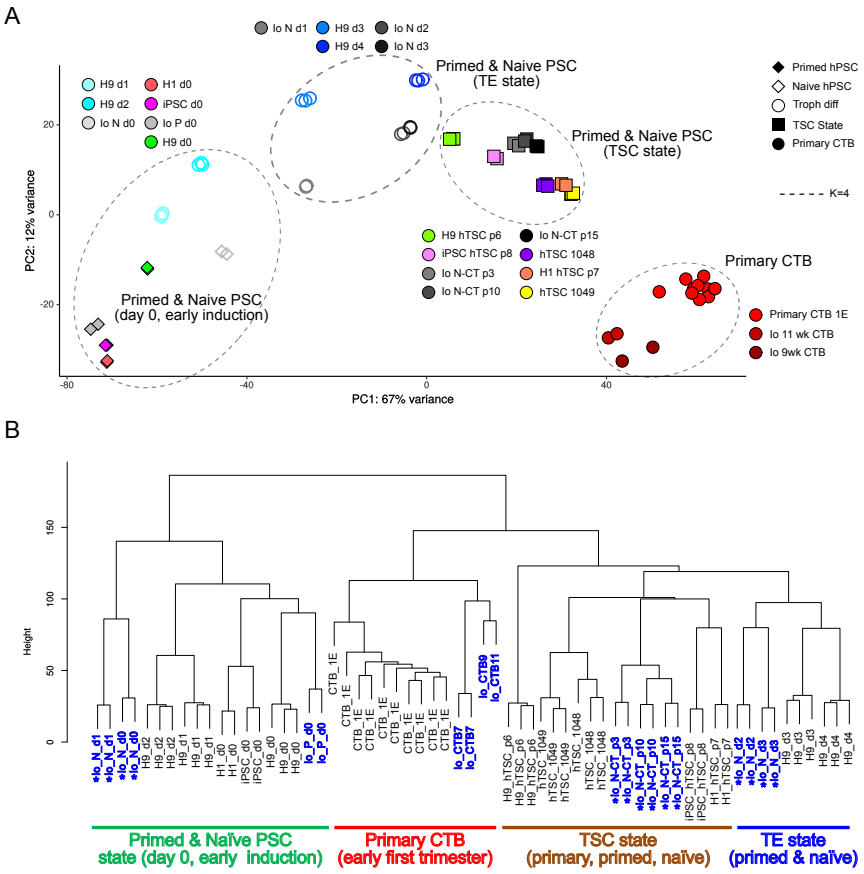


Figure 3.8: Comparison of trophectoderm and TSC derived from naïve and primed hPSC (part 1).

(A) Principal component analysis of our cells, combined with those published by Io et al. [84], including, from left to right: 1) undifferentiated naïve and primed hPSC (day 0) and their early trophoblast derivatives (days 1-2 in our BMP4/IWP2 treatment, and day 1 of Io et al.'s induction); 2) naïve and primed hPSC induced into a trophectoderm/TE-like fate (days 3-4 in our BMP4/IWP2 treatment, and days 2-3 of Io et al.'s induction); 3) primary (placenta-derived) TSC (1048 and 1049 lines), our primed hPSC-derived TSC (H9-TSC at passage 6, H1-TSC at passage 7, iPSC- TSC at passage 8), and Io et al.'s naïve hPSC-derived TSC (naïve cytotrophoblast or N-CT at passage 3-15); and 4) primary cytotrophoblast (CTB) (CTB-1E include our preps from 10 different 5-8 week-gestation placentae; other four samples include two 9-week and two 11-week gestation CTB preps from Io et al.). For hPSC lines, each dot represents a sample from an independent experiment (n = 2 for H1 and iPSC, day 0 and TSC, and all samples from Io et al.; n = 3 for all H9 samples); for CTB samples, each dot represents a biological replicate. Dotted lines show k-means clustering (k=4). (B) Dendrogram of the same samples, showing primed and naïve hPSC and their early derivatives clustering separately (left) from primary CTB, hPSC-derived TSC (naïve and primed), as well as hPSC-derived TE (naïve and primed).

We next evaluated the progression from primed pluripotency to TSC by single cell RNA sequencing (scRNA-seq) of H9 hESC, both undifferentiated and treated with BMP4/IWP2 for 12 hours, 24 hours, and 4 days, as well as H9-TSC, using 10X Genomics. A total of 11,259 cells across all timepoints were analyzed by Seurat, which separated the cells into 4 clusters (**Fig. 3.9A**). Cluster 0, despite containing mostly cells at early stages of trophoblast induction with BMP4/IWP2 (12h and 24h), showed a separate transcriptional program from undifferentiated hPSC in cluster 1. Feature plots of pluripotency- and trophoblast-associated genes showed that both clusters 0 and 1 have a strong pluripotency signature (POU5F1 and NANOG), with a subset of cluster 0 cells initiating the TE program, as noted by GATA3 expression, as early as 12 hours from induction. Cluster 3 cells (those treated with BMP4/IWP2 for 4 days) had completely lost NANOG expression, but had begun to express TE-associated genes, including KRT19, ENPEP, TACSTD2, and NR2F2, similar to Io's naïve PSC-derived TE [84] (**Fig. 3.9A**). As TSCs, cells gained expression of CTB-associated genes, including EGFR, VGLL1, and SIGLEC6, similar to Io's naïve PSC-derived CTB [84] (**Fig. 3.9A**). Naïve pluripotency markers were expressed at low levels in rare undifferentiated primed hPSC but (with the exception of DNMT3L) they were absent in BMP4/IWP2-treated hPSC (**Fig. 3.9A**). When the transcriptomic signatures of hPSC at day 4 of BMP4/IWP2 treatment as well as both primary and hPSC-derived TSC were compared to undifferentiated hPSC and analyzed by the PlacentaCellEnrich tool [86], all 3 groups showed highest enrichment of genes from trophoblast cells in the placenta, with similar enrichment for syncytiotrophoblast (STB)-, extravillous trophoblast (EVT)-, and cytotrophoblast (CTB)-associated genes (**Fig. 3.9B**). These data suggest that primed hPSC have the potential to give rise to TSC, following induction into a TE-like phase with 4 days of treatment with BMP4/IWP2, providing a simple protocol for obtaining TSC from a vast array of available embryonic and induced pluripotent stem cells.

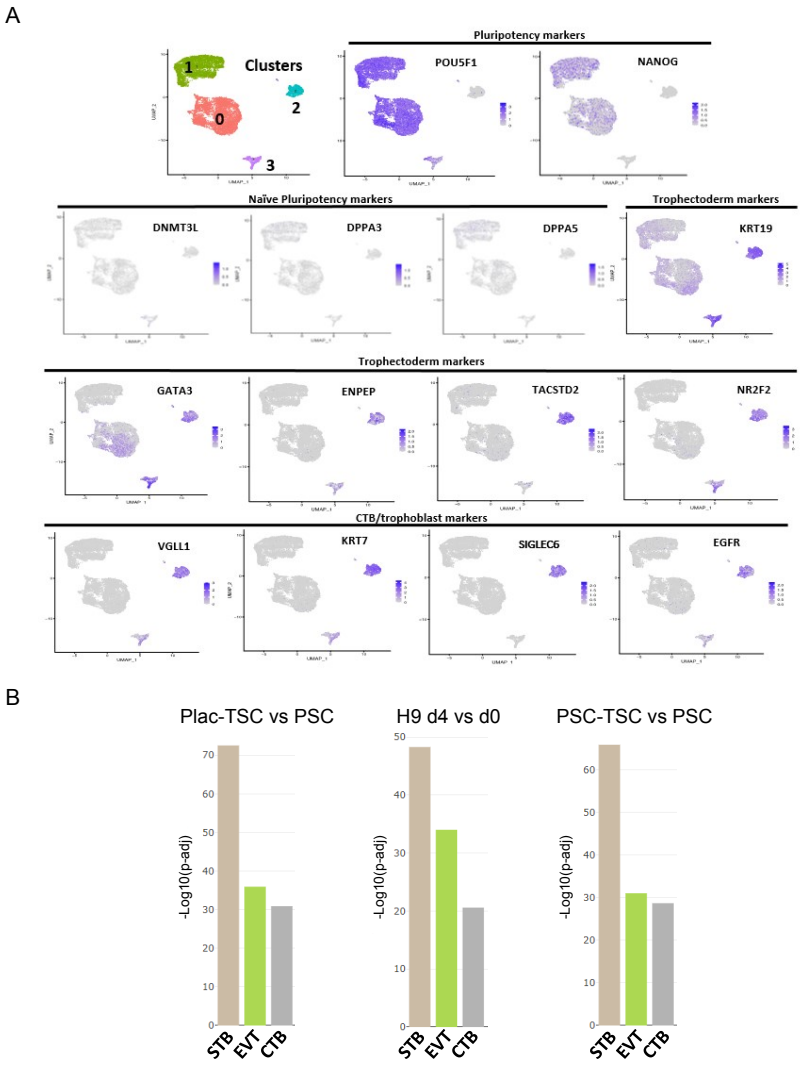


Figure 3.9: Comparison of trophectoderm and TSC derived from naïve and primed hPSC (part 2).
(A) Single cell RNA-seq analysis of H9-ESC at day 0, at 12-hours, 24-hours and 4 days post-BMP4/IWP2 treatment, as well as H9-TSC, separated into 4 clusters by Seurat. Feature plots of pluripotency-, TE-, and CTB-associated markers show induction of TE- (and not naïve pluripotency-) associated genes, prior to finally transitioning into TSC state, with expression of CTB-associated markers. (B) Analysis by PlacentaCellEnrich tool of transcriptomic signatures of BMP4/IWP2-treated hPSC, primary TSC, and hPSC-derived TSC, compared to undifferentiated hPSCs. Note similar enrichment proportions for syncytiotrophoblast (STB)-, extravillous trophoblast (EVT)-, and cytotrophoblast (CTB)-associated genes across all 3 comparisons.

3.4 Acknowledgements

Chapter 3 is an adapted reprint of material submitted for publication in Stem Cell Reports by Francesca Soncin*, Robert Morey*, Tony Bui, Daniela F. Requena, Virginia Chu Cheung, Sampada Kallol, Ryan Kittle, Omar Farah, Morgan Meads, Donald Pizzo, Mariko Horii, Kathleen M. Fisch, Mana M. Parast (*These authors contributed equally to this work). The dissertation author was one of the primary investigators and authors of this paper.

This work was supported by funds from the National Institutes of Health (NIH)/National Institute of Child Health and Human Development (NICHD) (R01HD-089537 to MMP; K99-R00-HD091452 to MH; and R01-HD096260 to FS), and the UC San Diego Stem Cell Program. RM and VCC were supported by a grant from the National Institutes of Health, USA (NIH grant T32GM8806). We thank Elsa Molina, PhD, Director of the UC San Diego Stem Cell Genomics and Microscopy Core for her expert assistance with the single-cell RNA-seq experiments. The manuscript also includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health SIG grant (S10 OD026929). The work was also partially supported by the National Institutes of Health, Grant 2UL1TR001442-06 of CTSA; the content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Finally, this publication used the Extreme Science and Engineering Discovery Environment (XSEDE) Comet for computational analysis, which is supported by National Science Foundation grant number ACI-1548562 (allocation ID: TG-MCB140074).

Chapter 4

Transcriptomic Drivers of Differentiation, Maturation, and Polyploidy in Human Extravillous Trophoblast

4.1 Introduction

The human placenta is essential for successful pregnancy and unique in its transitory nature. It performs a multitude of functions, including gas and nutrient exchange, synthesis of pregnancy-specific signaling molecules, and induction of maternal immunological tolerance. To realize the potential of creating an hPSC *in vitro* model that accurately represents human placentae, it is not only important to create trophoblast stem cells that could be differentiated into mature villous trophoblast cells but create *in vitro* trophoblast cells that were as close to primary trophoblast as possible. To accomplish this objective, it is therefore critical to characterize primary trophoblast and uncover the drivers of their differentiation and maturation in order to better model these cell types.

The placenta is also unique in its invasive nature. Early in development, the placenta

displays tumor-like properties as one of its component cell types, the extravillous trophoblast (EVT), aggressively invades the endomyometrium of the maternal uterus and remodels the spiral arteries [150]. EVTs arise from the proliferative epithelial stem cells of the placenta, the cytotrophoblast (CTB), and exit the cell cycle as they differentiate and invade. EVT share many of the molecular hallmarks of cancer cells [53], one of which is the frequent occurrence of structural genomic rearrangements and aneuploidy [163]. Trophoblast giant cells (TGCs), the mouse equivalent to EVTs, are known to be highly polyploid [17, 231], meaning they possess more than two sets of chromosomes, and undergo endoreduplication, a process by which cells undergo DNA replication in the absence of subsequent cell division [56]. Further studies have shown that TGCs harbor consistent regions of copy number variation (CNV) that may function as an important mode of genome regulation [67, 66]. Compared to rodents, there are only a small number of previous studies examining human trophoblast polyploidy or genomic CNVs [230, 229, 211, 126]. One such study in the human placenta showed an enrichment of CNVs, suggesting that, as in the mouse, the human placenta contains an atypical genome architecture that is important for the normal function of the organ [93]. Recently, however, another study focused on invasive EVT in first trimester human placenta, and reported that, unlike mouse TGCs, these cells did not contain CNVs but were predominantly tetraploid, and potentially undergo senescence and endoreduplication [197]. To date, similar analyses of term EVT has not been done.

Several groups, including ours, have characterized the transcriptome of human first trimester EVT using microarray-based profiling [3, 181, 185, 201]. These studies have shown that, compared to their CTB precursors, EVT downregulate pathways involving cell cycle, oxidative phosphorylation, p53, and fatty acid metabolism, while upregulating those involved in immune response, hypoxia- and hypoxia-inducible factor (HIF), mTOR signaling, and epithelial-mesenchymal transition (EMT), of which the latter has been extensively studied during EVT differentiation [44, 3, 181, 185, 201]. Nevertheless, there is a paucity of data, both on the gene

expression profile of term EVT, as well as on gene regulatory networks associated with EVT differentiation, maturation, and polyploidy.

In this chapter, we aim to extend the recent studies discussed above by performing comprehensive profiling of the genomic organization and transcriptome of first trimester and term EVT. To this end, we have used single-cell and bulk whole genome sequencing data, along with SNP genotyping, to investigate CNVs in first trimester and term EVT, compared to CTB and umbilical cord mesenchymal stem cells. We also used RNA sequencing to characterize the transcriptomes of both first trimester and term CTB and EVT, in order to identify pathways involved in EVT differentiation and maturation, as well as those that play a role in establishment of polyploidy in these cells. Additionally, we used hTSC derivation and differentiation [142] to evaluate development of polyploidy in *in vitro*-differentiated EVT, and to validate the unfolded protein response as a newly-identified pathway involved in EVT function. Finally, we also analyze our RNA-seq data to identify the TF networks involved in normal EVT formation and function.

4.2 Methods

4.2.1 Placenta samples, cell isolation, and EVT differentiation

Human placental tissues were collected under a UCSD Human Research Protections Program Committee Institutional Review Board-approved protocol; all patients provided informed consent for collection and use of these tissues. Cells were isolated from a total of 46 normal placentae, 27 first trimester and 19 term (**Table 4.1**). Gestational age (GA) is stated in weeks and days since the last menstrual period. For first trimester placentae, "normal" refers to elective termination of pregnancy in the absence of structural fetal abnormalities; for third trimester placentae (term), "normal" is defined by a non-hypertensive, non-diabetic singleton pregnancy, where the placenta is normally grown and shows no gross or histological abnormalities.

Table 4.1: Details of all samples used in study. CTB, cytotrophoblast; EVT, extravillous trophoblast; MSC, umbilical cord-derived mesenchymal stem cells; and TSC, trophoblast stem cells (derived from first trimester placentae).

Patient Number	Cell Type	Trimester	Gestational Age (wks^{day})	Fetal Sex	RNA-seq	scCNV	Ploidy Analysis by Flow Cytometry	FISH	Genotyping Array
716	CTB	1	11^{4}	Female	X				
788	CTB	1	12^{6}	Female	X				
793	CTB	1	10^{0}	Female	X				
804	CTB	1	10^{6}	Female	X				
805	CTB	1	11^{2}	Female	X				
822	CTB	1	13^{3}	Male	X				
825	CTB	1	10^{0}	Male	X				
830	CTB	1	14^{4}	Male	X				
831	CTB	1	10^{2}	Male	X				
832	CTB	1	13^{3}	Male	X				
2239	CTB	3	37^{0}	Male	X				
2256	CTB	3	39^{0}	Male	X				
2418	CTB	3	39^{0}	Female	X				
2430	CTB	3	39^{0}	Male	X				
2453	CTB	3	39^{0}	Male	X				
2504	CTB	3	39^{0}	Male	X				
2536	CTB	3	39^{0}	Female	X				
2538	CTB	3	39^{0}	Female	X				
2554	CTB	3	39^{0}	Female	X				
2567	CTB	3	39^{0}	Female	X				
2757	CTB	3	39^{2}	Male		X			
1026	EVT	1	10^{6}	Male	X				
1087	EVT	1	13^{5}	Male	X				
1089	EVT	1	12^{1}	Male	X				
1094	EVT	1	14^{6}	Male	X				
1096	EVT	1	11^{4}	Female	X				
1097	EVT	1	11^{2}	Female	X				
1099	EVT	1	14^{0}	Female	X				
951	EVT	1	11^{1}	Female	X				
970	EVT	1	10^{6}	Female	X				
977	EVT	1	11^{4}	Male	X				
2342	EVT	3	39^{2}	Male	X				
2656	EVT	3	39^{3}	Female	X				
2667	EVT	3	39^{2}	Female	X				
2700	EVT	3	39^{4}	Female	X				
2701	EVT	3	39^{0}	Male	X				
2706	EVT	3	40^{6}	Male	X				
2757	EVT	3	39^{2}	Male		X			
2757	UC-MSC	3	39^{2}	Male		X			
1386	EVT	1	8^{6}	Female					X
1386	CTB	1	8^{6}	Female					X
1391	EVT	1	10^{0}	Female					X
1391	CTB	1	10^{0}	Female					X (poor quality – dropped)
2754	EVT	3	39^{1}	Female			X	X	X
2754	CTB	3	39^{1}	Female			X	X	X
2754	UC-MSC	3	39^{1}	Female			X	X	X
2771	EVT	3	39^{1}	Male			X	X	X
2771	CTB	3	39^{1}	Male			X	X	X
2771	UC-MSC	3	39^{1}	Male			X	X	X
2834	EVT	3	37^{3}	Female			X	X	X
2834	CTB	3	37^{3}	Female			X	X	X
2834	UC-MSC	3	37^{3}	Female			X	X	X
1389	EVT	1	13^{6}	Unknown			X		
1389	CTB	1	13^{6}	Unknown			X		
1383	EVT	1	9^{0}	Unknown			X		
1383	CTB	1	9^{0}	Unknown			X		
1048	TSC	1	6^{2}	Female			Day 0 & Day 5/EVT-differentiated		
1049	TSC	1	6^{1}	Male			Day 0 & Day 5/EVT-differentiated		
1270-C	TSC	1	6^{0}	Male			Day 0 & Day 5/EVT-differentiated		

Isolation of cytotrophoblast (CTB) and extravillous trophoblast (EVT) from 20 term placentae (37- to 41-weeks GA) was performed as previously described in [112]. Briefly, the placentae were obtained immediately after C-section and placed on ice. Tissue from the basal plate (for EVT) and chorionic portion (for CTB) was dissected and minced, rinsed in 1X PBS

(Corning), and digested for 20 minutes in 1X Ca/Mg-free HBSS (Gibco), 1X trypsin (Gibco), collagenase, and DNase (Roche) three times, discarding the supernatant after each digestion. A Percoll® gradient (Sigma-Aldrich) centrifugation separation was then performed. Cells were then subjected to positive selection using magnetic activated cell sorting (MACS) (Miltenyi Biotec) and a PE-conjugated antibody against HLA-G (EXBIO MEM-G/9). The bound fraction was collected, tested for purity using flow cytometry. EVT preparations that contained greater than 90% HLA-G^+ cells were considered as adequate and used in downstream experiments. The unbound fraction was collected and CTB were selected using APC-conjugated antibody against EGFR (Biolegend 352906) and tested for purity using flow cytometry. CTB preparations yielding greater than 90% EGFR positivity were considered adequate and used in downstream experiments. Isolation of primary first trimester (9- to 14-week gestational age) trophoblast cells from 27 placenta samples was performed as previously described in [201, 168]. Briefly, chorionic villi were minced, washed in Hanks' balanced salt solution (Gibco) and digested three times with DNase I (Roche) and trypsin (Gibco). The cells were then pelleted and separated on a Percoll gradient (Sigma-Aldrich) and subjected to sequential MACS selection similar to the term placental samples.

Human Trophoblast Stem Cell (hTSC) lines were derived from early first trimester (6-7 week GA) placentae as previously described by [142]. Briefly, placental villi were minced, enzymatically digested, and then filtered. After Percoll® separation, the cells in the trophoblast fraction were MACS-purified with a PE-conjugated anti-ITGA6 antibody (Biolegend 313612; cell line 1049) or an APC-conjugated anti-EGFR antibody (Biolegend 352906; cell line 1048). The 1270C hTSC line was derived directly from the trophoblast fraction (after Percoll gradient) of the chorionic side of the placental tissue (manually separated from the basal side). Cells were then plated on collagen IV coated 6-well plates for at least 1 hour on TS media as described previously in [142]. Cells were first grown in modified basal media (advanced DMEM/F12, N2/B27 supplements, 2mM-glutamine, 10 μM 1-thioglycerol, 0.05% BSA, and 1% KSR). The media was

then changed to modified complete media (basal media with the addition of 2 μM CHIR99021, 500 nM A83-01, 1μM SB431542, 5 μM Y-27632, 0.8 mM valproic acid sodium, 100 ng/ml FGF2, 50 ng/ml EGF, 20 ng/ml Noggin, 50 ng/ml HGF) and grown to 80% confluency. Cells were passaged using TrypLE incubated for 15 minutes at 37°C. To characterize the trophoblast stem cell identity of these cells, their transcriptome was compared to that of cells derived by Okae et al. [142], as well as to primary CTB and EVT; our TSC were found to cluster with both the embryo- and placenta-derived Okae TSC (**Fig. 4.9**). EVT differentiation was performed by plating 0.75x 105 cells in a six-well plate precoated with 20 μg/ml fibronectin using the EVT differentiation media described in Okae et al., [142] (DMEM/F12 supplemented with 0.1 mM 2-mercaptoethanol, 0.3% BSA, 1% ITS-X supplement, 100 ng/ml NRG1, 7.5 μM A83-01, 2.5 μM Y27632, 2% Matrigel and 4% KnockOut Serum Replacement). On day 3, the medium was replaced with EVT medium lacking NRG1, and Matrigel was added to a final concentration of 0.5%. On day 5, the cells reached 80% confluency and were dissociated using TrypLE for 13 minutes at 37°C. The cells were assessed for differentiation efficiency by flow cytometric analysis using antibodies against HLA-G (EXBIO MEM-G/9) and EGFR (Biolegend 352906). For experiments evaluating the unfolded protein response pathway during EVT differentiation, the media was supplemented with 30 μM 4u8C (Sigma-Aldrich) or equivalent (vol/vol) amount of DMSO carrier.

Human umbilical cord (UC) was collected aseptically under a protocol approved by the Human Research Protections Program Committee of the UCSD Institutional Review Board (IRB number: 181917X). All patients provided informed consent for collection and use of these tissues, and all experiments were performed within guidelines and regulations set forth by the IRB. Umbilical cord mesenchymal stem cells (UC-MSCs) were derived from minced umbilical cord tissue per a published protocol [85, 221] Umbilical cord (UC) samples were collected and processed within 24 hours of delivery. Briefly, UC's were minced and washed to remove blood and then cultured in basal medium (aMEM with nucleosides (ThermoFisher), containing 10%

MSC-qualified FBS (OmegaScientific)). Cultures were maintained in a humidified atmosphere with 5% CO2 at 37°C. Approximately three weeks after plating, adherent fibroblast-like cells were detached using TrypLE Express for 5 minutes at 37°C (ThermoFisher) and filtered to remove any tissue fragments. The collected cells were then reseeded and maintained in growth medium containing b-FGF. After two weeks of growth with medium replacement every three days, the cells were checked for purity by flow cytometry analysis (BD FACS Canto 2 HTS). Cells were assessed for the expression of CD73 (FITC Mouse Anti-Human CD73 – BD 561254) and the absence of CD31 (APC-Cy™7 Mouse Anti-Human CD31 BD 563653) and CD45 (APC Mouse Anti-Human CD45 – BD 560973). UC-MSC samples displayed CD73 expression in over 90% of cells and lacked expression of CD31 and CD45. Sex of first trimester samples was determined based on PCR for SRY.

4.2.2 Whole Genome Sequencing Reanalysis

Whole genome sequencing fastq files from matched, isolated EGFR$^+$ and HLA-G$^+$ trophoblast from two patients (11-12 weeks gestational age) were downloaded from BioProject (Accession: PRJNA445189) [197]. The fastq files were trimmed (trim_galore v. 0.4.1) using a quality score cut-off of 30. The samples were then mapped to GRCh38 using Bowtie2 (v. 2.2.7) [105]. ERDS 1.1 was used to call CNVs on patient 1 using the erds_pipeline.pl script [228] with default parameters. CNV's found in both the EGFR$^+$ and HLA-G$^+$ sample were filtered out. Variants were called using GATK (v 4.0.11.0) on EGFR$^+$ and HLA-G$^+$ samples from both patients. Briefly, after merging replicate samples, duplicates were marked using Picard (v. 2.18.15), base quality scores were recalibrated, and variants were called using HaplotypeCaller. Joint genotyping followed by SNP and InDel recalibration was then performed according to GATK best practices. Low quality variants (GQ $<$ 20.0) were then removed and resulting vcf files were used to run PURPLE (PURity and PLoidy Estimator) [28]. To run PURPLE, Amber3 (v. 3.1) and Colbalt (v. 1.8) were first run in "reference/tumor" mode with the EGFR$^+$ sample

used as the reference sample and the matched HLA-G$^+$ sample used as the "tumor" sample.

4.2.3 Single cell RNA-seq reanalysis and InferCNV

Single cell data was downloaded from the European Genome-Phenome Archive hosted by the European Bioinformatics Institute (EBI; accession no. EGAS00001002449). Data from only the "normal" placenta samples (PN1, PN2, PN3C, PN3P, and PN4) (38 weeks gestational age) was used in the InferCNV (inferCNV of the Trinity CTAT Project) analysis [147]. PN2 was determined to be an outlier and was removed prior to cell cycle analysis. Prior to running InferCNV, data was analyzed using Scanpy (v. 1.4.3) [214]. Briefly, quality control was performed, and data were filtered, and batch corrected before dimensionality reduction and Louvain clustering. Cell cycle analysis was done using Scanpy's "score_genes_cell_cycle" command and gene sets determined previously [119]. Clusters were then annotated by ranking marker genes obtained by performing a modified t-test between each cluster and the remaining cells. Sub-clustering was performed on clusters that were not readily identifiable. Two EVT clusters were annotated and used downstream in the inferCNV algorithm as the "tumor" cells. All other annotated cells were considered part of the "normal" reference cells. Anscombe_transform normalization was used before running inferCNV to remove noisy variation at low counts, and the parameter HMM_type "i3" was used to perform inferCNV.

4.2.4 SNP genome-wide genotyping and CNV detection

DNA was isolated from placental cell pellets from two first trimester and three term placentae (Qiagen DNeasy Blood and Tissue kit) and quantified (Qubit dsDNA BR Assay Kits, Thermo Fisher Scientific) according to the manufacturer's protocol. DNA was genotyped using Illumina InfiniumOmni2-5-8v1-4 BeadChips (approx. 2,381,000 markers with a median spacing of 0.65 kb) at the IGM Genomics Center, at UC San Diego. Samples were called in GenomeStudio

(Illumina) with an average overall call rate of 99.4%. The CTB sample from one patient (1391) was removed from the analysis due to a low call rate (91.3%). CNVs were identified using the cnvPartition Plug-in (v. 3.2.0) in GenomeStudio. The cnvPartition confidence threshold was set at 100, with a minimum number of SNPs per CNV region of 10. The R (v. 3.6.1) package, allele-specific copy number analysis of tumors (ASCAT) (v. 2.5.2) [196], was used to estimate the ploidy of EVT samples. LogR ratios and BAF values were exported from GenomeStudio and no GC wave correction was performed. EVT samples were considered "tumor" samples and matched UC-MSCs were used as "reference" samples.

4.2.5 Single cell CNV detection

Matched EVT, CTB, and UC-MSCs from three term placentae were obtained as detailed above and following isolation were flash frozen. Cells were thawed and immediately resuspended into a single-cell suspension with 1X PBS and 0.04% BSA and filtered through a Flowmi cell strainer (Belart) before beginning the 10X Genomics single-cell DNA library prep. Briefly, between 100 to 500 cells in each sample were encapsulated in a hydrogel matrix, lysed, and then the genomic DNA (gDNA) was denatured and captured on a second microfluidic chip in Gel Beads containing unique cell indexes. After creation of amplified barcoded DNA fragments, sequencing libraries were created and sequenced on a NovaSeq at the IGM Genomics Center at UC San Diego. Cell Ranger (v. 1.1.0) (10X Genomics) DNA CNV pipeline was run to associate individual reads back to the individual cell. The reads were mapped to GRCh38 and downstream analysis was performed in LoupeBrowser. Each sample had on average over 1 million mapped and deduplicated reads per cell and a median estimated CNV resolution of less than 1 Mb.

4.2.6 Flow cytometric-based ploidy analysis

Matched EVT and CTB, and matched EVT, CTB, and UC-MSC were isolated from two first trimester and three term placentae, respectively; three hTSC lines were collected at day 0 and day 5 of EVT differentiation. Cells were washed first with PBS and then cold ethanol. Following the ethanol wash, the cells were allowed to rehydrate in PBS before pelleting and resuspending in 1 mL of a 3 μM DAPI in staining buffer (100 mM Tris, pH 7.4, 150 mM NaCl, 1mM CaCl2, 0.5 mM MgCl2, 0.1% Nonidet P-40) solution. Cells were incubated in the DAPI solution for 15 minutes at room temperature before filtering and running on a BD FACSCanto II Cell Analyzer.

4.2.7 FISH

Isolated term placental cell samples from three patients (CTB, EVT, and UC-MSCs from each patient) were sent to the Cytogenomics Laboratory at UCSD's Center for Advanced Laboratory Medicine. Fluorescence in situ hybridization (FISH) was performed using enumeration probes for chromosomes 2, 6, 18, and 20 (D2Z2, D6Z1, D18Z1, D20Z1: Abbott Molecular, Inc.). Each probe was examined in 200 interphase nuclei.

4.2.8 RNA isolation and RNA-seq library construction and analysis

RNA from ten first trimester CTB, ten term CTB, ten first trimester EVT, and six term EVT (split evenly between male and female), were isolated using the mirVana miRNA isolation kit (Ambion). RNA concentration was measured by Qubit RNA BR assay kit (ThermoFisher) and the quality of isolated RNA was checked using a bioanalyzer (Agilent). All samples were found to have a RIN above 7.5. RNA-seq libraries were prepared using the TruSeq Stranded mRNA sample preparation kit (Illumina) at the IGM Genomics Center at UC San Diego. Libraries were pooled and sequenced on NovaSeq 6000 S1Flow Cell (Illumina) to an average depth of 41 million uniquely mapped reads. Quality control was performed using FastQC (v.

0.11.8) and multiQC (v. 1.6). Reads were mapped to GRCh38.p10 (GENCODE release 26) using STAR (v. 2.7.3a) [47] and annotated using featureCounts (subread v.1.6.3, GENCODE release 26 primary assembly annotation) [113]. The STAR parameters used were: –runMode alignReads –outSAMmode Full –outSAMattributes Standard –genomeLoad LoadAndKeep –clip3pAdapterSeq AGATCGGAAGAGC –clip3pAdapterMMp 1 – outFilterScoreMinOverLread 0.3 –outFilterMatchNminOverLread 0.3. The featureCounts parameters were: -s 2 -p -t exon -T 13 -g gene_id. Ensembl genes without at least three samples with 5 or more reads were removed from analysis. Normalization and differential expression analysis were performed using the R (v. 3.6.3) package DESeq2 (v. 1.28.1) [117]. Sample sex was accounted for in the DESeq2 design and, unless otherwise stated, genes with an adjusted p-value < 0.05 and Log_2 fold change > 1 were considered differentially expressed. BiomaRt (v. 2.42.1) was used to convert Ensembl gene ID's to HUGO gene names, and gene set enrichment analysis was done with the R (v. 3.6.3) package FGSEA (v. 1.14.0) using 10,000 permutations and the hallmark (v. 7.0) pathways gene set, the GO term C5 (v. 7.0) gene set, and the transcription factor c3.tft (v. 7.2) gene set downloaded from MSigDB. Genes were ranked based on their Wald test statistic after performing differential expression. Additionally, where indicated, founder gene sets for the hallmark pathway gene sets were downloaded from MSigDB (v. 7.2). The cell senescence signature was downloaded from the Human Ageing Genomic Resources (https://genomics.senescence.info/download.html/cellage) [176]. Cell cycle related genes for each phase of the cell cycle were previously determined [119]. Characterization using principal component analysis (PCA) of the three hTSC lines derived for this study and two previously reported hTSCs [142], was done by merging the raw counts from six placental samples (three EVT and three CTB) and four hTSC samples (duplicates of blastocyst derived hTSCs and placental derived hTSCs) from Okae et al. [142], with the nine hTSC samples (triplicates of each hTSC line) and 36 placental samples from this study, filtered as specified above. The combined RNA-seq data was then processed and transformed using DESeq2's variance stabilizing transformation method before performing PCA. Gene list enrichment analysis

was done with Enrichr [101]. Visualization was performed with the R package ggplot2 or with the python packages seaborn, matplotlib, or plotly.

Gene regulatory networks were created by first performing GSEA using the transcription factor prediction gene set c3.tft (v. 7.2) from the Molecular Signatures Database. Transcription factors used as input into the gene regulatory network inference algorithm were selected based on adjusted p-value (< 0.05). Arboreto (v 0.1.5) [130] was run using the GRNBoost2 algorithm. The input into the algorithm consisted of the differentially expressed genes (adjusted p-value < 0.05 and Log_2 fold change > 1) from a given comparison and the significantly enriched transcription factors for the same comparison. For each target gene, the algorithm uses a tree-based regression model to predict its expression profile using the expression values of the set of input transcription factors. The algorithm outputs transcription factor targets with a calculated importance score. The top 1500 genes by Log_2 fold change were then used to create a protein-protein interaction network using the stringApp (v. 1.5.1, confidence (score) cutoff = 0.4, max additional interactors = 0, use smart delimiters) application in Cytoscape (v. 3.8.0). The networks were then clustered using MCL clustering with the clusterMaker2 application (v. 1.3.1, inflation value = 2.0, assumption that edges were undirected, and loops were adjusted before clustering). The importance scores from the genes in each cluster were then summed to find the transcription factors with the highest importance for each subcluster.

4.2.9 RNA isolation for qPCR of hTSC and EVT derivatives

RNA was isolated using NucleoSpin® (Macherey-Nagel, USA) kit and 300ng of RNA was reverse transcribed to prepare cDNA using PrimeScript™ RT reagent kit (TAKARA, USA) following the manufacturer's instructions. qRT-PCR was performed using Power SYBR® Green RT-PCR Reagents Kit (Applied Biosystems, USA). Data were normalized to beta-actin and shown as fold-change over day 0 (undifferentiated hTSC). Statistical analysis was performed using t-test. Data is expressed as mean ± SD of 2-ddCt values. The level of statistical significance was set at p

< 0.05.

4.2.10 Western Blot

hTSC 1049 cells were differentiated into EVT in 10cm dishes for 5 days. Cell lysate was collected every day for 5 days using RIPA buffer (Fisher Scientific, USA) containing protease and phosphatase inhibitors (Roche Applied Science, USA), according to the manufacturer's protocol. Protein concentration was quantified by BCA protein assay (Thermo Scientific, USA). 30μg of total protein was loaded onto a 10% denaturing polyacrylamide gel for separation and then transferred to PVDF membranes by electrophoresis. Membranes were blocked with 5% nonfat dried milk in Tris-buffered saline containing 0.1% (v/v) Tween 20 (Sigma-Aldrich) for 1 hour and then incubated overnight with primary antibodies: rabbit anti-STAT1 (Cell Signaling Technology or CST 9175) or mouse anti-ACTB (Sigma-Aldrich A5441). Followed by 1-hour incubation with HRP-conjugated secondary antibodies (donkey anti-rabbit IgG, CST 7074S, or anti-mouse IgG, CST 7076S), signals were developed using SuperSignal West Dura Extended Duration substrate (Thermo Fisher, USA) and captured on film.

4.2.11 Immunostaining and In-situ hybridization

First trimester placental tissues were fixed in 4% paraformaldehyde in phosphate-buffered saline for 10 minutes, then permeabilized with 0.5% Triton X-100 for 2 minutes. Tissues were stained with mouse anti-HLAG antibody (clone 4H84; Abcam) and rabbit anti-STAT1 (CST 9175), using Alexa Fluor-conjugated secondary antibodies (Thermo Fisher), and counterstained with DAPI (Invitrogen), then visualized using a Leica DM IRE2 inverted fluorescence microscope.

Term placenta samples were fixed in neutral-buffered formalin and embedded in paraffin wax. IHC and ISH was performed on 5-μm sections of these tissues on a Ventana Discovery Ultra automated stainer (Ventana Medical Systems) at the UC San Diego Advanced Tissue

Technology Core lab. For immunohistochemistry, standard antigen retrieval was performed for 40 min at 95°C as per the manufacturer's protocol (Ventana Medical Systems), and the section was stained using mouse anti-HLAG antibody (clone 4H84; Abcam). Staining was visualized using 3,3'- diaminobenzidine (DAB, Ventana Medical Systems) and slides were counterstained with Hematoxylin. For in-situ hybridization, we used the RNAscope method with probes specific to human GCM1 from ACD-Bio. Following amplification steps, the probes were visualized using DAB and slides counterstained with Hematoxylin. IHC and ISH slides were analyzed by conventional light microscopy on an Olympus BX43 microscope (Olympus).

4.3 Results

4.3.1 Term EVT lack recurrent copy number variations at specific genomic regions

To address the question whether human EVT are characterized by specific copy number variations (CNVs) or contain whole genome amplifications (polyploidy), we first reanalyzed whole genome sequencing (WGS) data from recently published $EGFR^+$ (CTB, n=2) and $HLA\text{-}G^+$ (EVT, n=2) trophoblast isolated from first trimester placentae (11- and 12-week GA) [197]. To identify CNVs, we applied the Estimation by Read Depth with Single-nucleotide variants (ERDS) algorithm [228], which was recently found to have high sensitivity and accuracy [188] and is an orthogonal method to those previously published [197]. Our reanalysis found fewer duplications in the $HLA\text{-}G^+$ samples compared to $EGFR^+$ samples; in addition, no duplications encompassed genes previously identified to be contained within amplified genomic regions of murine trophoblast giant cells [66]. Of the 35 genes found to have a duplication unique to the $HLA\text{-}G^+$ sample, three (TTC34, PKP1, and MBD5) were identified in similar previously-published data from second trimester whole human placental tissue [93] (data not shown). Read

depth CNV-detection algorithms use intra-chromosomal comparisons and therefore do not provide aneuploidy/polyploidy detection. Therefore, to determine the ploidy of the HLA-G$^+$ samples, we applied the PURPLE (PURity and PLoidy Estimator) algorithm to the WGS data [28, 156], which reported both HLA-G$^+$ samples to be diploid, with no evidence of significant duplications (**Fig. 4.1A**). We do note that the Patient 2 and Patient 1 samples were properly identified as female (2 copies of the X chromosome) and male (1 copy of the X chromosome), respectively.

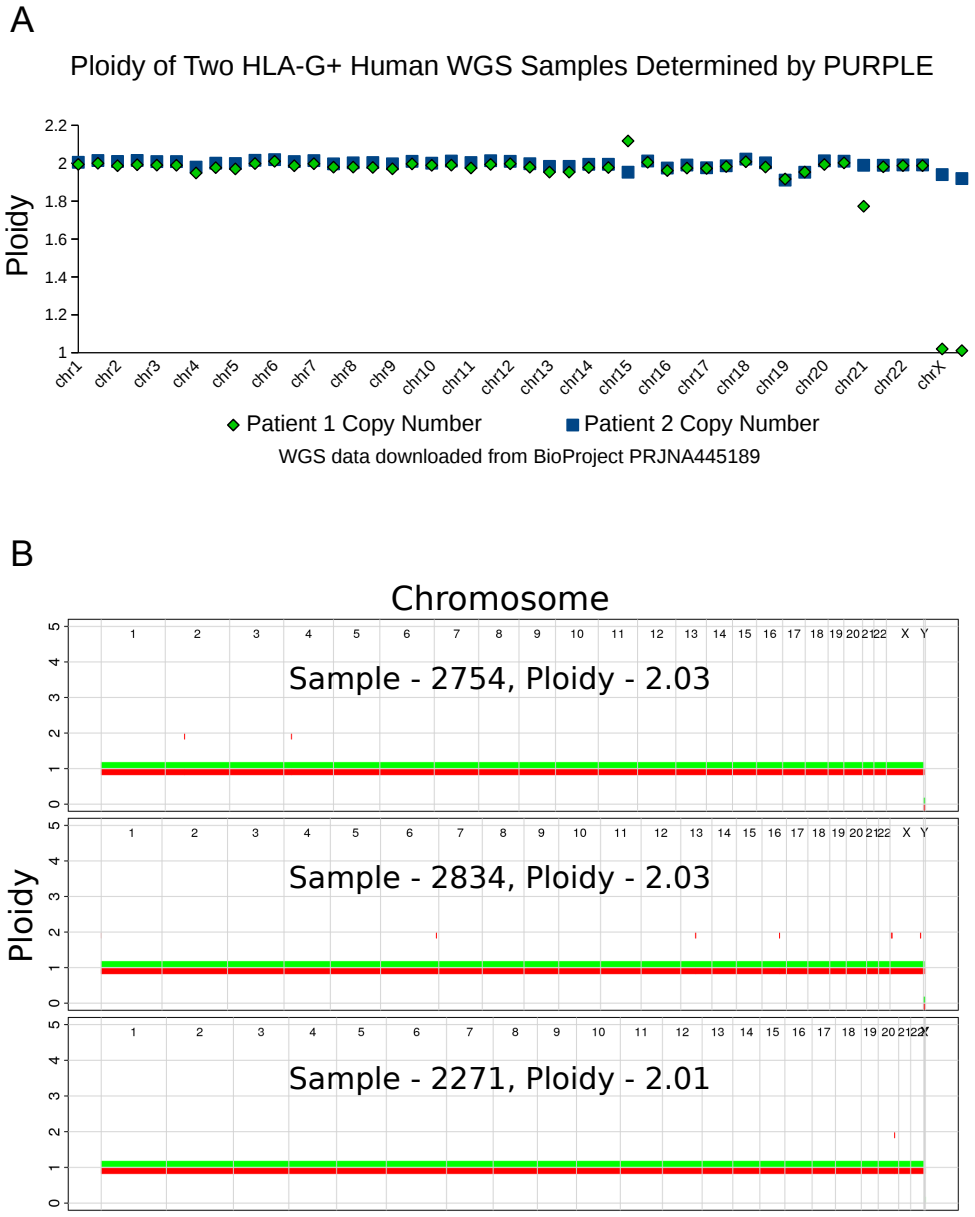


Figure 4.1: Ploidy determination of whole genome sequencing data and SNP genotyping array data.
(A) Ploidy of two HLAG^{+} samples (11- and 12-weeks gestational age) from BioProject accession PRJNA445189 [197] as determined by PURPLE reanalysis. (B) Ploidy of SNP genotyping array of our term EVT samples as determined by ASCAT.

Next, to find potential CNV genomic hotspots in term EVT, we isolated matched EVT, CTB, and UC-MSCs from three term placentae (**Table 4.1**) and performed CNV calling on data from genome-wide SNP genotyping arrays. For comparison, we performed a similar analysis with matched CTB and EVT from two first trimester placentae. After removing one of the first trimester CTB samples due to inadequate data quality, we found 9 CNVs unique to EVT samples (not found in either the matched CTB or the matched UC-MSC samples) (**Fig. 4.2A**). Of these 9 CNVs, 6 were found in the first trimester EVT sample that lacked a matched CTB sample (numbered in **Fig. 4.2A**), and 3 were duplications unique to the term EVT samples (green lines marked by * in (**Fig. 4.2A**), although not common between term EVT; none of the nine CNVs overlapped with previously-identified CNV regions [93]. We next sought to determine the ploidy of our EVT samples by running ASCAT (allele-specific copy number analysis of tumors) [196] on the data from genome-wide SNP genotyping arrays. We compared the EVT samples to their matched diploid UC-MSCs and again found no evidence of polyploidy in our term EVT samples (**Fig. 4.1B**).

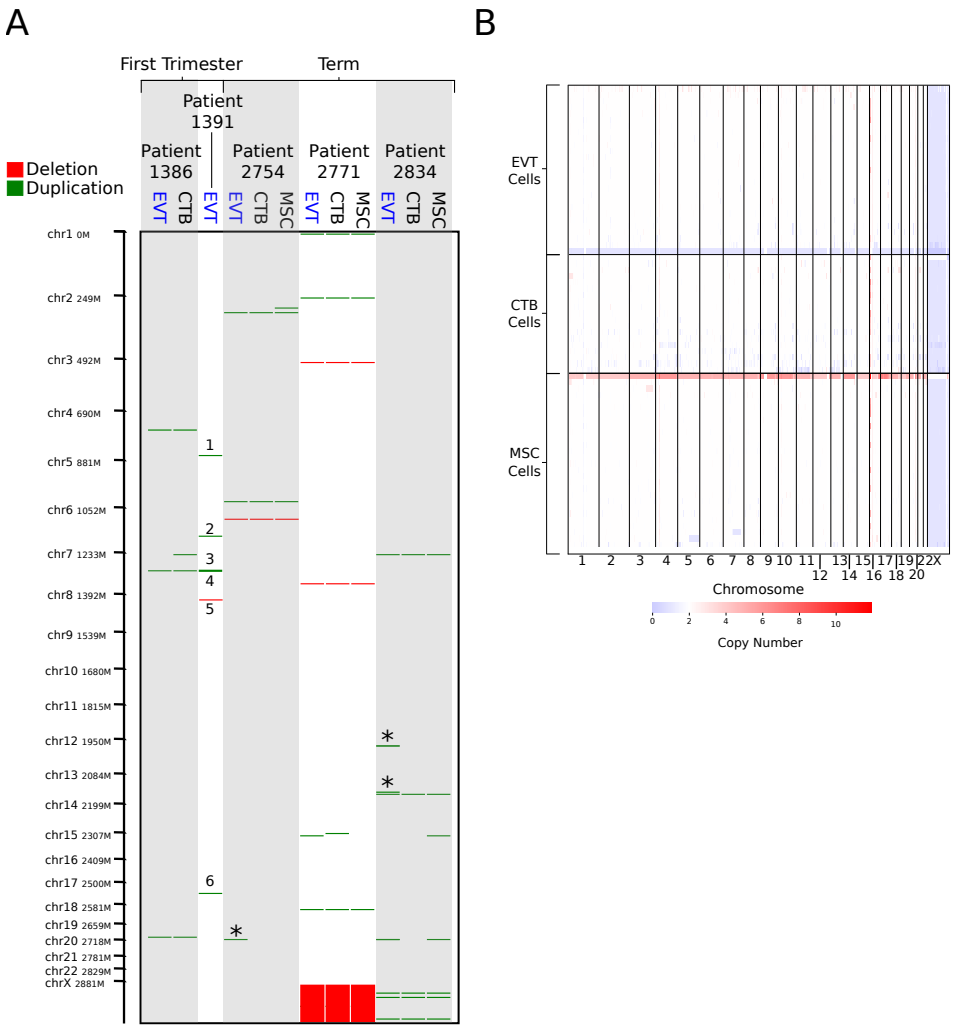


Figure 4.2: Copy number variation (CNV) analysis using genome-wide SNP genotyping array and single cell CNV data.
(A) CNV analysis on matched cytotrophoblast (CTB), extravillous trophoblast (EVT) and umbilical cord-derived mesenchymal stem cells (MSC) from three term placentae, and matched CTB and EVT from two first trimester placentae, using genome-wide SNP genotyping array (CTB sample from Patient 1391 was removed due to poor data quality). Deletions are shown in red and duplications in green. "*" indicate duplications that are unique to term EVT. Numbered CNVs are those unique to first trimester EVT. Patient 2771 is female, and all other patients are male. (B) CNV calls from single cell CNV analysis on matched CTB, EVT, and MSC from a single term placenta sample (Patient 2757 - male). Red on the heatmap represents increase in copy number and blue a loss in copy number. White represents a copy number of two.

To exclude the presence of a subpopulation of EVT cells showing polyploidy or a significantly higher load of CNVs, we performed single cell CNV analysis on over 600 matched EVT, CTB, and UC-MSC cells isolated from one term placenta. The estimated ploidy in each of the three samples was the same, at 1.95. Although some cells contained duplications, we were not able to verify the existence of a polyploid subpopulation (**Fig. 4.2B**). Finally, we interrogated a recently-published term placentae single-cell RNA-seq dataset [190] using InferCNV. After identifying EVT cells based on expression of HLA-G, we used InferCNV to compare their expression intensity to a set of reference "normal" cells in the same placental samples. We again found no evidence of polyploidy. We did find cells with numerous smaller CNVs in many of the EVT cells, but none that were common across multiple EVT (data not shown). We note that this experiment was performed on one placenta (2757), and that placenta was not the same as those used for other analyses.

These data suggest that, while EVT may display CNVs, high frequency CNVs common across EVT within and between different individual placentae might not exist. Moreover, based on reanalysis of existing trophoblast whole genome sequencing and single-cell RNA-seq data, as well as newly generated bulk SNP genotyping array and single-cell CNV data, we did not observe evidence of polyploidy in term EVT. However, as these techniques are not optimally designed to detect polyploidy, we proceeded with additional analyses to more directly assess this feature.

4.3.2 Flow cytometry and cytogenetic analysis confirm the presence of polyploid EVT at term

Given that the techniques used thus far were performed on bulk cell preparations (and thus might miss genetic alterations present in a subpopulation of component cells), and/or were not designed to detect polyploidy (e.g., single-cell CNV analysis), we next sought to evaluate term EVT using approaches that can reliably detect polyploidy on the single-cell level. First, to confirm the presence of a population of tetraploid first trimester EVT cells, as shown in Velicky et

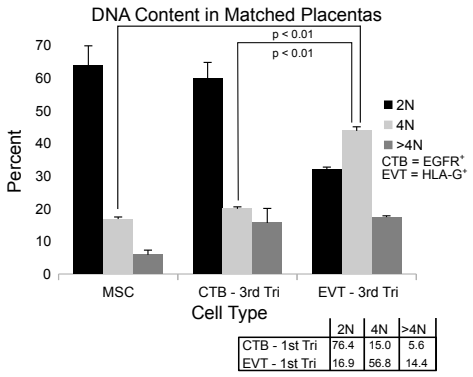
al. [197], we performed DNA ploidy analysis by flow cytometry on matched CTB and EVT from two first trimester placentae. As reported previously [230, 197], the CTB were predominantly diploid (76% of cells) compared to EVT, of which the majority (57%) were tetraploid (**Fig. 4.3A**). We also noted a small subpopulation of cells that contained a DNA content above 4N (5.6% in CTB and 14.4% in EVT). We next asked if isolated EVT from term placentae contain similar proportions of hyper-diploid cells. We found that, although, on average, a lower percentage of term EVT were tetraploid (44%), this was still a significantly larger proportion of tetraploidy compared to matched CTB and UC-MSC from the same placentae (p-value < 0.01) (**Fig. 4.3A and 4.3B**). Interestingly, compared to first trimester CTB and term MSC, term CTB showed an almost 3-fold increase in the proportion of >4N cells, and this proportion (15%) remained stable in matched term EVT (17%) (**Fig. 4.3A and 4.3B**). Additionally, we performed *in vitro* differentiation of three hTSC lines into HLA-G^+ EVT-like cells to assess how closely our *in vitro* differentiated EVT-like cells recapitulated the increased DNA content in primary EVT samples. We found no increase in the percentage of hyper-diploid cells following EVT differentiation of all three hTSC lines; in one cell line (1049), there was a decrease in the number of diploid cells (and thus an increase in the ratio of polyploid to diploid cells) following differentiation (**Fig. 4.3A**).

Figure 4.3: (*next page*): DNA content flow cytometry and cytogenetics analysis.
(A) Table showing results of DNA content flow cytometry analysis in umbilical cord-derived mesenchymal stem cells (MSC) from term placentae, cytotrophoblast (CTB) and extravillous trophoblast (EVT) from first trimester or term placentae CTB and EVT, and human trophoblast stem cells (hTSC) that are either undifferentiated (day 0) or differentiated (through a 5-day protocol) into EVT *in vitro*. (B) DNA content as determined by flow cytometry from matched CTB, EVT and MSC from three term placentae. Box in bottom right corner shows the mean percentage of cells in each ploidy group of matched CTB and EVT from two first trimester placentae as determined by flow cytometry. (C) Example images from FISH analysis of matched EVT and CTB from one term placenta. Images show probes targeting chromosome 2 (red) and chromosome 6 (green). (D) Bar graph showing the percentage of cells determined by FISH to be tetraploid (chromosomes 2, 6, 18, and 20) in MSC, CTB, and EVT from three term placentae.

A

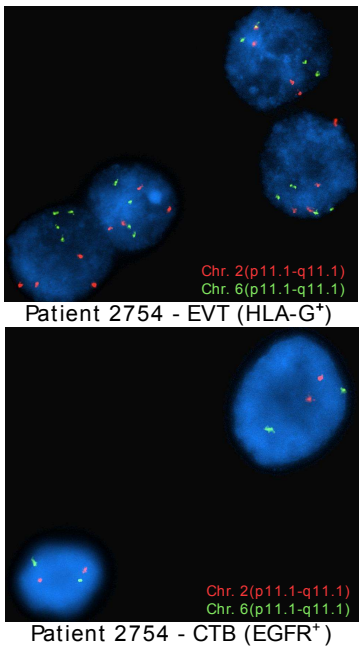
	MSC			CTB (EGFR$^+$)			EVT (HLA-G$^+$)		
Patient (First Trimester)	2N	4N	>4N	2N	4N	>4N	2N	4N	>4N
1389	-	-	-	77.2	13.7	5.2	15.3	65.9	14.3
1383	-	-	-	75.6	16.2	6.0	18.4	47.7	14.4
Average (1st Trimester)	-	-	-	76.4	15.0	5.6	16.9	56.8	14.4
Standard Error (1st Trimester)	-	-	-	0.8	1.3	0.4	1.6	9.1	0.05
Patient (Term)									
2754	54.5	18.1	8.7	50.5	21.1	24.0	29.8	43.9	18.1
2771	62.4	16.2	4.9	62.3	19.8	14.1	32.6	38.4	18.8
2834	75.0	16.2	4.0	67.0	19.0	8.8	33.5	49.3	15.4
Average (3rd Trimester)	64.0	16.8	5.9	59.9	20.0	15.6	32.0	43.9	17.4
Standard Error (3rd Trimester)	6.0	0.6	1.4	4.9	0.6	4.5	1.1	3.1	1.0
In Vitro Differentiated hTSC									
1270 day 0/CTB	-	-	-	34.6	29.7	21.4	-	-	-
1270 day 5/EVT	-	-	-	-	-	-	33.6	29.2	18.5
1049 day 0/CTB	-	-	-	45.1	25.7	17.6	-	-	-
1049 day 5/EVT	-	-	-	-	-	-	25.3	26.2	20.3
1048 day 0/CTB	-	-	-	27.7	28.3	27.7	-	-	-
1048 day 5/EVT	-	-	-	-	-	-	22.6	23.8	19.4
Average (Differentiated day 0)	-	-	-	35.8	27.9	22.2	-	-	-
Standard Error (Differentiated day 0)	-	-	-	5.1	1.2	3.0	-	-	-
Average (Differentiated day 5)	-	-	-	-	-	-	27.2	26.4	19.4
Standard Error (Differentiated day 5)	-	-	-	-	-	-	3.3	0.8	2.9

B

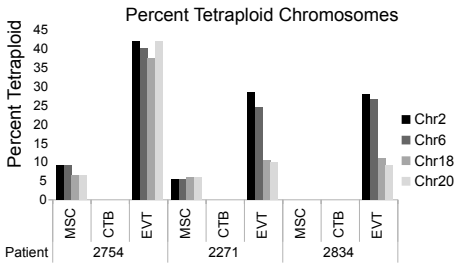


	2N	4N	>4N
CTB - 1st Tri	76.4	15.0	5.6
EVT - 1st Tri	16.9	56.8	14.4

C



D



Finally, using the same matched term placental cell isolates, we subjected CTB, EVT, and UC-MSCs to FISH analysis using enumeration probes for chromosomes 2, 6, 18, and 20. We again found that our EVT samples had a much higher percentage of tetraploid cells compared to their matched CTB and UC-MSC samples (**Fig. 4.3C and 4.3D**). We also noted that about 7% of EVT cells were called triploid but no triploid cells were found in any of the CTB or UC-MSC samples (data not shown). Taken together, these data suggest that similar to first trimester EVT, and in contrast to first trimester CTB and third trimester CTB and MSC, term EVT contain a large subpopulation of polyploid cells. Additionally, the similar proportion of >4N polyploid CTB and EVT at term suggests that this may be a shared feature among term trophoblasts.

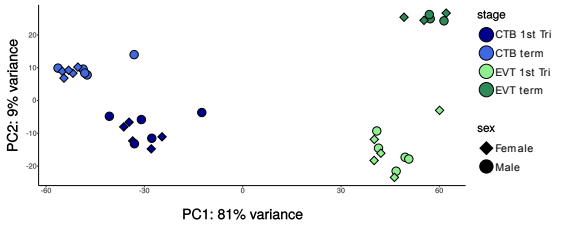
4.3.3 Global gene expression analysis reveals unique and common pathways involved in EVT differentiation and maturation

To further probe the differences between diploid CTB and polyploid EVT, we profiled the transcriptomes of both first trimester and term CTB and EVT. We isolated CTB and EVT from 10 first trimester placentae, CTB from 10 term placentae, and EVT from 6 term placentae (**Table 4.1**), with equal numbers of male and female placentae, and performed RNA-seq. Principal component analysis showed that samples clustered into the four expected groups (first trimester CTB, term CTB, first trimester EVT, and term EVT), based on cell-type along the first principal component, and on gestational age along the second principal component (**Fig. 4.4A**). There did not appear to be any obvious transcriptional differences based on sex. However, because we had sequenced a sufficient number of patient samples and had equal numbers of male and female placentae, we performed differential gene expression analysis between male and female EVT (adjusted p-value < 0.05, Log_2 fold change > 1, and normalized mean expression in group > 100) from both gestational ages. We found a small number of sex-specific differentially expressed genes (DEGs) in first trimester EVT, with just five genes upregulated in female EVT, including XIST and EPPK1, the latter a negative regulator of epithelial cell migration, and 11 genes upregulated

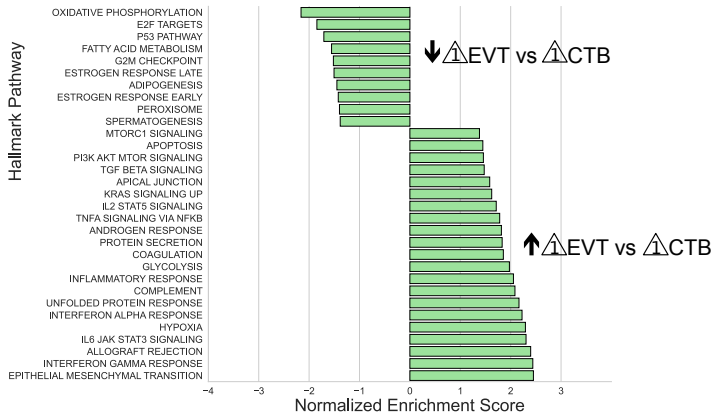
in male EVT, all of which were located on the Y chromosome except for PLXDC2 (**Fig. 4.5A**). Interestingly, term EVT showed a larger number of sex-specific DEGs, with 27 genes upregulated in the female samples and 24 in the male samples (**Fig. 4.5B**). Several of the genes on the Y chromosome were upregulated in both male term EVT and male first trimester EVT, but no significant gene ontology enrichment was identified in any groups.

Figure 4.4: (*next page*): Principal component analysis and Gene Set Enrichment Analysis (GSEA) of RNA-seq data.
(A) Principal component analysis showing the first two components using all genes post-filtering of all 36 placenta samples (see **Table 4.1**). Each sample group contained equal numbers of males and females. (B) GSEA using the Hallmark Pathway gene set of first trimester EVT compared to first trimester CTB. Genes were ranked based on their Wald test statistic after performing differential expression on first trimester EVT and first trimester CTB. Normalized enrichment scores (NES) indicate pathways either upregulated (NES > 0) or downregulated (NES < 0) in first trimester EVT vs. first trimester CTB. Only pathways with an adjusted p-value < 0.05 are shown. (C) GSEA using the Hallmark Pathway gene set of term EVT compared to term CTB. Genes were ranked based on their Wald test statistic after performing differential expression on term EVT and term CTB. Normalized enrichment scores (NES) indicate pathways either upregulated (NES > 0) or downregulated (NES < 0) in term EVT vs term CTB. Only pathways with an adjusted p-value < 0.05 are shown. Pathway names with an "*" are those that were also found to be significantly enriched in first trimester EVT vs first trimester CTB. (D) GSEA using the Hallmark Pathway gene set of term EVT compared to first trimester EVT. Genes were ranked based on their Wald test statistic after performing differential expression on term EVT and first trimester EVT. Normalized enrichment scores (NES) indicate pathways either upregulated (NES > 0) or downregulated (NES < 0) in term EVT vs first trimester EVT. Only pathways with an adjusted p-value < 0.05 are shown. Pathway names with an "*" are those that were also found to be significantly enriched in term EVT vs term CTB.

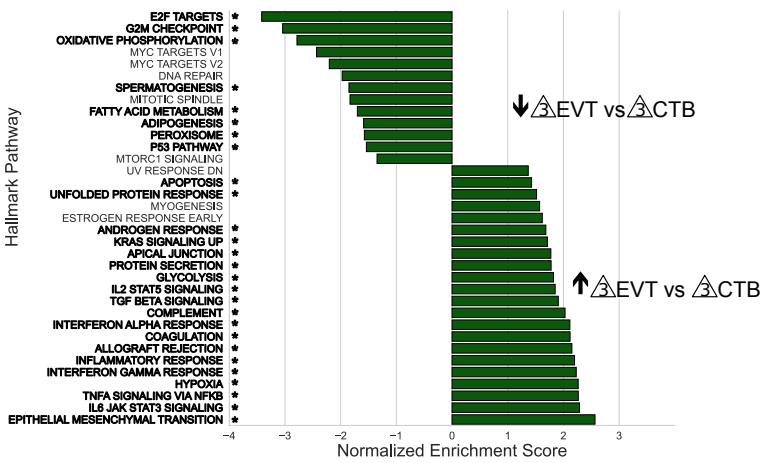
A


stage
CTB 1st Tri
CTB term
EVT 1st Tri
EVT term
sex
Female
Male
PC2: 9% variance
PC1: 81% variance

B


OXIDATIVE PHOSPHORYLATION
E2F TARGETS
P53 PATHWAY
FATTY ACID METABOLISM
G2M CHECKPOINT
ESTROGEN RESPONSE LATE
ADIPOGENESIS
ESTROGEN RESPONSE EARLY
PEROXISOME
SPERMATOGENESIS
MTORC1 SIGNALING
APOPTOSIS
PI3K AKT MTOR SIGNALING
TGF BETA SIGNALING
APICAL JUNCTION
KRAS SIGNALING UP
IL2 STAT5 SIGNALING
TNFA SIGNALING VIA NFKB
ANDROGEN RESPONSE
PROTEIN SECRETION
COAGULATION
GLYCOLYSIS
INFLAMMATORY RESPONSE
COMPLEMENT
UNFOLDED PROTEIN RESPONSE
INTERFERON ALPHA RESPONSE
HYPOXIA
IL6 JAK STAT3 SIGNALING
ALLOGRAFT REJECTION
INTERFERON GAMMA RESPONSE
EPITHELIAL MESENCHYMAL TRANSITION
Hallmark Pathway
↓1EVT vs 1CTB
↑1EVT vs 1CTB
Normalized Enrichment Score

C


E2F TARGETS *
G2M CHECKPOINT *
OXIDATIVE PHOSPHORYLATION *
MYC TARGETS V1
MYC TARGETS V2
DNA REPAIR
SPERMATOGENESIS *
MITOTIC SPINDLE
FATTY ACID METABOLISM *
ADIPOGENESIS *
PEROXISOME *
P53 PATHWAY *
MTORC1 SIGNALING
UV RESPONSE DN
APOPTOSIS *
UNFOLDED PROTEIN RESPONSE *
MYOGENESIS
ESTROGEN RESPONSE EARLY
ANDROGEN RESPONSE *
KRAS SIGNALING UP *
APICAL JUNCTION *
PROTEIN SECRETION *
GLYCOLYSIS *
IL2 STAT5 SIGNALING *
TGF BETA SIGNALING *
COMPLEMENT *
INTERFERON ALPHA RESPONSE *
COAGULATION *
ALLOGRAFT REJECTION *
INFLAMMATORY RESPONSE *
INTERFERON GAMMA RESPONSE *
HYPOXIA *
TNFA SIGNALING VIA NFKB *
IL6 JAK STAT3 SIGNALING *
EPITHELIAL MESENCHYMAL TRANSITION *
Hallmark Pathway
↓3EVT vs 3CTB
↑3EVT vs 3CTB
Normalized Enrichment Score

D

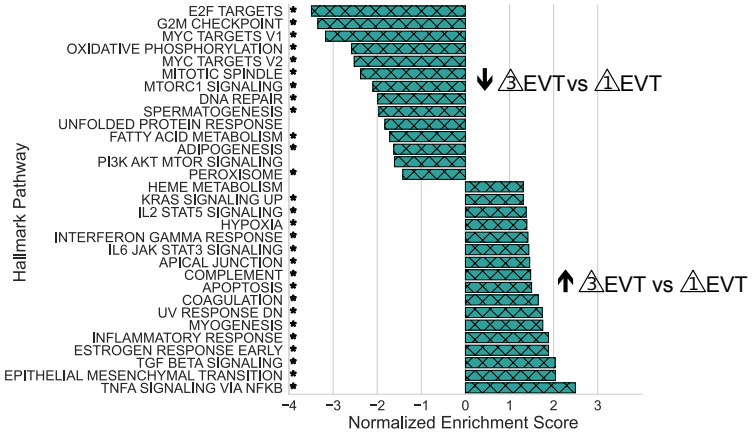
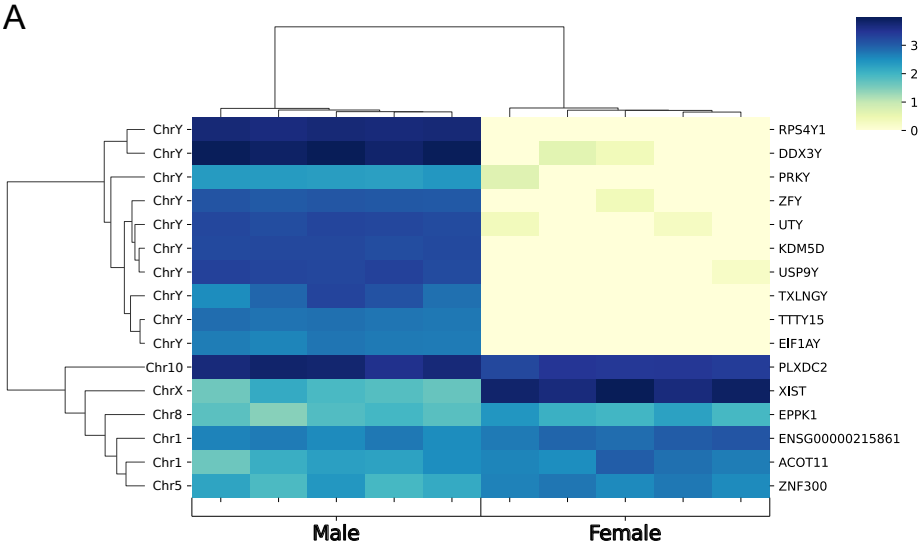

E2F TARGETS *
G2M CHECKPOINT *
MYC TARGETS V1 *
OXIDATIVE PHOSPHORYLATION *
MYC TARGETS V2 *
MITOTIC SPINDLE *
MTORC1 SIGNALING *
DNA REPAIR *
SPERMATOGENESIS *
UNFOLDED PROTEIN RESPONSE
FATTY ACID METABOLISM *
ADIPOGENESIS *
PI3K AKT MTOR SIGNALING
PEROXISOME *
HEME METABOLISM
KRAS SIGNALING UP *
IL2 STAT5 SIGNALING *
HYPOXIA *
INTERFERON GAMMA RESPONSE *
IL6 JAK STAT3 SIGNALING *
APICAL JUNCTION *
COMPLEMENT *
APOPTOSIS *
COAGULATION *
UV RESPONSE DN *
MYOGENESIS *
INFLAMMATORY RESPONSE *
ESTROGEN RESPONSE EARLY *
TGF BETA SIGNALING *
EPITHELIAL MESENCHYMAL TRANSITION *
TNFA SIGNALING VIA NFKB *
Hallmark Pathway
↓3EVT vs 1EVT
↑3EVT vs 1EVT
Normalized Enrichment Score

Figure 4.5: (*next page*): Differential gene expression between male and female samples in first trimester and term EVT.
(A) Heatmap of all differentially expressed genes between male and female samples in first trimester EVT. Blue- and pink- highlighted dendrogram rows are genes upregulated in male or female samples, respectively. (B) Heatmaps of differentially expressed genes between male and female samples in term EVT, with genes upregulated in the female samples shown in top heatmap and those upregulated in male samples in bottom heatmap. For both analyses, differentially expressed genes were determined by adj p-value < 0.05, Log_2 fold change > 1, mean normalized expression in group > 100; values were log transformed to create heatmap.

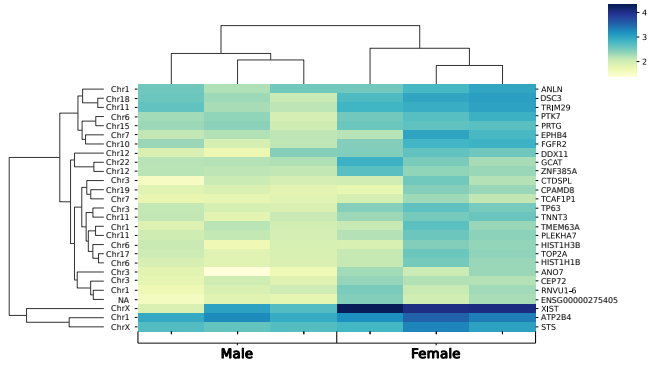
A



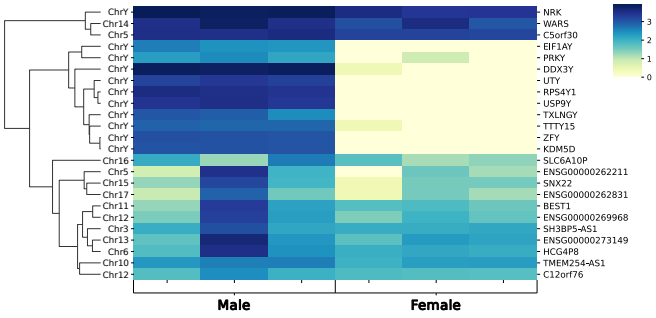
Upregulated in Male

Upregulated in Female

B



Upregulated in Female



Upregulated in Male

To identify differences among these four groups of cells (first trimester CTB and EVT, and term CTB and EVT), we first performed differential expression analysis, ranked genes based on their Wald test statistic, and conducted Gene Set Enrichment Analysis (GSEA). CTB are the proliferative epithelial cells of the placenta and differentiate early in pregnancy into EVT; therefore, we first sought to identify pathways which were significantly enriched between first trimester CTB and EVT (**Fig. 4.4B**). Similar to previously published microarray data [3, 181, 185], all seven of the pathways that make up the immune process category in the Hallmark gene set [114] were significantly upregulated in EVT vs CTB in the first trimester (adjusted p-value $<$ 0.05) (**Fig. 4.4B**). Furthermore, similar to what we previously found using gene expression microarrays [201], pathways such as hypoxia, unfolded protein response, and mTOR signaling were significantly upregulated, and pathways such as oxidative phosphorylation, P53, fatty acid metabolism, and those related to cell cycle control were significantly downregulated, in first trimester EVT (**Fig. 4.4B**).

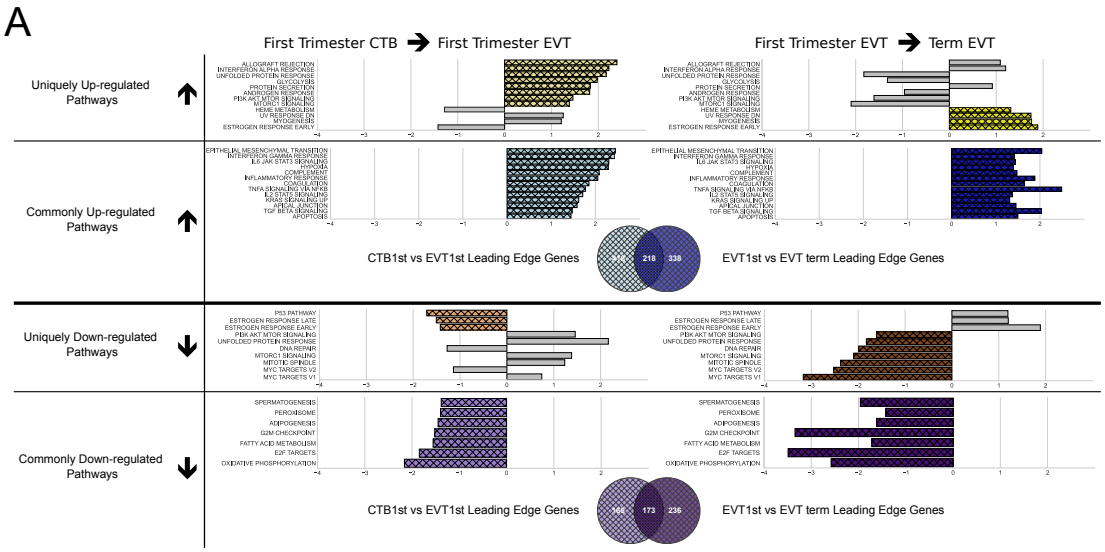
We next analyzed the pathways that were significantly enriched in term EVT compared to term CTB (**Fig. 4.4C**). Perhaps unexpectedly, nearly all the pathways upregulated in term EVT were also significantly upregulated in first trimester EVT (**Fig. 4.4C**, highlighted by *). Likewise, most pathways downregulated in term EVT were also similarly altered in the first trimester comparison (**Fig. 4.4C**, highlighted by *). However, in addition to the E2F targets, G2M checkpoint, and P53 pathways, term EVT also showed downregulation of the three remaining pathways in the proliferation process category (namely: MYC Targets V1, MYC Targets V2, and Mitotic Spindle) in the Hallmark gene set [114].

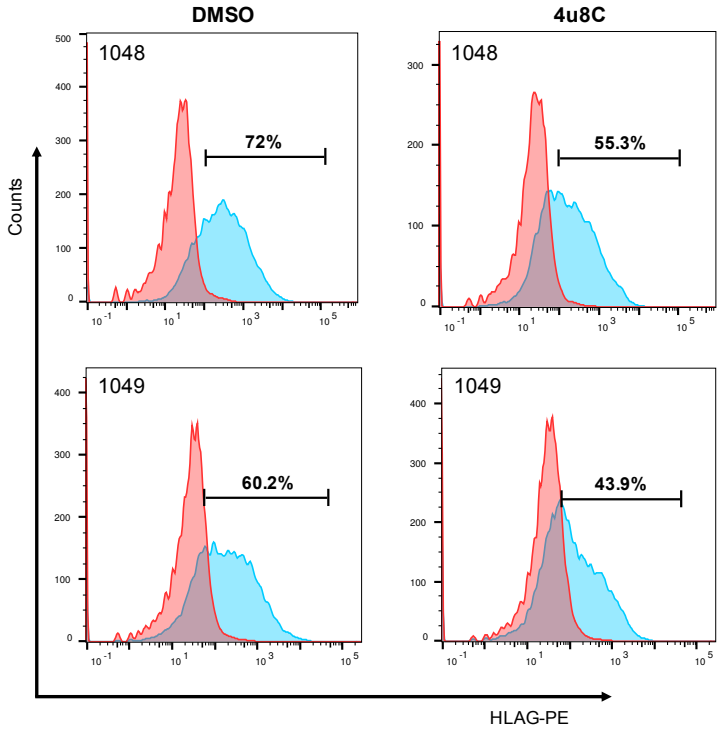
Next, we examined the EVT maturation process by comparing gene expression between first trimester and term EVT samples (**Fig. 4.4D**). We again saw many of the same pathways enriched in term EVT, compared to first trimester EVT, as in the comparison with term CTB (**Fig. 4.4D**, highlighted by *). We noted that compared to first trimester EVT, term EVT downregulated all of the proliferation process category pathways (E2F targets, G2M checkpoint, MYC targets,

mitotic spindle), except the P53 pathway gene set (**Fig. 4.4D**).

To better understand how specific pathways were regulated during EVT development, we next looked at the pathways that were unique or common through both steps of the EVT maturation process (first trimester CTB -> first trimester EVT -> term EVT) (**Fig. 4.6A**). Of the common downregulated pathways in EVT, the cell proliferation pathways E2F targets and G2M checkpoint had the lowest scores. We therefore repeated GSEA using just the founder gene sets [114] for these two pathways. We found that in both comparisons, term EVT showed significant downregulation for the neighborhood of CCNA2 (Cyclin A2), PCNA, and RRM2 in the GNF2 expression compendium (**Fig. 4.7A and Fig. 4.8**) (all comparisons with NES $>$ 3.34 and padj $<$ 0.006). These three genes are expressed just before the onset or during the S phase of the cell cycle, consistent with the absence of cells in the S phase in term EVT.

Figure 4.6: (*next page*): Pathways enriched in the EVT differentiation and maturation process. (A) Pathways enriched either in one (unique) or both (common) steps of EVT differentiation (first trimester CTB $\rightarrow$ first trimester EVT) and maturation (first trimester EVT $\rightarrow$ term EVT). Colored bars in grid show significantly enriched pathways (adjusted p-value < 0.05) that are either uniquely up or downregulated in first trimester EVT when compared to first trimester CTB (left side), or uniquely up or downregulated in term EVT when compared to first trimester EVT (right side). Likewise, common up and downregulated pathways are shown in both comparisons. The grey bars on each unique pathway bar graphs represent the enrichment score of the comparison on the opposite side of the graph to highlight the similarity or difference of the corresponding comparison. The grey bars may show enrichment in the same direction as the colored bars; however, the pathway is considered uniquely regulated because the grey bars represent enrichment that did not achieve statistical significance. (B) Validation of the IRE1-alpha arm of the Unfolded Protein Response (UPR) pathway regulating surface HLA-G expression in EVT. Two hTSC lines were differentiated to EVT in vitro in presence of either the IRE1-alpha inhibitor, 4u8c, or DMSO carrier alone. Graph shows the percentage of HLA-G^+ cells at the end of the 5-day treatment, with about 25% decrease in these cells in the presence of the inhibitor.


A
First Trimester CTB
First Trimester EVT
First Trimester EVT
Term EVT
Uniquely Up-regulated Pathways
Commonly Up-regulated Pathways
Uniquely Down-regulated Pathways
Commonly Down-regulated Pathways
CTB1st vs EVT1st Leading Edge Genes
EVT1st vs EVT term Leading Edge Genes
218
338
173
236
Normalized Enrichment Score


B
DMSO
4u8C
1048
1049
72%
55.3%
60.2%
43.9%
Counts
HLAG-PE

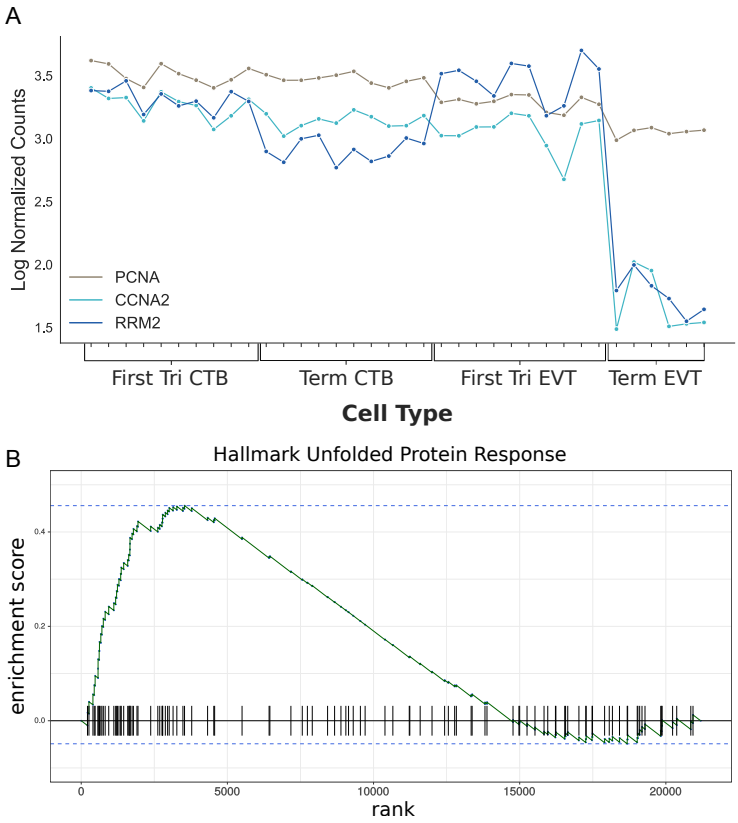


Figure 4.7: Common and unique pathways involved in EVT differentiation and maturation. (A) GSEA using only founder gene sets of the two common downregulated pathways during EVT maturation (first trimester EVT $\rightarrow$ term EVT) (E2F targets and G2M checkpoint pathways) showed downregulation in term EVT for the neighborhood of CCNA2 (Cyclin A2), PCNA, and RRM2 in the GNF2 expression compendium. The three genes shown in the figure are representative genes from each of these gene sets. (B) GSEA enrichment score plot showing Hallmark pathway Unfolded Protein Response (UPR) gene set in first trimester EVT compared to term EVT. The UPR pathway is also enriched in first trimester EVT compared to first trimester CTB.

Two of the Hallmark pathways that were found to be uniquely upregulated only in the first trimester EVT (compared to term EVT) were the PI3K/AKT/mTOR Signaling and Unfolded Protein Response (UPR) pathways (**Fig. 4.4B and Fig. 4.6A**). These two pathways appeared to switch directions during this two-step process: first upregulated during the initial conversion of first trimester CTB to EVT, and then downregulated in the subsequent maturation step from first trimester EVT to term EVT. The PI3K/AKT/mTOR pathway has been shown to be involved in the initial transition of CTB to EVT (reviewed in [154] and [53]), specifically by promoting

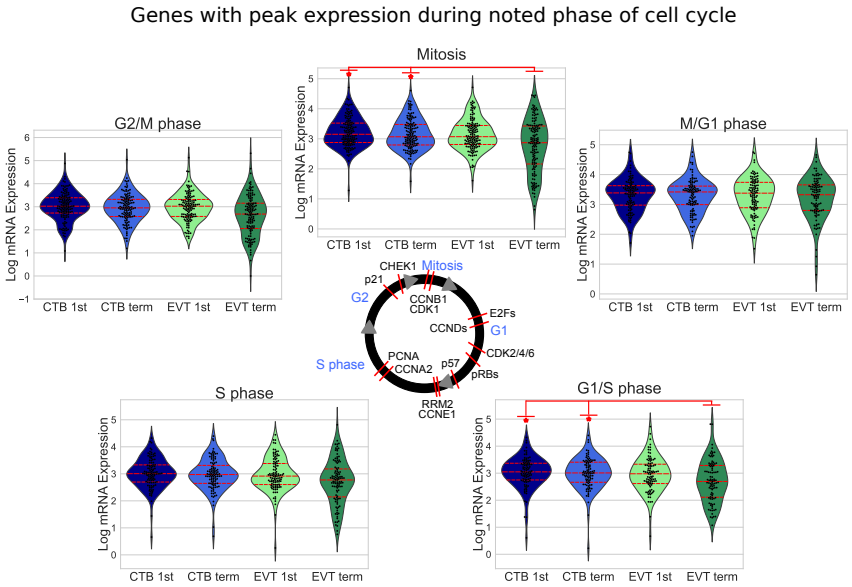


Figure 4.8: Cell cycle related gene expression in first trimester and term CTB and EVT. Expression of known cell cycle phase-related genes [119] are shown. The median value of each gene in each group was used to create swarm/violin plots. Inner dotted lines in swarm/violin plots represent the quartiles of the distribution. Red lines on cell cycle figure represent location during cell cycle with reported peak expression of the labeled gene (protein names have been used in certain cases to represent multiple genes or because they are more commonly used).

epithelial-to-mesenchymal transition. The other significantly upregulated pathway, Unfolded Protein Response (UPR, **Fig. 4.7B**), is mediated by ER stress, and is a method used by cells to detect, eliminate, and avoid further accumulation of misfolded proteins in the ER lumen, which build up due to several environmental cues, including hypoxia, a known EVT differentiation cue [201]. Moreover, UPR is a stress response phenotype, triggered by similar inducers to senescence, and recently found to be present in all types of senescence [153]. We have previously identified this pathway among those upregulated during the transition from CTB to proximal column EVT [201]; however, it has not been further validated in EVT differentiation and/or function. We therefore repeated GSEA using just the founder gene sets [114] for the UPR pathway and found that the Reactome activation of chaperone genes by XBP1S (adj p-value < 0.02) was significantly enriched. Furthermore, we found the GO term IRE1 mediated UPR to be significantly upregulated in first trimester EVT compared to first trimester CTB (adj p-value < 0.04). Therefore, to assess

the importance of the UPR pathway in EVT, we asked whether suppression of the IRE1-alpha arm, which is responsible for activating XBP1(S), would affect EVT formation *in vitro*. We derived two separate hTSC lines from early first trimester placentas; these lines appeared transcriptionally very similar to hTSC lines previously derived from early gestation placentas [142] (**Fig. 4.9**). We then applied the IRE1-alpha arm inhibitor 4u8c to both hTSC lines during differentiation into EVT and found a 25% decrease in the percentage of HLA-G$^+$ cells at the end of the protocol in two separate hTSC lines (**Fig. 4.6B**). However, qPCR did not show alteration of expression of any other EVT marker with 4u8C suppression (**Fig. 4.10A and 4.10B**). In addition, only total (and not spliced) XBP1 was increased with EVT differentiation (**Fig. 4.10C**). These results suggest that, at least the IRE1-alpha arm of the UPR pathway is needed for proper surface expression of HLA-G but is not required for EVT differentiation *per se*. Overall, these data provide, for the first time, a global look at pathways involved in EVT differentiation and maturation, identifying pathways that are uniquely and commonly up- or down-regulated in either step.

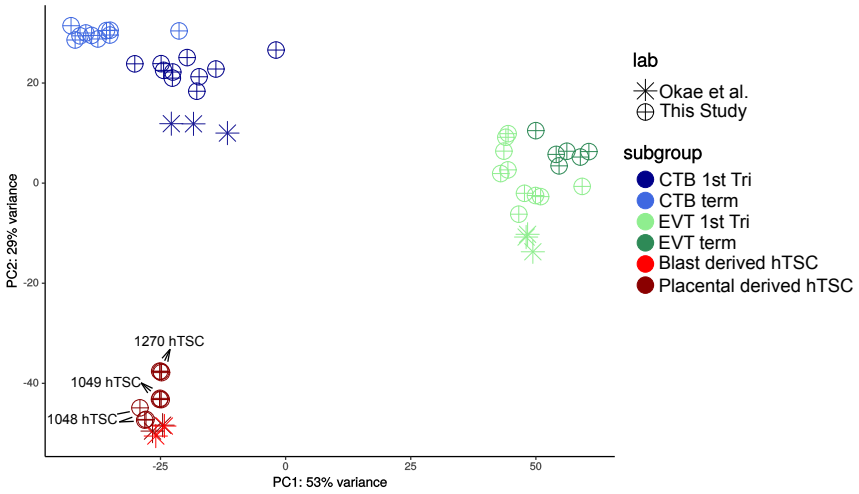


Figure 4.9: Principal component analysis of RNA-seq data comparing our three hTSC lines and those previously reported by Okae et al.
Principal component analysis shows the first two components using all common genes between the two datasets post-filtering. The plot shows all 36 placenta samples from this study and 6 from Okae et al. [142], triplicates of three hTSC lines (1048, 1049, and 1270) from this study, and duplicates of blastocyst derived hTSC and placental derived hTSC from Okae et al. [142] (see Table 4.1).

4.3.4 Transcriptome analysis suggests cell cycle arrest, cellular senescence, and endoreduplication as key features of EVT

The induction of cellular senescence leads to irreversible growth arrest and has been proposed as a ploidy-limiting mechanism [70, 91]. Our initial transcriptomic analysis using GSEA (**Fig. 4.4B-D**), along with the evaluation of expression of mitosis and cellular proliferation-associated genes (**Fig. 4.8**), suggested that term EVT are not cycling. We found that genes associated with all phases of the cell cycle exhibited decreased expression in term EVT, with the largest difference seen in genes associated with mitosis (**Fig. 4.8**). An active cell cycle is characterized by the expression of cyclins and cyclin-dependent kinases (CDKs). G1 phase cyclins and cyclin-dependent kinases were differentially expressed between EVT and CTB samples, with the lowest overall expression in term EVT, suggesting G1 cell cycle arrest (**Fig. 4.11A**). Additionally, two of the three retinoblastoma family genes, RB1 and RBL2, both shown to play pivotal roles in the negative control of the cell cycle by binding to E2F transcription factors and thus preventing S-phase entry [60], were significantly upregulated (adj p-value < 0.01) in EVT compared to CTB (**Fig. 4.11B**). Additionally, the genes encoding the "activating" E2Fs, which are known to interact with RB proteins to restrict cell cycle advancement [165] were significantly lower in term EVT compared to CTB (adj. p-value < 0.01) (**Fig. 4.11C**). Interestingly, mitosis-associated cyclin B (CCNB1) was highly expressed in first trimester EVT [197], but was roughly 100-fold lower in term EVT, suggesting absence of mitosis in term EVT (**Fig. 4.12A**). Moreover, expression of the mitosis-linked genes, CDK1, MKI67, and AURKA [229], was $>$10-fold lower in term EVT than the other three groups (**Fig. 4.12A**). Additionally, we performed cell cycle scoring on term single-cell RNA-seq data [190] and found that the EGFR$^+$ cluster contained a much higher fraction of cells in G1 and S phase compared to the HLA-G$^+$ cluster, which was contained predominantly in the G2/M phase, as previously reported [197] (**Fig. 4.12B**). Next, to investigate whether polyploid EVT displayed a senescence-like transcriptomic profile, we

performed principal component analysis using genes (n=1225) reported to comprise a human senescence transcriptomic signature [176]. We found that term EVT samples were uniquely clustered away from the other cell types, with first trimester EVT samples closer to both CTB groups than to the term EVT group (**Fig. 4.11D**). A similar clustering was not present in a PCA plot using a random set of 1225 genes (data not shown). Furthermore, the gene (GLB1) encoding the senescence-associated marker Beta-Galactosidase (SAβG) (adj. p-value < 0.001), along with several other senescence-associated secretory phenotype (SASP) and metabolic genes [18] were most highly expressed in either first trimester or term EVT (**Fig. 4.11E**). Taken together, these results suggest that EVT are undergoing cell cycle arrest and senescence.

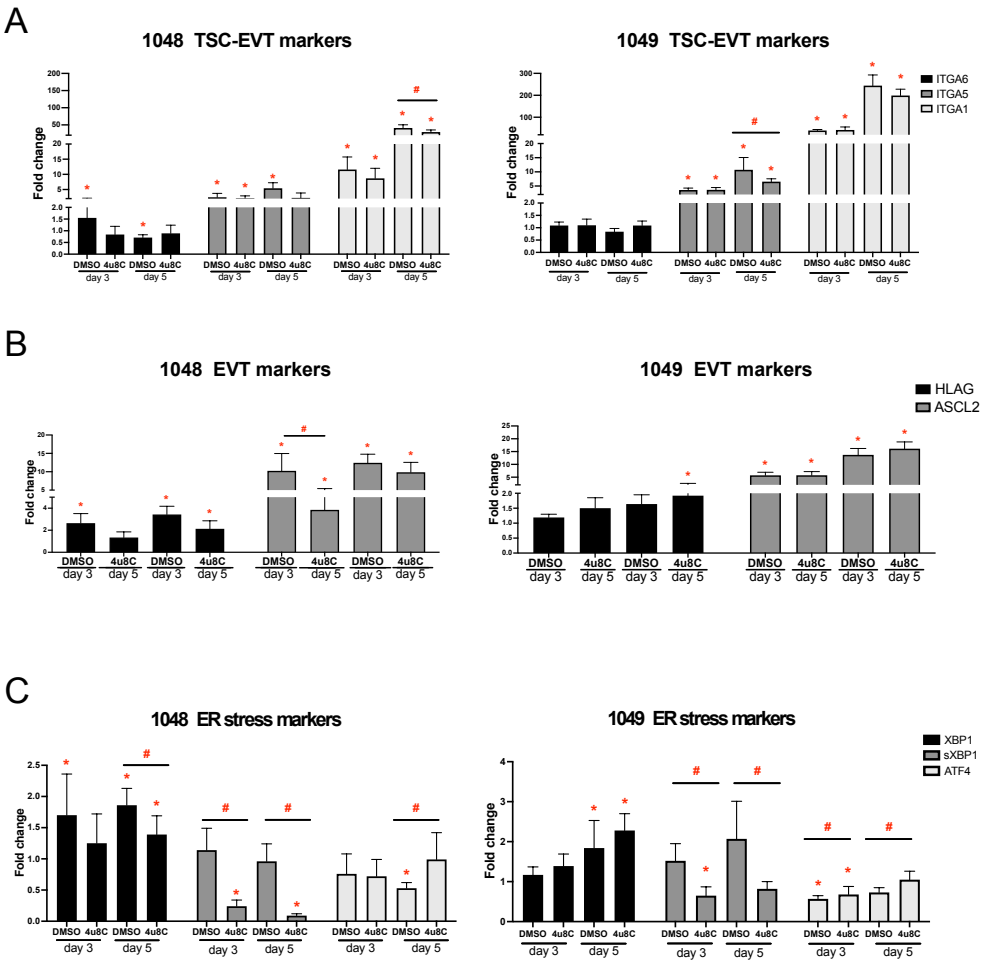


Figure 4.10: The role of IRE1-alpha arm of the Unfolded Protein Response (UPR) pathway in EVT differentiation of hTSCs.
Two different hTSC lines (1048 and 1049) were differentiated into EVT over 5 days in the presence or absence of the IRE1-alpha inhibitor, 4u8C. (A) qPCR for markers of CTB (ITGA6) or EVT (ITGA5 and ITGA1). (B) qPCR for markers of EVT (HLAG and ASCL2). (C) qPCR for UPR pathway genes, XBP1 (total and spliced) and ATF4. The decrease in spliced XBP1 following 4u8C treatment confirms inhibition of the IRE1-alpha arm of UPR. ddCT values were normalized to beta-actin and shown as fold change over day 0. "*" shows statistically significant difference from day 0, while # shows statistically significant difference from DMSO carrier alone treatment on the same day, based on t-test ($p<0.05$).

Figure 4.11: (*next page*): Cell cycle and endoreduplication gene expression.
(A) Heatmap of G1 phase cyclins and cyclin-dependent kinases using log transformed normalized gene counts. (B) Bar graph showing retinoblastoma family genes' (RB1, RBL1, and RBL2) normalized gene expression. (C) Bar graph displaying E2F activators' (E2F1, E2F2, and E2F3) normalized gene expression. (D) Three-dimensional principal component analysis using genes (n = 1225) associated with human cellular senescence signature. (E) Heatmap of previously identified [18] senescence-associated secretory phenotype (SASP)-associated genes) using log transformed normalized gene counts. (F) Normalized expression of p57 (CDKN1C) and CDK1 (p-value: *, < 0.01 by t-test). (G) Normalized expression of p21(CDKN1A) and CHEK1 (p-value: *, < 0.05 by t-test).

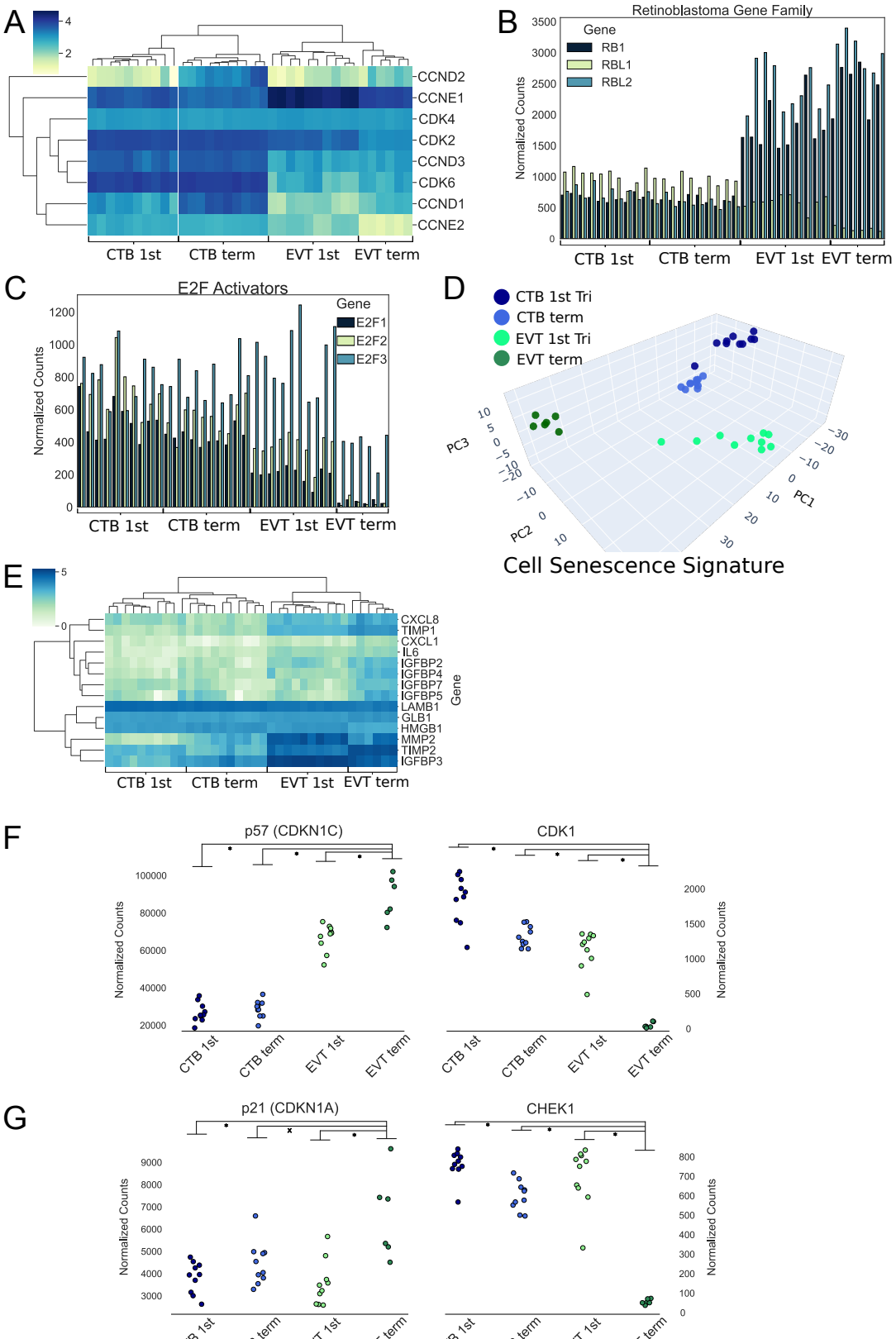
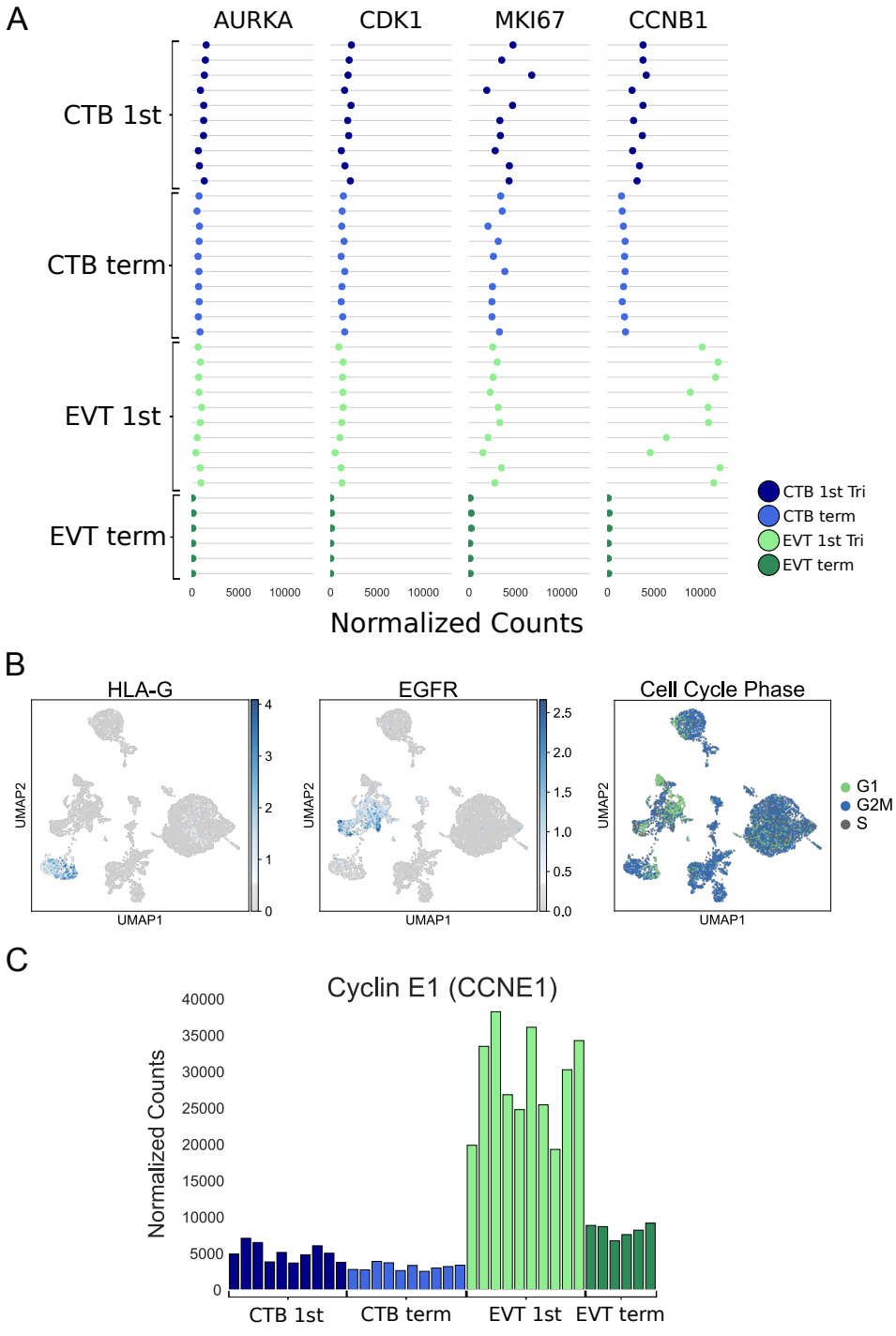

A
CCND2
CCNE1
CDK4
CDK2
CCND3
CDK6
CCND1
CCNE2
CTB 1st
CTB term
EVT 1st
EVT term
B
Retinoblastoma Gene Family
Gene
RB1
RBL1
RBL2
Normalized Counts
C
E2F Activators
Gene
E2F1
E2F2
E2F3
Normalized Counts
D
CTB 1st Tri
CTB term
EVT 1st Tri
EVT term
PC1
PC2
PC3
Cell Senescence Signature
E
CXCL8
TIMP1
CXCL1
IL6
IGFBP2
IGFBP4
IGFBP7
IGFBP5
LAMB1
GLB1
HMGB1
MMP2
TIMP2
IGFBP3
Gene
F
p57 (CDKN1C)
CDK1
Normalized Counts
G
p21 (CDKN1A)
CHEK1
Normalized Counts

Figure 4.12: (*next page*): Expression of cell cycle-associated genes in first trimester and term CTB and EVT.
(A) Normalized gene expression in all samples for AURKA, CDK1, MKI67, and CCNB1. (B) Cell cycle scoring based on previously-identified cell cycle phase-specific gene expression [119] on term single-cell RNA-seq data [190] visualized in a UMAP representation. HLAG and EGFR expression are shown to identify EVT and CTB cell clusters, respectively. (C) Cell cycle and endoreduplication-associated gene expression (normalized counts) of Cyclin E1 (CCNE1).


A
AURKA
CDK1
MKI67
CCNB1
CTB 1st
CTB term
EVT 1st
EVT term
0
5000
10000
Normalized Counts
CTB 1st Tri
CTB term
EVT 1st Tri
EVT term
B
HLA-G
EGFR
Cell Cycle Phase
UMAP1
UMAP2
G1
G2M
S
C
Cyclin E1 (CCNE1)
Normalized Counts
40000
35000
30000
25000
20000
15000
10000
5000
0
CTB 1st
CTB term
EVT 1st
EVT term

A recent study has suggested that first trimester EVT induce endocycles and enter a senescent state [197]. Endoreduplication consists of DNA replication without cell or nuclear division. It is thought to be triggered by inhibition of CDK1 by p57 (CDKN1C), and suppression of checkpoint protein kinase (CHEK1) by p21 (CDKN1A), preventing induction of apoptosis [192]. In our data, we noted the strongest reciprocal expression of p57 and CDK1, as well as p21 and CHEK1, in term EVT (p-value < 0.05) (**Fig. 4.11F and 4.11G**). Endoreduplication is also characterized by downregulation of CDK1, Cyclin A, and Cyclin B, with simultaneous persistence of Cyclin E expression [193]. In our data, CDK1 was uniquely decreased in term EVT (**Fig. 4.11F**); however, both Cyclin B (CCNB1) and Cyclin E1 (CCNE1) were highly expressed in first trimester EVT, with CCNB1 expression plummeting and CCNE1 persisting, albeit at a lower level, in term EVT (**Fig. 4.12A and 4.12C**). Although Cyclin A (CCNA1) had a similar expression profile to Cyclin E1, it was expressed at an extremely low level throughout (data not shown). This pattern of gene expression is most consistent with endoreduplication occurring in some first trimester EVT but becoming more ubiquitous/pronounced in term EVT.

4.3.5 Transcription factor drivers characteristic of EVT

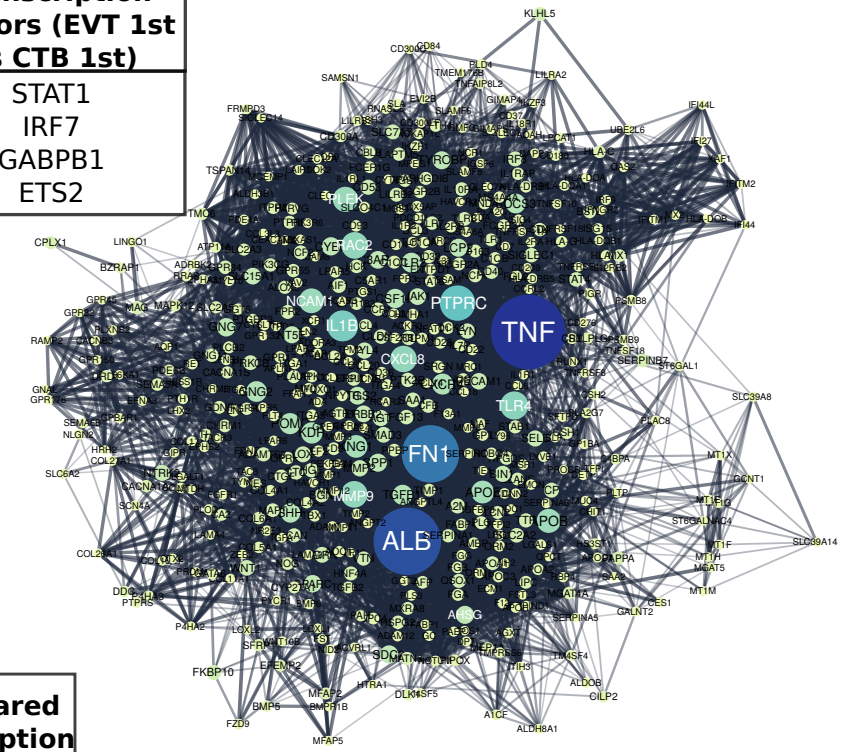
To better understand the transcription factor (TF) regulatory drivers of first trimester and term EVT, we performed GSEA using the TF prediction gene sets from the Molecular Signatures Database. To determine which enriched TFs were critical in each set of differentially expressed genes, and to infer gene regulatory networks, we used a tree-based regression model to calculate an "importance score" for each TF gene target pair using GRNBoost2 in the Arboreto software library [130]. Next, we created STRING networks for each of the top 1500 differentially expressed genes in each of the different cell type comparisons and clustered the networks into subnetworks. We were then able to use the TF gene target importance scores to infer which TFs were critical to each of the subnetworks. In this process, more than one subnetwork may be assigned to a given TF. We first asked which TFs had the highest importance scores when

assessing the differentiation of CTB to EVT, initially focusing on the paired first trimester cells and evaluating genes upregulated in EVT over CTB. Following network clustering, these genes clustered into several subnetworks, with the largest containing close to 1000 genes with TNF, FN1, and ALB as the genes with the highest centrality scores (**Fig. 4.13, top**), and the top four TFs being STAT1, IRF7, GABPB1, and ETS2 (**Fig. 4.13, top**).

Next, we evaluated the TF network upregulated in EVT compared to CTB at term. Following clustering, we compared the largest subnetwork in this comparison to the largest subnetwork upregulated in first trimester EVT (compared to first trimester CTB) and found that about 25% of the genes were the same. The top four TFs unique to the largest term EVT subnetwork were MEF2A, RREB1, NFATC3, and FOXO4 (**Fig. 4.13, bottom**). Additionally, FN1, the highest expressed gene in our EVT samples, and TNF, appeared to again have the highest centrality scores in this subnetwork (**Fig. 4.13, bottom**). The top ranked TFs in terms of importance scores shared in these two largest subnetworks in both first trimester and term EVT were ZNF436, IRF2, IRF8, and TFEB (**Fig. 4.13, center**).

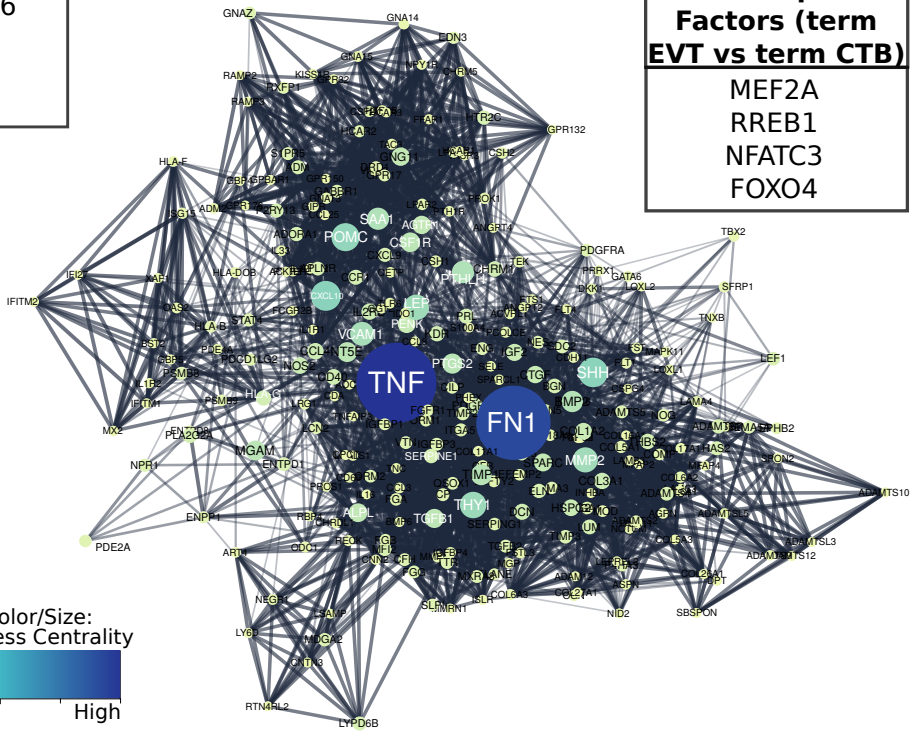
Figure 4.13: (*next page*): Comparison of gene regulatory networks involved in development of first trimester and term EVT.
Networks were created by first generating protein to protein interaction STRING networks of upregulated genes, either in first trimester EVT compared to first trimester CTB (top), or in term EVT compared to term CTB (bottom), then clustering networks into subnetworks. The largest subnetwork from each comparative analysis is shown. Displayed "unique" transcription factors (TFs) (top left and bottom right tables) were those found to have the highest summed importance scores in terms of the labeled subnetwork. Displayed "top shared" TFs (middle table) were the top four transcription factors in terms of importance for both subnetworks. Subnetwork node sizes and colors are determined by the nodes calculated betweenness centrality scores.

Unique Transcription Factors (EVT 1st vs CTB 1st)
STAT1
IRF7
GABPB1
ETS2



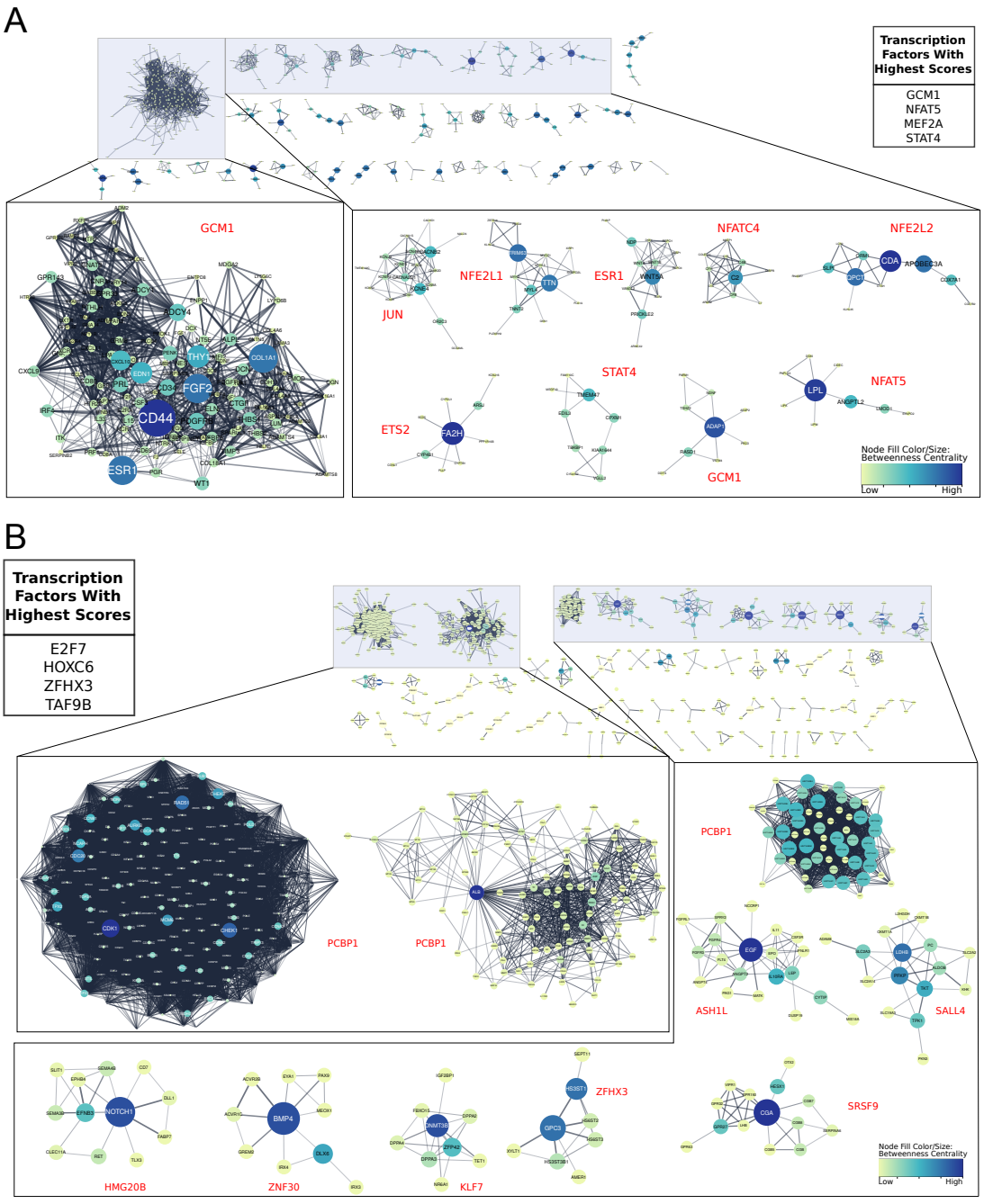
Top Shared Transcription Factors
ZNF436
IRF2
IRF8
TFEB

Unique Transcription Factors (term EVT vs term CTB)
MEF2A
RREB1
NFATC3
FOXO4



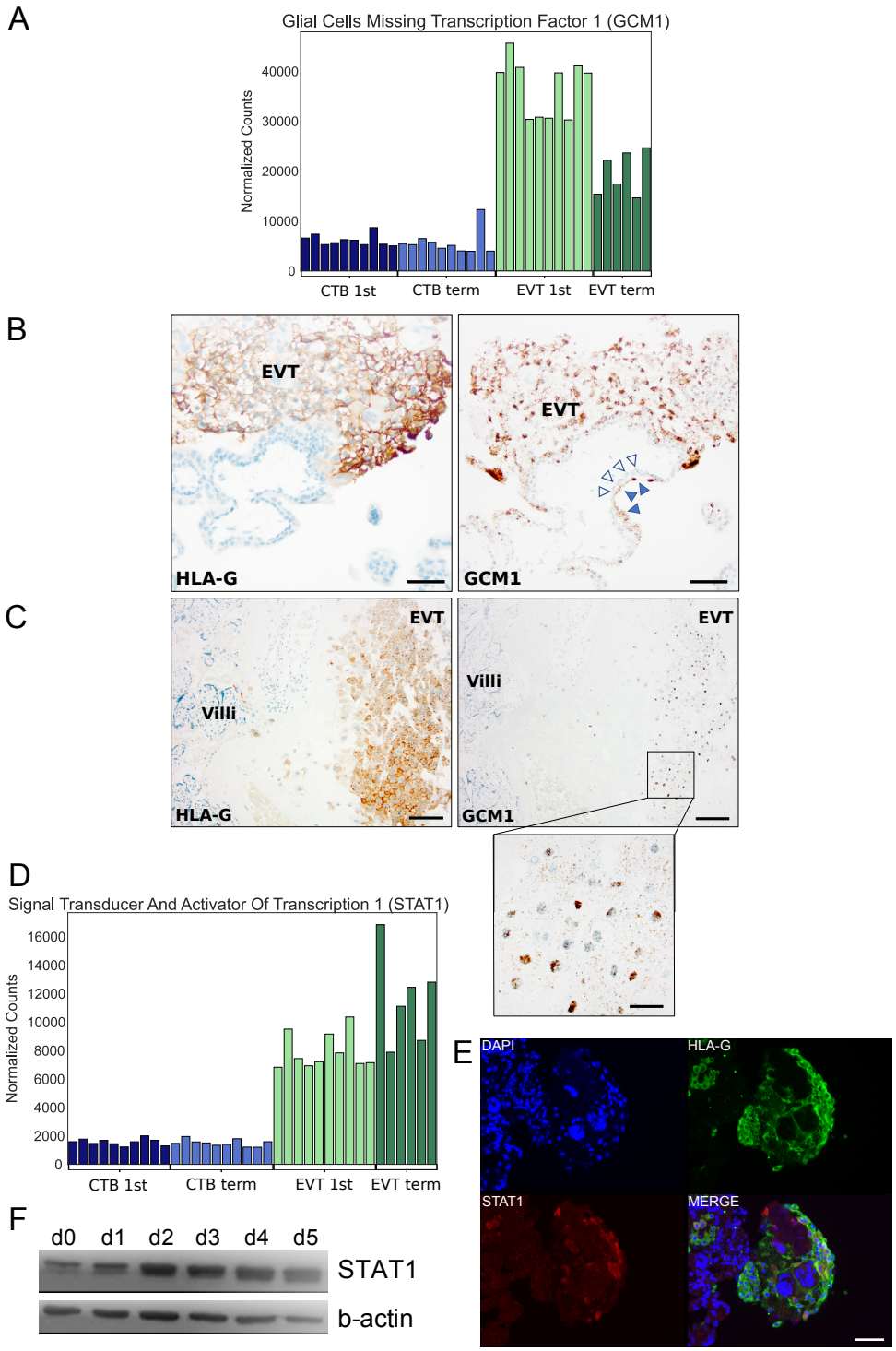
We next asked which TFs were important for EVT maturation. Using the genes that were upregulated in term EVT, compared to first trimester EVT, we found that the four TFs with the highest importance scores were GCM1, NFAT5, MEF2A, and STAT4 (**Fig. 4.14A**). GCM1 had decreased expression in term EVT compared to first trimester EVT but was also the TF with the highest importance score in the largest subnetwork (**Fig. 4.14A**). This subnetwork was enriched for genes in the PI3K/AKT/mTOR pathway and extracellular matrix organization (adj p-value $<$ 0.01). We then analyzed which TFs had the highest importance scores when comparing genes downregulated in term EVT, compared to first trimester EVT. The top four TFs were E2F7, HOXC6, ZFHX3, and TAF9B (**Fig. 4.14B**). Additionally, we found that PCBP1 was the TF with the highest importance score in the three largest subnetworks in this comparison (**Fig. 4.14B**). The identification of TFs involved in EVT differentiation and maturation provides the first step toward the ability to model this important placental cell type *in vitro* and to begin to decipher the various functions these cells serve at the maternal-fetal interface.

Figure 4.14: (*next page*): Gene regulatory networks involved in EVT maturation.
All gene regulatory subnetworks were created by clustering STRING networks of up and downregulated genes in term EVT compared to first trimester EVT. (A) All subnetworks created from genes upregulated in term EVT compared to first trimester EVT. The top ten largest subclusters are enlarged. The top four transcription factors (TFs) with highest summed scores for all genes in the comparison are shown in the top right table. (B) All subnetworks created from downregulated genes in term EVT compared to first trimester EVT. The top 10 largest subclusters are enlarged. The top four TFs with highest summed scores for all genes in the comparison are shown in the top left table. For both analyses, the red TF names alongside each subnetwork are those with the highest importance scores in that subnetwork. Subnetwork node sizes and colors are determined by the nodes calculated betweenness centrality scores.


A
Transcription Factors With Highest Scores
GCM1
NFAT5
MEF2A
STAT4
GCM1
NFE2L1
ESR1
NFATC4
NFE2L2
JUN
STAT4
NFAT5
ETS2
GCM1
Node Fill Color/Size: Betweenness Centrality
Low
High
B
Transcription Factors With Highest Scores
E2F7
HOXC6
ZFHX3
TAF9B
PCBP1
PCBP1
PCBP1
ASH1L
SALL4
ZFHX3
SRSF9
HMG20B
ZNF30
KLF7
Node Fill Color/Size: Betweenness Centrality
Low
High

To validate some of these findings, we chose to focus on two TFs: STAT1, because it was identified in the top four TFs of the largest TF subnetwork in first trimester EVT (**Fig. 4.13, top**), and GCM1, because it was within the top four TFs with the highest importance scores in term, compared to first trimester EVT, and the TF with the highest importance score in the largest subnetwork (**Fig. 4.14A**). We first evaluated GCM1 expression and confirmed that this gene is most enriched in first trimester EVT with levels decreased at term (**Fig. 4.15A**). We performed in-situ hybridization on first trimester and term placental sections and confirmed enrichment of this gene to be highest in first trimester HLA-G$^+$ EVT (**Fig. 4.15B-C**). We next confirmed STAT1 gene expression in first trimester and term CTB and EVT and found that in fact it is enriched in EVT, with similar levels at the different gestational timepoints (**Fig. 4.15D**). We stained first trimester placental tissues with antibodies to HLA-G and STAT1 and found that STAT1 expression was confined to HLA-G$^+$ cells (**Fig. 4.15E**). We also differentiated one of our primary hTSC lines into EVT and found that STAT1 expression significantly increases over this differentiation time course (**Fig. 4.15F**). These data confirm that our analyses have indeed identified novel TF drivers of EVT differentiation and/or function. Future studies are needed to further validate the numerous additional findings from our TF network analyses, and to functionally assess the role of each of these TF's in EVT.

Figure 4.15: (*next page*): Localization and expression of GCM1 and STAT1 transcription factors.
(A) Bar graph depicting GCM1 expression in all 36 placental samples using RNA-seq. (B) In-situ hybridization of GCM1 (right-side) and immunohistochemistry for HLA-G (left-side) in adjacent sections of first trimester EVT. Empty arrowheads point to CTB and filled arrowheads to syncytiotrophoblast, which also expresses GCM1. HLA-G staining highlights EVT. Scale bars = 50 μm. (C) In-situ hybridization of GCM1 (right-side, with further magnification in inset) and immunohistochemistry for HLA-G (left-side) in adjacent sections of term EVT. HLA-G again highlights EVT at the basal plate. Scale bar = 100 μm in main panels and 25 μm in inset. (D) Bar graph depicting STAT1 expression in all 36 placental samples using RNA-seq. (E) Immunostaining of first trimester placenta with antibodies against HLA-G and STAT1, and counterstained with DAPI. Scale bars = 50 μm. (F) Western blot of STAT1 and beta-actin (control) during EVT differentiation of 1049 hTSC line.


A
Glial Cells Missing Transcription Factor 1 (GCM1)
Normalized Counts
40000
30000
20000
10000
0
CTB 1st
CTB term
EVT 1st
EVT term
B
EVT
HLA-G
EVT
GCM1
C
EVT
Villi
HLA-G
EVT
Villi
GCM1
D
Signal Transducer And Activator Of Transcription 1 (STAT1)
Normalized Counts
16000
14000
12000
10000
8000
6000
4000
2000
0
CTB 1st
CTB term
EVT 1st
EVT term
E
DAPI
HLA-G
STAT1
MERGE
F
d0
d1
d2
d3
d4
d5
STAT1
b-actin

4.4 Acknowledgements

Chapter 4 is an adapted reprint of material "Transcriptomic Drivers of Differentiation, Maturation, and Polyploidy in Human Extravillous Trophoblast" by Robert Morey, Omar Farah, Sampada Kallol, Daniela F. Requena, Morgan Meads, Matteo Moretto-Zita, Francesca Soncin, Louise C. Laurent, and Mana M. Parast published in *Frontiers in Cell and Developmental Biology* 2021. The dissertation author was one of the primary investigators and authors of this paper.

This study was supported by funds from the National Institutes of Health (NIH)/National Institute of Child Health and Human Development (NICHD, R01-HD089537 and R21-HD094618 to MP). RM was also supported by a grant from the National Institutes of Health, United States (NIH grant T32GM8806). OF was also supported by a fellowship from the Lalor Foundation. SK, LL, and FS were also supported by the NIH/NICHD R01-HD096260 to FS. This publication includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health SIG grant (S10 OD026929), and analysis was done using the Extreme Science and Engineering Discovery Environment (XSEDE) Comet for computational analysis, which is supported by the National Science Foundation grant number ACI-1548562 (allocation ID: TG-MCB140074).

The authors would like to thank all patients who donated placental tissues for this study, as well as the UC San Diego Health Clinical Laboratories and specifically Marie Dell' Aquila and Graciela Resolme for their help with the cytogenetics experiment. We would also like to thank Donald Pizzo for help with the in situ hybridization experiment.

Chapter 5

Conclusion

5.0.1 Pluripotency exists along a continuum

Human developmental biology owes much of its success to experiments conducted on model organisms. However, model organisms have obvious biological, ethical, and economic limitations. Fundamental differences in placental development between humans and model organisms, along with the ethical and safety concerns of experimental access to early human embryos, have necessitated the use of *in vitro* model systems [168]. In comparison to human embryonic development, there have been relatively few studies that have examined the use of *in vitro* systems to model human placental development. Human pluripotent stem cells (hPSCs) and trophoblast stem cells (hTSCs) have the potential to be valuable *in vitro* modeling tools for investigating early placental development. To realize this potential, a thorough characterization of the genomic, epigenomic, and transcriptomic landscape of hPSCs, hTSCs, and primary trophoblast must be undertaken. Here, using both bioinformatic tools and experimental methods, we have systematically characterized these cell types creating a holistic framework in which to model placental development. This dissertation provides a detailed description of the precursor cells and primary trophoblast cells needed to create an accurate model of early placental development.

The ability of hPSCs to recapitulate and model development depends on an understanding

of both their pluripotent substate and stability. In chapter 2, we systematically and quantitatively compared the cellular phenotype and genetic, epigenetic, and transcriptional changes in isogenic naïve and primed hESCs across time from derivation through long-term passage. In contrast to the large majority of previous studies on naïve hPSCs, in which hPSCs were established in primed conditions and later converted to naïve, all four of our hESC lines were derived in naïve conditions, and then each line was split and subjected to long-term culture in naïve and primed conditions in parallel, with longitudinal samples collected to enable phenotypic and genomic comparisons between matched naïve and primed cultures, as well as between derivation (p20, naïve only), early (p30, naïve and primed), and late (p50, naïve and primed) passages. We had unique access not only to blastocyst-stage preimplantation embryos, but also parental DNA, which enabled definitive identification of both *de novo* genetic aberrations and informative loci for evaluation of allelic expression.

Previous studies that compared the genetic stability of naïve and primed hESCs have reported conflicting results, with some reporting a higher frequency of genetic changes in naïve cells [8, 12, 64], while others showed the opposite [1, 110]. To our knowledge, a well-controlled molecular examination of genetic stability in different pluripotent substates has not yet been reported. To address this gap in the literature, we performed detailed genomic analysis of our unique model of virgin naïve hESCs and their isogenic primed counterparts. Karyotype and CMA analyses of early virgin naïve and their primed counterparts were normal, while WGS revealed several small aberrations in predominantly non-coding regions. In addition to variant calling using WGS, we used our higher coverage RNA-seq to detect SNVs of low allelic fraction in transcribed regions, but as with our WGS, found no substantial differences between our naïve and primed cultures, which is consistent with a recently reported study [171].

Since the phenotypic changes observed in our late passage cultures could not be explained by genetic changes alone, we explored the possibility that some differences might be the result of culture adaptations driven by epigenetic and/or transcriptomic changes, which has previously

been reported in hESCs [213], mESCs [137], primary mammalian cell lines [57], and human cancer cells [200]. Our naïve and primed cultures exhibited differential expression of mRNAs and miRNAs, as well as differences in DNA methylation. Both the transcriptional and DNA methylation profiles of our naïve lines showed that they are more similar to recently reported intermediate pluripotent cells [224] than to the majority of naïve hPSCs. Interestingly, the DNA methylation profiles of our naïve cultures were markedly different from hPSCs converted to naïve pluripotency using 5i/t2iL+PKCi-like protocols [182, 178, 64, 224], but were quite similar to those derived in NHSM naïve media [58, 146], which is the media most similar to that used in our study.

For all but one of our virgin naïve hESC lines, culture in naïve culture conditions results in a rapid transition (by passage 30) to a primed-like DNA methylation signature, increasing methylation not only globally, but also at known naïve pluripotency-associated genes and imprinted regions. We note that over time in culture, even the cultures maintained in primed conditions experienced epigenetic drift, with increases in global DNA methylation. Similarly, our mRNA and miRNA profiling results, particularly when examined in the context of previously published naïve and primed hPSC datasets (**Fig. 2.5B-D**), suggest that our p20 virgin naïve hESC lines occupy a naïve-like substate, and drift toward a more primed-like substate over time in culture. By passage 30, our naïve cultures appear stabilize in a state that remains slightly more naïve than the recently described intermediate substate [224].

Rather than any one of these substates being universally "better" than the others, the ability to stabilize hPSCs in a range of substates might prove to be advantageous by providing researchers with multiple developmental options depending on desired downstream applications. The utility of hESCs to model placental development largely depends on their potential to differentiate efficiently into functional trophoblast stem cells. Despite recent successes [39, 49] in differentiating naïve hPSCs to hTSCs, it might still be necessary to optimize differentiation protocols using naïve hPSCs as a starting material, in order to confirm their efficacy. Additionally,

our study confirmed the suspected risks of extended culture, which resulted in phenotypic changes and genetic instability of both naïve and primed hESCs, and highlights the importance of monitoring cultures for both changes in the pluripotency state and genetic stability. Our results support the hypothesis that pluripotency exists along a continuum, with the pluripotent substate being heavily influenced by the specific media used. However, we also observed accumulation of genetic aberrations and both epigenetic and transcriptomic drift over time in both primed and naïve conditions, which indicates that further research is needed to identify conditions that will allow for stable long-term propagation of hPSCs at specific desired substates of pluripotency.

5.0.2 Trophoblast stem cells derived from primed human pluripotent stem cells

The modeling of human trophoblast differentiation has proven difficult for multiple reasons, including significant divergence of factors regulating early embryonic development [21], as well as the establishment and maintenance of human trophoblast stem cells (hTSCs) [168]. Recently, hTSCs have been derived from human embryos, and early first trimester placenta. Aside from ethical challenges, the unknown disease potential of these cells limits their scientific utility necessitating an approach that enables the derivation of hTSCs from patients with known birth outcomes. Although several groups have recently derived hTSCs from naïve hPSCs [39, 49], these cells have potential limitations such as the loss of imprinting as discussed in chapter 2. In chapter 3, we show that hTSCs can be derived from not only naïve pluripotent stem cells but also from the primed substate following a short induction period with BMP4-containing media.

In chapter 3, we showed that our hPSC-derived TSC functionally recapitulate primary TSC. They are able to differentiate into both hCG-producing syncytiotrophoblast (STB) and surface HLAG-expressing extravillous trophoblast (EVT). This differentiation potential was recapitulated *in vivo*, by their ability to form trophoblastic tumors in immunocompromised mice, with the typical necrotic center and expression of STB- and EVT- associated markers. Such

features of the hPSC- derived TSC functionally define their true identity as trophoblast, and exclude other identities, including amnion. In addition to transcriptomic differences, as described in chapter 3, normal amniotic epithelium is a simple cuboidal epithelium, lacking expression of TP63 [112]. Rare reports of amnion expressing HLA-G do not cite specific antibodies used, and in fact, report co-expression of class I HLA molecules in these cells [172]; similarly, the single report of hCG expression in early amnion within the developing human embryo [215] shows staining in a simple, non-stratified epithelium, rather than the secretory ability of a stem cell, following forskolin- induced syncytialization, as occurs in primary TSC. When morphologic features and marker expression are considered together, hPSC-derived TSC most closely resemble primary, placenta- derived TSC, and are clearly distinct from amniotic epithelium.

Perhaps the most difficult-to-explain aspect of the trophoblast differentiation ability of primed hPSC is the developmental trajectory of these cells. The initial segregation of the TE lineage follows the transition from totipotency (2-cell/2C state) to naïve pluripotency, which is the defining feature of pre-implantation inner cell mass/ICM or epiblast [160, 12, 50]. It has been suggested that, at least in mouse, this process of lineage segregation into TE and ICM is in fact epigenetically irreversible, precluding mouse ESC transdifferentiated into TSC-like cells (e.g. by overexpression of Cdx2) to significantly contribute to the trophoblast compartment [27]. Primed hPSC are considered equivalent to post-implantation epiblast, a stage of development far beyond TE development, and thus, by definition, should not have the potential to generate TE [45]. In fact, in mouse, both naïve and primed ESC exclusively give rise to epiblast-derived cells when injected into proper stage embryos [223, 80]. However, unlike naïve mouse ESC, naïve human ESC have recently been shown by multiple groups to readily give rise to trophoblast [84, 63, 39, 49], suggesting greater plasticity of human ESC. In addition, it has been shown that media containing BMP4 induces chromatin remodeling in primed mouse ESC [98, 224], inducing some cells toward a naïve, and even 2C-like, fate, while producing TE-like features in other cells [98, 186]. Our own analysis of primed hPSC treated with BMP4-containing media points to a

direct conversion of these cells to a TE-like fate, following 4 days of treatment with BMP4 and IWP2, with induction of only one naïve pluripotency marker noted. Comparison to Io et al.'s [84] naïve hPSC-derived cells also showed that primed hPSC treated with BMP4/IWP2 clustered with naïve TE, not naïve hPSC, thus suggesting differentiation into TSC through a TE intermediate. Finally, recent reports have shown that direct reprogramming of human fibroblasts to a TSC-like fate can occur when Yamanaka pluripotency factors are applied to the cells, and the media switched from one used for culture of hPSC (E7) to Okae's TSC media during reprogramming [31]. Analysis of the reprogramming process suggests that iTSC can be captured from a TE-like subpopulation, without initial formation of naïve hPSC intermediates [116], suggesting perhaps greater plasticity, and fewer epigenetic barriers, between early human stem cell lineages.

Our protocol offers a relatively simple way to convert the many existing hPSC lines, including induced pluripotent stem cells representing many diseases, into TSC-like cells, thus allowing researchers broad access to platforms for human placental research, even where this research may be limited by lack of access to first trimester placental tissues. Nevertheless, much work, and many questions, remain, including how primed hPSC-derived TSC compare to similar cells derived from naïve hPSC, not just in their DNA methylation patterns, but in other aspects of their epigenome (including their chromatin and miRNA landscape) as well as their ability to recapitulate trophoblast-based disorders of the placenta. In addition, the different pluripotent states of human ESC, as well as TE and TSC, deserve further study in context of the human embryo, including the extent of their lineage segregation, as well as the role of BMP4 in potentially enhancing inter-conversion.

5.0.3 Characterization of pathways and transcription factor networks involved in both initial differentiation and maturation of extravillous trophoblast

To accurately model placental development, it is critical to comprehensively characterize the terminal differentiation point of the model and highlight the important pathways and factors driving the maturation process. Not only is a well-defined map of trophoblast differentiation necessary to build a good model but it is also imperative in order to decipher what abnormal placental development looks like. Abnormal placental development has been linked to numerous pregnancy complications, including pre-eclampsia, intrauterine growth restriction, miscarriage, and stillbirth [95, 96, 82, 25, 54, 99]. The placenta develops by forming primary villi consisting of rapidly-proliferating cytotrophoblast (CTB). These cells fuse to form an outer layer of villous syncytiotrophoblast (STB). At the same time, the CTB start to differentiate into EVT within the trophoblast columns of the early gestation placenta, anchoring the placenta to the uterine wall [191]. EVT mature as they move distally within the trophoblast column, and subsequently invade into the decidua and myometrium as interstitial EVT, or remodel decidual arterioles as endovascular EVT [150, 151]. While much has been done to characterize early gestation EVT, fewer studies have focused on mature EVT. In chapter 4, we set out to characterize these cells from normal term placentae, including their genome and transcriptome; given this tissue source, the majority of these cells are likely mature interstitial (and not endovascular) EVT. Additionally, by comparison to both first trimester and term CTB, as well as first trimester EVT, we assembled gene regulatory networks to better understand the pathways and transcription factors involved in the maturation and unique genomic architecture of EVT.

EVT share numerous cellular characteristics with tumor cells, including epithelial-mesenchymal transition [199]. Specifically, earlier studies [230, 229] suggested that EVT show moderate genome amplification (up to 8N) but are not highly polyploid, unlike mouse

TGCs; the latter not only show significant polyploidy (with some cells >900N), but also contain functionally-relevant under- or over-represented genomic regions [67, 66]. In chapter 4, to gain a better understanding of the genomes of normal human EVT, we applied multiple cellular and bioinformatic methods. First, we reanalyzed a previously-published whole genome sequencing dataset from CTB and EVT purified from first trimester placentae [197], applying CNV detection algorithms not used in the original analysis. We did not find CNVs that had been reported in mouse TGCs but did find three duplications previously reported in bulk second trimester human placental samples [93, 41]. Additionally, we performed genome-wide CNV analysis of our own samples using high-resolution SNP genotyping arrays and identified three duplications in our term EVT samples, none of which were common between samples or previously reported [93, 41]. Although we could not identify any common EVT-specific CNVs among preparations from different placentae, including those from a previous publication, definitive assessment of this observation will require a substantially larger sample size. We also applied ploidy-detection algorithms on Velicky et al.'s whole genome sequencing data [197] as well as our SNP genotyping data but found no evidence of polyploidy. Lastly, to rule out the presence of a subpopulation of EVT cells showing a high polyploidy rate or widespread CNV's, we performed single-cell CNV analysis on a set of isolated samples and again found no evidence of polyploidy or cells with a large number of CNVs. Nevertheless, we suspect that the algorithms used in our analysis are poorly suited to calling polyploidy.

To validate previous reports of polyploidy in EVT, and to confirm the limitations of the algorithms applied to our genomic data, we determined the DNA content of our isolated placental cells by flow cytometry and found the majority (57%) of first trimester EVT, and a lower percentage (44%) of term EVT to be tetraploid. FISH confirmed our flow cytometry results. As previously reported [230], we also noted a population of EVT with > 4N status, with a slightly larger proportion of such cells at term. Additionally, we also differentiated primary hTSC cells to EVT, using established protocols [142] to assess how well current *in vitro* models recapitulate the

polyploid phenotype seen in primary EVT. We found that *in vitro* differentiation did not increase the proportion of hyper-diploid cells, suggesting that perhaps *in vivo*, EVT receive additional signals from their environment that lead to polyploidization. Additional work is thus needed to better recapitulate the *in vivo* EVT state *in vitro*.

The biological significance of polyploidization remains unclear, particularly in the placenta. In the liver, polyploidization has been hypothesized to be a hallmark of terminal differentiation, a mechanism through which a cell may shift energy usage from cell division to more important functions, and/or as a way to protect cells against oxidative stress and genotoxic damage [207]. Oxygen tension in blood surrounding the placental villi has been reported to increase threefold during pregnancy, causing oxidative stress [88] and recently, [41] reported to cause a substantial mutational burden in placental tissue. Thus, the onset of endoreduplication and senescence, which requires replication arrest in a previously proliferative cell type such as EVT, would lead to the acquisition of multiple sets of chromosomes, and could function to buffer cells against harmful mutations. Our evaluation of the EVT transcriptome, discussed below, may shed some light on this question; however, future studies examining transcriptomes of EVT subpopulations, separated based on different levels of ploidy, along with delineation of the spatial distribution of these subpopulations, are needed to more precisely define the function(s) of polyploid EVT.

To help characterize the differences between the largely diploid CTB and the majority polyploid EVT, we profiled the transcriptomes of isolated first trimester and term CTB and EVT, using RNA-seq. Our RNA-seq dataset was well powered, consisting of 10 first trimester CTB, 10 first trimester EVT, 10 term CTB, and 6 term EVT samples at an average depth of over 40 million uniquely mapped reads, offering a detailed look at how these two trophoblast cell types differed at two different gestational ages. With respect to first trimester EVT, our GSEA results confirmed previously published microarray datasets [3, 181, 185, 201] and added further evidence that EVT differentiation entails upregulation of epithelial-mesenchymal transition and hypoxia

signaling, along with many inflammatory- and immune-mediated processes, and eventually a downregulation of proliferation and cell cycle pathways suggesting terminal differentiation. Perhaps surprisingly, many of the same pathways were up- or down-regulated when comparing CTB and EVT from term placenta, suggesting that, despite a previous report [125], given the right conditions, it may be possible to differentiate term CTB to EVT. Interestingly, two pathways that were found to be upregulated in first trimester EVT and not term EVT were the PI3K/AKT/mTOR signaling and Unfolded Protein Response (UPR) pathways. By inhibiting one arm of the UPR pathway during EVT differentiation of primary hTSC, we found that this pathway is important for surface expression of HLA-G, a molecule required for EVT crosstalk with maternal natural killer cells [155]. ER stress has been studied in context of oxidative stress-induced placental dysfunction (i.e. in the setting of preeclampsia and intrauterine growth restriction) [26, 129]. It is also known that enhanced induction of this pathway disrupts placental development [226]; but, until now, it had not been studied specifically in context of EVT. Interestingly, hypoxic conditions, known to promote EVT differentiation [201], induce adaptive cellular responses including the UPR pathway. It is tempting to speculate that hypoxia-induced EVT differentiation is partially mediated through the IRE1-alpha arm of UPR; further studies are needed to test this hypothesis.

At the same time, we also noted a large overlap in pathways that were significantly different in first trimester CTB vs. EVT and first trimester vs. term EVT. This suggests that these overlapping pathways may be essential for both differentiation and maturation of, or simply characteristic of both immature and mature, EVT. Many such pathways are likely to involve signals from the decidua and decidual immune cells, of which natural killer (NK) and macrophages are the most abundant [198, 152]. Given the lack of polyploidy in our *in vitro*-differentiated EVT, it is worth exploring which, if any, of these pathways require further manipulation in order to optimize EVT differentiation of hTSC lines *in vitro*.

Polyploidy has long been intricately linked with cellular senescence [29] and a recent report has suggested that first trimester EVT exhibit both endoreduplication-induced polyploidy

and senescence [197]. Therefore, we examined our RNA-seq data for genes involved in the cell cycle, endoreduplication, and cellular senescence. We found that term EVT did not express mitosis-linked genes, such as cyclin B, Ki67, and Aurora B, but all three genes were expressed in first trimester EVT, in contrast to previous reports [197]. As discussed above, endoreduplication consists of DNA replication without cell or nuclear division, and is triggered by p57 (CDKN1C)'s inhibition of CDK1, and p21 (CDKN1A)'s suppression of checkpoint protein kinase (CHEK1), preventing induction of apoptosis [192]. Our term EVT showed marked decrease in CDK1 and CHEK1 and corresponding increases in p57 and p21, suggesting that, as in mouse TGCs, human term EVT undergo endoreduplication. Interestingly, the G1-S transition promoting cyclin E1 was highly expressed in first trimester EVT, but nearly 3-fold lower in term EVT, whereas cyclin E2 progressively decreased between both CTB, first trimester EVT, and term EVT. Additionally, the mitosis-linked cyclin B gene was also highly expressed in first trimester EVT but almost undetectable in term EVT. However, despite Velicky et al.'s claim that Cyclin $A^{+}/p57^{-}$ expression could be used as a marker for endoreduplicating HLA-G^{+} trophoblast, we found very low levels of cyclin A in all of our samples despite relatively deep sequencing. We did observe the genes encoding the RB protein and its E2F transcription factor family targets also drop precipitously in term EVT, whereas senescence-associated secretory phenotype (SASP)-associated genes were most highly expressed in term EVT. In the context of our FISH and flow cytometry data, these results suggest that cells may begin to undergo endoreduplication and senescence in the first trimester, but progress into a more fully senescent phenotype only at term. However, further studies, including evaluation of protein and phosphorylation levels (including RB protein phosphorylation) [60] should be conducted to more precisely evaluate the cell cycle during EVT formation.

Using our RNA-seq data, we sought to identify TFs needed to drive EVT differentiation and maturation. We found that many TFs identified as being important in our protein-to-protein interaction networks were well-known regulators of processes previously identified as being vital

in normal EVT development. During differentiation of first trimester CTB to EVT, we found the top four TFs with highest importance scores in our largest clustered subnetwork to be TFEB, IRF7, IRF8, and STAT1, all of which are involved in immune response [122, 24, 89, 92], correlating well with our GSEA findings [227]. We validated STAT1 as uniquely expressed in EVT within first trimester placental tissue, with its expression increasing during *in vitro* differentiation of hTSC into EVT; similar findings were recently reported by [33], who have suggested that the signals leading to STAT1 induction are derived from decidual stromal cells. Additionally, we noted that this clustered subnetwork, and the largest subnetwork in the genes up-regulated in term EVT vs. CTB, was centered around FN1, which, we found to be linked with XBP1, a key part of the UPR pathway in mammals [74]. XBP1 was recently reported to initiate FN1 expression in colon cancer cells [216] and thus deserves further study as a possible link between hypoxia, UPR, and ECM remodeling in the context of EVT differentiation.

In our analysis of TFs involved in EVT maturation, we identified GCM1 as the TF with the highest importance score in the largest subnetwork of genes upregulated in term, compared to first trimester, EVT, although its expression appeared to decrease in term, compared to first trimester, EVT. GCM1 is a TF known best as a master regulator of labyrinthine or villous trophoblast differentiation in both mice and human [43, 9]. However, we and others have shown that GCM1 is also highly expressed in human EVT [10, 37, 201]. GCM1 has been shown to play a role in trophoblast invasion, acting through HTRA4, a serine protease that facilitates fibronectin cleavage, to suppress cell-cell fusion and promote invasion [205]. However, while this describes a clear role for GCM1 in a basic function (invasion) of all EVT, including those in first trimester, its identification as a key TF involved in regulation of term EVT transcriptome requires further study.

Other TFs identified as controlling up-regulated genes during EVT maturation included NFAT5 and STAT4, both of which are involved in immune response, with the latter regulating response to IL-12 signaling [133, 109]. The IL-12 cytokine family, produced by EVT, is important in establishment of maternal–fetal tolerance through modulation of naïve conventional T cells

and their conversion into induced regulatory T cells [115]. Papuchova et al. [145] has pointed to heterogeneity within term EVT, with subtypes showing differing capacities for modulating resident immune cells, including regulatory T cells. Future studies, including single-cell analysis, are warranted to further study EVT heterogeneity, based not just on ploidy and gene expression, but on functional capacities, in order to better understand the role of these cells in establishment and maintenance of the maternal–fetal interface.

In summary, chapter 4 builds on earlier reports characterizing first trimester EVT, extending such genomic and transcriptomic studies to term EVT, and defining pathways and transcription factor networks involved in both initial differentiation and maturation of this important trophoblast lineage at the maternal-fetal interface. Our results suggest that term EVT lack high rates of copy number variations, though studies using whole genome sequencing with substantially larger samples sizes are needed to definitively identify or rule-out the presence of functionally-relevant under- or over-represented genomic regions. Additionally, we have highlighted senescence and polyploidy-related genes, pathways, networks, and transcription factors that appeared to be important in EVT differentiation and maturation, and have validated a critical role for the unfolded protein response in formation of functional EVT. Lastly, our results highlight the need for more optimized *in vitro* models of EVT differentiation, further research into functional differences among EVT subpopulations with different ploidy levels, and studies of placental diseases that may be associated with changes in cellular ploidy or dysfunctional EVT differentiation or maturation.

5.0.4 Acknowledgements

Chapter 5 contains portions of adapted material currently being prepared for submission for publication by Chen Dekel*, Robert Morey*, Jacob Hanna, Louise Laurent, Dalit Ben-Yosef, and Hadar Amir (*These authors contributed equally to this work). The dissertation author was one of the primary investigators and authors of this paper. Additionally, chapter five contains portions of adapted material submitted for publication in Stem Cell Reports by Francesca Soncin*,

Robert Morey*, Tony Bui, Daniela F. Requena, Virginia Chu Cheung, Sampada Kallol, Ryan Kittle, Omar Farah, Morgan Meads, Donald Pizzo, Mariko Horii, Kathleen M. Fisch, Mana M. Parast (*These authors contributed equally to this work). The dissertation author was one of the primary investigators and authors of this paper. Lastly, chapter 5 contains portions of adapted reprint of material "Transcriptomic Drivers of Differentiation, Maturation, and Polyploidy in Human Extravillous Trophoblast" by Robert Morey, Omar Farah, Sampada Kallol, Daniela F. Requena, Morgan Meads, Matteo Moretto-Zita, Francesca Soncin, Louise C. Laurent, and Mana M. Parast published in *Frontiers in Cell and Developmental Biology* 2021. The dissertation author was one of the primary investigators and authors of this paper.

Bibliography

[1] Hadar Amir, Thomas Touboul, Karen Sabatini, Divya Chhabra, Ibon Garitaonandia, Jeanne F Loring, Robert Morey, and Louise C Laurent. Spontaneous Single-Copy Loss of TP53 in Human Embryonic Stem Cells Markedly Increases Cell Proliferation and Survival. *Stem cells (Dayton, Ohio)*, 35(4):872–885, apr 2017.

[2] Katherine Amps, Peter W Andrews, George Anyfantis, Lyle Armstrong, Stuart Avery, Hossein Baharvand, Julie Baker, Duncan Baker, Maria B Munoz, Stephen Beil, Nissim Benvenisty, Dalit Ben-Yosef, Juan-Carlos Biancotti, Alexis Bosman, Romulo Martin Brena, Daniel Brison, Gunilla Caisander, María V Camarasa, Jieming Chen, Eric Chiao, Young Min Choi, Andre B H Choo, Daniel Collins, Alan Colman, Jeremy M Crook, George Q Daley, Anne Dalton, Paul A De Sousa, Chris Denning, Janet Downie, Petr Dvorak, Karen D Montgomery, Anis Feki, Angela Ford, Victoria Fox, Ana M Fraga, Tzvia Frumkin, Lin Ge, Paul J Gokhale, Tamar Golan-Lev, Hamid Gourabi, Michal Gropp, Guangxiu Lu, Ales Hampl, Katie Harron, Lyn Healy, Wishva Herath, Frida Holm, Outi Hovatta, Johan Hyllner, Maneesha S Inamdar, Astrid Kresentia Irwanto, Tetsuya Ishii, Marisa Jaconi, Ying Jin, Susan Kimber, Sergey Kiselev, Barbara B Knowles, Oded Kopper, Valeri Kukharenko, Anver Kuliev, Maria A Lagarkova, Peter W Laird, Majlinda Lako, Andrew L Laslett, Neta Lavon, Dong Ryul Lee, Jeoung Eun Lee, Chunliang Li, Linda S Lim, Tenneille E Ludwig, Yu Ma, Edna Maltby, Ileana Mateizel, Yoav Mayshar, Maria Mileikovsky, Stephen L Minger, Takamichi Miyazaki, Shin Yong Moon, Harry Moore, Christine Mummery, Andras Nagy, Norio Nakatsuji, Kavita Narwani, Steve K W Oh, Sun Kyung Oh, Cia Olson, Timo Otonkoski, Fei Pan, In-Hyun Park, Steve Pells, Martin F Pera, Lygia V Pereira, Ouyang Qi, Grace Selva Raj, Benjamin Reubinoff, Alan Robins, Paul Robson, Janet Rossant, Ghasem H Salekdeh, Thomas C Schulz, Karen Sermon, Jameelah Sheik Mohamed, Hui Shen, Eric Sherrer, Kuldip Sidhu, Shirani Sivarajah, Heli Skottman, Claudia Spits, Glyn N Stacey, Raimund Strehl, Nick Strelchenko, Hirofumi Suemori, Bowen Sun, Riitta Suuronen, Kazutoshi Takahashi, Timo Tuuri, Parvathy Venu, Yuri Verlinsky, Dorien Ward-van Oostwaard, Daniel J Weisenberger, Yue Wu, Shinya Yamanaka, Lorraine Young, and Qi Zhou. Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. *Nature biotechnology*, 29(12):1132–1144, nov 2011.

[3] R Apps, A Sharkey, L Gardner, V Male, M Trotter, N Miller, R North, S Founds, and

A Moffett. Genome-wide expression profile of first trimester villous and extravillous human trophoblast cells. *Placenta*, 32(1):33–43, jan 2011.

[4] Richard Apps, Shawn P Murphy, Raymond Fernando, Lucy Gardner, Tashmeeta Ahad, and Ashley Moffett. Human leucocyte antigen (HLA) expression of primary trophoblast cells and placental cell lines, determined using single antigen beads to characterize allotype specificities of anti-HLA antibodies. *Immunology*, 127(1):26–39, may 2009.

[5] Sorin Armeanu, Michael Bitzer, Ulrich M Lauer, Sascha Venturelli, Anita Pathil, Matthias Krusch, Stephan Kaiser, Jürgen Jobst, Irina Smirnow, Annika Wagner, Alexander Steinle, and Helmut R Salih. Natural killer cell-mediated lysis of hepatoma cells via specific induction of NKG2D ligands by the histone deacetylase inhibitor sodium valproate. *Cancer research*, 65(14):6321–6329, jul 2005.

[6] Martin J Aryee, Andrew E Jaffe, Hector Corrada-Bravo, Christine Ladd-Acosta, Andrew P Feinberg, Kasper D Hansen, and Rafael A Irizarry. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, 30(10):1363–1369, may 2014.

[7] Stuart Avery, Adam J Hirst, Duncan Baker, Chin Yan Lim, Sharmini Alagaratnam, Rolf I Skotheim, Ragnhild A Lothe, Martin F Pera, Alan Colman, Paul Robson, Peter W Andrews, and Barbara B Knowles. BCL-XL mediates the strong selective advantage of a 20q11.21 amplification commonly found in human embryonic stem cell cultures. *Stem cell reports*, 1(5):379–386, 2013.

[8] Yishai Avior, Kevin Eggan, and Nissim Benvenisty. Cancer-Related Mutations Identified in Primed and Naive Human Pluripotent Stem Cells. *Cell Stem Cell*, 25(4):456–461, 2019.

[9] D Baczyk, S Drewlo, L Proctor, C Dunk, S Lye, and J Kingdom. Glial cell missing-1 transcription factor is required for the differentiation of the human trophoblast. *Cell death and differentiation*, 16(5):719–727, may 2009.

[10] D Baczyk, A Satkunaratnam, B Nait-Oumesmar, B Huppertz, J C Cross, and J C P Kingdom. Complex patterns of GCM1 mRNA and protein in villous and extravillous trophoblast cells of the human placenta. *Placenta*, 25(6):553–559, jul 2004.

[11] Tao Bai, Chian Yu Peng, Ivy Aneas, Noboru Sakabe, Daniela F. Requena, Christine Billstrand, Marcelo Nobrega, Carole Ober, Mana Parast, and John A. Kessler. Establishment of human induced trophoblast stem-like cells from term villous cytotrophoblasts. *Stem Cell Research*, 56, 2021.

[12] Duncan Baker, Adam J Hirst, Paul J Gokhale, Miguel A Juarez, Steve Williams, Mark Wheeler, Kerry Bean, Thomas F Allison, Harry D Moore, Peter W Andrews, and Ivana Barbaric. Detecting Genetic Mosaicism in Cultures of Human Pluripotent Stem Cells. *Stem cell reports*, 7(5):998–1012, nov 2016.

[13] Duncan E C Baker, Neil J Harrison, Edna Maltby, Kath Smith, Harry D Moore, Pamela J Shaw, Paul R Heath, Hazel Holden, and Peter W Andrews. Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nature biotechnology*, 25(2):207–215, feb 2007.

[14] Bradley P. Balaton, Allison M. Cotton, and Carolyn J. Brown. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biology of Sex Differences*, 6(1):1–11, 2015.

[15] Shiran Bar, Maya Schachter, Talia Eldar-Geva, and Nissim Benvenisty. Large-Scale Analysis of Loss of Imprinting in Human Pluripotent Stem Cells. *Cell Reports*, 19(5):957–968, 2017.

[16] Tahsin Stefan Barakat, Florian Halbritter, Man Zhang, André F. Rendeiro, Elena Perenthaler, Christoph Bock, and Ian Chambers. Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells. *Cell Stem Cell*, 23(2):276–288.e8, 2018.

[17] P W Barlow and M I Sherman. The biochemistry of differentiation of mouse trophoblast: studies on polyploidy. *Journal of embryology and experimental morphology*, 27(2):447–465, apr 1972.

[18] Nathan Basisty, Abhijit Kale, Ok Hee Jeon, Chisaka Kuehnemann, Therese Payne, Chirag Rao, Anja Holtz, Samah Shah, Vagisha Sharma, Luigi Ferrucci, Judith Campisi, and Birgit Schilling. A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS biology*, 18(1):e3000599, jan 2020.

[19] Leanne Bellamy, Juan-Pablo Casas, Aroon D Hingorani, and David J Williams. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *BMJ (Clinical research ed.)*, 335(7627):974, nov 2007.

[20] Uri Ben-David, Yoav Mayshar, and Nissim Benvenisty. Large-scale analysis reveals acquisition of lineage-specific chromosomal aberrations in human adult stem cells. *Cell stem cell*, 9(2):97–102, aug 2011.

[21] Paul Blakeley, Norah M E Fogarty, Ignacio del Valle, Sissy E Wamaitha, Tim Xiaoming Hu, Kay Elder, Philip Snell, Leila Christie, Paul Robson, and Kathy K Niakan. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development (Cambridge, England)*, 142(18):3151–3165, sep 2015.

[22] M Borges, P Bose, H-G Frank, P Kaufmann, and A J G Pötgens. A two-colour fluorescence assay for the measurement of syncytial fusion between trophoblast-derived cell lines. *Placenta*, 24(10):959–964, nov 2003.

[23] A Bradley, M Evans, M H Kaufman, and E Robertson. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, 309(5965):255–256, may 1984.

[24] Owen A Brady, José A Martina, and Rosa Puertollano. Emerging roles for TFEB in the immune response and inflammation. *Autophagy*, 14(2):181–189, 2018.

[25] Jan J Brosens, Robert Pijnenborg, and Ivo A Brosens. The myometrial junctional zone spiral arteries in normal and abnormal pregnancies: a review of the literature. *American journal of obstetrics and gynecology*, 187(5):1416–1423, nov 2002.

[26] G J Burton, H-W Yung, T Cindrova-Davies, and D S Charnock-Jones. Placental endoplasmic reticulum stress and oxidative stress in the pathophysiology of unexplained intrauterine growth restriction and early onset preeclampsia. *Placenta*, 30 Suppl A(Suppl):S43–8, mar 2009.

[27] Francesco Cambuli, Alexander Murray, Wendy Dean, Dominika Dudzinska, Felix Krueger, Simon Andrews, Claire E Senner, Simon J Cook, and Myriam Hemberger. Epigenetic memory of the first cell fate decision prevents complete ES cell reprogramming into trophoblast. *Nature communications*, 5:5538, nov 2014.

[28] Daniel L Cameron, Jonathan Baber, Charles Shale, Anthony T Papenfuss, Jose Espejo Valle-Inclan, Nicolle Besselink, Edwin Cuppen, and Peter Priestley. GRIDSS, PURPLE, LINX: Unscrambling the tumor genome via integrated analysis of structural variation and copy number. *bioRxiv*, page 781013, jan 2019.

[29] Judith Campisi and Fabrizio d'Adda di Fagagna. Cellular senescence: when bad things happen to good cells. *Nature reviews. Molecular cell biology*, 8(9):729–740, sep 2007.

[30] A M Carter. Animal models of human placentation–a review. *Placenta*, 28 Suppl A:S41–7, apr 2007.

[31] Gaël Castel, Dimitri Meistermann, Betty Bretin, Julie Firmin, Justine Blin, Sophie Loubersac, Alexandre Bruneau, Simon Chevolleau, Stéphanie Kilens, Caroline Chariau, Anne Gaignerie, Quentin Francheteau, Harunobu Kagawa, Eric Charpentier, Léa Flippe, Valentin François-Campion, Sandra Haider, Bianca Dietrich, Martin Knöfler, Takahiro Arima, Jérémie Bourdon, Nicolas Rivron, Damien Masson, Thierry Fournier, Hiroaki Okae, Thomas Fréour, and Laurent David. Induction of Human Trophoblast Stem Cells from Somatic Cells and Pluripotent Stem Cells. *Cell reports*, 33(8):108419, nov 2020.

[32] Yun-Shen Chan, Jonathan Göke, Jia-Hui Ng, Xinyi Lu, Kevin Andrew Uy Gonzales, Cheng-Peow Tan, Wei-Quan Tng, Zhong-Zhi Hong, Yee-Siang Lim, and Huck-Hui Ng. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell stem cell*, 13(6):663–675, dec 2013.

[33] Chao Chen, Xiaomin Kang, Congcong Li, Feng Guo, Qiaohong Wang, and Aimin Zhao. Involvement of signal transducers and activators of transcription in trophoblast differentiation. *Placenta*, 105:94–103, feb 2021.

[34] Edward Y Chen, Christopher M Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz Meirelles, Neil R Clark, and Avi Ma'ayan. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14:128, apr 2013.

[35] Hongwei Chen, Irène Aksoy, Fabrice Gonnot, Pierre Osteil, Maxime Aubry, Claire Hamela, Cloé Rognard, Arnaud Hochard, Sophie Voisin, Emeline Fontaine, Magali Mure, Marielle Afanassieff, Elouan Cleroux, Sylvain Guibert, Jiaxuan Chen, Céline Vallot, Hervé Acloque, Clémence Genthon, Cécile Donnadieu, John De Vos, Damien Sanlaville, Jean-François Guérin, Michael Weber, Lawrence W Stanton, Claire Rougeulle, Bertrand Pain, Pierre-Yves Bourillot, and Pierre Savatier. Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency. *Nature communications*, 6:7095, may 2015.

[36] Connie G Chiu, Pascal St-Pierre, Ivan R Nabi, and Sam M Wiseman. Autocrine motility factor receptor: a clinical review. *Expert review of anticancer therapy*, 8(2):207–217, feb 2008.

[37] Yueh Ho Chiu and Hungwen Chen. GATA3 inhibits GCM1 activity and trophoblast cell invasion. *Scientific reports*, 6:21630, feb 2016.

[38] Chih-Hung Chou, Sirjana Shrestha, Chi-Dung Yang, Nai-Wen Chang, Yu-Ling Lin, Kuang-Wen Liao, Wei-Chi Huang, Ting-Hsuan Sun, Siang-Jyun Tu, Wei-Hsiang Lee, Men-Yee Chiew, Chun-San Tai, Ting-Yen Wei, Tzi-Ren Tsai, Hsin-Tzu Huang, Chung-Yu Wang, Hsin-Yi Wu, Shu-Yi Ho, Pin-Rong Chen, Cheng-Hsun Chuang, Pei-Jung Hsieh, Yi-Shin Wu, Wen-Liang Chen, Meng-Ju Li, Yu-Chun Wu, Xin-Yi Huang, Fung Ling Ng, Waradee Buddhakosai, Pei-Chun Huang, Kuan-Chun Lan, Chia-Yen Huang, Shun-Long Weng, Yeong-Nan Cheng, Chao Liang, Wen-Lian Hsu, and Hsien-Da Huang. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic acids research*, 46(D1):D296–D302, jan 2018.

[39] Jessica K. Cinkornpumin, Sin Young Kwon, Yixin Guo, Ishtiaque Hossain, Jacinthe Sirois, Colleen S. Russett, Hsin Wei Tseng, Hiroaki Okae, Takahiro Arima, Thomas F. Duchaine, Wanlu Liu, and William A. Pastor. Naive Human Embryonic Stem Cells Can Give Rise to Cells with a Trophoblast-like Transcriptome and Methylome. *Stem Cell Reports*, 15(1):198–213, 2020.

[40] Amanda J. Collier and Peter J. Rugg-Gunn. Identifying Human Naïve Pluripotent Stem Cells Evaluating State-Specific Reporter Lines and Cell-Surface Markers. *BioEssays*, 40(5):1–12, 2018.

[41] Tim H H Coorens, Thomas R W Oliver, Rashesh Sanghvi, Ulla Sovio, Emma Cook, Roser Vento-Tormo, Muzlifah Haniffa, Matthew D Young, Raheleh Rahbari, Neil Sebire, Peter J Campbell, D Stephen Charnock-Jones, Gordon C S Smith, and Sam Behjati. Inherent mosaicism and extensive mutation of human placentas. *Nature*, 592(7852):80–85, apr 2021.

[42] J C Cross, D Baczyk, N Dobric, M Hemberger, M Hughes, D G Simmons, H Yamamoto, and J C P Kingdom. Genes, development and evolution of the placenta. *Placenta*, 24(2-3):123–130, 2003.

[43] James C Cross, Haruo Nakano, David R C Natale, David G Simmons, and Erica D Watson. Branching morphogenesis during development of placental villi. *Differentiation; research in biological diversity*, 74(7):393–401, sep 2006.

[44] Sonia Dasilva-Arnold, Joanna L. James, Abdulla Al-Khan, Stacy Zamudio, and Nicholas P. Illsley. Differentiation of first trimester cytotrophoblast to extravillous trophoblast involves an epithelial-mesenchymal transition. *Placenta*, 36(12):1412–1418, 2015.

[45] Alejandro De Los Angeles. The Pluripotency Continuum and Interspecies Chimeras. *Current protocols in stem cell biology*, 50(1):e87, sep 2019.

[46] Manon Desgres and Philippe Menasché. Clinical Translation of Pluripotent Stem Cell Therapies: Challenges and Considerations. *Cell stem cell*, 25(5):594–606, nov 2019.

[47] Alexander Dobin, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1):15–21, jan 2013.

[48] Benjamin T. Dodsworth, Klas Hatje, Maria Rostovskaya, Rowan Flynn, Claas A. Meyer, and Sally A. Cowley. Profiling of naïve and primed human pluripotent stem cells reveals state-associated miRNAs. *Scientific Reports*, 10(1):1–9, 2020.

[49] Chen Dong, Mariana Beltcheva, Paul Gontarz, Bo Zhang, Pooja Popli, Laura A. Fischer, Shafqat A. Khan, Kyoung Mi Park, Eun Ja Yoon, Xiaoyun Xing, Ramakrishna Kommagani, Ting Wang, Lilianna Solnica-Krezel, and Thorold W. Theunissen. Derivation of trophoblast stem cells from naïve human pluripotent stem cells. *eLife*, 9:1–26, 2020.

[50] Chen Dong, Laura A. Fischer, and Thorold W. Theunissen. Recent insights into the naïve state of human pluripotency and its applications. *Experimental Cell Research*, 385(1):111645, 2019.

[51] Jonathan S Draper, Kath Smith, Paul Gokhale, Harry D Moore, Edna Maltby, Julie Johnson, Lorraine Meisner, Thomas P Zwaka, James A Thomson, and Peter W Andrews. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nature biotechnology*, 22(1):53–54, jan 2004.

[52] Galbha Duggal, Sharat Warrier, Sabitri Ghimire, Dorien Broekaert, Margot Van der Jeught, Sylvie Lierman, Tom Deroo, Luc Peelman, Ann Van Soom, Ria Cornelissen, Björn Menten, Pieter Mestdagh, Jo Vandesompele, Matthias Roost, Roderick C Slieker, Bastiaan T Heijmans, Dieter Deforce, Petra De Sutter, Susana Chuva De Sousa Lopes, and Björn Heindryckx. Alternative Routes to Induce Naïve Pluripotency in Human Embryonic Stem Cells. *Stem cells (Dayton, Ohio)*, 33(9):2686–2698, sep 2015.

[53] C Ferretti, L Bruni, V Dangles-Marie, A P Pecking, and D Bellet. Molecular circuits shared by placental and cancer cells, and their implications in the proliferative, invasive and migratory capacities of trophoblasts. *Human reproduction update*, 13(2):121–141, 2007.

[54] Susan J Fisher. Why is placentation abnormal in preeclampsia? *American journal of obstetrics and gynecology*, 213(4 Suppl):S115–22, oct 2015.

[55] Simon A Forbes, David Beare, Harry Boutselakis, Sally Bamford, Nidhi Bindal, John Tate, Charlotte G Cole, Sari Ward, Elisabeth Dawson, Laura Ponting, Raymund Stefancsik, Bhavana Harsha, Chai Yin Kok, Mingming Jia, Harry Jubb, Zbyslaw Sondka, Sam Thompson, Tisham De, and Peter J Campbell. COSMIC: somatic cancer genetics at high-resolution. *Nucleic acids research*, 45(D1):D777–D783, jan 2017.

[56] Donald T Fox and Robert J Duronio. Endoreplication and polyploidy: insights into development and disease. *Development (Cambridge, England)*, 140(1):3–12, jan 2013.

[57] Julia Franzen, Theodoros Georgomanolis, Anton Selich, Chao-Chung Kuo, Reinhard Stöger, Lilija Brant, Melita Sara Mulabdić, Eduardo Fernandez-Rebollo, Clara Grezella, Alina Ostrowska, Matthias Begemann, Miloš Nikolić, Björn Rath, Anthony D Ho, Michael Rothe, Axel Schambach, Argyris Papantonis, and Wolfgang Wagner. DNA methylation changes during long-term in vitro cell culture are caused by epigenetic drift. *Communications Biology*, 4(1):598, 2021.

[58] Ohad Gafni, Leehee Weinberger, Abed Alfatah Mansour, Yair S. Manor, Elad Chomsky, Dalit Ben-Yosef, Yael Kalma, Sergey Viukov, Itay Maza, Asaf Zviran, Yoach Rais, Zohar Shipony, Zohar Mukamel, Vladislav Krupalnik, Mirie Zerbib, Shay Geula, Inbal Caspi, Dan Schneir, Tamar Shwartz, Shlomit Gilad, Daniela Amann-Zalcenstein, Sima Benjamin, Ido Amit, Amos Tanay, Rada Massarwa, Noa Novershtern, and Jacob H. Hanna. Derivation of novel human ground state naive pluripotent stem cells. *Nature*, 504(7479):282–286, 2013.

[59] R L Gardner. Origin and differentiation of extraembryonic tissues in the mouse. *International review of experimental pathology*, 24:63–133, 1983.

[60] C Giacinti and A Giordano. RB and cell cycle progression. *Oncogene*, 25(38):5220–5227, aug 2006.

[61] Kai Li Gu, Qiang Zhang, Ying Yan, Ting Ting Li, Fei Fei Duan, Jing Hao, Xi Wen Wang, Ming Shi, Da Ren Wu, Wen Ting Guo, and Yangming Wang. Pluripotency-associated miR-290/302 family of microRNAs promote the dismantling of naive pluripotency. *Cell Research*, 26(3):350–366, 2016.

[62] Chao-Wan Guo, Miho Kawakatsu, Marie Idemitsu, Yoshishige Urata, Shinji Goto, Yusuke Ono, Kimikazu Hamano, and Tao-Sheng Li. Culture under low physiological oxygen conditions improves the stemness and quality of induced pluripotent stem cells. *Journal of cellular physiology*, 228(11):2159–2166, nov 2013.

[63] Ge Guo, Giuliano Giuseppe Stirparo, Stanley E Strawbridge, Daniel Spindlow, Jian Yang, James Clarke, Anish Dattani, Ayaka Yanagida, Meng Amy Li, Sam Myers, Buse Nurten Özel, Jennifer Nichols, and Austin Smith. Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell*, 2021.

[64] Ge Guo, Ferdinand von Meyenn, Fatima Santos, Yaoyao Chen, Wolf Reik, Paul Bertone, Austin Smith, and Jennifer Nichols. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem cell reports*, 6(4):437–446, apr 2016.

[65] T Haaf. The effects of 5-azacytidine and 5-azadeoxycytidine on chromosome structure and function: implications for methylation-associated cellular processes. *Pharmacology therapeutics*, 65(1):19–46, jan 1995.

[66] Roberta L Hannibal and Julie C Baker. Selective Amplification of the Genome Surrounding Key Placental Genes in Trophoblast Giant Cells. *Current biology : CB*, 26(2):230–236, jan 2016.

[67] Roberta L. Hannibal, Edward B. Chuong, Juan Carlos Rivera-Mulia, David M. Gilbert, Anton Valouev, and Julie C. Baker. Copy Number Variation Is a Fundamental Aspect of the Placental Genome. *PLoS Genetics*, 10(5), 2014.

[68] Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M 3rd Mauck, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, Paul Hoffman, Marlon Stoeckius, Efthymia Papalexi, Eleni P Mimitou, Jaison Jain, Avi Srivastava, Tim Stuart, Lamar M Fleming, Bertrand Yeung, Angela J Rogers, Juliana M McElrath, Catherine A Blish, Raphael Gottardo, Peter Smibert, and Rahul Satija. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587.e29, jun 2021.

[69] Seyedeh-Nafiseh Hassani, Mehdi Totonchi, Ali Sharifi-Zarchi, Sepideh Mollamohammadi, Mohammad Pakzad, Sharif Moradi, Azam Samadian, Najmehsadat Masoudi, Shahab Mirshahvaladi, Ali Farrokhi, Boris Greber, Marcos J Araúzo-Bravo, Davood Sabour, Mehdi Sadeghi, Ghasem Hosseini Salekdeh, Hamid Gourabi, Hans R Schöler, and Hossein Baharvand. Inhibition of TGFβ signaling promotes ground state pluripotency. *Stem cell reviews and reports*, 10(1):16–30, feb 2014.

[70] L HAYFLICK. THE LIMITED IN VITRO LIFETIME OF HUMAN DIPLOID CELL STRAINS. *Experimental cell research*, 37:614–636, mar 1965.

[71] Sven Heinz, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C Lin, Peter Laslo, Jason X Cheng, Cornelis Murre, Harinder Singh, and Christopher K Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38(4):576–589, may 2010.

[72] Myriam Hemberger, Courtney W. Hanna, and Wendy Dean. Mechanisms of early placental development in mouse and humans. *Nature Reviews Genetics*, 21(1):27–43, 2020.

[73] Myriam Hemberger, Ramya Udayashankar, Paul Tesar, Harry Moore, and Graham J Burton. ELF5-enforced transcriptional networks define an epigenetically regulated trophoblast stem cell compartment in the human placenta. *Human molecular genetics*, 19(12):2456–2467, jun 2010.

[74] Claudio Hetz and Feroz R Papa. The Unfolded Protein Response and Cell Fate Control. *Molecular cell*, 69(2):169–181, jan 2018.

[75] Mariko Horii, Tony Bui, Ojeni Touma, Hee Young Cho, and Mana M Parast. An Improved Two-Step Protocol for Trophoblast Differentiation of Human Pluripotent Stem Cells. *Current protocols in stem cell biology*, 50(1):e96, sep 2019.

[76] Mariko Horii, Robert Morey, Tony Bui, Ojeni Touma, Katharine K Nelson, Hee-Young Cho, Hannah Rishik, Louise C Laurent, and Mana M Parast. Modeling preeclampsia using human induced pluripotent stem cells. *Scientific reports*, 11(1):5877, mar 2021.

[77] Sheng-Da Hsu, Feng-Mao Lin, Wei-Yun Wu, Chao Liang, Wei-Chih Huang, Wen-Ling Chan, Wen-Ting Tsai, Goun-Zhou Chen, Chia-Jung Lee, Chih-Min Chiu, Chia-Hung Chien, Ming-Chia Wu, Chi-Ying Huang, Ann-Ping Tsou, and Hsien-Da Huang. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic acids research*, 39(Database issue):D163–9, jan 2011.

[78] Zhixing Hu, Hanqin Li, Houbo Jiang, Yong Ren, Xinyang Yu, Jingxin Qiu, Aimee B Stablewski, Boyang Zhang, Michael J Buck, and Jian Feng. Transient inhibition of mTOR in human pluripotent stem cells enables robust formation of mouse-human chimeric embryos. (May):1–17, 2020.

[79] Hsi-Yuan Huang, Yang-Chi-Dung Lin, Jing Li, Kai-Yao Huang, Sirjana Shrestha, Hsiao-Chin Hong, Yun Tang, Yi-Gang Chen, Chen-Nan Jin, Yuan Yu, Jia-Tong Xu, Yue-Ming Li, Xiao-Xuan Cai, Zhen-Yu Zhou, Xiao-Hang Chen, Yuan-Yuan Pei, Liang Hu, Jin-Jiang Su, Shi-Dong Cui, Fei Wang, Yue-Yang Xie, Si-Yuan Ding, Meng-Fan Luo, Chih-Hung Chou, Nai-Wen Chang, Kai-Wen Chen, Yu-Hsiang Cheng, Xin-Hong Wan, Wen-Lian Hsu, Tzong-Yi Lee, Feng-Xiang Wei, and Hsien-Da Huang. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic acids research*, 48(D1):D148–D154, jan 2020.

[80] Zhen Huang, Neng Zhang, Lang Zha, Hong-Chao Mao, Xuan Chen, Ji-Feng Xiang, Hua Zhang, and Zi-Wei Wang. Aberrant expression of the autocrine motility factor receptor correlates with poor prognosis and promotes metastasis in gastric carcinoma. *Asian Pacific journal of cancer prevention : APJCP*, 15(2):989–997, 2014.

[81] Tim Hulsen, Jacob de Vlieg, and Wynand Alkema. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC genomics*, 9:488, oct 2008.

[82] J Hustin, E Jauniaux, and J P Schaaps. Histological study of the materno-embryonic interface in spontaneous abortion. *Placenta*, 11(6):477–486, 1990.

[83] Kassie J Hyde and Danny J Schust. Genetic considerations in recurrent pregnancy loss. *Cold Spring Harbor perspectives in medicine*, 5(3):a023119, feb 2015.

[84] Shingo Io, Mio Kabata, Yoshiki Iemura, Katsunori Semi, Nobuhiro Morone, Atsutaka Minagawa, Bo Wang, Ikuhiro Okamoto, Tomonori Nakamura, Yoji Kojima, Chizuru Iwatani, Hideaki Tsuchiya, Belinda Kaswandy, Eiji Kondoh, Shin Kaneko, Knut Woltjen, Mitinori Saitou, Takuya Yamamoto, Masaki Mandai, and Yasuhiro Takashima. Capturing human trophoblast development with naive pluripotent stem cells ¡em¿inxa0;vitro¡/em¿. *Cell Stem Cell*, 28(6):1023–1039.e13, jun 2021.

[85] Ikuo Ishige, Tokiko Nagamura-Inoue, Masaki J Honda, Ratanakanit Harnprasopwat, Michiko Kido, Mitsuhiro Sugimoto, Hiromitsu Nakauchi, and Arinobu Tojo. Comparison of mesenchymal stem cells derived from arterial, venous, and Wharton's jelly explants of human umbilical cord. *International journal of hematology*, 90(2):261–269, sep 2009.

[86] Ashish Jain and Geetu Tuteja. PlacentaCellEnrich: A tool to characterize gene sets using placenta cell-specific gene enrichment analysis. *Placenta*, 103:164–171, jan 2021.

[87] Jean-Louis Janneau, Juan Maldonado-Estrada, Gérard Tachdjian, Isabelle Miran, Nelly Motté, Patrick Saulnier, Jean-Christophe Sabourin, Jean-François Coté, Bénédicte Simon, René Frydman, Gérard Chaouat, and Dominique Bellet. Transcriptional expression of genes involved in cell invasion and migration by normal and tumoral trophoblast cells. *The Journal of clinical endocrinology and metabolism*, 87(11):5336–5339, nov 2002.

[88] E Jauniaux, A L Watson, J Hempstock, Y P Bao, J N Skepper, and G J Burton. Onset of maternal arterial blood flow and placental oxidative stress. A possible factor in human early pregnancy failure. *The American journal of pathology*, 157(6):2111–2122, dec 2000.

[89] Caroline A Jefferies. Regulating IRFs in IFN Driven Disease. *Frontiers in immunology*, 10:325, 2019.

[90] Wen G Jiang, Avraham Raz, Anthony Douglas-Jones, and Robert E Mansel. Expression of autocrine motility factor (AMF) and its receptor, AMFR, in human breast cancer. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 54(2):231–241, feb 2006.

[91] Yoshikazu Johmura and Makoto Nakanishi. Multiple facets of p53 in senescence induction and maintenance. *Cancer Science*, 107(11):1550–1555, 2016.

[92] So Ri Jung, Thomas M Ashhurst, Phillip K West, Barney Viengkhou, Nicholas J C King, Iain L Campbell, and Markus J Hofer. Contribution of STAT1 to innate and adaptive immunity during type I interferon-mediated lethal virus infection. *PLoS pathogens*, 16(4):e1008525, apr 2020.

[93] Laura Kasak, Kristiina Rull, Pille Vaas, Pille Teesalu, and Maris Laan. Extensive load of somatic CNVs in the human placenta. *Scientific reports*, 5:8342, feb 2015.

[94] Tim Kehl, Fabian Kern, Christina Backes, Tobias Fehlmann, Daniel Stöckel, Eckart Meese, Hans-Peter Lenhof, and Andreas Keller. miRPathDB 2.0: a novel release of the miRNA Pathway Dictionary Database. *Nucleic acids research*, 48(D1):D142–D147, jan 2020.

[95] T Y Khong, F De Wolf, W B Robertson, and I Brosens. Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and by small-for-gestational age infants. *British journal of obstetrics and gynaecology*, 93(10):1049–1059, oct 1986.

[96] T Y Khong, H S Liddell, and W B Robertson. Defective haemochorial placentation as a cause of miscarriage: a preliminary study. *British journal of obstetrics and gynaecology*, 94(7):649–655, jul 1987.

[97] N K Khoo, J F Bechberger, T Shepherd, S L Bond, K R McCrae, G S Hamilton, and P K Lala. SV40 Tag transformation of the normal invasive trophoblast results in a premalignant phenotype. I. Mechanisms responsible for hyperinvasiveness and resistance to anti-invasive action of TGFbeta. *International journal of cancer*, 77(3):429–439, jul 1998.

[98] Cody Kime, Masayo Sakaki-Yumoto, Leeanne Goodrich, Yohei Hayashi, Salma Sami, Rik Derynck, Michio Asahi, Barbara Panning, Shinya Yamanaka, and Kiichiro Tomoda. Autotaxin-mediated lipid signaling intersects with LIF and BMP signaling to promote the naive pluripotency transcription factor program. *Proceedings of the National Academy of Sciences of the United States of America*, 113(44):12478–12483, nov 2016.

[99] Martin Knöfler, Sandra Haider, Leila Saleh, Jürgen Pollheimer, Teena K J B Gamage, and Joanna James. Human placenta and trophoblast development: key molecular mechanisms and model systems. *Cellular and molecular life sciences : CMLS*, 76(18):3479–3496, sep 2019.

[100] Gennady Korotkevich, Vladimir Sukhov, Nikolay Budin, Boris Shpak, Maxim N Artyomov, and Alexey Sergushichev. Fast gene set enrichment analysis. *bioRxiv*, 2021.

[101] Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, Michael G McDermott, Caroline D Monteiro, Gregory W Gundersen, and Avi Ma'ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1):W90–W97, may 2016.

[102] Dorota Kurek, Alex Neagu, Melodi Tastemel, Nesrin Tüysüz, Johannes Lehmann, Harmen J G van de Werken, Sjaak Philipsen, Reinier van der Linden, Alex Maas, Wilfred F J van IJcken, Micha Drukker, and Derk Ten Berge. Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem cell reports*, 4(1):114–128, jan 2015.

[103] Irina Kuznetsova, Artur Lugmayr, Stefan J Siira, Oliver Rackham, and Aleksandra Filipovska. CirGO: an alternative circular way of visualising gene ontology terms. *BMC bioinformatics*, 20(1):84, feb 2019.

[104] Erika M Kwon, John P Connelly, Nancy F Hansen, Frank X Donovan, Thomas Winkler, Brian W Davis, Halah Alkadi, Settara C Chandrasekharappa, Cynthia E Dunbar, James C Mullikin, and Paul Liu. iPSCs and fibroblast subclones from the same fibroblast population contain comparable levels of sequence variations. *Proceedings of the National Academy of Sciences of the United States of America*, 114(8):1964–1969, feb 2017.

[105] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, 2012.

[106] Louise C Laurent, Igor Ulitsky, Ileana Slavin, Ha Tran, Andrew Schork, Robert Morey, Candace Lynch, Julie V Harness, Sunray Lee, Maria J Barrero, Sherman Ku, Marina Martynova, Ruslan Semechkin, Vasiliy Galat, Joel Gottesfeld, Juan Carlos Izpisua Belmonte, Chuck Murry, Hans S Keirstead, Hyun-Sook Park, Uli Schmidt, Andrew L Laslett, Franz-Josef Muller, Caroline M Nievergelt, Ron Shamir, and Jeanne F Loring. Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell stem cell*, 8(1):106–118, jan 2011.

[107] Cheryl Q.E. Lee, Lucy Gardner, Margherita Turco, Nancy Zhao, Matthew J. Murray, Nicholas Coleman, Janet Rossant, Myriam Hemberger, and Ashley Moffett. What Is Trophoblast? A Combination of Criteria Define Human First-Trimester Trophoblast. *Stem Cell Reports*, 6(2):257–272, 2016.

[108] Jong Hee Lee, Sarah Laronde, Tony J. Collins, Zoya Shapovalova, Borko Tanasijevic, Jamie D. McNicol, Aline Fiebig-Comyn, Yannick D. Benoit, Jung Bok Lee, Ryan R. Mitchell, and Mickie Bhatia. Lineage-Specific Differentiation Is Influenced by State of Human Pluripotency. *Cell Reports*, 19(1):20–35, 2017.

[109] Naeun Lee, Donghyun Kim, and Wan-Uk Kim. Role of NFAT5 in the Immune System and Pathogenesis of Autoimmune Diseases. *Frontiers in immunology*, 10:270, 2019.

[110] Nathalie Lefort, Maxime Feyeux, Cécile Bas, Olivier Féraud, Annelise Bennaceur-Griscelli, Gerard Tachdjian, Marc Peschanski, and Anselme L Perrier. Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. *Nature biotechnology*, 26(12):1364–1366, dec 2008.

[111] Monkol Lek, Konrad J Karczewski, Eric V Minikel, Kaitlin E Samocha, Eric Banks, Timothy Fennell, Anne H O'Donnell-Luria, James S Ware, Andrew J Hill, Beryl B Cummings, Taru Tukiainen, Daniel P Birnbaum, Jack A Kosmicki, Laramie E Duncan, Karol Estrada, Fengmei Zhao, James Zou, Emma Pierce-Hoffman, Joanne Berghout, David N Cooper, Nicole Deflaux, Mark DePristo, Ron Do, Jason Flannick, Menachem Fromer, Laura Gauthier, Jackie Goldstein, Namrata Gupta, Daniel Howrigan, Adam Kiezun, Mitja I

Kurki, Ami Levy Moonshine, Pradeep Natarajan, Lorena Orozco, Gina M Peloso, Ryan Poplin, Manuel A Rivas, Valentin Ruano-Rubio, Samuel A Rose, Douglas M Ruderfer, Khalid Shakir, Peter D Stenson, Christine Stevens, Brett P Thomas, Grace Tiao, Maria T Tusie-Luna, Ben Weisburd, Hong-Hee Won, Dongmei Yu, David M Altshuler, Diego Ardissino, Michael Boehnke, John Danesh, Stacey Donnelly, Roberto Elosua, Jose C Florez, Stacey B Gabriel, Gad Getz, Stephen J Glatt, Christina M Hultman, Sekar Kathiresan, Markku Laakso, Steven McCarroll, Mark I McCarthy, Dermot McGovern, Ruth McPherson, Benjamin M Neale, Aarno Palotie, Shaun M Purcell, Danish Saleheen, Jeremiah M Scharf, Pamela Sklar, Patrick F Sullivan, Jaakko Tuomilehto, Ming T Tsuang, Hugh C Watkins, James G Wilson, Mark J Daly, Daniel G MacArthur, and Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616):285–291, 2016.

[112] Yingchun Li, Matteo Moretto-Zita, Francesca Soncin, Anna Wakeland, Lynlee Wolfe, Sandra Leon-Garcia, Raj Pandian, Donald Pizzo, Li Cui, Kristopher Nazor, Jeanne F Loring, Christopher P Crum, Louise C Laurent, and Mana M Parast. BMP4-directed trophoblast differentiation of human embryonic stem cells is mediated through a ΔNp63+ cytotrophoblast stem cell state. *Development (Cambridge, England)*, 140(19):3965–3976, oct 2013.

[113] Yang Liao, Gordon K Smyth, and Wei Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, apr 2014.

[114] Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P Mesirov, and Pablo Tamayo. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems*, 1(6):417–425, dec 2015.

[115] Jia Liu, Shengnan Hao, Xi Chen, Hui Zhao, Lutao Du, Hanxiao Ren, Chuanxin Wang, and Haiting Mao. Human placental trophoblast cells contribute to maternal-fetal tolerance through expressing IL-35 and mediating iT(R)35 conversion. *Nature communications*, 10(1):4601, oct 2019.

[116] Xiaodong Liu, John F. Ouyang, Fernando J. Rossello, Jia Ping Tan, Kathryn C. Davidson, Daniela S. Valdes, Jan Schröder, Yu B.Y. Sun, Joseph Chen, Anja S. Knaupp, Guizhi Sun, Hun S. Chy, Ziyi Huang, Jahnvi Pflueger, Jaber Firas, Vincent Tano, Sam Buckberry, Jacob M. Paynter, Michael R. Larcombe, Daniel Poppe, Xin Yi Choo, Carmel M. O'Brien, William A. Pastor, Di Chen, Anna L. Leichter, Haroon Naeem, Pratibha Tripathi, Partha P. Das, Alexandra Grubman, David R. Powell, Andrew L. Laslett, Laurent David, Susan K. Nilsson, Amander T. Clark, Ryan Lister, Christian M. Nefzger, Luciano G. Martelotto, Owen J.L. Rackham, and Jose M. Polo. Reprogramming roadmap reveals route to human induced trophoblast stem cells. *Nature*, 586(7827):101–107, 2020.

[117] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, 2014.

[118] Pavlo Lutsik, Lars Feuerbach, Julia Arand, Thomas Lengauer, Jörn Walter, and Christoph Bock. BiQ Analyzer HT: locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing. *Nucleic acids research*, 39(Web Server issue):W551–6, jul 2011.

[119] Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, John J Trombetta, David A Weitz, Joshua R Sanes, Alex K Shalek, Aviv Regev, and Steven A McCarroll. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, 161(5):1202–1214, may 2015.

[120] A Malassiné, J L Frendo, and D Evain-Brion. A comparison of placental development and endocrine functions between the human and mouse model. *Human reproduction update*, 9(6):531–539, 2003.

[121] Sarah A. Marshall, Natalie J. Hannan, Maria Jelinic, Thy P.H. Nguyen, Jane E. Girling, and Laura J. Parry. Animal models of preeclampsia: translational failings and why. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 314(4):R499–R508, 2018.

[122] Jose A Martina, Yong Chen, Marjan Gucek, and Rosa Puertollano. MTORC1 functions as a transcriptional regulator of autophagy by preventing nuclear transport of TFEB. *Autophagy*, 8(6):903–914, jun 2012.

[123] Kristen Martins-Taylor, Benjamin S Nisler, Seth M Taapken, Tiwanna Compton, Leann Crandall, Karen Dyer Montgomery, Marc Lalande, and Ren-He Xu. Recurrent copy number variations in human induced pluripotent stem cells., jun 2011.

[124] Yoav Mayshar, Uri Ben-David, Neta Lavon, Juan-Carlos Biancotti, Benjamin Yakir, Amander T Clark, Kathrin Plath, William E Lowry, and Nissim Benvenisty. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell stem cell*, 7(4):521–531, oct 2010.

[125] M T McMaster, C L Librach, Y Zhou, K H Lim, M J Janatpour, R DeMars, S Kovats, C Damsky, and S J Fisher. Human placental HLA-G expression is restricted to differentiated cytotrophoblasts. *Journal of immunology (Baltimore, Md. : 1950)*, 154(8):3771–3778, apr 1995.

[126] G Meinhardt, S Kaltenberger, C Fiala, M Knöfler, and J Pollheimer. ERBB2 gene amplification increases during the transition of proximal EGFR(+) to distal HLA-G(+) first trimester cell column trophoblasts. *Placenta*, 36(8):803–808, aug 2015.

[127] Florian T Merkle, Sulagna Ghosh, Nolan Kamitaki, Jana Mitchell, Yishai Avior, Curtis Mello, Seva Kashin, Shila Mekhoubad, Dusko Ilic, Maura Charlton, Genevieve Saphier, Robert E Handsaker, Giulio Genovese, Shiran Bar, Nissim Benvenisty, Steven A McCarroll, and Kevin Eggan. Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature*, 545(7653):229–233, may 2017.

[128] Tobias Messmer, Ferdinand von Meyenn, Aurora Savino, Fátima Santos, Hisham Mohammed, Aaron Tin Long Lun, John C. Marioni, and Wolf Reik. Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution. *Cell Reports*, 26(4):815–824.e4, 2019.

[129] Masahito Mizuuchi, Tereza Cindrova-Davies, Matts Olovsson, D Stephen Charnock-Jones, Graham J Burton, and Hong Wa Yung. Placental endoplasmic reticulum stress negatively regulates transcription of placental growth factor via ATF4 and ATF6β: implications for the pathophysiology of human pregnancy complications. *The Journal of pathology*, 238(4):550–561, mar 2016.

[130] Thomas Moerman, Sara Aibar Santos, Carmen Bravo González-Blas, Jaak Simm, Yves Moreau, Jan Aerts, and Stein Aerts. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics (Oxford, England)*, 35(12):2159–2161, jun 2019.

[131] María de Lourdes Mora-García, Alfonso Duenas-González, Jorge Hernández-Montes, Erick De la Cruz-Hernández, Enrique Pérez-Cárdenas, Benny Weiss-Steider, Edelmiro Santiago-Osorio, Vianney Francisco Ortíz-Navarrete, Víctor Hugo Rosales, David Cantú, Marcela Lizano-Soberón, Martha Patricia Rojo-Aguilar, and Alberto Monroy-García. Up-regulation of HLA class-I antigen expression and antigen-specific CTL response in cervical cancer cells by the demethylating agent hydralazine and the histone deacetylase inhibitor valproic acid. *Journal of translational medicine*, 4:55, dec 2006.

[132] Robert Morey, Omar Farah, Sampada Kallol, Daniela F. Requena, Morgan Meads, Matteo Moretto-Zita, Francesca Soncin, Louise C. Laurent, and Mana M. Parast. Transcriptomic Drivers of Differentiation, Maturation, and Polyploidy in Human Extravillous Trophoblast. *Frontiers in Cell and Developmental Biology*, 9(September):1–24, 2021.

[133] Akio Morinobu, Massimo Gadina, Warren Strober, Roberta Visconti, Albert Fornace, Cristina Montagna, Gerald M Feldman, Ryuta Nishikomori, and John J O'Shea. STAT4 serine phosphorylation is critical for IL-12-induced IFN-gamma production but not for cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America*, 99(19):12281–12286, sep 2002.

[134] Carla Mulas, Tüzer Kalkan, and Austin Smith. NODAL Secures Pluripotency upon Embryonic Stem Cell Progression from the Ground State. *Stem Cell Reports*, 9(1):77–91, 2017.

[135] A Nagy, J Rossant, R Nagy, W Abramow-Newerly, and J C Roder. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 90(18):8424–8428, sep 1993.

[136] Kosei Nakajima and Avraham Raz. Autocrine motility factor and its receptor expression in musculoskeletal tumors. *Journal of bone oncology*, 24:100318, oct 2020.

[137] Colm E Nestor, Raffaele Ottaviano, Diana Reinhardt, Hazel A Cruickshanks, Heidi K Mjoseng, Rhoanne C McPherson, Antonio Lentini, John P Thomson, Donncha S Dunican, Sari Pennings, Stephen M Anderton, Mikael Benson, and Richard R Meehan. Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems. *Genome Biology*, 16(1):11, 2015.

[138] H T Nguyen, M Geens, A Mertzanidou, K Jacobs, C Heirman, K Breckpot, and C Spits. Gain of 20q11.21 in human embryonic stem cells improves cell survival by increased expression of Bcl-xL. *Molecular human reproduction*, 20(2):168–177, feb 2014.

[139] Kathy K Niakan, Jinnuo Han, Roger A Pedersen, Carlos Simon, and Renee A Reijo Pera. Human pre-implantation embryo development. *Development (Cambridge, England)*, 139(5):829–841, mar 2012.

[140] J Nichols and R L Gardner. Heterogeneous differentiation of external cells in individual isolated early mouse inner cell masses in culture. *Journal of embryology and experimental morphology*, 80:225–240, apr 1984.

[141] Jennifer Nichols and Austin Smith. Pluripotency in the embryo and in culture. *Cold Spring Harbor perspectives in biology*, 4(8):a008128, aug 2012.

[142] Hiroaki Okae, Hidehiro Toh, Tetsuya Sato, Hitoshi Hiura, Sota Takahashi, Kenjiro Shirane, Yuka Kabayama, Mikita Suyama, Hiroyuki Sasaki, and Takahiro Arima. Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell*, 22(1):50–63.e6, 2018.

[143] Ikuhiro Okamoto, Catherine Patrat, Dominique Thépot, Nathalie Peynot, Patricia Fauque, Nathalie Daniel, Patricia Diabangouaya, Jean-Philippe Wolf, Jean-Paul Renard, Véronique Duranthon, and Edith Heard. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature*, 472(7343):370–374, apr 2011.

[144] Anna Osnato, Stephanie Brown, Christel Krueger, Simon Andrews, Amanda J. Collier, Shota Nakanoh, Mariana Quiroga Londoño, Brandon T. Wesley, Daniele Muraro, A. Sophie Brumm, Kathy K. Niakan, Ludovic Vallier, Daniel Ortmann, and Peter J. Rugg-Gunn. Tgfβ signalling is required to maintain pluripotency of human naïve pluripotent stem cells. *eLife*, 10:1–28, 2021.

[145] Henrieta Papuchova, Sarika Kshirsagar, Lily Xu, Hannah A Bougleux Gomes, Qin Li, Vidya Iyer, Errol R Norwitz, Jack L Strominger, and Tamara Tilburgs. Three types of HLA-G+ extravillous trophoblasts that have distinct immune regulatory properties. *Proceedings of the National Academy of Sciences of the United States of America*, 117(27):15772–15777, jul 2020.

[146] William A. Pastor, Di Chen, Wanlu Liu, Rachel Kim, Anna Sahakyan, Anastasia Lukianchikov, Kathrin Plath, Steven E. Jacobsen, and Amander T. Clark. Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell Stem Cell*, 18(3):323–329, 2016.

[147] Anoop P Patel, Itay Tirosh, John J Trombetta, Alex K Shalek, Shawn M Gillespie, Hiroaki Wakimoto, Daniel P Cahill, Brian V Nahed, William T Curry, Robert L Martuza, David N Louis, Orit Rozenblatt-Rosen, Mario L Suvà, Aviv Regev, and Bradley E Bernstein. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (New York, N.Y.)*, 344(6190):1396–1401, jun 2014.

[148] Sophie Petropoulos, Daniel Edsgärd, Björn Reinius, Qiaolin Deng, Sarita Pauliina Panula, Simone Codeluppi, Alvaro Plaza Reyes, Sten Linnarsson, Rickard Sandberg, and Fredrik Lanner. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell*, 165(4):1012–1026, may 2016.

[149] Ruth Pidsley, Elena Zotenko, Timothy J Peters, Mitchell G Lawrence, Gail P Risbridger, Peter Molloy, Susan Van Djik, Beverly Muhlhausler, Clare Stirzaker, and Susan J Clark. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology*, 17(1):208, 2016.

[150] R Pijnenborg, G Dixon, W B Robertson, and I Brosens. Trophoblastic invasion of human decidua from 8 to 18 weeks of pregnancy. *Placenta*, 1(1):3–19, 1980.

[151] R Pijnenborg, L Vercruysse, and M Hanssens. The uterine spiral arteries in human pregnancy: facts and controversies. *Placenta*, 27(9-10):939–958, 2006.

[152] Roger Pique-Regi, Roberto Romero, Adi L. Tarca, Edward D. Sendler, Yi Xu, Valeria Garcia-Flores, Yaozhu Leng, Francesca Luca, Sonia S. Hassan, and Nardhy Gomez-Lopez. Single cell transcriptional signatures of the human placenta in term and preterm parturition. *eLife*, 8:1–22, 2019.

[153] Olivier Pluquet, Albin Pourtier, and Corinne Abbadie. The unfolded protein response and cellular senescence. A review in the theme: cellular mechanisms of endoplasmic reticulum stress signaling in health and disease. *American journal of physiology. Cell physiology*, 308(6):C415–25, mar 2015.

[154] J Pollheimer and M Knöfler. Signalling pathways regulating the invasive differentiation of human trophoblasts: a review. *Placenta*, 26 Suppl A:S21–30, apr 2005.

[155] Jürgen Pollheimer, Sigrid Vondra, Jennet Baltayeva, Alexander Guillermo Beristain, and Martin Knöfler. Regulation of Placental Extravillous Trophoblasts by the Maternal Uterine Environment. *Frontiers in immunology*, 9:2597, 2018.

[156] Peter Priestley, Jonathan Baber, Martijn P Lolkema, Neeltje Steeghs, Ewart de Bruijn, Charles Shale, Korneel Duyvesteyn, Susan Haidari, Arne van Hoeck, Wendy Onstenk, Paul Roepman, Mircea Voda, Haiko J Bloemendal, Vivianne C G Tjan-Heijnen, Carla M L van Herpen, Mariette Labots, Petronella O Witteveen, Egbert F Smit, Stefan Sleijfer, Emile E Voest, and Edwin Cuppen. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature*, 575(7781):210–216, nov 2019.

[157] Han Qin, Miroslav Hejna, Yanxia Liu, Michelle Percharde, Mark Wossidlo, Laure Blouin, Jens Durruthy-Durruthy, Priscilla Wong, Zhongxia Qi, Jingwei Yu, Lei S Qi, Vittorio Sebastiano, Jun S Song, and Miguel Ramalho-Santos. YAP Induces Human Naive Pluripotency. *Cell reports*, 14(10):2301–2312, mar 2016.

[158] Leonardo Romorini, Ximena Garate, Gabriel Neiman, Carlos Luzzani, Verónica Alejandra Furmento, Alejandra Sonia Guberman, Gustavo Emilio Sevlever, María Elida Scassa, and Santiago Gabriel Miriuka. AKT/GSK3β signaling pathway is critically involved in human pluripotent stem cell survival. *Scientific reports*, 6:35660, oct 2016.

[159] Matthias S Roost, Liesbeth van Iperen, Yavuz Ariyurek, Henk P Buermans, Wibowo Arindrarto, Harsha D Devalla, Robert Passier, Christine L Mummery, Françoise Carlotti, Eelco J P de Koning, Erik W van Zwet, Jelle J Goeman, and Susana M Chuva de Sousa Lopes. KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas. *Stem cell reports*, 4(6):1112–1124, jun 2015.

[160] Janet Rossant. Human embryology: Implantation barrier overcome. *Nature*, 533(7602):182–183, may 2016.

[161] Joel Rozowsky, Robert R Kitchen, Jonathan J Park, Timur R Galeev, James Diao, Jonathan Warrell, William Thistlethwaite, Sai L Subramanian, Aleksandar Milosavljevic, and Mark Gerstein. exceRpt: A Comprehensive Analytic Platform for Extracellular RNA Profiling. *Cell systems*, 8(4):352–357.e3, apr 2019.

[162] Anna Sahakyan, Kathrin Plath, and Claire Rougeulle. Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 372(1733), nov 2017.

[163] Laurent Sansregret and Charles Swanton. The Role of Aneuploidy in Cancer Evolution. *Cold Spring Harbor perspectives in medicine*, 7(1), jan 2017.

[164] Andrew J Sharp, Elisavet Stathaki, Eugenia Migliavacca, Manisha Brahmachary, Stephen B Montgomery, Yann Dupre, and Stylianos E Antonarakis. DNA methylation profiles of human active and inactive X chromosomes. *Genome research*, 21(10):1592–1600, oct 2011.

[165] J W Shay, O M Pereira-Smith, and W E Wright. A role for both RB and p53 in the regulation of human cellular senescence. *Experimental cell research*, 196(1):33–39, sep 1991.

[166] Austin Smith. Formative pluripotency: The executive phase in a developmental continuum. *Development (Cambridge)*, 144(3):365–373, 2017.

[167] Zachary D Smith, Michelle M Chan, Kathryn C Humm, Rahul Karnik, Shila Mekhoubad, Aviv Regev, Kevin Eggan, and Alexander Meissner. DNA methylation dynamics of the human preimplantation embryo. *Nature*, 511(7511):611–615, jul 2014.

[168] Francesca Soncin, Marwa Khater, Cuong To, Donald Pizzo, Omar Farah, Anna Wakeland, Kanaga Arul Nambi Rajan, Katharine K Nelson, Ching-Wen Chang, Matteo Moretto-Zita, David R Natale, Louise C Laurent, and Mana M Parast. Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development. *Development (Cambridge, England)*, 145(2), jan 2018.

[169] Francesca Soncin, David Natale, and Mana M Parast. Signaling pathways in mouse and human trophoblast differentiation: a comparative review. *Cellular and molecular life sciences : CMLS*, 72(7):1291–1302, apr 2015.

[170] Henrik Sperber, Julie Mathieu, Yuliang Wang, Amy Ferreccio, Zhuojin Xu, Karin A Fischer, Arikketh Devi, Damien Detraux, Stephanie L Battle, Megan Showalter, Cristina Valensisi, Jason H Bielas, Nolan G Ericson, Lilyana Margaretha, Aaron M Robitaille, Daciana Margineantu, Oliver Fiehn, David Hockenbery, C Anthony Blau, Daniel Raftery, R David Hawkins, Randall T Moon, Carol B Ware, Hannele Ruohola, Oregon Health, and Hutchinson Cancer. HHS Public Access. 17(12):1523–1535, 2016.

[171] Giuliano Giuseppe Stirparo, Austin Smith, and Ge Guo. Cancer-Related Mutations Are Not Enriched in Naive Human Pluripotent Stem Cells. *Cell stem cell*, 28(1):164–169.e2, jan 2021.

[172] Stephen C Strom and Roberto Gramignoli. Human amnion epithelial cells expressing HLA-G as novel cell-based treatment for liver disease. *Human immunology*, 77(9):734–739, sep 2016.

[173] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, and Jill P Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545 LP – 15550, oct 2005.

[174] Fran Supek, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one*, 6(7):e21800, 2011.

[175] Seth M Taapken, Benjamin S Nisler, Michael A Newton, Tori L Sampsell-Barron, Kimberly A Leonhard, Erik M McIntire, and Karen D Montgomery. Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells., apr 2011.

[176] Robi Tacutu, Daniel Thornton, Emily Johnson, Arie Budovsky, Diogo Barardo, Thomas Craig, Eugene Diana, Gilad Lehmann, Dmitri Toren, Jingwei Wang, Vadim E Fraifeld, and João P de Magalhães. Human Ageing Genomic Resources: new and updated databases. *Nucleic acids research*, 46(D1):D1083–D1090, jan 2018.

[177] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–676, aug 2006.

[178] Yasuhiro Takashima, Ge Guo, Remco Loos, Jennifer Nichols, Gabriella Ficz, Felix Krueger, David Oxley, Fatima Santos, James Clarke, William Mansfield, Wolf Reik, Paul Bertone, and Austin Smith. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*, 158(6):1254–1269, sep 2014.

[179] Eric Talevich and A. Hunter Shain. CNVkit-RNA: Copy number inference from RNA-Sequencing data. *bioRxiv*, page 408534, 2018.

[180] Walfred W C Tang, Sabine Dietmann, Naoko Irie, Harry G Leitch, Vasileios I Floros, Charles R Bradshaw, Jamie A Hackett, Patrick F Chinnery, and M Azim Surani. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell*, 161(6):1453–1467, jun 2015.

[181] B P Telugu, K Adachi, J M Schlitt, T Ezashi, D J Schust, R M Roberts, and L C Schulz. Comparison of extravillous trophoblast cells derived from human embryonic stem cells and from first trimester human placentas. *Placenta*, 34(7):536–543, jul 2013.

[182] Thorold W. Theunissen, Marc Friedli, Yupeng He, Evarist Planet, Ryan C. O'Neil, Styliani Markoulaki, Julien Pontis, Haoyi Wang, Alexandra Iouranova, Michaël Imbeault, Julien Duc, Malkiel A. Cohen, Katherine J. Wert, Rosa Castanon, Zhuzhu Zhang, Yanmei Huang, Joseph R. Nery, Jesse Drotar, Tenzin Lungjangwa, Didier Trono, Joseph R. Ecker, and Rudolf Jaenisch. Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell*, 19(4):502–515, 2016.

[183] Thorold W Theunissen, Benjamin E Powell, Haoyi Wang, Maya Mitalipova, Dina A Faddah, Jessica Reddy, Zi Peng Fan, Dorothea Maetzel, Kibibi Ganz, Linyu Shi, Tenzin Lungjangwa, Sumeth Imsoonthornruksa, Yonatan Stelzer, Sudharshan Rangarajan, Ana D'Alessio, Jianming Zhang, Qing Gao, Meelad M Dawlaty, Richard A Young, Nathanael S Gray, and Rudolf Jaenisch. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell stem cell*, 15(4):471–487, oct 2014.

[184] J A Thomson, J Itskovitz-Eldor, S S Shapiro, M A Waknitz, J J Swiergiel, V S Marshall, and J M Jones. Embryonic stem cell lines derived from human blastocysts. *Science (New York, N.Y.)*, 282(5391):1145–1147, nov 1998.

[185] Tamara Tilburgs, Ângela C Crespo, Anita van der Zwan, Basya Rybalov, Towfique Raj, Barbara Stranger, Lucy Gardner, Ashley Moffett, and Jack L Strominger. Human HLA-G+ extravillous trophoblasts: Immune-activating cells that interact with decidual leukocytes. *Proceedings of the National Academy of Sciences of the United States of America*, 112(23):7219–7224, jun 2015.

[186] Kiichiro Tomoda, Haiming Hu, Yoshiki Sahara, Hashimita Sanyal, Minoru Takasato, and Cody Kime. Reprogramming epiblast stem cells into pre-implantation blastocyst cell-like cells. *Stem cell reports*, 16(5):1197–1209, may 2021.

[187] T Touboul, S Chen, C C To, S Mora-Castilla, K Sabatini, R H Tukey, and L C Laurent. Stage-specific regulation of the WNT/beta-catenin pathway enhances differentiation of hESCs into hepatocytes. *J Hepatol*, 64(6):1315–1326, 2016.

[188] Brett Trost, Susan Walker, Zhuozhi Wang, Bhooma Thiruvahindrapuram, Jeffrey R MacDonald, Wilson W L Sung, Sergio L Pereira, Joe Whitney, Ada J S Chan, Giovanna Pellecchia, Miriam S Reuter, Si Lok, Ryan K C Yuen, Christian R Marshall, Daniele Merico, and Stephen W Scherer. A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. *American journal of human genetics*, 102(1):142–155, jan 2018.

[189] Yien Che Tsai, Arnulfo Mendoza, Jennifer M Mariano, Ming Zhou, Zlatka Kostova, Bo Chen, Timothy Veenstra, Stephen M Hewitt, Lee J Helman, Chand Khanna, and Allan M Weissman. The ubiquitin ligase gp78 promotes sarcoma metastasis by targeting KAI1 for degradation. *Nature medicine*, 13(12):1504–1509, dec 2007.

[190] Jason C H Tsang, Joaquim S L Vong, Lu Ji, Liona C Y Poon, Peiyong Jiang, Kathy O Lui, Yun-Bi Ni, Ka Fai To, Yvonne K Y Cheng, Rossa W K Chiu, and Yuk Ming Dennis Lo. Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proceedings of the National Academy of Sciences*, 114(37):E7786—-E7795, 2017.

[191] Margherita Y. Turco and Ashley Moffett. Development of the human placenta. *Development (Cambridge)*, 146(22):1–14, 2019.

[192] Zakir Ullah, Matthew J Kohn, Rieko Yagi, Lyubomir T Vassilev, and Melvin L DePamphilis. Differentiation of trophoblast stem cells into giant cells is triggered by p57/Kip2 inhibition of CDK1 activity. *Genes development*, 22(21):3024–3036, nov 2008.

[193] Zakir Ullah, Chrissie Y Lee, Mary A Lilly, and Melvin L DePamphilis. Developmentally programmed endoreduplication in animals. *Cell cycle (Georgetown, Tex.)*, 8(10):1501–1509, may 2009.

[194] Ludovic Vallier, Thomas Touboul, Stephanie Brown, Candy Cho, Bilada Bilican, Morgan Alexander, Jessica Cedervall, Siddharthan Chandran, Lars Ahrlund-Richter, Anne Weber, and Roger A Pedersen. Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. *Stem cells (Dayton, Ohio)*, 27(11):2655–2666, nov 2009.

[195] Geraldine A Van der Auwera, Mauricio O Carneiro, Christopher Hartl, Ryan Poplin, Guillermo Del Angel, Ami Levy-Moonshine, Tadeusz Jordan, Khalid Shakir, David Roazen, Joel Thibault, Eric Banks, Kiran V Garimella, David Altshuler, Stacey Gabriel, and Mark A DePristo. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics*, 43(1110):11.10.1–11.10.33, 2013.

[196] Peter Van Loo, Silje H Nordgard, Ole Christian Lingjærde, Hege G Russnes, Inga H Rye, Wei Sun, Victor J Weigman, Peter Marynen, Anders Zetterberg, Bjørn Naume, Charles M Perou, Anne-Lise Børresen-Dale, and Vessela N Kristensen. Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 107(39):16910–16915, sep 2010.

[197] Philipp Velicky, Gudrun Meinhardt, Kerstin Plessl, Sigrid Vondra, Tamara Weiss, Peter Haslinger, Thomas Lendl, Karin Aumayr, Mario Mairhofer, Xiaowei Zhu, Birgit Schütz, Roberta L. Hannibal, Robert Lindau, Beatrix Weil, Jan Ernerudh, Jürgen Neesen, Gerda Egger, Mario Mikula, Clemens Röhrl, Alexander E. Urban, Julie Baker, Martin Knöfler, and Jürgen Pollheimer. Genome amplification and cellular senescence are hallmarks of human placenta development. *PLoS Genetics*, 14(10):1–25, 2018.

[198] Roser Vento-Tormo, Mirjana Efremova, Rachel A. Botting, Margherita Y. Turco, Miquel Vento-Tormo, Kerstin B. Meyer, Jong Eun Park, Emily Stephenson, Krzysztof Polański, Angela Goncalves, Lucy Gardner, Staffan Holmqvist, Johan Henriksson, Angela Zou, Andrew M. Sharkey, Ben Millar, Barbara Innes, Laura Wood, Anna Wilbrey-Clark, Rebecca P. Payne, Martin A. Ivarsson, Steve Lisgo, Andrew Filby, David H. Rowitch, Judith N. Bulmer, Gavin J. Wright, Michael J.T. Stubbington, Muzlifah Haniffa, Ashley Moffett, and Sarah A. Teichmann. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature*, 563(7731):347–353, 2018.

[199] L Vićovac and J D Aplin. Epithelial-mesenchymal transition during trophoblast differentiation. *Acta anatomica*, 156(3):202–216, 1996.

[200] Elanor N Wainwright and Paola Scaffidi. Epigenetics and Cancer Stem Cells: Unleashing, Hijacking, and Restricting Cellular Plasticity. *Trends in cancer*, 3(5):372–386, may 2017.

[201] Anna K Wakeland, Francesca Soncin, Matteo Moretto-Zita, Ching-Wen Chang, Mariko Horii, Don Pizzo, Katharine K Nelson, Louise C Laurent, and Mana M Parast. Hypoxia Directs Human Extravillous Trophoblast Differentiation in a Hypoxia-Inducible Factor-Dependent Manner. *The American journal of pathology*, 187(4):767–780, apr 2017.

[202] Sissy E Wamaitha and Kathy K Niakan. Human Pre-gastrulation Development. *Current topics in developmental biology*, 128:295–338, 2018.

[203] Jichang Wang, Manvendra Singh, Chuanbo Sun, Daniel Besser, Alessandro Prigione, Zoltán Ivics, Laurence D Hurst, and Zsuzsanna Izsvák. Isolation and cultivation of naive-like human pluripotent stem cells based on HERVH expression. *Nature protocols*, 11(2):327–346, feb 2016.

[204] Kai Wang, Mingyao Li, and Hakon Hakonarson. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16):e164–e164, sep 2010.

[205] Liang-Jie Wang, Mei-Leng Cheong, Yun-Shien Lee, Ming-Ting Lee, and Hungwen Chen. High-temperature requirement protein A4 (HtrA4) suppresses the fusogenic activity of syncytin-1 and promotes trophoblast invasion. *Molecular and cellular biology*, 32(18):3707–3717, sep 2012.

[206] Liang-Jie Wang, Hsiao-Fan Lo, Cheng-Fu Lin, Pui-Sze Ng, Yi-Hung Wu, Yun-Shien Lee, Mei-Leng Cheong, and Hungwen Chen. SFRP3 negatively regulates placental extravillous trophoblast cell migration mediated by the GCM1-WNT10B-FZD7 axis. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 33(1):314–326, jan 2019.

[207] Min Jun Wang, Fei Chen, Joseph T.Y. Lau, and Yi Ping Hu. Hepatocyte polyploidization and its association with pathophysiological processes. *Cell death disease*, 8(5):e2805, 2017.

[208] Carol B Ware, Angelique M Nelson, Brigham Mecham, Jennifer Hesson, Wenyu Zhou, Erica C Jonlin, Antonio J Jimenez-Caliani, Xinxian Deng, Christopher Cavanaugh, Savannah Cook, Paul J Tesar, Jeffrey Okada, Lilyana Margaretha, Henrik Sperber, Michael Choi, C Anthony Blau, Piper M Treuting, R David Hawkins, Vincenzo Cirulli, and Hannele Ruohola-Baker. Derivation of naive human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 111(12):4484–4489, mar 2014.

[209] S. Warrier, M. Van Der Jeught, G. Duggal, L. Tilleman, E. Sutherland, J. Taelman, M. Popovic, S. Lierman, S. Chuva De Sousa Lopes, A. Van Soom, L. Peelman, F. Van Nieuwerburgh, D. I.M. De Coninck, B. Menten, P. Mestdagh, J. Van De Sompele, D. Deforce, P. De Sutter, and B. Heindryckx. Direct comparison of distinct naive pluripotent states in human embryonic stem cells. *Nature Communications*, 8:1–10, 2017.

[210] Ting Wei, Jinfu Nie, Nicholas B Larson, Zhenqing Ye, Jeanette E Eckel-Passow, Keith D Robertson, Jean-Pierre A Kocher, and Liguo Wang. CpGtools: a python package for DNA methylation analysis. *Bioinformatics (Oxford, England)*, 37(11):1598–1599, jul 2021.

[211] Jingly F Weier, Heinz-Ulrich G Weier, Christine J Jung, Matthew Gormley, Yan Zhou, Lisa W Chu, Olga Genbacev, Alexi A Wright, and Susan J Fisher. Human cytotrophoblasts acquire aneuploidies as they differentiate to an invasive phenotype. *Developmental biology*, 279(2):420–432, mar 2005.

[212] Leehee Weinberger, Muneef Ayyash, Noa Novershtern, and Jacob H. Hanna. Dynamic stem cell states: Naive to primed pluripotency in rodents and humans. *Nature Reviews Molecular Cell Biology*, 17(3):155–169, 2016.

[213] Uri Weissbein, Omer Plotnik, Dan Vershkov, and Nissim Benvenisty. Culture-induced recurrent epigenetic aberrations in human pluripotent stem cells. *PLoS genetics*, 13(8):e1006979, aug 2017.

[214] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):15, feb 2018.

[215] Lifeng Xiang, Yu Yin, Yun Zheng, Yanping Ma, Yonggang Li, Zhigang Zhao, Junqiang Guo, Zongyong Ai, Yuyu Niu, Kui Duan, Jingjing He, Shuchao Ren, Dan Wu, Yun Bai, Zhouchun Shang, Xi Dai, Weizhi Ji, and Tianqing Li. A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature*, 577(7791):537–542, jan 2020.

[216] Yinghui Xie, Cui Liu, Yanqing Qin, Jianfeng Chen, and Jing Fang. Knockdown of IRE1 suppresses metastatic potential of colon cancer cells through inhibiting FN1-Src/FAK-GTPases signaling. *The international journal of biochemistry cell biology*, 114:105572, sep 2019.

[217] Zhuorui Xie, Allison Bailey, Maxim V Kuleshov, Daniel J B Clarke, John E Evangelista, Sherry L Jenkins, Alexander Lachmann, Megan L Wojciechowicz, Eryk Kropiwnicki, Kathleen M Jagodnik, Minji Jeon, and Avi Ma'ayan. Gene Set Knowledge Discovery with Enrichr. *Current protocols*, 1(3):e90, mar 2021.

[218] Zhuojin Xu, Aaron M Robitaille, Jason D Berndt, Kathryn C Davidson, Karin A Fischer, Julie Mathieu, Jennifer C Potter, Hannele Ruohola-Baker, and Randall T Moon. Wnt/β-catenin signaling promotes self-renewal and inhibits the primed state transition in naïve human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 113(42):E6382–E6390, oct 2016.

[219] Koji Yamanegi, Junko Yamane, Kenta Kobayashi, Nahoko Kato-Kogoe, Hideki Ohyama, Keiji Nakasho, Naoko Yamada, Masaki Hata, Toshihiro Nishioka, Satoru Fukunaga, Hiroyuki Futani, Haruki Okamura, and Nobuyuki Terada. Sodium valproate, a histone deacetylase inhibitor, augments the expression of cell-surface NKG2D ligands, MICA/B, without increasing their soluble forms to enhance susceptibility of human osteosarcoma cells to NK cell-mediated cytotoxicity. *Oncology reports*, 24(6):1621–1627, dec 2010.

[220] Liying Yan, Mingyu Yang, Hongshan Guo, Lu Yang, Jun Wu, Rong Li, Ping Liu, Ying Lian, Xiaoying Zheng, Jie Yan, Jin Huang, Ming Li, Xinglong Wu, Lu Wen, Kaiqin Lao, Ruiqiang Li, Jie Qiao, and Fuchou Tang. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature structural molecular biology*, 20(9):1131–1139, sep 2013.

[221] Penghua Yang, Aihua Dai, Andrei P Alexenko, Yajun Liu, Amanda J Stephens, Laura C Schulz, Danny J Schust, R Michael Roberts, and Toshihiko Ezashi. Abnormal oxidative stress responses in fibroblasts from preeclampsia infants. *PloS one*, 9(7):e103110, 2014.

[222] Wei Yang, Yanyan Li, Ruoling Gao, Zenghe Xiu, and Ting Sun. MHC class I dysfunction of glioma stem cells escapes from CTL-mediated immune response via activation of Wnt/β-catenin signaling pathway. *Oncogene*, 39(5):1098–1111, jan 2020.

[223] Qi-Long Ying, Jason Wray, Jennifer Nichols, Laura Batlle-Morera, Bradley Doble, James Woodgett, Philip Cohen, and Austin Smith. The ground state of embryonic stem cell self-renewal. *Nature*, 453(7194):519–523, may 2008.

[224] Leqian Yu, Yulei Wei, Hai Xi Sun, Ahmed K. Mahdi, Carlos A. Pinzon Arteaga, Masahiro Sakurai, Daniel A. Schmitz, Canbin Zheng, Emily D. Ballard, Jie Li, Noriko Tanaka, Aoi Kohara, Daiji Okamura, Adrian A. Mutto, Ying Gu, Pablo J. Ross, and Jun Wu. Derivation of Intermediate Pluripotent Stem Cells Amenable to Primordial Germ Cell Specification. *Cell Stem Cell*, 28(3):550–567.e12, 2021.

[225] Yi Yu, Shuchen Gu, Wenjian Li, Chuang Sun, Fenfang Chen, Mu Xiao, Lei Wang, Dewei Xu, Ye Li, Chen Ding, Zongping Xia, Yi Li, Sheng Ye, Pinglong Xu, Bin Zhao, Jun Qin, Ye Guang Chen, Xia Lin, and Xin Hua Feng. Smad7 enables STAT3 activation and promotes pluripotency independent of TGF-β signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 114(38):10113–10118, 2017.

[226] Hong Wa Yung, Myriam Hemberger, Erica D Watson, Claire E Senner, Carolyn P Jones, Randal J Kaufman, D Stephen Charnock-Jones, and Graham J Burton. Endoplasmic reticulum stress disrupts placental morphogenesis: implications for human intrauterine growth restriction. *The Journal of pathology*, 228(4):554–564, dec 2012.

[227] Xiao-Jing Zhang, Peng Zhang, and Hongliang Li. Interferon regulatory factor signalings in cardiometabolic diseases. *Hypertension (Dallas, Tex. : 1979)*, 66(2):222–247, aug 2015.

[228] Mingfu Zhu, Anna C Need, Yujun Han, Dongliang Ge, Jessica M Maia, Qianqian Zhu, Erin L Heinzen, Elizabeth T Cirulli, Kimberly Pelak, Min He, Elizabeth K Ruzzo, Curtis Gumbs, Abanish Singh, Sheng Feng, Kevin V Shianna, and David B Goldstein. Using ERDS to infer copy-number variants in high-coverage genomes. *American journal of human genetics*, 91(3):408–421, sep 2012.

[229] T G Zybina, H G Frank, S Biesterfeld, and P Kaufmann. Genome multiplication of extravillous trophoblast cells in human placenta in the course of differentiation and invasion into endometrium and myometrium. II. Mechanisms of polyploidization. *Tsitologiia*, 46(7):640–648, 2004.

[230] T G Zybina, P Kaufmann, H G Frank, J Freed, M Kadyrov, and S Biesterfeld. Genome multiplication of extravillous trophoblast cells in human placenta in the course of differentiation and invasion into endometrium and myometrium. I. Dynamics of polyploidization. *Tsitologiia*, 44(11):1058–1067, 2002.

[231] Tatiana G Zybina, Grigori I Stein, and Eugenia V Zybina. Endopolyploid and proliferating trophoblast cells express different patterns of intracellular cytokeratin and glycogen localization in the rat placenta. *Cell biology international*, 35(7):649–655, jul 2011.