**Title**
Modeling Early Human Development with Human Pluripotent Stem Cells



**Author**
Chan, Ngam Fung David

**Publication Date**
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Modeling Early Human Development with Human Pluripotent Stem Cells

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular, Cell and Developmental Biology

by

Ngam Fung David Chan

2012

ABSTRACT OF THE DISSERTATION

Modeling Early Human Development with Human Pluripotent Stem Cells

by

Ngam Fung David Chan

Doctor of Philosophy in Molecular, Cell and Developmental Biology

University of California, Los Angeles, 2012

Professor William E Lowry, Chair

Human embryonic stem cells (hESC) and induced pluripotent stem cells (hiPSC) are pluripotent stem cells (hPSCs) that have the capacity to differentiate into a panoply of various cell types and to serve as a model for human development. Nevertheless, it is unknown whether *in vitro* development of hPSCs mirrors *in vivo* human development. We performed comprehensive transcriptome profiling to compare hPSC-derived progeny with tissue-derived counterparts. While hESC and hiPSC make nearly identical progeny for the neural, hepatic and mesenchymal lineages, these hPSC-derived progeny maintained, regardless of the cell lineage, the expression of genes normally associated with early development, including *LIN28A*, *LIN28B* and *DPPA4*. These data and expression patterns in early human fetal tissue suggested the hPSC-derivatives were reflective of very early human development. The LIN28/let-7 pathway has been known to regulate development. We hypothesize that high LIN28 and low level let-7 activities may be inhibiting the developmental maturity in hPSC-derivatives. In addition, such a LIN28/let-7 expression pattern is also thought to contribute to tumor aggressiveness. We have

established a human cell-based system to screen for small molecules that could modulate LIN28/let-7 activity. This screening could potentially unveil powerful small molecules that could modulate developmental maturity and therapeutically combat cancer. Previous reports and our own transcriptional profiling indicate that hPSCs are suitable for modeling early human development. Epithelial to mesenchymal transitions (EMTs) are thought to be essential to generate diversity of tissues during early fetal development, but these events are impossible to study at the molecular level *in vivo* in humans. The first EMT event in human development occurs just prior to generation of the primitive streak. Because hPSCs are thought to most closely resemble cells found in epiblast stage embryos prior to formation of the primitive streak, we sought to model this first human EMT in vitro with hPSCs. We have demonstrated the EMT progression in hPSCs with embryoid bodies. We also identified PTK7 as a marker of this EMT population, which was used to purify these cells for identification of novel markers of human germ layer development.

The dissertation of Ngam Fung David Chan is approved.

Luisa M Iruela-Arispe

Harley I Kornblum

Alvaro Sagasti

William E Lowry, Committee Chair

University of California, Los Angeles

2012

This dissertation is dedicated to my parents Ivy Lin and Terence Chan

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor and mentor, Dr. William Lowry for his continuous support over the years. I have learned to work independently and challenge myself. I would also like to thank my committee members, Dr. Luisa Iruela-Arispe, Dr. Harley Kornblum, and Dr. Alvaro Sagasti for their time and efforts. They are always supportive and helpful with their suggestions and constructive criticisms.

I am also grateful to the past and present members of the Lowry lab for making it a fun and friendly place to work in, and especially to Yan Cui, Sarah Gomez, Soheila Azghadi and Joan Khuu for excellent technical support. I would also like to thank Dr. Andrew White and Dr. Peiyee Lee for constructive conversations and suggestions. I would like to thank Dr. Hanna Mikkola, Dr. Denis Evseenko and Dr. Ben Van Handel for their help in various experiments. I would also like to acknowledge the core facilities at UCLA including: FACS (Felicia Cordea and Jessica Scholes); the Clinical Microarray Core (Xinmin Li and Jian Zhou); and the hiPSC and hESC Cores. In addition, I would like to thank all my collaborators throughout the years for sharing ideas and broadening my scope of research.

Most of all, I would like to thank my family and friends. I would like to thank my parents, Ivy and Terence, for their guidance and unconditional support in life. I would like to thank my sister Rita and brother-in-law Jason for their support and wisdom. I would also like to thank my brother Anthony for the continuous encouragement. Lastly, I would like to thank my girlfriend Jun, who is not only a great collaborator and scientist, but also a wonderful person in life.

I would also like to acknowledge the California Institute of Regenerative Medicine (pre-doctoral grant) that provided my financial support for three years (TG2-01169), and the UCLA Dissertation Year Fellowship, which funded me for one year.

Chapter 2 is a version of the following paper: Defining the nature of human pluripotent stem cell progeny published by *Cell Research* in 2012 and authored by MP Patterson*, DN Chan*, I Ha, D Case, Y Cui, B Van Handel and WE Lowry. The authors are grateful to Eric Wexler and Heather Martin for assistance in the acquisition of fetal tissues; to Margaret Baron for sharing the AFP-GFP reporter; to Soheila Azghadi, Hung Trinh, Otaren Aimiuwu, Kimberly Loo and Adelaja Akinlolou for technical support; and to Mark Chin and Raj Sasidharan for assistance with bioinformatic analyses. The authors would like to acknowledge the core facilities at the UCLA including: FACS (Felicia Cordea and Jessica Scholes); the Clinical Microarray Core (Xinmin Li and Jian Zhou); and the hiPSC and hESC Cores.

VITA

| | |
|---|---|
| 2006 | B.S., Biological Sciences<br>University of California, Irvine<br>Irvine, California |
| 2008 | Advancement to Candidacy for Ph.D.<br>University of California, Los Angeles<br>Los Angeles, California |

## HONORS and AWARDS

**UCLA**

| | |
|---|---|
| 2011-2012 | Dissertation Year Fellowship |
| 2011 | Broad Stem Cell Center Travel Award for ISSCR, |
| 2009 | MCDB Annual Retreat Poster Competition: 2nd runner up |
| 2007-2010 | California Institute of Regenerative Medicine Pre-Doctoral Training Grant |
| 2006-2007 | Chancellor's Award |
| 2006 | Shapiro International Fellowship |

**UC-Irvine**

| | |
|---|---|
| 2006 | Honors in Biological Sciences |
| 2006 | B.S. in Biological Sciences, Summa cum laude |
| 2005 | Excellence in Research Award |
| 2004-2005 | Dr. John R. Miltner/ Irvine Health Foundation Scholarship |
| 2004 | Summer Undergraduate Research Program Fellowship |
| 2003-2005 | Undergraduate Research Opportunity Program Fellowship |
| 2002-2005 | UCI Bio Sci Dean's Honor List |

## PUBLICATIONS AND PRESENTATIONS

**Chan DN**, Cohen J, Naito J, Mott KR, Osorio N, Jin L, Fraser NJ, Jones C, Wechsler SL, Perng GC (2006). A mutant deleted for most of the herpes simplex virus type 1 (HSV-1) UOL gene does not affect the spontaneous reactivation phenotype in rabbits. Journal of Neurovirology. 12(1), 5-16

**Chan DN**, Patterson M, Cui Y, Azghadi S and Lowry WE (October 2011). "The Fidelity of Hepatic Differentiation from Human Pluripotent Stem Cells." Poster Presentation. 4th NIGMS Workshop on Human Embryonic Stem Cell Research. Bethesda, Maryland, United States.

**Chan DN**, Patterson M, Cui Y, Azghadi S and Lowry WE (June 2011). "Improving the Fidelity of Hepatic Differentiation from Human Pluripotent Stem Cells." Poster Presentation. International Society of Stem Cell Research 9th Annual Meeting. Toronto, Ontario, Canada.

**Chan DN**, Patterson M, Cui Y, Azghadi S and Lowry WE (June 2010). "Assessing the Fidelity of Hepatic Differentiation from Human Pluripotent Stem Cells." Poster Presentation. International Society of Stem Cell Research 8th Annual Meeting. San Francisco, California, United States.

Patterson M*, **Chan DN***, Ha IC, Case D, Cui Y, Van Handel, B, Mikkola HK, and Lowry WE (2012). Defining the Nature of Human Pluripotent Stem Cell Progeny. Cell Research, 22(1), 178-193 (* indicates equal contribution)

Rhodes KE, Wang Y, Gekas C, Lux CT, Francis CS, **Chan DN**, Conway S, Orkin SH, Yoder MC, Mikkola HK (2008). Hematopoietic Stem Cells Emerge in the Placental Vasculature in the Absence of Circulation. Cell Stem Cell, 2(3), 252-263

# CHAPTER 1

## Introduction

### 1.1 Definition of Stem Cells

All organisms are made up of building blocks called cells. Similar to any machinery, the organismal body is subject to wear and tear, and the building blocks need to be replaced continuously throughout the life cycle. To date, stem cells have been discovered in most organs. By definition, stem cells have the ability to self-renew and differentiate into various cell types of a tissue. The best and most stringent example to define stemness is perhaps demonstrated in hematopoietic stem cell research, where any putative hematopoietic stem cell population needs to be serially transplanted in irradiated immunodeficient mice and functionally analyzed to fully prove its stem cell properties [1-3].

Stem cells are further stratified into different classes mostly based on their potency, i.e. their ability to differentiate. The most "potent" stem cell is the totipotent stem cell, such as spores and zygotes, which can differentiate into all cell types of the organism, including any extraembryonic tissues. Pluripotent stem cells, on the other hand, can generate any fetal or adult cell type but not extraembryonic structures, hence limiting their ability to develop fully into a fetal or adult organism. In human, the two kinds of pluripotent stem cells commonly studied are embryonic stem cells and induced pluripotent stem cells [4, 5]. In adult bodies, however, the more common stem cells are unipotent or multipotent, meaning their differentiation ability is restricted to one or a few types of progeny cells. For example, adult neural stem cells can differentiate into neurons, glial cells and oligodendrocytes [6, 7].

### 1.2 Human Pluripotent Stem Cells

The study of human pluripotent stem cells originate from teratocarcinoma studies. Teratocarcinoma is a malignant germ cell tumor that arises in gonads, and is morphologically stunning and scientifically intriguing because of the formation of structures such as teeth, hair and even fingers [8, 9]. Tertaocarcinoma contain stem cells that perpetuate in an

undifferentiated state, which contribute to pockets of pluripotent embryonal carcinoma cells nested within complex, differentiated derivatives. These embryonal carcinoma cells were adapted for culture first in mouse, then in human. Because of their ability to maintain pluripotency as well as to differentiate into all three germ layers (ectoderm, mesoderm, and endoderm), embryonal carcinoma cells have been suggested as a possible *in vitro* system for studying mammalian development [8, 10]. But this hypothesis is based on the assumption that embryonal carcinoma cells can accurately capture the developmental state in the inner cell mass (ICM). The shortcomings of embryonal carcinoma cells are several fold: (1) embryonal carcinoma cells can lose its pluripotency when kept in culture [9, 10]; (2) embryonal carcinoma cells derived from different tumors represent different embryonic cell types [8, 11, 12]; (3) embryonal carcinoma can harbor genetic mutations and abnormal karyotype; and (4) protein expression profile of embyonal carcinoma is inconsistent with that of inner cell mass [13]. Nevertheless, the similar developmental potential to inner cell mass, the ability to form all three germ layers, the ability to aggregate into embryoid bodies and the partial absence of X-inactivation in embryonic carcinoma cells all contribute to its intriguing potential in studying human development [10, 12, 14-16].

### 1.3 Human Embryonic Stem Cells

The most-studied pluripotent stem cells to date are embryonic stem cells. Embryonic stem cells were initially derived from mouse, and later on from primates and human [4, 17-19]. Typically, embryonic stem cells are derived from the inner cell mass of the developing blastocyst. In animals, the early embryos are harvested from the donor mother; whereas in human, the main source of embryos is surplus embryos from in vitro fertilization facilities. Embryonic stem cells are cultured in the presence of necessary growth factors, some of which are provided by feeder cells or in conditioned media. Human embryonic stem cells can be perpetually cultured in vitro

as pluripotent stem cells. They can also be differentiated into any of the three germ layers, as demonstrated by embryoid body formation and teratoma assays [4]. To date, human embryonic stem cells have been shown to differentiate into function neurons, cardiomyocytes, liver, blood, endothelial cells, among many others. The potential of embryonic stem cells in therapeutics and developmental studies is enormous, but the traditional methods of generating embryonic stem cells necessitate the destruction of live human embryos, a highly debatable ethical issue. Recently, it has been shown that embryonic stem cells can be generated from a single blastomere without the destruction of human embryos, making it a keystone towards ES-based individualized medicine [20]. Nevertheless, such derivation techniques would still require invasive procedures to the early embryo. In addition, the generation of embryonic stem cell lines is time and resource intensive, and faces high scrutiny from federal and local laws.

### 1.4 Human Induced Pluripotent Stem Cells

There has long been immense interest in reprogramming somatic cells into a pluripotent state. One of the means to achieve this is by somatic-cell nuclear transfer (SCNT), which entails the replacement of the egg nucleus with a somatic nucleus [21-23]. Others have shown that cell fusion with a pluripotent stem cell can reprogram a somatic cell back to a pluripotent state [24, 25]. SCNT is highly inefficient, and is not perfect because the mitochondria in the egg are not replaced by the somatic cell. In addition, it requires human egg donation, which is a highly controversial matter. Induced pluripotency by cell fusion is useful for studying the mechanism behind reprogramming, but it is again inefficient. Also, the product of cell fusion is by default polyploid and unsuitable for therapeutic uses [26, 27].

In 2006, the Yamanaka group has successfully reprogrammed mouse fibroblasts into pluripotent stem cells by forcibly inducing the expression of four transcription factors (*Oct4, Sox2, Klf4, c-Myc*). Like embryonic stem cells, these reprogrammed cells can be perpetually cultured as

pluripotent cells, and can differentiate into cells of all lineages, as demonstrated by teratoma formation and tetraploid complementation [28]. These cells are termed induced pluripotent stem cells. In subsequent years, the Yamanaka group and other independent groups have reprogrammed human somatic cells into human induced pluripotent stem cells using similar cocktails of transcription factors [5, 29, 30]. Following this breakthrough discovery, the research on induced pluripotency has exploded, with increasingly advanced methods of reprogramming, making use of small molecules, proteins and miRNAs [31, 32]. In general, the development of iPS cell technology has opened the door not only for personalized medicine, but also for patient-specific disease modeling in vitro [33, 34].

### 1.5 Discrepancies between ES and iPS cells

The original goal of induced pluripotency is to reprogram any somatic cell to a pluripotent state similar to the gold standard-the embryonic stem cells. Since the advent of induced pluripotent stem cells, there have been intense studies on the global differences between ES and iPS cells. ES cells and iPS cells were shown to have distinct global transcriptional and epigenetic profiles [35-37]. Whereas it is believed that continuous passaging in pluripotent culture condition can 'correct' some of the differences exhibited by iPS cells, such an erasure is not complete [35, 38]. It is widely believed that the iPS cells retain an 'epigenetic memory' that lingers from the initial cell type [39-42]. Large-scale global transcriptome and methylome studies have illustrated larger variability among different iPS lines, effectively placing a significant portion of them away from ES cells [43-45]. In addition, recent research has demonstrated in iPS cells genomic aberrations that have not been captured with traditional karyotyping as well as acquisitions of mutations during the process of reprogramming [46-48]. These discoveries call into the question the readiness of clinical application of iPS cell technology because of the causative effects in cancer [49, 50].

### 1.6 Therapeutic Uses

One of the biggest obstacles for human pluripotent stem cells to be employed in cell-based therapy is their propensity to form tumors [51, 52]. Much effort has been allocated to address the issue in induced pluripotent stem cells, with emphasis on the combination of transcription factors, the mode of delivery, genomic integration, and even the starting cell type [53]. With the development of new methods of reprogramming *sans* any oncogenic factors or genomic integration (such as small molecules, proteins, or RNAs) [31, 32, 52-61], the concerns for iPS-related tumorigenicity may one day be allayed.

As opposed to the issue of tumorigenicity, there has generally been less concern about the developmental state of pluripotent stem cell-derivatives. In most studies describing the derivation of a certain cell type, investigators are primarily interested in the expression of a handful of key markers and the functional analysis of these target cell types. While these are undoubtedly important analysis, it does not address whether these pluripotent stem cell-derivatives are adequate replacements for mature adult cells in therapeutic uses. Though there are comparisons of a pluripotent stem cell derivative vis-à-vis the natural counterpart cell type, differences observed may be due to specific differentiation protocol of a certain lineage [62-68]. To thoroughly understand the developmental state of pluripotent stem cell-derivatives, it would be advisable to compare with natural counterparts in multiple well-studied differentiated cell types representing different lineages.

### 1.7 Definitive Endoderm and Hepatocytes

In normal mammalian development, the embryo develops into three germ layers: Ectoderm, Mesoderm, and Endoderm before further development into respective lineages. The Ectoderm gives rise to neural tissues and the skin; the mesoderm gives rise to muscles, blood, cartilage

connective tissues (such as fibroblasts), dermis, kidneys and gonads; the endoderm gives rise to internal organs such as liver and lungs.

The liver is an essential organ in the human body, serving roles such as detoxification, plasma protein synthesis, glycogen storage, red blood cell decomposition and bile production. Because of the diverse functions and the workload that liver bears, it is prone to various diseases. For terminal liver diseases exhibiting liver failure, transplant is the only viable long-term option, but this has been hampered by a chronic shortage of donors and acute transplant rejection [69]. Autologous hiPSC-derived hepatocytes provide hope for solving these problems. In addition, hESC and hiPSC-derived hepatocytes would be invaluable source for pharmacological screenings and could potentially provide a platform for studying the basic biology of hepatic development and diseases [63, 67].

### 1.8 As a Model of Development

Because pluripotent stem cells bear resemblance to cells in the developing blastocyst [70], it has long been regarded as a viable strategy to study human development. Many groups have devised strategies to differentiate pluripotent stem cells into various lineages and target cell types to tease apart the molecular underpinnings of natural human development [62, 64-68]. Further investigation needs to be conducted on how closely development *in vitro* mirrors that *in vivo*. In addition, the directed differentiation of a highly pure derivative of a certain lineage may not have taken into account the intricate micro-environment generally provided by a dynamically developing embryo.

### 1.9 Embryoid Bodies

Embryoid bodies are three-dimensional cell aggregates formed from pluripotent stem cells. Embryoid bodies are spontaneous differentiating and have the ability to differentiate into all

three germ layers [71]. Because of the origin of pluripotent stem cells and the three-dimensional nature of embryoid bodies, some have suggested that human embryoid bodies parallel a developing embryo and can be a model for studying human embryonic development [71, 72]. Nevertheless, such a hypothesis has not been fully explored.

### 1.10 Markers of Germ Layer Formation

Due to ethical concerns, performing developmental studies in developing human embryos is impossible. Our understanding of germ layer formation and early development mostly comes from animal studies such as chick, Xenopus, and mouse. While the general molecular underpinnings are shared, there are many differences [73]. It is highly possible that unknown regulators may play significant roles in human development that cannot be deduced from animal models. The lack of a human developmental model has severely restricted discoveries relevant to early development and perhaps congenital diseases. The advent of human pluripotent stem cells appears to be a solution to these limitations; however, this potential has not been fully explored. Embryoid bodies and teratoma formation have routinely shown the ability to differentiate to all three germ layers [4, 5, 71], but the process and mechanism of germ layer development has not been thoroughly examined.

Similar to the epiblast, pluripotent stem cells are epithelial in nature, expressing classical epithelial markers such as E-CADHERIN and EPCAM. In vertebrates, gastrulation is the first important morphogenetic movement after implantation, resulting in the formation of mesendodermal progenitors from the epiblast. The underlying epithelial-to-mesenchymal transition has been intensely studied in multiple model organisms. In amniotes, multiple signaling pathways such as Nodal. Wnt, EGF and FGF help establish the organizing center and initiate the EMT program [74, 75]. The TGF-beta family members activate SNAIL genes (SNAIL and SLUG) which are essential for cell shape and movements [76, 77]. There continues to be

new discoveries on important regulators for gastrulation and other developmental events. As mentioned earlier, embryoid bodies have routinely been used to demonstrate pluripotency by formation of three germ layers [71], but there has not been in-depth studies on the molecular events governing germ layer commitment.

### 1.11 PTK7, a Planar Cell Polarity Gene, Regulates Important Developmental Events

Current research on mesoderm commitment from pluripotent stem cells put heavy emphasis on the identification of populations expressing primitive streak markers such as SNAIL, TWIST, and BRACHYURY [72]. Less effort has been spent on other regulators of cell polarity and the resulting large-scale restructuring of tissues. In vertebrate development, the regulation of planar cell polarity is crucial for morphogenetic events such as gastrulation and neural tube folding, via the mechanism of convergent extension [78, 79]. Recent research has uncovered the roles of planar cell polarity genes, such as PTK7 and VANGL2 in these events [80-85]. For example, loss of function in PTK7 leads to defects in gastrulation and mesoderm commitment in mouse, as well as failure in neural tube closure in zebrafish and mouse. As previously described, gastrulation is the first EMT event following implantation. Since human pluripotent stem cells most closely resemble cells found in the epiblast stage embryo [70], it is intriguing whether any known cell polarity regulators coincide or even play a role in the spontaneous EMT in pluripotent stem cells. In addition, the three-dimensional nature of embryoid bodies may be advantageous for visualizing any large-scale tissue restructuring potentially discernible by cell polarity markers/regulators [71, 72]. These studies can lead to important functional characterization of polarity genes in human development. If PTK, a member of the protein tyrosine kinase family [86], proves to be a specific marker for a certain morphogenetic event such as the EMT pertaining to gastrulation, it is also possible to use FACS to capture a specific population earmarked for this developmental event.

### 1.12 Undertaking the challenges in studying human development

Here we describe our efforts in modeling human development with human pluripotent stem cells. We have compared the differences in transcriptional landscapes between hPSC-derivatives and natural counterparts, and explored mechanism to eliminate these differences. We also employed hPSCs and embryoid bodies to delineate molecular events in human embryonic development. We hope to contribute to the mapping of human development with these studies.

### 1.13 Bibliography


1. Iscove, N.N. and K. Nawa, *Hematopoietic stem cells expand during serial transplantation in vivo without apparent exhaustion.* Curr Biol, 1997. **7**(10): p. 805-8.
2. Keller, G. and R. Snodgrass, *Life span of multipotential hematopoietic stem cells in vivo.* J Exp Med, 1990. **171**(5): p. 1407-18.
3. Wu, A.M., et al., *Cytological evidence for a relationship between normal hemotopoietic colony-forming cells and cells of the lymphoid system.* J Exp Med, 1968. **127**(3): p. 455-64.
4. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.
5. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
6. Snyder, E.Y., et al., *Multipotent neural cell lines can engraft and participate in development of mouse cerebellum.* Cell, 1992. **68**(1): p. 33-51.
7. Temple, S., *Division and differentiation of isolated CNS blast cells in microculture.* Nature, 1989. **340**(6233): p. 471-3.

8. Martin, G.R., *Teratocarcinomas and mammalian embryogenesis.* Science, 1980. **209**(4458): p. 768-76.

9. Stevens, L.C., *The biology of teratomas.* Adv Morphog, 1967. **6**: p. 1-31.

10. Martin, G.R., *Teratocarcinomas as a model system for the study of embryogenesis and neoplasia.* Cell, 1975. **5**(3): p. 229-43.

11. Sherman, M.I. and R.A. Miller, *F9 embryonal carcinoma cells can differentiate into endoderm-like cells.* Dev Biol, 1978. **63**(1): p. 27-34.

12. Martin, G.R. and M.J. Evans, *Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro.* Proc Natl Acad Sci U S A, 1975. **72**(4): p. 1441-5.

13. Martin, G.R., S. Smith, and C.J. Epstein, *Protein synthetic patterns in teratocarcinoma stem cells and mouse embryos at early stages of development.* Dev Biol, 1978. **66**(1): p. 8-16.

14. Nicolas, J.F., et al., *Cell lines derived from teratocarcinomas.* Cancer Res, 1976. **36**(11 Pt. 2): p. 4224-31.

15. Martin, G.R., et al., *X-chromosome inactivation during differentiation of female teratocarcinoma stem cells in vitro.* Nature, 1978. **271**(5643): p. 329-33.

16. McBurney, M.W., *Clonal lines of teratocarcinoma cells in vitro: differentiation and cytogenetic characteristics.* J Cell Physiol, 1976. **89**(3): p. 441-55.

17. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. **78**(12): p. 7634-8.

18. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.

19. Thomson, J.A., et al., *Isolation of a primate embryonic stem cell line.* Proc Natl Acad Sci U S A, 1995. **92**(17): p. 7844-8.

20. Klimanskaya, I., et al., *Human embryonic stem cell lines derived from single blastomeres.* Nature, 2006. **444**(7118): p. 481-5.

21. Hochedlinger, K. and R. Jaenisch, *Monoclonal mice generated by nuclear transfer from mature B and T donor cells.* Nature, 2002. **415**(6875): p. 1035-8.

22. Egli, D., et al., *Developmental reprogramming after chromosome transfer into mitotic mouse zygotes.* Nature, 2007. **447**(7145): p. 679-85.

23. Yang, X., et al., *Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning.* Nat Genet, 2007. **39**(3): p. 295-302.

24. Blau, H.M., C.P. Chiu, and C. Webster, *Cytoplasmic activation of human nuclear genes in stable heterocaryons.* Cell, 1983. **32**(4): p. 1171-80.

25. Tada, M., et al., *Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells.* Curr Biol, 2001. **11**(19): p. 1553-8.

26. Yamanaka, S. and H.M. Blau, *Nuclear reprogramming to a pluripotent state by three approaches.* Nature. **465**(7299): p. 704-12.

27. Gurdon, J.B. and D.A. Melton, *Nuclear reprogramming in cells.* Science, 2008. **322**(5909): p. 1811-5.

28. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.

29. Lowry, W.E., et al., *Generation of human induced pluripotent stem cells from dermal fibroblasts.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 2883-8.

30. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.

31. Ichida, J.K., et al., *A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog.* Cell Stem Cell, 2009. **5**(5): p. 491-503.

32. Shi, Y., et al., *A combined chemical and genetic approach for the generation of induced pluripotent stem cells.* Cell Stem Cell, 2008. **2**(6): p. 525-8.

33. Hanna, J., et al., *Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin.* Science, 2007. **318**(5858): p. 1920-3.

34. Kazuki, Y., et al., *Complete genetic correction of ips cells from Duchenne muscular dystrophy.* Mol Ther. **18**(2): p. 386-93.

35. Chin, M.H., et al., *Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures.* Cell Stem Cell, 2009. **5**(1): p. 111-23.

36. Ghosh, Z., et al., *Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells.* PLoS One. **5**(2): p. e8975.

37. Marchetto, M.C., et al., *Transcriptional signature and memory retention of human-induced pluripotent stem cells.* PLoS One, 2009. **4**(9): p. e7076.

38. Chin, M.H., et al., *Molecular analyses of human induced pluripotent stem cells and embryonic stem cells.* Cell Stem Cell. **7**(2): p. 263-9.

39. Bar-Nur, O., et al., *Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells.* Cell Stem Cell. **9**(1): p. 17-23.

40. Polo, J.M., et al., *Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells.* Nat Biotechnol. **28**(8): p. 848-55.

41. Kim, K., et al., *Epigenetic memory in induced pluripotent stem cells.* Nature. **467**(7313): p. 285-90.

42. Ohi, Y., et al., *Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells.* Nat Cell Biol. **13**(5): p. 541-9.

43. Bock, C., et al., *Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines.* Cell. **144**(3): p. 439-52.

44. Lister, R., et al., *Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.* Nature. **471**(7336): p. 68-73.

45. Narsinh, K.H., et al., *Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells.* J Clin Invest. **121**(3): p. 1217-21.

46. Hussein, S.M., et al., *Copy number variation and selection during reprogramming to pluripotency.* Nature. **471**(7336): p. 58-62.

47. Gore, A., et al., *Somatic coding mutations in human induced pluripotent stem cells.* Nature. **471**(7336): p. 63-7.

48. Laurent, L.C., et al., *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): p. 106-18.

49. Sun, N., M.T. Longaker, and J.C. Wu, *Human iPS cell-based therapy: considerations before clinical applications.* Cell Cycle. **9**(5): p. 880-5.

50. Sancho-Martinez, I., E. Nivet, and J.C. Izpisua Belmonte, *The labyrinth of nuclear reprogramming.* J Mol Cell Biol. **3**(6): p. 327-9.

51. Yamanaka, S., *A fresh look at iPS cells.* Cell, 2009. **137**(1): p. 13-7.

52. Okita, K., T. Ichisaka, and S. Yamanaka, *Generation of germline-competent induced pluripotent stem cells.* Nature, 2007. **448**(7151): p. 313-7.

53. Miura, K., et al., *Variation in the safety of induced pluripotent stem cell lines.* Nat Biotechnol, 2009. **27**(8): p. 743-5.

54. Kim, D., et al., *Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins.* Cell Stem Cell, 2009. **4**(6): p. 472-6.

55. Anokye-Danso, F., et al., *Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency.* Cell Stem Cell. **8**(4): p. 376-88.

56. Liao, B., et al., *MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition.* J Biol Chem. **286**(19): p. 17359-64.

57. Subramanyam, D., et al., *Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells.* Nat Biotechnol. **29**(5): p. 443-8.

58. Nakagawa, M., et al., *Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts.* Nat Biotechnol, 2008. **26**(1): p. 101-6.

59. Kim, J.B., et al., *Generation of induced pluripotent stem cells from neural stem cells.* Nat Protoc, 2009. **4**(10): p. 1464-70.

60. Zhao, X.Y., et al., *iPS cells produce viable mice through tetraploid complementation.* Nature, 2009. **461**(7260): p. 86-90.

61. Giorgetti, A., et al., *Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2.* Cell Stem Cell, 2009. **5**(4): p. 353-7.

62. Dimos, J.T., et al., *Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons.* Science, 2008. **321**(5893): p. 1218-21.

63. Si-Tayeb, K., et al., *Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells.* Hepatology. **51**(1): p. 297-305.

64. Pfannkuche, K., et al., *Cardiac myocytes derived from murine reprogrammed fibroblasts: intact hormonal regulation, cardiac ion channel expression and development of contractility.* Cell Physiol Biochem, 2009. **24**(1-2): p. 73-86.

65. Karumbayaram, S., et al., *Directed differentiation of human-induced pluripotent stem cells generates active motor neurons.* Stem Cells, 2009. **27**(4): p. 806-11.

66. Zhang, J., et al., *Functional cardiomyocytes derived from human induced pluripotent stem cells.* Circ Res, 2009. **104**(4): p. e30-41.

67. Song, Z., et al., *Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells.* Cell Res, 2009. **19**(11): p. 1233-42.

68. Choi, K.D., et al., *Hematopoietic and endothelial differentiation of human induced pluripotent stem cells.* Stem Cells, 2009. **27**(3): p. 559-67.

69. Kmiec, Z., *Cooperation of liver cells in health and disease.* Adv Anat Embryol Cell Biol, 2001. **161**: p. III-XIII, 1-151.

70. Nichols, J. and A. Smith, *Naive and primed pluripotent states.* Cell Stem Cell, 2009. **4**(6): p. 487-92.

71. Itskovitz-Eldor, J., et al., *Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers.* Mol Med, 2000. **6**(2): p. 88-95.

72. Sharon, N., et al., *Molecular and functional characterizations of gastrula organizer cells derived from human embryonic stem cells.* Stem Cells. **29**(4): p. 600-8.

73. Winston, N.J. and M.H. Johnson, *Can the mouse embryo provide a good model for the study of abnormal cellular development seen in human embryos?* Hum Reprod, 1992. **7**(9): p. 1291-6.

74. Thiery, J.P., et al., *Epithelial-mesenchymal transitions in development and disease.* Cell, 2009. **139**(5): p. 871-90.

75. Heisenberg, C.P. and L. Solnica-Krezel, *Back and forth between cell fate specification and movement during vertebrate gastrulation.* Curr Opin Genet Dev, 2008. **18**(4): p. 311-6.

76. Oda, H., S. Tsukita, and M. Takeichi, *Dynamic behavior of the cadherin-based cell-cell adhesion system during Drosophila gastrulation.* Dev Biol, 1998. **203**(2): p. 435-50.

77. Carver, E.A., et al., *The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition.* Mol Cell Biol, 2001. **21**(23): p. 8184-8.

78. Wallingford, J.B., S.E. Fraser, and R.M. Harland, *Convergent extension: the molecular control of polarized cell movement during embryonic development.* Dev Cell, 2002. **2**(6): p. 695-706.

79. Roszko, I., A. Sawada, and L. Solnica-Krezel, *Regulation of convergence and extension movements during vertebrate gastrulation by the Wnt/PCP pathway.* Semin Cell Dev Biol, 2009. **20**(8): p. 986-97.

80. Murdoch, J.N., et al., *Circletail, a new mouse mutant with severe neural tube defects: chromosomal localization and interaction with the loop-tail mutation.* Genomics, 2001. **78**(1-2): p. 55-63.

81. Montcouquiol, M., et al., *Identification of Vangl2 and Scrb1 as planar polarity genes in mammals.* Nature, 2003. **423**(6936): p. 173-7.

82. Lu, X., et al., *PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates.* Nature, 2004. **430**(6995): p. 93-8.

83. Yen, W.W., et al., *PTK7 is essential for polarized cell motility and convergent extension during mouse gastrulation.* Development, 2009. **136**(12): p. 2039-48.

84. Curtin, J.A., et al., *Mutation of Celsr1 disrupts planar polarity of inner ear hair cells and causes severe neural tube defects in the mouse.* Curr Biol, 2003. **13**(13): p. 1129-33.

85. Hamblet, N.S., et al., *Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure.* Development, 2002. **129**(24): p. 5827-38.

86. Grassot, J., et al., *Origin and molecular evolution of receptor tyrosine kinases with immunoglobulin-like domains.* Mol Biol Evol, 2006. **23**(6): p. 1232-41.

# CHAPTER 2

## Defining the nature of human pluripotent stem cell progeny

## 2.1 Abstract

While it is clear that human pluripotent stem cells (hPSC) can differentiate to generate a panoply of various cell types, it is unknown how closely *in vitro* development mirrors that which occurs *in vivo*. To determine whether human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) make equivalent progeny, and whether either makes cells that are analogous to tissue derived cells, we performed comprehensive transcriptome profiling of purified PSC derivatives and their tissue derived counterparts. Expression profiling demonstrated that hESCs and hiPSCs make nearly identical progeny for the neural, hepatic and mesenchymal lineages, and absence of re-expression from exogenous reprogramming factors in hiPSC-progeny. However, when compared to a tissue derived counterpart, the progeny of both hESCs and hiPSCs maintained expression of a subset of genes normally associated with early mammalian development, regardless of the type of cell generated. While pluripotency genes (*OCT4, SOX2, REX1, and NANOG)* appeared to be silenced immediately upon differentiation from hPSCs, genes normally unique to early embryos (*LIN28A*, *LIN28B, DPPA4 and others)* were not fully silenced in hPSC derivatives. These data and evidence from expression patterns in early human fetal tissue (3-16 weeks of development) suggests that the differentiated progeny of hPSCs are reflective of very early human development (<6 weeks). These findings provide support for the idea that hPSCs can serve as useful *in vitro* models of early human development, but also raise important issues for disease modeling and the clinical application of hPSC derivatives.

## 2.2 Introduction

hiPSCs are similar to hESCs in that they share the same pluripotency markers, display self-renewal, and are capable of differentiation into all three germ layers. The rapid development of hiPSC technology has delivered new hope for personalized medicine. In addition, hiPSCs may present a better tool for modeling disease as they can be derived from patients with various genetic diseases at virtually any age [1].

We and others have demonstrated that undifferentiated hiPSCs at early passages display a somewhat unique gene expression pattern from hESCs [2-4]. However, the functional relevance of any gene expression differences between hiPSCs and hESCs remains unclear. Recent evidence suggests that these differences may be explained by the fact that hiPSCs are more variant than hESCs at the epigenetic and transcriptional level [5-7]. It is also possible that the differences between hiPSCs and hESCs persist upon differentiation and that these differences influence the derivation of specified progeny, but this has not been formally tested. Recent work in murine and human iPSCs has shown that a residual epigenetic memory of the starting cell type is retained after reprogramming, and this memory appears to influence the efficiency of differentiation [8-11]. Furthermore, epigenetic differences between mouse iPSC lines diminish as the cells are passaged [8], consistent with what was shown for gene expression between hiPSCs and hESCs [2, 12]. So far, differentiation efficiency appears to be as variable amongst hESCs lines as it is for hiPSCs lines [5, 13-15].

Despite varying efficiencies, all hPSC lines appear to have some potential to generate an array of functional cell types, including motor neurons, cardiomyocytes, dopaminergic neurons, hematopoietic cells, hepatocytes, etc. [16-21]. While it is clear that hiPSCs are capable of generating similar cell types as hESCs as determined by expression of a select number of markers, it remains to be seen if, on a global level, these derivatives are truly analogous. Furthermore, for hESCs or hiPSCs to be employed in therapy or as models of human

development, it is also important to determine how well their differentiation mirrors natural development. Because both hiPSCs and hESCs undergo development *in vitro* instead of *in vivo*, it is not clear if current differentiation protocols accurately recapitulate the development that occurs in the human embryo. Nor is it clear whether progeny cells generated from hPSCs display mature adult cell phenotypes.

To address these issues, we generated and purified derivatives of all three germ layers from hESCs and hiPSCs using established protocols. We compared the global gene expression pattern between the progeny of hESCs and hiPSCs, and primary tissue-derived equivalents from various stages of development. Our results show that the gene expression differences found between hiPSCs and hESCs in the undifferentiated state mostly dissipated after differentiation. Also, derivatives from hiPSCs and hESCs were very similar to each other transcriptionally. On the other hand, a significant number of genes were different between PSC derivatives and their respective natural counterpart regardless of the cell type generated. While *OCT4, SOX2, REX1* and *NANOG* were effectively silenced, an alternative set of genes normally associated with the pluripotent state including *LIN28A, LIN28B*, and *DPPA4* remained expressed in pluripotent progeny. We hypothesize and provide significant evidence that these cells derived from hPSCs are representative of cells found during human development prior to 6 week post-fertilization. Our data are consistent with the notion that ESC-derived cells represent early stages of development [22-25], but we demonstrate here that this also applies to hiPSC-derived cells. Furthermore, we attempt to place the PSC derivatives into a more precise developmental timeframe than has been established previously. These findings underscore the need to consider the maturity of cells produced from hPSCs for disease modeling or regenerative medicine.

## 2.3 Materials and Methods

*Human tissue sourcing*

Primary human hepatocytes and fibroblasts were acquired from Lonza (Switzerland). Fetal tissues were generated from discarded anonymized material obtained from elective terminations of first and second trimester pregnancies performed by Family Planning Associates or UCLA Medical Center. The fetal ages discussed throughout the paper represent the developmental age established two weeks after the first day of the last menstral cycle. Tissues were harvested directly into sterile containers with PBS and transported on ice in PBS containing 5% FBS (Hyclone), 0.1% Ciprofloxacin HCl (10 μg/mL, Sigma), 1% amphotericin B (250 μg/mL, Invitrogen) and 1% penicillin-streptomycin (10,000 U/mL-10,000 μg/mL, Gibco) and processed the same day. Fetal spinal cord NPC lines were generated from 6-16 week spinal cords treated with collagenase and dispase for 1 hour at 37°C and plated on poly-ornithine/Laminin coated plates.

*Cell culture*

hESC and hiPSC were cultured as described previously [2, 26] in accordance with UCLA ESCRO. Neural rosette derivation, NPC purification, and further differentiation to neurons and glia were performed as described [27]. Hepatocytes were differentiated using the 4-stage protocol as described [20], with the exception of using FBS during the first 3 days of endoderm derivation. The AFP-GFP reporter [28] was transfected during differentiation using Lipofectamine2000 (Invitrogen) and labeled cells were isolated using FACSARIA (BD biosciences). To generate fibroblasts, Embryoid Bodies (EBs) were cultured in adherent conditions using standard fibroblast culturing media [26] and passaged until adopting typical fibroblast morphology.

*Immunostaining and Western Blot*

Tissue acquired from embryonic spinal cord was fixed in 4% PFA for 1 hour followed by cryopreservation with sucrose and embedding in O.C.T. Compound (Sakura). Fetal livers were fresh embedded in O.C.T. Compound. Both tissues were sectioned on a CM3050S cryostat (Leica) at 6-10uM thickness. Liver sections were fixed in formalin for 10 minutes at room temperature prior to immunostaining. Immunostaining was performed as described [18, 29]. Coverslips were fixed in 4% PFA , blocked for 1hr in 10% serum + 0.1% Tween20 (or 0.1% Triton-X-100), then incubated overnight at 4°C with primary antibodies. Following primary antibody incubation, the coverslips/cryosections were incubated with Alexa Fluor secondary antibodies (Invitrogen) at room temperature for 1 hour and mounted in Prolong Gold with DAPI (Invitrogen). All imaging was performed on Zeiss Axio Imager A1. Antibodies used include the following: mouse anti-DPPA4 (Abnova), goat anti-ALB (Bethyl Laboratories) , rabbit anti-SOX2 (Biolegend), rabbit anti-LIN28A, rabbit anti-LIN28B, and rabbit anti-Nanog (Cell Signaling Technology), rabbit anti-PAX6, rabbit anti-Tuj1(Covance) , rabbit anti-GFAP (DAKO), anti-CD44 (Developmental Studies Hybridoma Bank), mouse anti-NESTIN (Neuromics), goat anti-COL3A1, mouse anti-AFP, mouse anti-OCT3/4 (Santa Cruz Biotechnology Inc.), rabbit anti-SERPINA1 (Sigma Aldrich), and rabbit anti-Ki67 (Abcam). Western Blot Analysis was performed using standard procedures as described [30].

*Assays for function*

For Periodic Acid-Schiff Assay (PAS Assay), cells were fixed with formalin-ethanol fixative and stained using standard protocol described in manufacturer's manual (Sigma Aldrich). For Albumin ELISA Assay, media were collected from confluent wells and subjected to ELISA Assay according to manufacturer's protocol (Bethyl Laboratories). For Alizarin red staining, natural and PSC-FBs were subjected to osteogenic induction and were stained with Alizarin Red as

described [31]. Student t-tests for the ELISA and NPC differentiation assays were performed in Excel.

*Reporter Assay*

Cells transfected with the psiCHECK2- let-7 8x luciferase reporter (Addgene) or psiCHECK2 control reporter (Promega) were lysed 72 hours post-transfection and subjected to dual-glo luciferase assay as described in manufacturer's protocol (Promega). The renilla luciferase gene was driven by SV40 promoter and contained 8 let-7 targeting sequences in the 3' UTR, and firefly luciferase driven by a constitutive promoter as a transfection control. Luciferase assays were carried out in a GloMax 96 Microplate Luminometer (Promega).

*Expression analysis*

RNA isolation, reverse transcription and real-time PCR were performed as described (Lowry et al., 2008). Microarray profiling was performed with Affymetrix Human HG-U133 2.0 Plus arrays as described [26, 32]. Data were normalized with Robust Multichip Algorithm (RMA) in Genespring. Probe sets that were not expressed at a raw value of >50 in at least 10% of samples were eliminated from further analysis. Note that outside data was normalized and analyzed separately, but with identical methods described above. Later analysis on the NPC lineage, was performed after separate normalization and filtering of only relevant samples. Outside data sets were collected from Gene Expression Omnibus (NCBI) and include: GSE19735 (Rafii) [55], comparing hESC-derived endothelial cell (EC) to human umbilical vein endothelial cells (HUVEC) and human smooth muscle cells (SMC); GSE14897 (Duncan) [16], comparing undifferentiated hESCs and hiPSCs to hepatocytes made from each; GSE20013 (Wilson) [54], comparing endothelial cells made from hESCs to endothelial cells made form tissue (HUVEC); and GSE18887 (Fang) [49], comparing entire human embryo samples from 3-

5 weeks of development. All outside data were analyzed separately, but by employing the same methods. Fang et al. [49] utilized a U133 chip, which had fewer probe sets than our U133 2.0 plus chip. Only those probe sets that made it past the filtering on both chips were included in this analysis. Hierarchical Euclidian clustering was performed with complete linkage. Gene expression differences were judged to be significant if the p-value of the fold change was < 0.01 and at least 1.54 fold different between indicated samples. Gene expression differences were also judged by Benjamini-Hochberg correction for false discovery rate (FDR) at a p-value of <0.05 and 1.54 fold change (Fig S2). Further statistical analysis for hypergeometric distribution and 3-way simulation was performed with *R*, package 2.9.2 as described [12]. Heat maps were generated by averaging the raw value of like samples and representing it as a ratio of the average of the raw value of all samples (Cluster 3.0, Java TreeView). Pearson correlations were generated in Excel.

### 2.4 Results

#### *hESC and hiPSC lines are capable of generating derivatives representing all three embryonic germ layers*

Using previously established protocols, we differentiated hESCs and hiPSCs into derivatives representing all three embryonic germ layers: neural progenitor cells (ectoderm) [18], hepatocytes (endoderm) [20], and fibroblasts (mesoderm). In our studies, the efficiency of derivation of various cell types from PSCs (including hESCs and fibroblast-derived hiPSCs) was highly variable amongst various lines and across multiple experiments (data not shown and Karumbayaram et al. [18]). However, it was clear that all the PSC lines used here were able to generate a reasonable number of cells of the indicated type for analysis of gene expression and functional capacity.

*Generation of Ectoderm*

When hESC and hiPSC were directed to generate neural progenitor cells (NPCs) and isolated based on rosette morphology (p1), well-established neural stem cell markers (MUSASHI, PROMININ, PAX6, SOX2, and NESTIN) were induced as measured by RT-PCR (Fig S1A). Immunocytochemistry confirmed the expression of these and other NPC markers at the protein level (Fig 1A and S1A'). By immunostaining, all the NPC markers labeled as least 80% of cells, demonstrating that the culture represented a homogenous pool of NPCs. While hESC and hiPSC-derived NPCs appeared to express these genes at a similar level to one another, there was variation in the level of protein expression on a per-cell basis observed between PSC-NPCs and those isolated from 16 week fetal brain. The transcription factor PAX6 was expressed at a lower level in PSC-NPCs when compared to 16 week fetal brain-derived NPCs (FNPC-16 Br) on a per cell basis, while NESTIN was higher (Fig 2-1A). Because Retinoic acid (RA) and smoothened agonist (SAG) were used in the neural specification, it is possible that we induced a more posterior/ventral fate than NPCs isolated from fetal brain. To explore this possibility, the PCS-NPCs were also compared with NPCs isolated from 15.5 week fetal spinal cord (FNPC-15.5SC) and expanded under the same conditions. In fact, NESTIN and PAX6 were expressed in FNPC-15.5SCs at a similar level as in our PCS-NPCs (Fig 2-1A and Fig S2-1A'). By immunostaining and morphology, both PSC-NPCs and 16 week tissue-derived NPCs had the capacity to generate both Tuj1+ neurons and GFAP+ glia (Fig 2-1A'); however, the p1 PSC-NPCs mostly generated neurons (Tuj1+), while the 16 week tissue derived NPCs mostly produced glia (GFAP+)(Fig 2-1A''). Because neurogenesis precedes gliogenesis during *in vivo* development [33], these data suggest that PSC-NPCs may represent earlier developmental time points than the NPCs derived from 16 week fetal tissue.

*Generation of Endoderm*

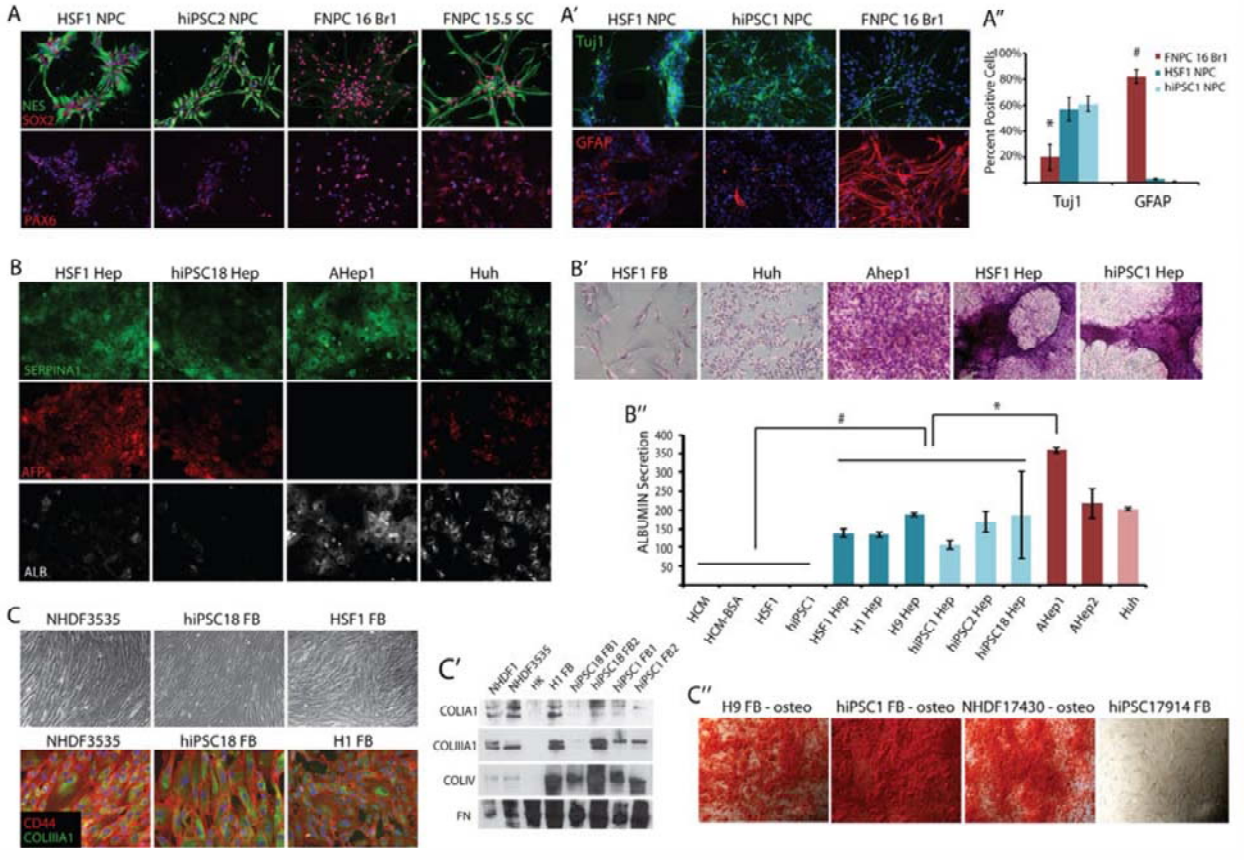



**Figure 2-1. hESC and hiPSC lines make cell types representing all three germ layers.**

hESC and hiPSC lines were directed to differentiate into either NPCs **(A)**, hepatocytes **(B)**, or fibroblasts **(C)**. **A,** Immunofluorescence staining for SOX2 (red, top panel), NESTIN (green, top panel), and DNA (blue) and PAX6 (red, bottom panel). **A',** Immunofluorescence staining demonstrating that NPCs derived from hESCs, hiPSC or natural sources could be differentiated into Tuj1+ neurons (green) and GFAP+ glia (red). **A'',** Quantification of the percent of cells undergoing neuronal (Tuj1+) or glial (GFAP+) differentiation. Error bars represent standard error over 5-8 fields of view. * indicates *P*<0.05; # indicates *P*<1.0E-06. **B,** Immunofluorescence staining of SERPINA1 (green), AFP (red), ALBUMIN (white) and DNA (blue). **B',** PAS-stain demonstrating glycogen storage in natural- and pluripotent-derived hepatocytes. **B'',** Elisa measuring albumin secretion on confluent plates. Error bars represent standard error over 2 replicates. * indicates *P*<0.01; # indicates *P*<0.05. **C, TOP,** Phase-contrast images of fusiform morphology displayed by pluripotent- and naturally-derived fibroblasts. **C, BOTTOM,** Immunofluorescence staining of CD44 (red), COLIIIA1 (green), and DNA (blue). **C',** Western blot for secreted collagen proteins (COLIA1, COLIIIA1, and COLIV) and FIBRONECTIN (FN). HK = Human Keratinocyte. **C'',** Alizarin Red stain following further differentiation of pluripotent- and tissue-derived fibroblasts down the osteogenic lineage.

For hepatocyte derivation, hESC and hiPSC were directed to undergo definitive endoderm formation, hepatic specification, hepatoblast expansion and finally hepatic maturation as described previously [16, 20]. After three days of differentiation, definitive endoderm markers HNF3B and SOX17 were induced, as demonstrated by RT-PCR (Fig S2-1B). Following hepatic induction and expansion, the SOX17 mRNA level declined, while various hepatic markers (AFP, ALB, SERPINA1, CYP3A4, CYP3A7) were induced over time, as found in mature adult tissue derived hepatocytes (AHep) or a hepatocarcinoma cell line (Huh) (Fig 2-1B and Fig S1B). By morphology and immunostaining, the hepatocytes produced from PSCs were more similar to hepatoblasts, or immature hepatocytes that populate the developing early fetal liver (Fig 2-1B and Fig S2-1B). These hepatic derivatives expressed a higher level of fetal hepatic genes such as *AFP* and *CYP3A7*, and a lower level of the more mature equivalents *ALBUMIN* and *CYP3A4*, when compared to counterparts made from adult liver (Fig 2-1B and Fig S2-1B). Again, this suggested that PSC-Heps could represent a younger developmental stage than those of adult liver. Regardless, the hepatocytes generated from PSCs were able to both store Glycogen (Fig 2-1B') and secrete albumin (Fig 2-1B'').

*Generation of Mesoderm*

For fibroblast (FB) generation, embryoid bodies (EBs) were first induced from hESCs and hiPSCs. The EBs were then plated in adherent conditions and grown in fibroblast culturing media. Following several passages, these cultures began to display a homogenous fusiform morphology typical of fibroblasts (Fig 2-1C). For comparison, fibroblasts were derived from human dermis or lung at various developmental time points and were grown *in vitro* under the same conditions. When characterized at the RNA level, fibroblast markers were expressed in PSC-FBs at a level comparable to fibroblasts derived from the dermis of skin (Fig S2-1C). Furthermore, using immunocytochemistry, two fibroblast markers, CD44 and COLIIIA1, were

expressed in PSC-FBs at levels comparable to that observed in a neonatal dermal fibroblast line (Fig 2-1C). Functionally, the PSC-FBs secreted a profile of collagens similar to that secreted by dermal fibroblasts, while human keratinocytes (HK) did not (Fig 2-1C'). In addition, the PSC-FBs, like their natural counterparts, were not necessarily terminally differentiated as they still retained the ability to undergo osteogenic specification (Fig 2-1C'').

***Global characterization of PSC derivatives by gene expression profiling***

To more precisely determine the identity of the derivatives of PSCs, each of the indicated cell types was purified and profiled for gene expression. We analyzed the transcriptional profiles of PSC-derivatives as well as natural counterparts, and performed unsupervised hierarchical clustering (Fig 2-2A). From this clustering analysis we made several important observations.

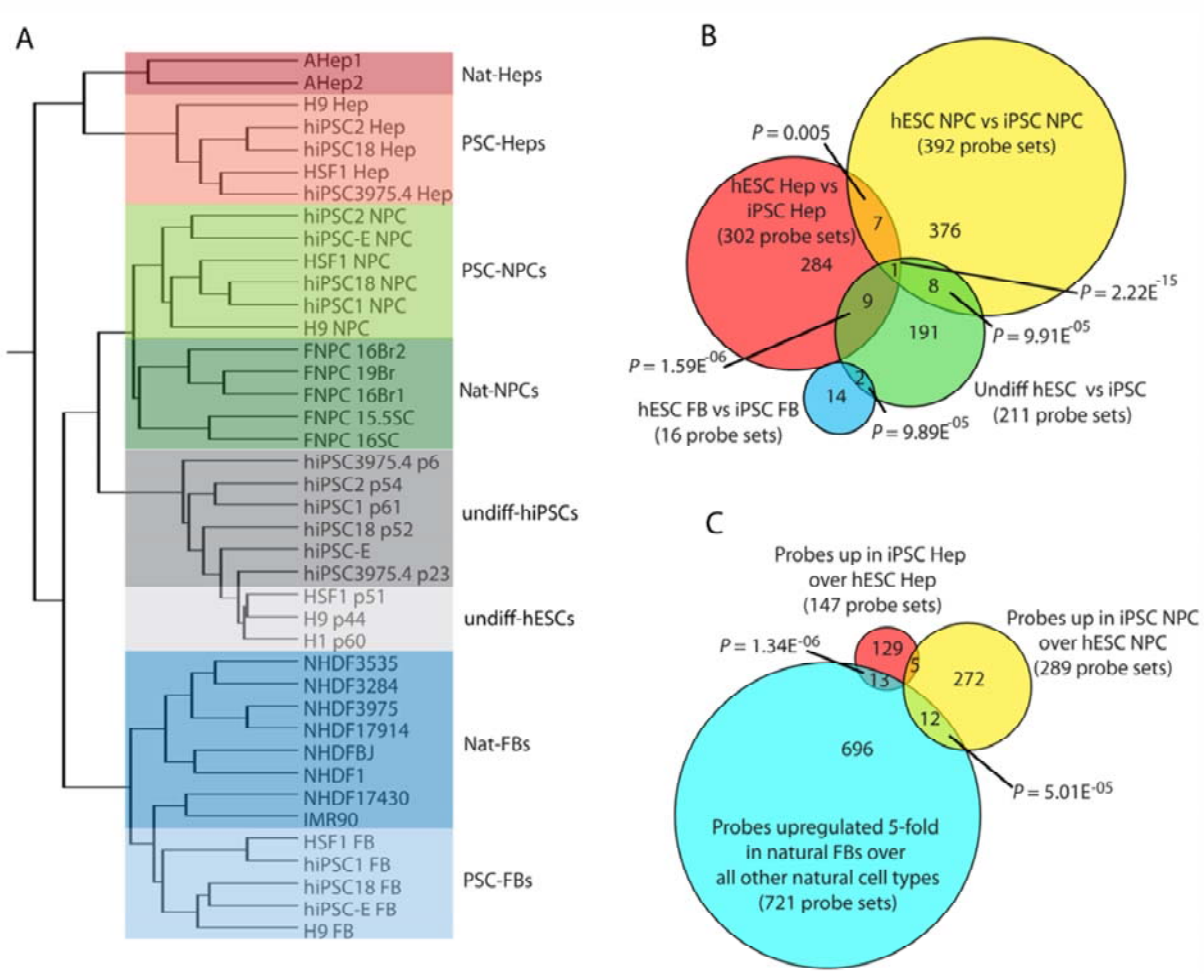
First, most of the gene expression differences observed between hiPSCs and hESCs in the undifferentiated state were not found when hESC and hiPSC were differentiated, evident by the fact that hESC- and hiPSC-derived progeny did not segregate from one another. Second, while a small number of significant differences was detected between hESC progeny and that of hiPSCs for each lineage (roughly 300), these differences were not detected when a more stringent statistical measure (FDR) was applied (Fig 2-2B and data not shown). Furthermore, very few of these differences overlapped with the list of genes differentially expressed between undifferentiated hESCs and hiPSCs (Fig 2-2B), and none of them could be categorized with gene ontological (GO) analysis. Whether these differences serve to functionally distinguish hESC from hiPSC derivatives will require extensive investigation of differentiated progeny from hiPSCs derived from a variety of different types of somatic cell types.

Recent evidence suggests that murine and human iPSCs retain an epigenetic signature from their cell of origin that influences their differentiation potential [8-11]. All of the hiPSCs used in our analysis were derived from fibroblasts, making similar types of analysis in our model system impossible. We did, however, look for lingering expression of fibroblast specific genes in hiPSC derivatives versus hESC derivatives and found a handful of genes that might represent residual gene expression from the target cell of reprogramming (Fig 2-2C). We also ruled out that any gene expression difference was due to expression of integrated reprogramming factors in the hiPSCs because the reprogramming factors were not significantly expressed in these progeny, except for SOX2, as would be expected for NPCs (Fig 2-2D).

The most striking observation made from our transcriptome analysis was that, while PSC derivatives representing the three germ layers clustered with their respective natural counterparts, within each germ layer cluster, the PSC progeny were always distinguished from their naturally derived counterparts (Fig 2-2A). This finding suggested that, with the methods employed here and commonly applied elsewhere, PSC derivatives are similar but not identical to tissue-derived counterparts. This led to further investigation of the transcriptional differences between hPSC derivatives and their tissue derived counterparts, and whether these differences are reflective of diversities in developmental maturity or arise due to development performed *in vitro*.

***PSC derivatives and tissue derived counterparts are distinguished by their gene expression***

The genes differentially expressed between PSC derivatives and their tissue-derived counterparts were compared using a t-test ($p<0.01$) and requiring at least a 1.54-fold expression difference. Out of 36749 probe sets that were expressed in at least 10% of samples, 2922 were




D

| Affy Probe Set | | | hESC (3) | hiPSC (3) | Nat-Hep (2) | hiPSC-Hep (2) | Nat-FB (8) | hiPSC-FB (2) | Nat-NPC (5) | hiPSC-NPC (3) |
|---|---|---|---|---|---|---|---|---|---|---|
| 208286 | *OCT4* | Total | 11303.2 | 8332.5 | 216.8 | 159.4 | 126.8 | 173.8 | 123.4 | 160.9 |
| 214532 | *OCT4* | Total | 7223.5 | 6755.6 | 196.9 | 160.5 | 143.9 | 195.5 | 130.9 | 164.3 |
| 202431 | *C-MYC* | Total | 961.2 | 1570.9 | 876.3 | 752.9 | 1203.9 | 2198.5 | 683.0 | 591.3 |
| 213721 | *SOX2* | Total | 1384.7 | 1471.8 | 55.1 | 87.9 | 43.1 | 60.5 | 2498.7 | 2085.6 |
| 228038 | *SOX2* | Endogenous | 7792.5 | 6009.9 | 17.2 | 227.5 | 17.4 | 9.9 | 7734.2 | 6608.8 |
| 213722 | *SOX2* | Endogenous | 276.8 | 201.5 | 37.6 | 43.7 | 43.3 | 47.3 | 393.3 | 372.6 |
| 220184 | *NANOG* | Endogenous | 5685.4 | 6432.9 | 44.2 | 32.6 | 28.5 | 25.6 | 48.7 | 61.6 |
| 221841 | *KLF4* | Endogenous | 357.3 | 250.8 | 886.7 | 776.1 | 877.2 | 162.2 | 143.3 | 76.1 |

**Figure 2-2. Global gene expression analysis.**

**A,** Hierarchical clustering analysis of global gene expression in undifferentiated hESCs, hiPSC, and their progeny compared to naturally derived cells. **B,** Venn diagram summarizing the probe sets that were differentially expressed (t-test $P<0.01$; fold change ≥ 1.54) between the progeny of hiPSCs versus the progeny of hESCs for each germ layer and the undifferentiated. **C,** Venn diagram overlapping Fibroblast signature probe sets (t-test between natural-FB and all other natural cell types; Upregulated in FBs ≥ 5.0) with probe sets upregulated in iPSC progeny over ESC progeny for the NPC and Hep lineages. *P*-values from B and C were measured by hypergeometric distribution or simulation as in [2]. **D,** Normalized values from microarray probe sets for the reprogramming factors used to make the hiPSCs used in this study.

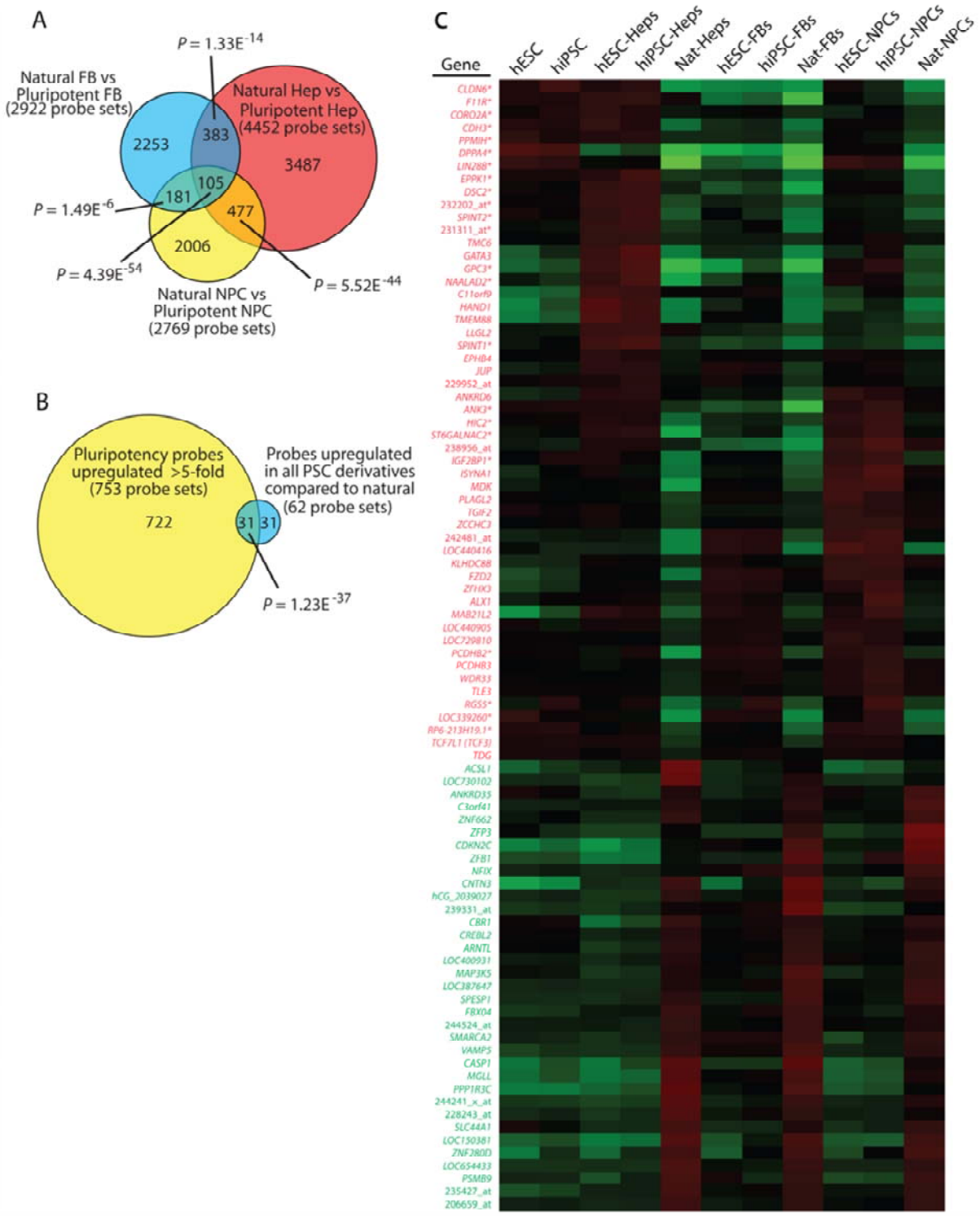

A
P = 1.33E-14
Natural FB vs
Pluripotent FB
(2922 probe sets)
Natural Hep vs
Pluripotent Hep
(4452 probe sets)
2253
383
3487
105
181
477
2006
P = 1.49E-6
P = 4.39E-54
P = 5.52E-44
Natural NPC vs
Pluripotent NPC
(2769 probe sets)
B
Pluripotency probes
upregulated >5-fold
(753 probe sets)
722
31
31
Probes upregulated
in all PSC derivatives
compared to natural
(62 probe sets)
P = 1.23E-37
C
Gene
hESC
hiPSC
hESC-Heps
hiPSC-Heps
Nat-Heps
hESC-FBs
hiPSC-FBs
Nat-FBs
hESC-NPCs
hiPSC-NPCs
Nat-NPCs
CLDN6*
F11R*
CORO2A*
CDH3*
PPMIH*
DPPA4*
LIN28B*
EPPK1*
DSC2*
232202_at*
SPINT2*
231311_at*
TMC6
GATA3
GPC3*
NAALAD2*
C11orf9
HAND1
TMEM88
LLGL2
SPINT1*
EPHB4
JUP
229952_at
ANKRD6
ANK3*
MIC2*
ST6GALNAC2*
238956_at
IGF2BP1*
ISYNA1
MDK
PLAGL2
TGIF2
ZCCHC3
242481_at
LOC440416
KLHDC8B
FZD2
ZFHX3
ALX1
MAB21L2
LOC440905
LOC729810
PCDHB2*
PCDHB3
WDR33
TLE3
RGS5*
LOC339260*
RP6-213H19.1*
TCF7L1 (TCF3)
TDG
ACSL1
LOC730102
ANKRD33
C3orf41
ZNF662
ZFP3
CDKN2C
ZFB1
NFIX
CNTN3
hCG_2039027
239331_at
CBR1
CREBL2
ARNTL
LOC400931
MAP3K5
LOC387647
SPESP1
FBXO4
244524_at
SMARCA2
VAMP5
CASP1
MGLL
PPP1R3C
244241_x_at
228243_at
SLC44A1
LOC150381
ZNF280D
LOC654433
PSMB9
235427_at
206659_at

**Figure 2-3. Expression profiling identifies a conserved list of probe sets differentially expressed between pluripotent derivatives and their natural counterparts.**

**A,** A T-test ($p<0.01$) was performed to identify probe sets differentially expressed between PSC derivatives (FB = Blue; Hep = Red; NPC = Yellow) and their respective natural counterparts (fold change ≥ 1.54). Venn diagram reveals the overlap of these differences across the different progeny. **B,** Overlap of the 62 probe sets specifically upregulated in 3A with probe sets that demonstrate a significant difference between pluripotent cells and naturally derived somatic cells (Fold upregulation ≥ 5). *P*-values from A and B measured by hypergeometric distribution or simulation as in [2]. **C,** A heat map was generated for the 88 unique genes or ESTs represented by the 105 probe sets shown in Fig 3A that are differentially expressed between PSC derivatives and tissue derived cells. Note that signal shown represents value divided by the average of all samples and genes in red were consistently found upregulated in PSC progeny versus tissue derived cells, while those in green were always downregulated. * indicates genes expressed highly in the pluripotent stem cells (identified in B).

differentially expressed between PSC-FBs and dermal/lung fibroblasts; 4452 were differentially expressed between PSC-Heps and adult hepatocytes; and 2769 were differentially expressed between PSC-NPCs and 16 week fetal NPCs. Gene Ontological (GO) analysis of the genes differentially expressed in each germ layer representative yielded many categories for hepatocytes, but few for NPCs and FBs (Fig S2-2D). Surprisingly, when superimposing these pools of differentially expressed probe sets and taking direction of differential expression into account, 105 were found to be differentially expressed between all PSC derivatives and their natural counterparts (Fig 2-3A), suggesting that all types of PSC-derivatives share common differences with tissue derived cells. No GO terms were conserved across these 105 probesets (Fig S2-2D). A more stringent analysis employing FDR correction produced a smaller list of probe sets, but the differences still overlapped significantly across the three germ layers and were entirely inclusive with the non-FDR corrected analysis (Fig S2-2A).

From this pool of 105 differentially regulated probe sets, 62 were upregulated in all PSC derivatives versus their tissue-derived counterparts. Notably, 31 of these 62 probe sets overlap with probe sets that are highly upregulated in undifferentiated PSCs versus specified somatic cells (Fig 2-3B). This indicates that the PSC progeny continued to express a significant subset of genes associated with either pluripotency or early embryonic development. Many of these same genes were also expressed at a higher level in PSC derivatives versus tissue-derived cells generated independently by other groups, suggesting that these observations were not specific to our methods or the particular cell lines used here (Fig S2-2C). *OCT4*, *SOX2*, and *NANOG* were not amongst the 31 probe sets related to pluripotency that remained high in PSC derivatives, demonstrating that these genes were silenced upon differentiation, as has been shown extensively. Instead, most of the 31 probe sets appeared to be not only expressed in PSCs, but also play roles in early embryonic development as judged by functional data from

lower organisms [34-36]. The expression pattern of all genes specifically expressed in PSCs can be found in Supplemental Table 2-1 (http://www.nature.com/cr/journal/v22/n1/extref/cr2011133x3.pdf.

The expression pattern across cell types for the 88 unique genes and unannotated probe sets differentially expressed between PSC derivatives and tissue-derived counterparts (represented by 105 probe sets) is represented in Figure 3C as a heat map. Of the 53 genes that were higher in PSC derivatives (in red), 22 were also strongly expressed in undifferentiated PSCs relative to somatic cells (indicated with asterisk). This list included *LIN28B, DPPA4,* and *TCF7L1 (TCF3)*, all of which are known to play role in ESCs and very early mammalian development [35, 37-40]. Furthermore, 35 genes were not induced upon differentiation to any cell type (in green), perhaps reflecting a state of incomplete specification regardless of the cell type generated.

*LIN28* was first discovered as a regulator of developmental timing in *C. elegans [36, 41]*. *LIN28A* and *LIN28B* are highly expressed in undifferentiated hPSCs, but are thought to be silenced as tissues are specified and mature. *LIN28A* has also been employed as a reprogramming factor in the generation of hiPSCs [42], suggesting it can play a functional role in maintaining or inducing immature cell fate. *LIN28B* was expressed in all hESC and hiPSC-derivatives, whereas *LIN28A* was found at a high level in PSC-NPCs and PSC–Heps, but not PSC-FBs. *LIN28A, LIN28B* and *DPPA4* were also frequently expressed at a high level in PSC derivatives generated by other groups (Fig S2C), indicating that expression of these genes are not confined to just the hESC and hiPSC derivatives produced in our lab, but is perhaps an attribute of PSC derivatives in general.

To determine whether the changes detected in the RNA profiling data was just due to residual stabilized RNA from the pluripotent state, the expression of LIN28A, LIN28B and DPPA4 in PSC derivatives was also examined at the protein level. In fact, DPPA4, LIN28A, LIN28B were all expressed at the protein level in PSC-NPCs (Fig 2-4A-B) and PSC–Heps (Fig 2-4C), but not in their tissue-derived counterparts. On the other hand, none of the other classic pluripotency factors (OCT4, NANOG, REX1/ZFP42) were expressed in any of the PSC derivatives at the RNA or protein levels (Figure 2-4A-C, and data not shown).

***The LIN28/let-7 circuit in PSC derivatives***

*LIN28* has been shown to act as an RNA binding protein that regulates miRNA maturation, particularly the let-7 family [34, 36-38, 43-47]. If *LIN28* activity is higher in PSC derivatives, these cells would be expected to have low levels of mature let-7. To determine if the increased *LIN28* expression in PSC-NPCs correlated with low levels of mature let-7, RT-PCR was used to probe the relative levels of mature let-7 miRNA family members. As expected, PSC-NPCs were found to have very low relative levels of mature let-7 family members compared to tissue derived NPCs (Fig 2-4D).

To determine if the low levels of let-7 expression in PSC-NPCs also correlated with low activity of these miRNAs, cells were transfected with a reporter that drives constitutive expression of the renilla luciferase gene with let-7 seed sequences added to its 3' UTR[48]. Therefore, higher luciferase activity is a result of decreased let-7 miRNA activity in the cells. NPCs generated from 16 week fetal brain or spinal cord showed very little reporter activity, indicating high let-7 activity, whereas PSC-NPCs displayed high reporter activity, indicating the opposite (Fig 2-4E).

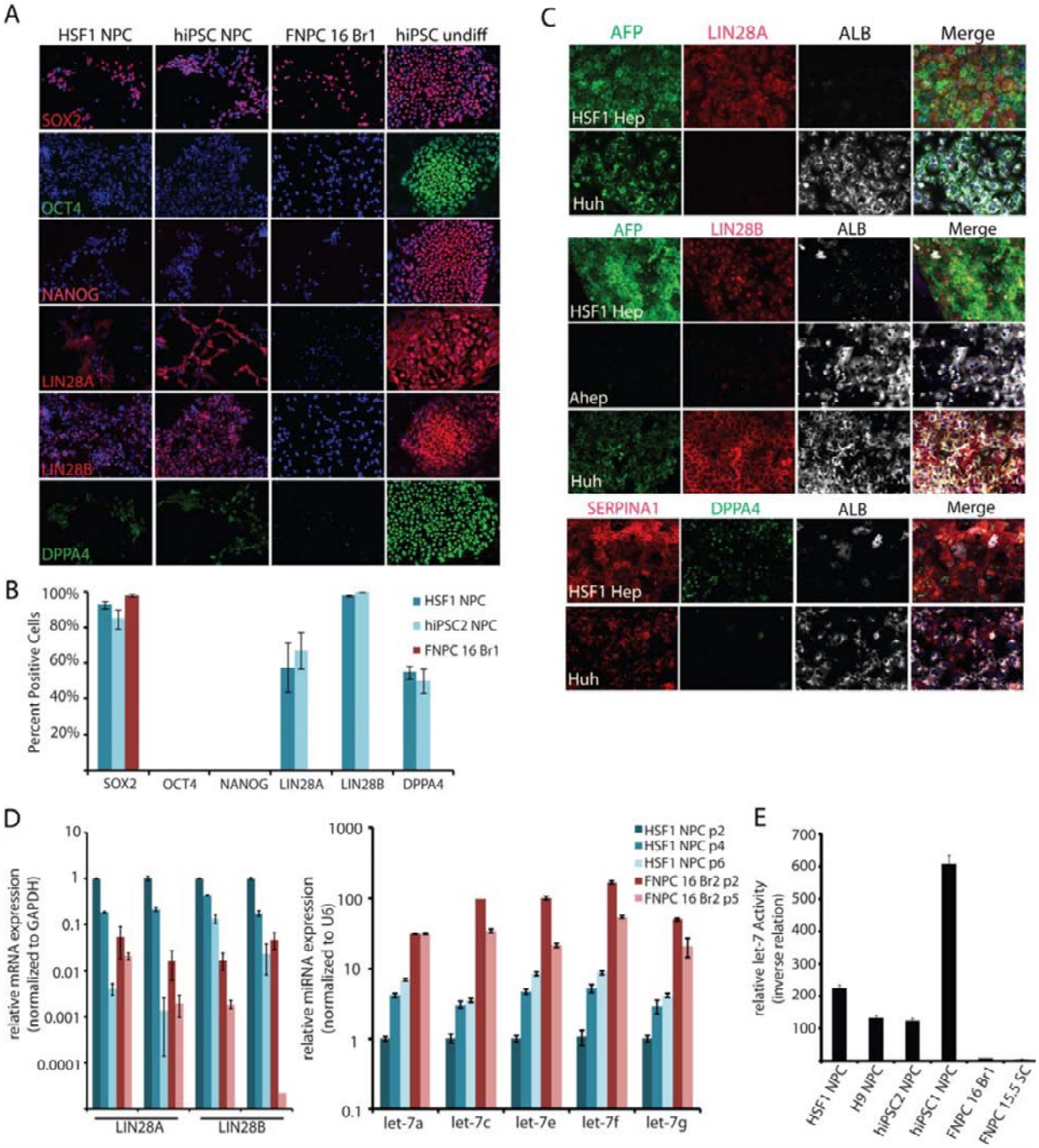



**Figure 2-4. Expression and activity of LIN28and DPPA4 in PSC derivatives.**

**A,** NPCs made from PSCs and brain were stained with pluripotency markers SOX2, OCT4, NANOG, LIN28A, LIN28B, and DPPA4. Undifferentiated hiPSCs were stained as a positive control for the pluripotency markers. **B,** Quantification of the percent of FNPCs and PSC-NPCs expressing the indicated pluripotency markers. **C,** HSF1 derived hepatocytes and control cells were immunostained with antibodies recognizing ALBUMIN, AFP, or SERPINA1 to highlight both immature and mature cells and

either LIN28A, LIN28B, DPPA4 to demonstrate that these pluripotency factors are not silenced immediately upon differentiation. Hepatocytes taken from adult human liver did not express any of these pluripotency genes, while Huh, a hepatocarcinoma cell line expressed LIN28B. **D,** Real time PCR for LIN28A and LIN28B mRNA (left) and let-7 miRNA family members (right). mRNA expression was normalized to GAPDH, while miRNA expression was normalized to U6. Error bars represent standard error over 3 or 4 replicates. **E,** to determine the relative let-7 activity in the indicated cell types, each was transfected with let-7 reporter and constitutive reporter as a transfection control. Dual luciferase assays were performed 48 hours after transfection in triplicate. Assay shown was representative of three independent experiments.

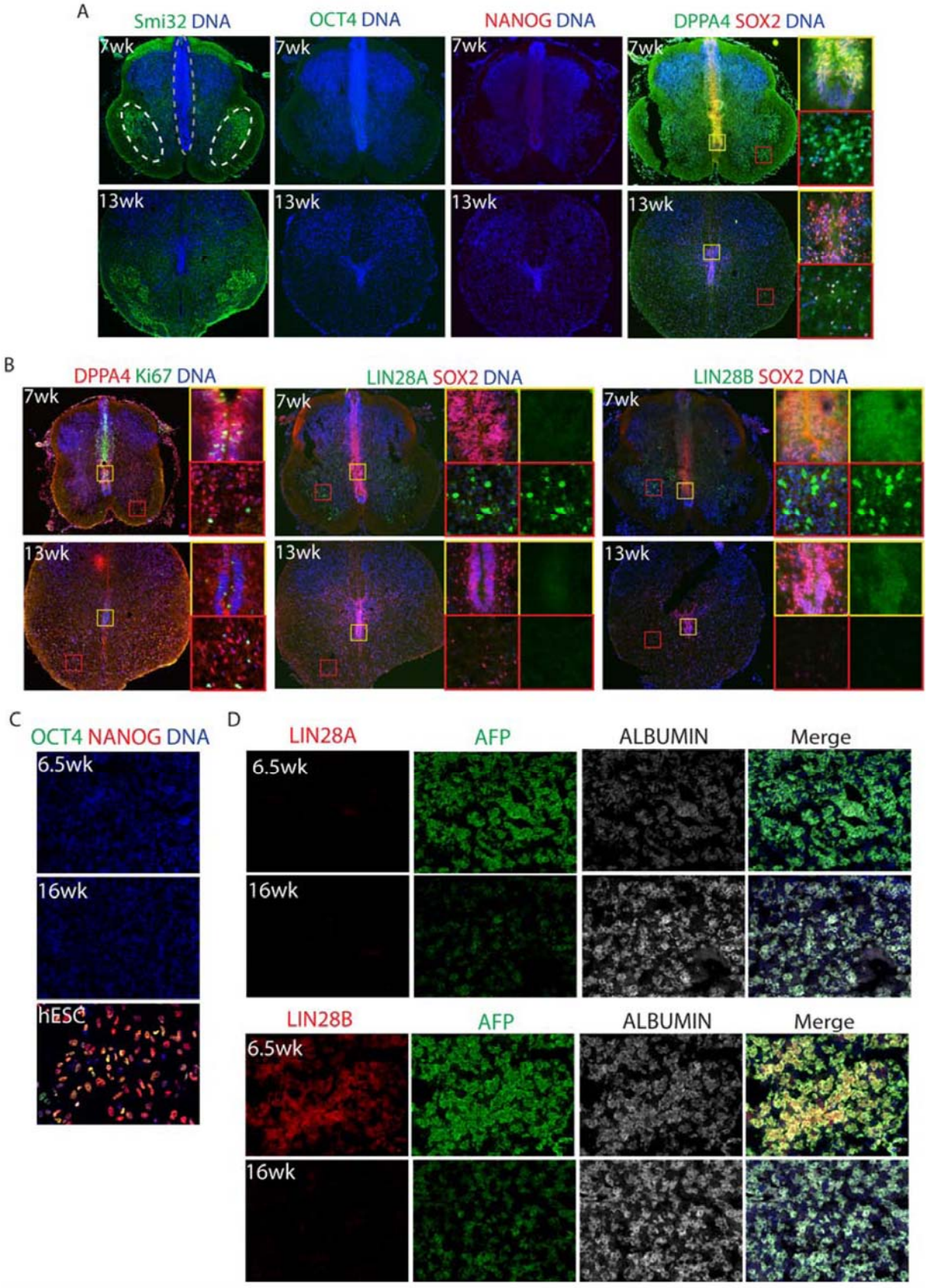

A
Smi32 DNA
OCT4 DNA
NANOG DNA
DPPA4 SOX2 DNA
7wk
13wk
B
DPPA4 Ki67 DNA
LIN28A SOX2 DNA
LIN28B SOX2 DNA
C
OCT4 NANOG DNA
6.5wk
16wk
hESC
D
LIN28A
AFP
ALBUMIN
Merge
LIN28B

**Figure 2-5. Differentially expressed genes, DPPA4, LIN28A and LIN28B are found in early fetal tissues.**

**A,** 7- and 13- week spinal cord tissues were fixed, sectioned and stained with the indicated markers. Smi32 was used to highlight the motor neuron pool (white circles). SOX2 labels the NPCs found along the midline (grey outline). **B,** DPPA4 co-localized with proliferation marker Ki67 along the midline (yellow inset) at 7 weeks, but not in the dispersed lateral stain (red inset). At 13 weeks, DPPA4 was more dispersed and the number of Ki67 positive cells was decreased. LIN28A had a cytosolic staining pattern and was located in cells outside of the midline (red and yellow insets) at 7 wks, but absent at 13 weeks. LIN28B had a nuclear staining pattern at 7 wks. It was often co-localized with SOX2 along the midline at 7 weeks and was weaker at 13 weeks (yellow inset). **C,** 6.5- and 16 week fetal liver were stained with the indicated pluripotency markers, showing the complete silencing of pluripotency genes. hESCs were used as a positive control. **D,** 6.5- and 16-week fetal liver were stained with LIN28A, LIN28B and the indicated liver markers.

Considering the *LIN28* expression pattern, and the let-*7* activity assay, PSC derivatives not only had high *LIN28* expression, but also high *LIN28* activity, which in turn led to low let-*7* activity.

While the *LIN28/let-7* pathway has been implicated in development of many species including mouse [34, 37], this appears to be the first demonstration that human PSC derivatives have high *LIN28* expression and low let-*7* activity. These data further suggest PSC-derivatives are developmentally immature compared to their natural counter parts and that caution is warranted in clinical application of these cells, as many human cancers are defined by high *LIN28* and low let-7 expression [34, 46].

***Expression of early embryonic genes in PSC derivatives suggests that they represent early stages of human development***

To determine whether expression of the *LIN28* genes and *DPPA4* in PSC derivatives was indicative of very early fetal development or simply an *in vitro* phenomemon, fetal tissues were also probed for expression of these genes. In the human fetal spinal cord, regional identity was established with Smi32, a marker of motor neurons, which was localized to bilateral pools on each side of the ventral cord (Fig 2-5A). The neural progenitor pool was localized at the midline and expressed SOX2 and Ki67 (Fig 2-5A and 2-5B). DPPA4 was also strongly expressed in the midline of the spinal cord along with SOX2 at 7 weeks of development, but was weaker by 13 weeks (Fig 2-5B). LIN28A was expressed in scattered cells in 7 week spinal cord, but was lost by 13 weeks. LIN28B, on the other hand, was strongly detected in 7 week human spinal cord cells outside of the midline and weakly expressed in the midline progenitor cells along with SOX2. Expression of LIN28B was significantly reduced, but not absent in the spinal cord by 13 weeks of development (Fig 2-5C). The fact that midline progenitor cells of the human spinal cord did express DPPA4 and LIN28B at 7 weeks could further suggest that PSC-NPCs are

more similar to an earlier stage of fetal development, or that multiple pools of diverse progenitors are present at this time point.

In the fetal liver, LIN28B but not LIN28A was detectable at 6.5 weeks (Fig 2-5D). Neither LIN28 protein was detectable in fetal liver by 16 weeks (Fig 2-5D). As expected, neither OCT4 nor NANOG were expressed in the spinal cord or fetal liver at any time point analyzed, consistent with the notion that these pluripotency genes are silenced very early in human development (Fig 2-5A and 2-5C). These data highlight the possibility that PSC derivatives differed from the tissue derived cells as shown in Figure 2 and 3 because the cells were taken from tissues that were at least 16 weeks of development. The presence of both LIN28 proteins and DPPA4 in 7 week spinal cord and liver, the fact that PSC-NPCs tended to be neurogenic as opposed to gliogenic, and that PSC-Heps express AFP instead of Albumin are all consistent with the notion that pluripotent derivatives are similar to cells found at 7 weeks of development or earlier.

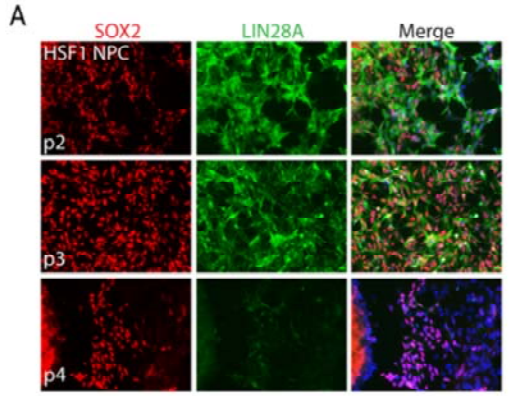
***Are PSC-NPCs distinguished from fetal NPCs by time in vitro or time in vivo?***

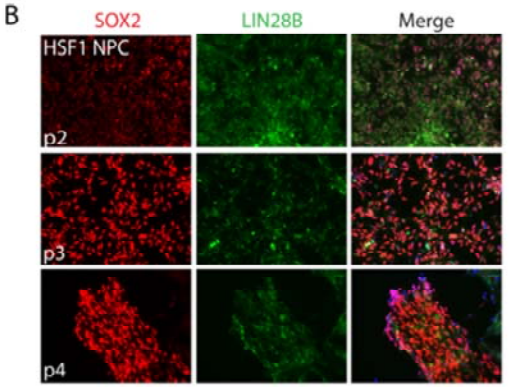
Following the observation that early embryonic genes were expressed in younger fetal samples, but significantly reduced at later stages, we were prompted to ask two additional questions: 1) Could additional time in culture bring the PSC derivatives closer to a natural counterpart; and 2) Would younger fetal samples more closely resemble our PSC derivatives on a transcriptome level? To answer these questions we performed a series of additional gene expression profiling experiments on the NPCs.

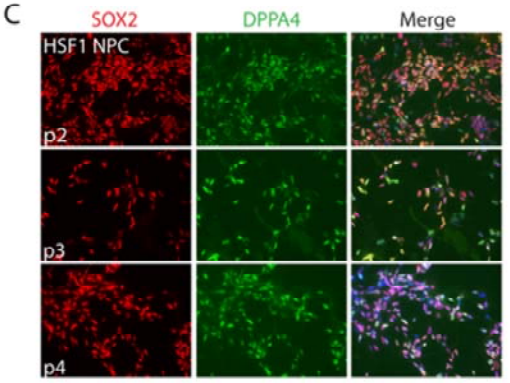
When PSC-NPCs were cultured for another month (each passage representing ~5-7 days in culture), both *LIN28A* and *LIN28B* mRNA were reduced as measured by RT-PCR analysis (Fig 2-4D). Furthermore, let-7 family members were upregulated with extended culture (Fig 2-4D).

The reduction of LIN28A and LIN28B was also observed at the protein level, while DPPA4 did not seem to change with passaging (Fig 2-6A-C). In order to determine if continued passaging brought PSC- NPCs globally closer to 16 week fetal spinal cord derived NPCs, we performed two pearson correlations comparing the transcriptomes of the indicated samples. When using only those probe sets identified by the original t-test (Fig 2-3A; 2769 less 46 probes that did not make it past the new filtering), we observed a small increase in similarity between PSC-NPCs and fetal derived NPCs (Fig 2-6E). In addition, we observed that a small but statistically significant number of the original 2723 probe sets were "corrected" upon extended passaging, including *LIN28A* and *LIN28B* (Fig 2-6F). Ultimately, when all probe sets were considered, we observed only a modest increase in global transcriptome similarity with extended passaging (Fig 2-6D), suggesting that simply culturing PSC-NPCs does not generate cells that are equivalent to their tissue derived counterparts.

To functionally determine whether passaging PSC-NPCs promotes their developmental maturity, we compared the differentiation potential of PSC-NPCs before and after continued culturing. This analysis indicated that culturing PSC-NPCs for an additional month did increase their gliogenic capacity from <1% to ~15% (Fig 2-6G), but not to levels typical of any tissue-derived NPCs we have tested to date (~50-80% gliogenic with NPCs derived from fetal tissue at 6.5-19 weeks of development (Fig 2-1A'' and data not shown). These data further suggest that while continued culture can bring PSC-NPCs closer to their tissue derived counterparts, this effect is small, and by itself is not sufficient to generate cells equivalent to tissue derived NPCs. In addition, continued culture of PSC-NPCs in these conditions beyond 2 months (passage 4) led to the subsequent loss of NPC markers and differentiation capacity (data not shown). This suggests that, under these conditions, one cannot simply passage the cells indefinitely and expect continued developmental maturation *in vitro*.



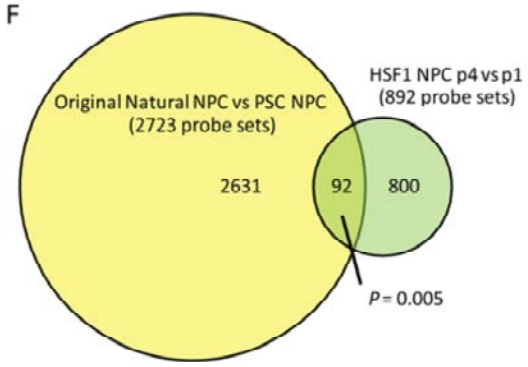




D

| | HSF1 NPC p1 (1) | HSF1 NPC p1 (2) | HSF1 NPC p2 (1) | HSF1 NPC p2 (2) | HSF1 NPC p4 (1) | HSF1 NPC p4 (2) |
|---|---|---|---|---|---|---|
| HSF1 NPC p1 (1) | 1.000 | | | | | |
| HSF1 NPC p1 (2) | 0.996 | 1.000 | | | | |
| HSF1 NPC p2 (1) | 0.974 | 0.977 | 1.000 | | | |
| HSF1 NPC p2 (2) | 0.970 | 0.976 | 0.989 | 1.000 | | |
| HSF1 NPC p4 (1) | 0.969 | 0.971 | 0.992 | 0.978 | 1.000 | |
| HSF1 NPC p4 (2) | 0.964 | 0.970 | 0.981 | 0.992 | 0.984 | 1.000 |
| FNPC 15.5 SC | 0.901 | 0.901 | 0.902 | 0.900 | 0.918 | 0.916 |
| FNPC 16 SC | 0.897 | 0.897 | 0.893 | 0.890 | 0.914 | 0.912 |

E

| | HSF1 NPC p1 (1) | HSF1 NPC p1 (2) | HSF1 NPC p2 (1) | HSF1 NPC p2 (2) | HSF1 NPC p4 (1) | HSF1 NPC p4 (2) |
|---|---|---|---|---|---|---|
| HSF1 NPC p1 (1) | 1.000 | | | | | |
| HSF1 NPC p1 (2) | 0.992 | 1.000 | | | | |
| HSF1 NPC p2 (1) | 0.873 | 0.878 | 1.000 | | | |
| HSF1 NPC p2 (2) | 0.873 | 0.887 | 0.979 | 1.000 | | |
| HSF1 NPC p4 (1) | 0.856 | 0.857 | 0.951 | 0.929 | 1.000 | |
| HSF1 NPC p4 (2) | 0.848 | 0.863 | 0.931 | 0.955 | 0.967 | 1.000 |
| FNPC 15.5 SC | 0.479 | 0.474 | 0.441 | 0.443 | 0.552 | 0.551 |
| FNPC 16 SC | 0.536 | 0.533 | 0.497 | 0.500 | 0.603 | 0.601 |



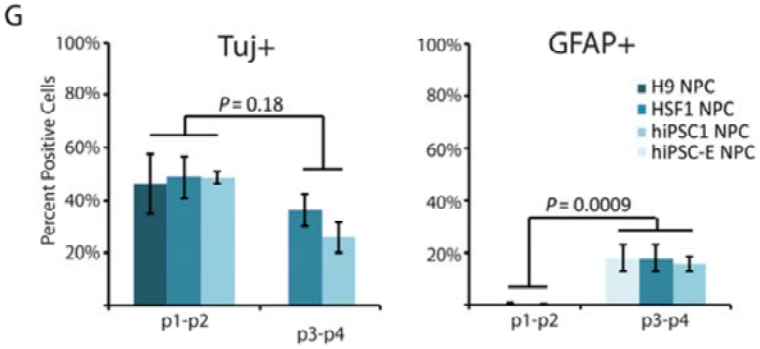



**Figure 2-6. Continued passaging of PSC-NPCs reduces LIN28 expression and corrects a small portion of the gene expression discrepancies.**

**A-C,** Immunofluorescence staining of HSF1 derived NPCs over 4 passages. Each passage represents ~5-7 days in culture. **D,** Pearson correlation comparing global gene expression between HSF1 NPCs over several passages and NPCs derived from 16 wk fetal spinal cord. **E,** Pearson correlation including only those probe sets identified as different between PSC-NPC and Nat-NPCs (analysis from Figure 3a). **F,** Venn diagram demonstrating the original differences identified in Figure 3a overlap significantly with gene expression differences between p1 and p4 PSC-NPCs (t-test $p<0.01$; Fold change $\geq 1.54$). Direction of differential expression was taken into account. Statistical analysis performed by hypergeometric distribution. Note: Later analyses were performed by normalizing and filtering only samples of the neural lineage. As a result the original 2769 probe sets identified by analysis in Figure 3a were reduced to 2723. **G,** Percent of PSC-NPCs at the indicated passage undergoing neuronal (Tuj1) or glial (GFAP) differentiation following three weeks of culture.

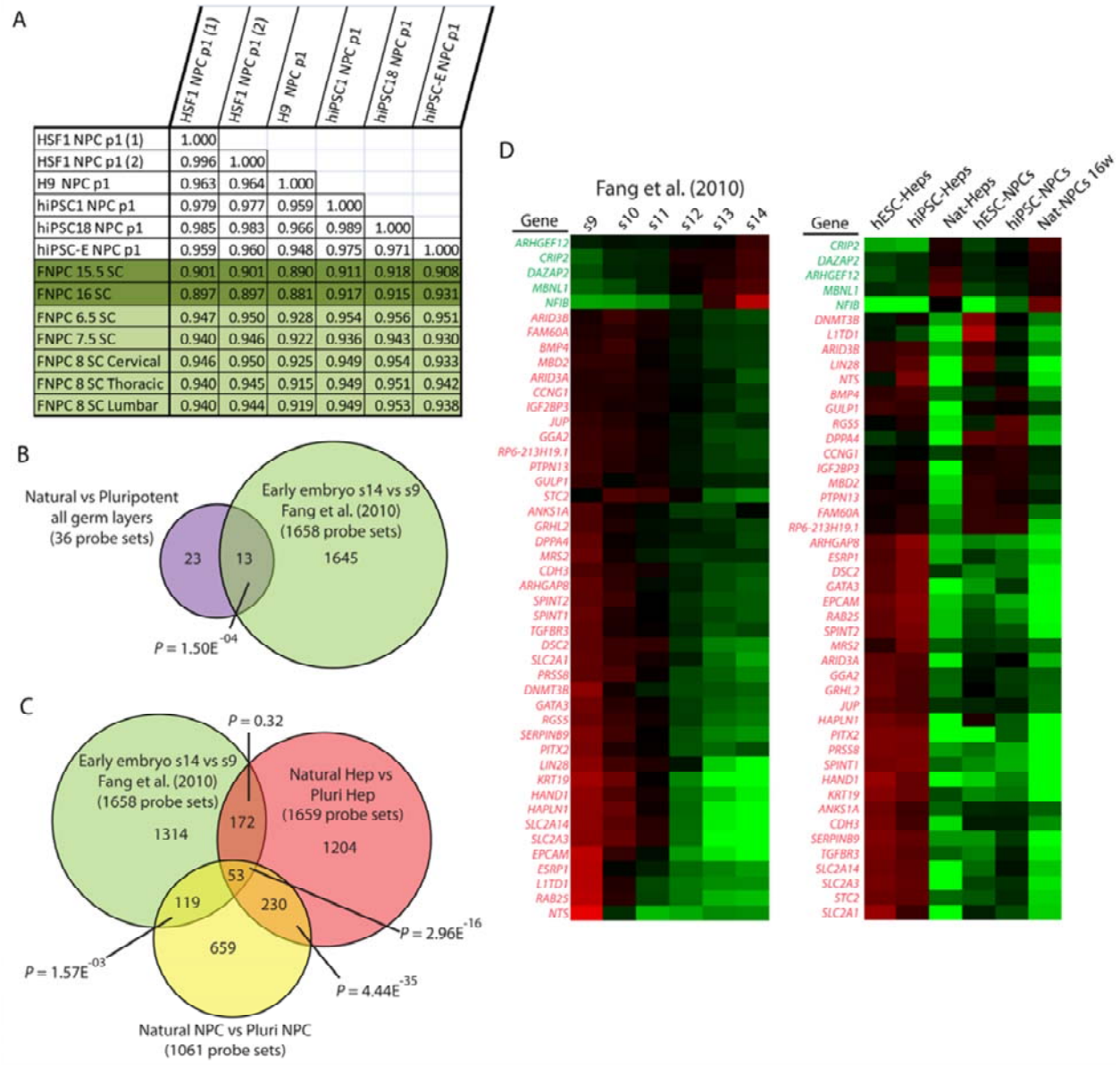



**Figure 2-7. Evidence that PSC derivatives reflect cell types found prior to 6 weeks of development.**

**A,** Pearson correlation comparing global gene expression between PSC-NPCs and fetal spinal cord NPCs. **B,** Comparison of the original probe sets identified as different between PSC derivatives and their natural counterpart (Figure 3a) and those differentially expressed between stage 9 and stage 14 embryos (Fang et al., 2010; t-test $P<0.01$, Fold change ≥ 1.54). **C,** Venn diagram comparing the probe sets different between PSC derivatives and their tissue derived counterparts for the Hep and NPC lineages and those differentially expressed between stage 9 and stage 14 embryos. **D,** Heat maps generated for the 46 unique genes represented by the 53 probe sets shown in Fig 7C. **Left,** Samples include the 6 stages of embryonic development represented in Fang et al. 2010. **Right,** Samples include PSC-derived NPCs and Heps and their respective natural counterparts. Green probe sets represent those genes upregulated over the course of development, while red probe sets are those downregulated over the course of development.

To determine if PSC- NPCs would more closely resemble an NPC isolated from younger fetal spinal cord, we profiled additional fetal samples from 6.5 to 8 weeks of development. Pearson correlation of the global transcriptome demonstrated a dramatic increase in similarity between PSC-derived NPCs and young fetal spinal cord compared to 16 week spinal cord (Fig 2-7A). Ideally, a comparison between cells derived from fetal tissue earlier than 6 weeks would determine whether PSC derivatives accurately reflect their counterparts found during very early fetal development, but because of the lack of access to such tissue, this question is difficult to answer. However, data on gene expression across whole human embryos from 3-5 weeks of development recently became available [49].

We identified a list of 1645 probe sets differentially expressed between 3 week (stage 9) and 5 week (stage 14) embryos. We overlaid these identified probe sets with the list of probe sets differentially expressed between PSC derivatives and their natural counterparts (Fig 3A; 105 is reduced to 36 due to the more restrictive chip used by Fang et al [49]. See Material and Methods for details) and found significant overlap. *LIN28A* was the most differentially expressed probe set between stage 9 and stage 14 embryos. Knowing that *LIN28A* was differentially expressed in both the NPC and hepatocyte lineages, but not the fibroblasts of our data, we repeated the analysis and excluded fibroblasts (Fig 2-7C). Amongst the 53 probe sets conserved in our data and the Fang et al. (2010) data are *DPPA4* and *LIN28A* (Note: *LIN28B* is not represented on the Fang et al. array). The expression pattern for the 46 unique genes represented by those 53 probe sets is represented by two heat maps, one including the samples from Fang et al. (2010) and the other including our own samples (Fig 2-7D). The fact that a significant number of genes that are normally downregulated between 3-5 weeks of development appear to distinguish PSC derivatives from their tissue derived counterparts

further suggests that PSC derivatives might accurately recapitulate cells found prior to 6 weeks of development.

### 2.5 Discussion

Our data have revealed several important insights about differentiation from hPSCs. First, our data showed that hESCs and hiPSCs make specified derivatives that are nearly equivalent transcriptionally. This was surprising considering the vastly different circumstances by which hESCs and hiPSCs are derived, and in light of well-documented differences between them at the epigenetic and transcriptional levels in the undifferentiated state [2-6, 8, 9, 50, 51]. This similarity might be attributed to the fact that only high passage hiPSCs (>40) were used. However, even at high passage, a small number of genes still appeared to distinguish the undifferentiated hiPSC and hESC lines used. The fact that these differences were largely undetectable in the differentiated state could suggest that the progeny of these cell types are more similar than their parental cells or that the genes differentially expressed between them are not expressed in the specified progeny.

Second, upon differentiation, we did not detect appreciable re-expression of the exogenous reprogramming factors. Because the expression of *OCT4* was undetectable in any PSC derivative, it is unlikely that the loci representing the retroviral reprogramming factors were re-activated upon differentiation or even continued culture. This finding could suggest that, at least in the contexts analyzed here, concerns over re-expression of oncogenic factors from hiPSCs lines generated by viral integration could be mitigated by specification of hiPSCs to even a progenitor state, such as NPC, hepatoblast or fibroblast.

Third, both hESCs and hiPSCs made progeny that continued to express a group of genes known to play roles in very early embryonic development. While the progeny of ESCs have been proposed to represent embryonic cell types because of the primitive nature of the starting cell types[23, 25, 52], to our knowledge human pluripotent progeny have not yet been placed into such a narrow developmental context. The most logical interpretation of our data is that the PSC derivatives generated here represent cells similar to those found earlier than 6 weeks of development.

It remains possible, however, that current protocols to make hPSC derivatives generate cells that do not completely represent cells found in tissue at any developmental stage. With limited access to tissues representing the earliest stages of development, this possibility will remain unexplored for now. It is also possible that new culture conditions could be defined that improve differentiation *in vitro* to better recapitulate that which occurs *in vivo*. In addition, it is possible that experimentally manipulating the expression of early embryonic genes described here could be used to accelerate development *in vitro*. Regardless, the fact that PSC derivatives produced in other labs also express some of the same early embryonic genes suggests that many protocols lead to a consistent result, so perhaps a shift in differentiation strategy or expectations is required.

Our data also suggest that simple approaches, such as continued culture, can further the development of PSC-derived cells, though this method also has its limitations as described above. Regardless, it is tempting to speculate that, upon specification, the process is more or less pre-determined by mechanisms that lock cells into a process that takes a specific amount of time, or number of cell divisions, regardless of the culture conditions employed. This idea has significant support from studies with *in vitro* murine development[52], where differentiation of

mESCs under minimal conditions allowed appropriate temporal and regional specification of neural tissue.

Recent work suggested that hPSC derivatives share hallmarks of gene expression with oncogenic cells[53]. Our data suggest that the presence of oncogenic hallmarks could be explained by persistent expression of genes in hPSC derivatives that are typical of the early embryo and known to be re-expressed in cancers. For example, *LIN28* is not normally expressed in post-natal mammalian tissue, it was recently shown to be re-expressed in 15% of human cancers; and cancers with poor prognosis that are high in *LIN28* expression are low in let-7 family expression [46]. This study also showed that overexpression of *LIN28* drove transformation of fibroblasts, consistent with its proposed role in reprogramming [42, 46]. Therefore, if the derivatives of PSCs are to be used clinically, it could be important to take into account the residual expression of the early embryonic genes, particularly, *LIN28*. Finally, one of the great benefits of iPS technology is the ability to model human diseases *in vitro* using patient derived cells. Our data would suggest that it could be difficult to model human diseases in this context unless a phenotype manifests very early in development.

### 2.6 Bibliography


1. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
2. Chin, M.H., et al., *Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures.* Cell Stem Cell, 2009. **5**(1): p. 111-23.
3. Ghosh, Z., et al., *Persistent Donor Cell Gene Expression among Human Induced Pluripotent Stem Cells Contributes to Differences with Human Embryonic Stem Cells.* PLoS One, 2010. **5**(2): p. e8975.

4. Marchetto, M.C., et al., *Transcriptional signature and memory retention of human-induced pluripotent stem cells.* PLoS One, 2009. **4**(9): p. e7076.

5. Bock, C., et al., *Reference Maps of Human ES and iPS Cell Variation Enable High-Throughput Characterization of Pluripotent Cell Lines.* Cell, 2011. **144**: p. 439-452.

6. Lister, R., et al., *Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.* Nature, 2011. **471**: p. 68-73.

7. Narsinh, K.H., et al., *Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells.* J Clin Invest, 2011. **121**(3): p.1217-1221.

8. Polo, J.M., et al., *Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells.* Nature Biotechnology, 2010. **8**: p. 848-55.

9. Kim, K., et al., *Epigenetic memory in induced pluripotent stem cells.* Nature, 2010. **467:** p.285-90.

10. Ohi, Y., et al., *Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells.* Nat Cell Biol, 2011. **13**(5): p. 541-9.

11. Bar-Nur, O., et al., *Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet Beta cells.* Cell stem cell, 2011. **9**(1): p. 17-23.

12. Chin, M.H., et al., *Molecular analyses of human induced pluripotent stem cells and embryonic stem cells.* Cell Stem Cell, 2010. 7(2): p.263-9

13. Osafune, K., et al., *Marked differences in differentiation propensity among human embryonic stem cell lines.* Nat Biotechnol, 2008. **26**(3): p. 313-5.

14. Feng, Q., et al., *Hemangioblastic Derivatives from Human Induced Pluripotent Stem Cells Exhibit Limited Expansion and Early Senescence.* Stem Cells, 2010. **28**(4): p.704-12.

15. Boulting, G.L., et al., *A functionally characterized test set of human induced pluripotent stem cells.* Nat Biotechnol, 2011. **29**(3): p.279-86.

16. Si-Tayeb, K., et al., *Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells.* Hepatology, 2010. **51**(1): p. 297-305.

17. Pfannkuche, K., et al., *Cardiac myocytes derived from murine reprogrammed fibroblasts: intact hormonal regulation, cardiac ion channel expression and development of contractility.* Cell Physiol Biochem, 2009. **24**(1-2): p. 73-86.

18. Karumbayaram, S., et al., *Directed differentiation of human-induced pluripotent stem cells generates active motor neurons.* Stem Cells, 2009. **27**(4): p. 806-11.

19. Zhang, J., et al., *Functional cardiomyocytes derived from human induced pluripotent stem cells.* Circ Res, 2009. **104**(4): p. e30-41.

20. Song, Z., et al., *Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells.* Cell Res, 2009. **19**(11): p. 1233-42.

21. Choi, K.D., et al., *Hematopoietic and endothelial differentiation of human induced pluripotent stem cells.* Stem Cells, 2009. **27**(3): p. 559-567.

22. Kennedy, M., et al., *Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures.* Blood, 2007. **109**(7): p. 2679-87.

23. Keller, G., *Embryonic stem cell differentiation: emergence of a new era in biology and medicine.* Genes Dev, 2005. **19**(10): p. 1129-55.

24. Pankratz, M.T., et al., *Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage.* Stem Cells, 2007. **25**(6): p. 1511-20.

25. Elkabetz, Y., et al., *Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage.* Genes Dev, 2008. **22**(2): p. 152-65.

26. Lowry, W.E., et al., *Generation of human induced pluripotent stem cells from dermal fibroblasts.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 2883-8.

27. Karumbayaram, S., et al., *Human embryonic stem cell-derived motor neurons expressing SOD1 mutants exhibit typical signs of motor neuron degeneration linked to ALS.* Dis Model Mech, 2009. **2**(3-4): p. 189-95.

28. Kwon, G.S., et al., *Tg(Afp-GFP) expression marks primitive and definitive endoderm lineages during mouse development.* Dev Dyn, 2006. **235**(9): p. 2549-58.

29. Lee, R., et al., *Genetic studies on the functional relevance of the protein prenyltransferases in skin keratinocytes.* Hum Mol Genet, 2010. **19**(8): p. 1603-17.

30. Lowry, W.E., et al., *Defining the impact of beta-catenin/Tcf transactivation on epithelial stem cells.* Genes Dev, 2005. **19**(13): p. 1596-611.

31. Tang, Z., et al., *Overexpression of the ZIP1 zinc transporter induces an osteogenic phenotype in mesenchymal stem cells.* Bone, 2006. **38**(2): p. 181-98.

32. Tchieu, J., et al., *Female human iPS cells retain an inactive X-chromosome.* Cell Stem Cell, 2010. **7**(3): p. 329-42.

33. Deneen, B., et al., *The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord.* Neuron, 2006. **52**(6): p. 953-68.

34. Viswanathan, S.R. and G.Q. Daley, *Lin28: A microRNA regulator with a macro role.* Cell, 2010. **140**(4): p. 445-9.

35. Madan, B., et al., *The pluripotency-associated gene Dppa4 is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis.* Mol Cell Biol, 2009. **29**(11): p. 3186-203.

36. Moss, E.G. and L. Tang, *Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites.* Dev Biol, 2003. **258**(2): p. 432-42.

37. Zhu, H., et al., *Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies.* Nat Genet, 2010. **42**(7): p. 626-30.

38. West, J.A., et al., *A role for Lin28 in primordial germ-cell development and germ-cell malignancy.* Nature, 2009. **460**(7257): p. 909-13.

39. Merrill, B.J., et al., *Tcf3: a transcriptional regulator of axis induction in the early embryo.* Development, 2004. **131**(2): p. 263-74.

40. Pereira, L., F. Yi, and B.J. Merrill, *Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal.* Mol Cell Biol, 2006. **26**(20): p. 7479-91.

41. Ambros, V. and H.R. Horvitz, *Heterochronic mutants of the nematode Caenorhabditis elegans.* Science, 1984. **226**(4673): p. 409-16.

42. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.

43. Lehrbach, N.J., et al., *LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans.* Nat Struct Mol Biol, 2009. **16**(10): p. 1016-20.

44. Hagan, J.P., E. Piskounova, and R.I. Gregory, *Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells.* Nat Struct Mol Biol, 2009. **16**(10): p. 1021-5.

45. Heo, I., et al., *TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation.* Cell, 2009. **138**(4): p. 696-708.

46. Viswanathan, S.R., et al., *Lin28 promotes transformation and is associated with advanced human malignancies.* Nat Genet, 2009. **41**(7): p. 843-8.

47. Heo, I., et al., *Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA.* Mol Cell, 2008. **32**(2): p. 276-84.

48. Iwasaki, S., T. Kawamata, and Y. Tomari, *Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression.* Mol Cell, 2009. **34**(1): p. 58-67.

49. Fang, H., et al., *Transcriptome analysis of early organogenesis in human embryos.* Dev Cell, 2010. **19**(1): p. 174-84.

50. Doi, A., et al., *Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.* Nat Genet, 2009. **41**(12): p.1350-3.

51. Hawkins, R.D., et al., *Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.* Cell Stem Cell, 2010. **6**(5): p. 479-91.

52. Gaspard, N., et al., *An intrinsic mechanism of corticogenesis from embryonic stem cells.* Nature, 2008. **455**(7211): p. 351-7.

53. Ghosh, Z., et al., *Dissecting the Oncogenic Potential of Human Embryonic and Induced Pluripotent Stem Cell Derivatives.* Cancer Res, 2011. **71**(14): p.5030-9.

54. Li, Z., et al., *Functional and transcriptional characterization of human embryonic stem cell-derived endothelial cells for treatment of myocardial infarction.* PLoS One, 2009. **4**(12):e8443

54. James, D., et al., *Expansion and maintenance of human embryonic stem cell-derived endothelial cells by TGFbeta inhibition is Id1 dependent.* Nat Biotechnol, 2010. **28**(2): p.161-6.

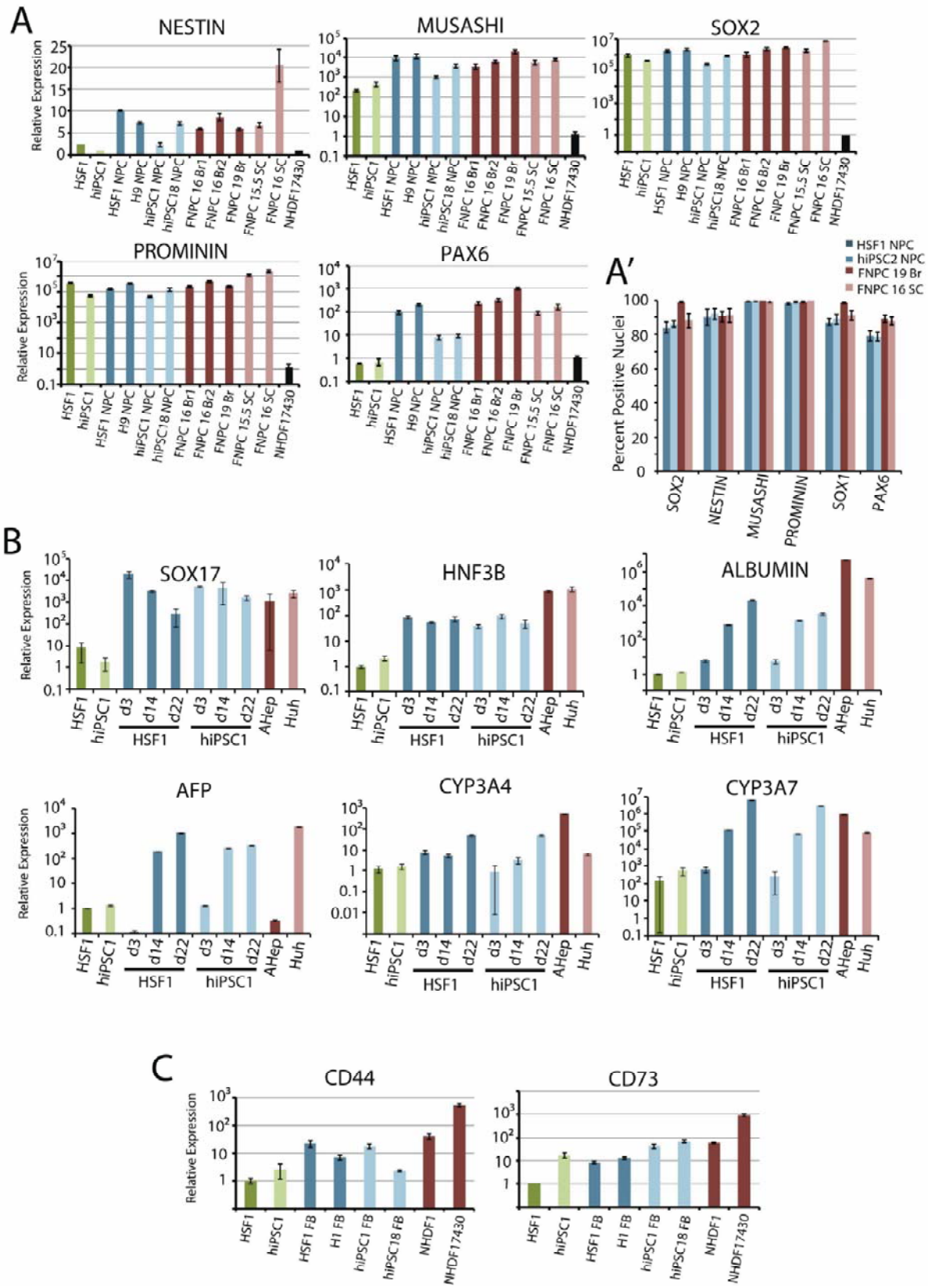
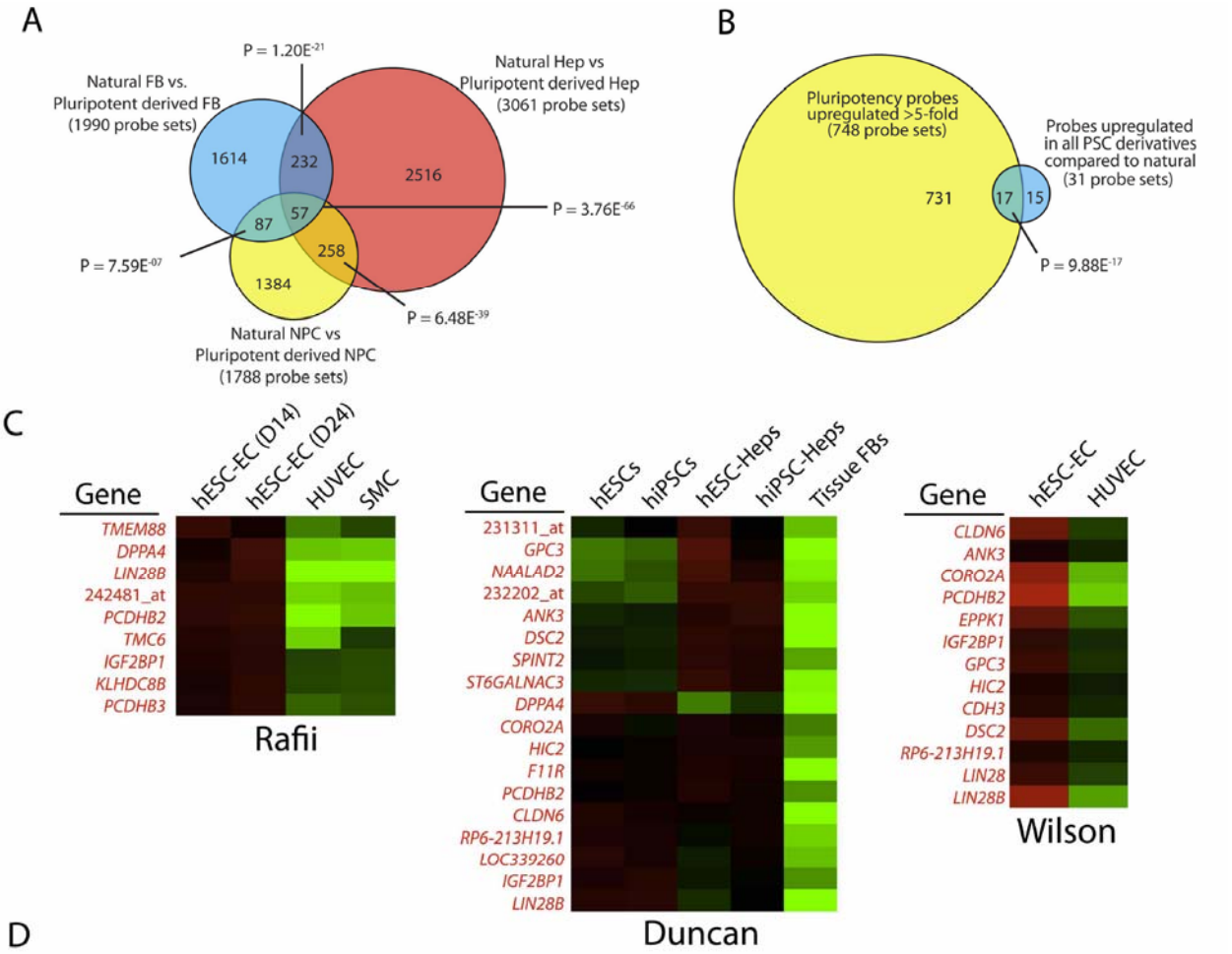

A
NESTIN
MUSASHI
SOX2
PROMININ
PAX6
Relative Expression
HSF1
hiPSC1
HSF1 NPC
H9 NPC
hiPSC1 NPC
hiPSC18 NPC
FNPC 16 Br1
FNPC 16 Br2
FNPC 19 Br
FNPC 15.5 SC
FNPC 16 SC
NHDF17430
A'
HSF1 NPC
hiPSC2 NPC
FNPC 19 Br
FNPC 16 SC
Percent Positive Nuclei
SOX2
NESTIN
MUSASHI
PROMININ
SOX1
PAX6
B
SOX17
HNF3B
ALBUMIN
AFP
CYP3A4
CYP3A7
d3
d14
d22
HSF1
hiPSC1
AHep
Huh
C
CD44
CD73
HSF1 FB
H1 FB
hiPSC1 FB
hiPSC18 FB
NHDF1

**Figure S2-1. hESC and hiPSC lines make cell types representing all three germ layers.**

**A,** RT-PCRs demonstrating mRNA expression of NPC markers (SOX2, NESTIN, MUSASHI, PROMININ, and PAX6) in NPCs derived from PSCs, fetal brain and fetal spinal cord. Error bars represent standard error over 4 replicates. Values set relative to NHDF17430. **A',** Quantification of the percent positive cells in NPC cultures expressing SOX2, PAX6, MMUSASHI, PROMININ, and NESTIN). Error bars represent standard error over 10 fields of view **B,** RT-PCRs of definitive endoderm markers (SOX17 and HNF3B), fetal hepatocyte markers (AFP and CYP3A7) and adult hepatocyte markers (ALBUMIN and CYP3A4) following 3, 14 and 22 days of directed-differentiation. Error bars represent standard error over 4 replicates. Values set relative to HSF1. **C,** RT-PCR of fibroblast markers (CD44 and CD73) following differentiation of pluripotent stem cells to fibroblasts. Error bars represent standard error over 4 replicates. Values set relative to HSF1.




| Go Terms: Biological Processes | | |
|---|---|---|
| **NPC** | **Hep** | **FB** |
| **homophilic cell adhesion**; **nervous system development**; developmental process; calcium-dependent cell-cell adhesion | **nervous system development**; oxidation reduction; acute inflammatory response; metabolic process; cofactor metabolic process; cellular lipid metabolic process; regulation of RNA metabolic process; regulation of nucleobase, nucleoside, nucleotide; regulation of transcription, DNA-dependent; regulation of transcription; aromatic compound metabolic process; regulation of gene expression; biphenyl metabolic process; bilirubin conjugation; humoral immune response mediated circulating immunoglobulin; activation of plasma proteins during accute inflammatory response; complement activation, classical pathway; complement activation; lipid metabolic process; regulation of macromolecule biosynthetic process; tetrapyrrole catabolic process; porphyrin catabolic process; cellular metabolic process; regulation of biosynthetic process; transcription; regulation of cellular metabolic process; benzene and derivative metabolic process; negative regulation of biosynthetic process; heterocycle metabolic process; regulation of metabolic process; steroid metabolic process; negative regulation of gene expression; biosynthetic process; negative regulation of transcription; negative regulation of nucleobase, nucleoside, nucleotide; regulation of macromolecule metabolic process; tetrapyrrole metabolic process; porphyrin metabolic process | **homophilic cell adhesion**; cell-cell adhesion |

**Figure S2-2. Gene expression differences between PSC and tissue derivatives are conserved regardless of statistical analyses employed or lab.**

**A and B,** A similar analysis as shown in Figure 3A and B, but instead using FDR ($p \leq 0.05$) and fold change $\geq 1.54$ as criteria to determine genes that were differentially expressed. **C,** The same analysis

shown in Figure 3C was performed with 3 outside datasets. Outside data sets included: GSE19735 (Rafii), comparing hESC-derived endothelial cell (EC) to human umbilical vein endothelial cells (HUVEC) and human smooth muscle cells (SMC); GSE14897 (Duncan), comparing undifferentiated hESCs and hiPSCs to hepatocytes made from each; and GSE20013 (Wilson), comparing endothelial cells made from hESCs to endothelial cells made form tissue (HUVEC). **D,** Gene Ontology Biological Processes terms identified as different between PSC derivatives and their natural counterparts ($p<0.01$). Underlined terms were identified in 2 of 3 germ layers. Note: no term was identified in all three germ layers.

# CHAPTER 3

## Small molecule screening of let-7 modulators in a human cancer cell line

### 3.1 Abstract

Previously, our group has demonstrated that hPSC-derived progeny maintained the expression of genes normally associated with early development, including *LIN28A*, *LIN28B* and *DPPA4*, which possibly limited the complete maturation of these hPSC-derivatives. The deregulation of LIN28/let7 circuit is particularly interesting because (1) it has been shown to play important roles in regulating developmental timing and maturity, (2) it has been employed to induce pluripotency, and (3) many aggressive, undifferentiated tumors displayed perturbed LIN28/let-7 expressions. The best strategies to modulate the LIN28/let-7 pathways involve reagents that are easy to deliver in differentiating hPSC cultures or in the case of therapeutics, low toxicity drugs with high bioavailability. In recent years, small molecules have not only shown promise in treatment of cancer, but also in the *in vitro* reprogramming of cell fates. We hypothesize that small molecules can also be employed to modulate the LIN28/let-7 circuit. To this end, we have established a human cell-based system to screen for small molecules that could modulate LIN28/let-7 activity. Here we present our findings of this high throughput screening. This screening could potentially unveil powerful small molecules that could modulate developmental maturity and therapeutically combat cancer.

### 3.2 Introduction

Human pluripotent stem cells are potentially promising model systems for studying human development and human diseases. However, such a promise is based on the assumption that human pluripotent stem cells can be induced to differentiate into cells that closely resemble or even replace naturally developed somatic cells found in a human body. It has been reported that while differentiation of human pluripotent cells can generate functional progeny, differentiation *in vitro* only approximates that which occurs *in vivo* [1]. Global transcriptome analysis shows that genes related to embryonic development are misexpressed. Notable among these is the higher expression of LIN28 and let-7 targets in pluripotent stem cell derivatives [1].

In mammals, LIN28A and LIN28B are RNA-binding proteins with two types of nucleic acid interacting domains: a cold shock domain (CSD) followed by two repeats of CCHC-type zinc-binding motifs. Structural analysis revealed that these domains bind to the stem loop and the GGAG domains of let-7 precursors respectively, allowing specific interactions with and repressions of various pre-let7 members [2-5]. Spatially, it has been suggested that LIN28B is localized in the nucleus and LIN28A resides mostly in the cytoplasm. LIN28B has been proposed to chaperone primary let-7s (pri-let-7) in the nucleolus and away from the processing machinery, thus inhibiting its maturation. In the cytoplasm, LIN28A recruits the TUTase ZCCHC11 to inhibit the maturation of precursor let-7s (pre-let-7) [6, 7].

Let-7 family members of mi-RNA are known to regulate developmental timing and cell-fate decisions in lower organisms [8]. Let-7 family members have identical seed sequences and divergent stem-loop regions [2]. Their targets include many oncogenes (*c-MYC, RAS, HMGA-2*), cell-cycle regulators (*CYCLIND1, D2*), as well as other developmental regulators such as *LIN28A* and *LIN28B* [9]. Their mutual inhibition with LIN28 forms a powerful regulatory loop that is thought to have broad effects on developmental maturity [10].

Overexpression of LIN28A has been shown to delay the onset of puberty in mice [11] while the LIN28/let-7 pathway affects developmental traits such as height [11-16] . LIN28 overexpression

and let-7 inhibition have also been employed for reprogramming somatic cells to pluripotency [17-19]. We hypothesize that, by modulating the LIN28/let-7 regulatory loop, one can alter the maturity/developmental state of pluripotent stem cell derivatives as well as other somatic cells. In fact, it has been shown that overexpression of LIN28B can reprogram adult bone marrow hematopoietic progenitors to mediate multilineage reconstitution that resemble fetal lymphopoiesis [20]. Currently, the experimental approaches employed to modulate LIN28 activity includes RNAi or overexpression; whereas let-7 activity can be induced by let-7 mimics or suppressed by antagomirs. These strategies, however, can be difficult to apply to differentiating cultures; for example, in hepatic differentiation, where hepatic precursors typically grow in three-dimensional clusters (data not shown). Even for easily transfectable cultures that grow as monolayer, it may be hard to consistently apply the same kind of modulation to the LIN28/let-7 pathway with these methods. Small molecules have already been employed to replace certain transcription factors in induced pluripotency with good reproducibility [21, 22]. We propose that it may also be possible to use small molecules to modulate activity of the LIN28/let-7 pathway. Identification of such small molecules could potentially improve the maturity of pluripotent stem cell derivatives or facilitate reprogramming of somatic cells to a younger developmental age.

In addition to the role in development, LIN28 has also been shown to be activated in 15% of human tumors and its expression correlates with tumor progression and poor prognosis [10, 23]. Based on its role in pluripotency, LIN28 may be promoting the growth of an aggressive and poorly differentiated tumor [23, 24]. Some studies have even shown decreased tumor malignancy and growth when LIN28 is inhibited [25, 26]. On the other hand, because the targets of let-7 include oncogenes (*c-MYC, RAS*) as well as genes frequently found upregulated in tumors (*LIN28* itself is a target of let-7), let-7 may have tumor suppressive effects. Indeed, loss of function of let-7 has been linked to cancer formation [27]. It is possible that, by downregulating *LIN28B* and/or upregulating let-7 activity, cancer progression can be reversed

[26]. Such a therapeutic potential can be best realized by the development/discovery of small molecules with high bioavailability.

Here we describe the initial rounds of small molecule screening for let-7 activity modulators, using a human cell line carrying a luciferase-based reporter of this pathway. We have demonstrated the dynamic range of detection of let-7 activity in this model, and have identified several possible modulators of let-7 activity.

### 3.3 Materials and Methods

*Cell Culture*

Huh7, Huh7.5.1, Huh7.5.1 L7L were cultured as previously described [28]. Briefly, these cells are cultured in DMEM Hi-Glu (Life Technologies) supplemented with 10% inactivated FBS (Hyclone), 1X P/S-Glu, and 1X HEPES (Life Technologies).

*Reporter assay*

Cells transfected with the psiCHECK2- let-7 8× luciferase reporter (Addgene #20932) [29] or psiCHECK2 control reporter (Promega) were lysed 72hr post-transfection and subjected to dual-glo luciferase assay as described in the manufacturer's protocol (Promega). The Renilla luciferase gene was driven by CMV promoter and contained eight let-7 targeting sequences in the 3′ UTR, and Firefly luciferase driven by a constitutive promoter as a transfection control. Luciferase assays were carried out in a GloMax 96 Microplate Luminometer (Promega).

*Selectable let-7 activity reporter*

The PBabe Neo plasmid (Addgene #1767) was linearized with the restriction enzyme BamHI (New England BioLabs). The Amplicin-resistance cassette in psiCHECK2-let-7 8X was digested with BamHI and BglII (New England Biolabs) and the 5000bp fragment containing the

luciferase reporters (but no $Amp^R$) was ligated with the linearized Pbabe Neo. The resulting plasmid was ligated, selected with Ampicilin, and sequenced.

*Stable reporter cell line*

Huh7.5.1 was transfected with the selectable let-7 activity reporter using Lipofectamine 2000 (Life Technologies) according to manufacturer's protocol. The transfected cells were selected with G418 (Life Technologies) at 500-1000ug/mL for 3 weeks. The stable reporter cell line was maintained with 600-800ug/mL of G418.

*siRNA and let-7 mimics transfection*

Huh7.5.1 let-7-luciferase line was transfected with Lipofectamine RNAiMax (Life Technologies) according to manufacturer's protocol. siRNA against Lin28B and let-7 mimics were purchased from Dharmacon.

*High Throughput Screen*

The high throughput screen (HST) measured let-7 activity as a function of Renilla Luciferase expression and cell concentration in let7-luc transfected Huh7.5.1 cells incubated with compounds from diverse chemical libraries (Biomol, Prestwick, Emerald, Microsource, NIH). The staff of UCLA's Molecular Screening Shared Resource (MSSR) plated compounds in Matrix 384-well tissue culture treated plates (Thermo Scientific) with a Biomek FX liquid handler. Huh 7.5.1 reporter line cells in Huh media [DMEM High Glucose (Invitrogen), 10% FBS (HyClone), 1% HEPES Buffer (Invitrogen), 1% NEAA (Invitrogen), 1% penicillin/streptomycin, L-Glutamine] were dispensed over the compounds using a Multidrop 384 (Thermo LabSystems) at a cell concentration of 2000/well and a final compound concentration of 10uM. After 48 hours of incubation, Renilla Luciferase expression was evaluated through addition by Multidrop 384 of the live-cell luciferase reagent, ViviRen (Promega) at a concentration of 60uM per well.

Luminescence was assayed using the luminometer function of an Acumen Explorer (TTP Labtech). Cell concentration per well was similarly measured using Cell Titer Glo (Promega). Data from the screen was uploaded to the Collaborative Drug Discovery platform www.collaborativedrug.com, where hit identity was established.

*Primary Hit Validation*

Compounds found in the HTS to significantly stimulate or inhibit Renilla luciferase expression, suggesting let-7 activity regulation, were procured from the MSSR and plated at 10uM on Huh7.5.1 reporter line cells in 48-well or 96-well plates. After 48 hour incubation, let-7-regulated Renilla luciferase and constitutively expressed Firefly luciferase were measured using Promega's Dual-Luciferase Reporter Assay System and a GloMax 96 Microplate Luminometer (Promega). Validated hits were purchased from vendors [Sigma-Aldrich (http://www.sigmaaldrich.com/chemistry.html), Asinex (http://www.asinex.com/), and Molport (http://www.molport.com/buy-chemicals/index)] for further testing.

## 3.4 Results

***Measuring the Lin28 expression level and let-7 expression in cancer cell lines***

We and others have shown that let-7 activity can be precisely assayed using a luciferase-based method (PSI-Check2 let-7 8X) [1, 29]. In short, the Renilla luciferase is flanked by 8 repeats of let-7 target sequence and its mRNA is subject to a higher rate of degradation in the presence of a higher let-7 activity. The control Firefly luciferase is constitutively driven by an HSV-TK promoter. We analyzed a handful of breast cancer and hepatocarcinoma cell lines (MCF7, MCF15, Huh7 and Huh7.5.1) to assay the detectable let-7 activity. In human hepatocarcinoma cell lines (Huh), we observed a high level of Lin28B expression at both the RNA and protein level; and as a result, a low level of let-7 activity, as shown by let-7 luciferase reporter assay (Fig 3-1).

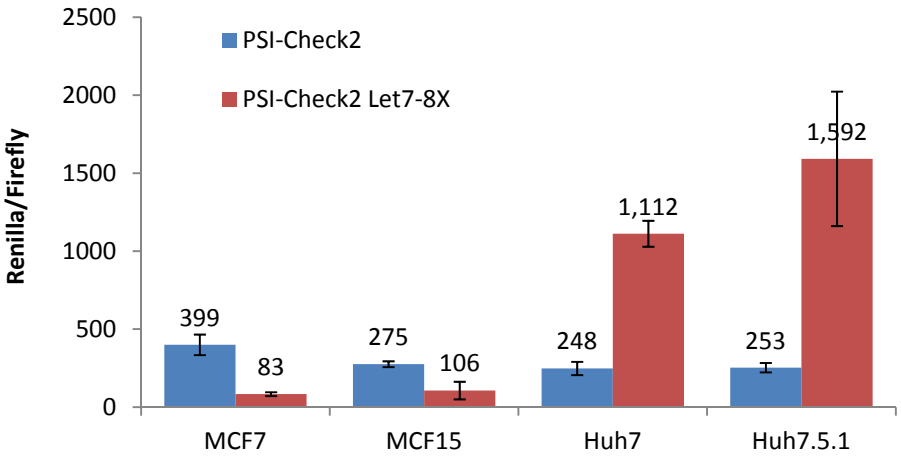



**Figure 3-1. Let-7 activity assay in cancer cell lines.**

Human breast cancer cell lines MCF7, MCF15 and human hepatocarcinoma cell lines Huh7, Huh7.5.1 were transfected with either PSI-Check2 or PSI-CHeck2 Let7-8X. Dual-glo luciferase assay was performed 48hrs after transfection in triplicates.

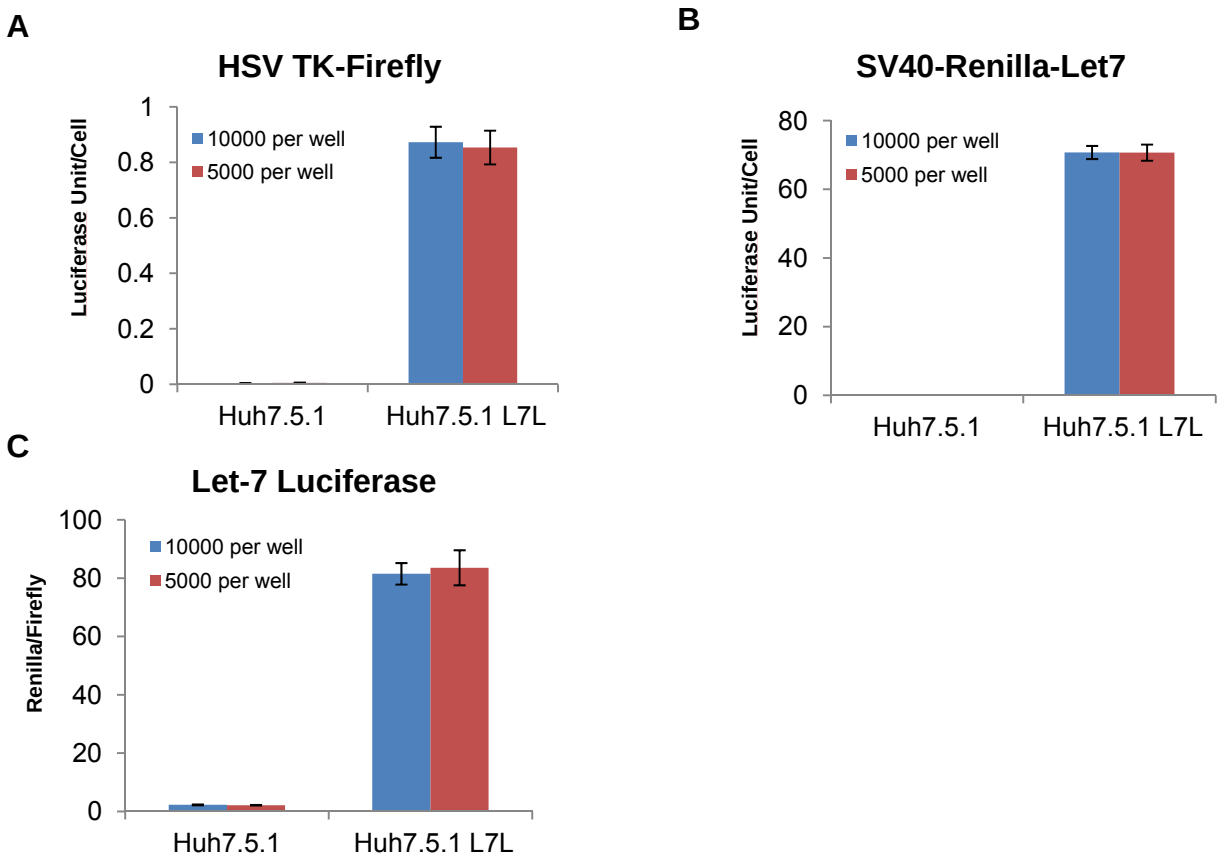


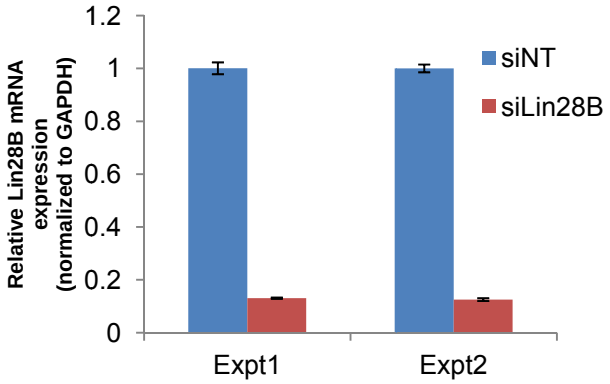

**Figure 3-2. Validation of Stable Huh Let-7 reporter line.**

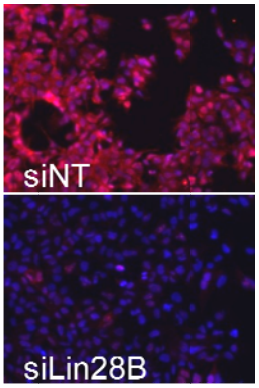
Huh7.5.1 and Huh7.5.1 let-7 luciferase reporter cell line were plated at 5000 or 10000 cells per well, and assayed with Dual-glo luciferase in triplicates 24hrs after plating.

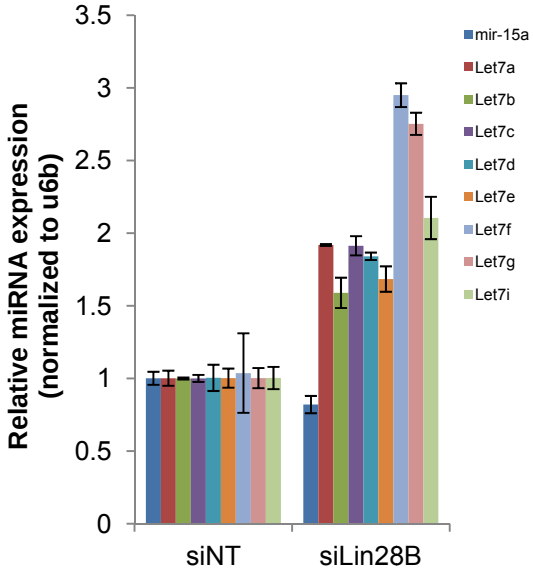
(A) Measurement of Raw Firefly Luciferase Unit in untrasnfected Huh7.5.1 and Huh7.5.1 let-7 luciferase reporter line at the two plating concentrations

(B) Measurement of Raw Renilla Luciferase Unit in untrasnfected Huh7.5.1 and Huh7.5.1 let-7 luciferase reporter line at the two plating concentrations

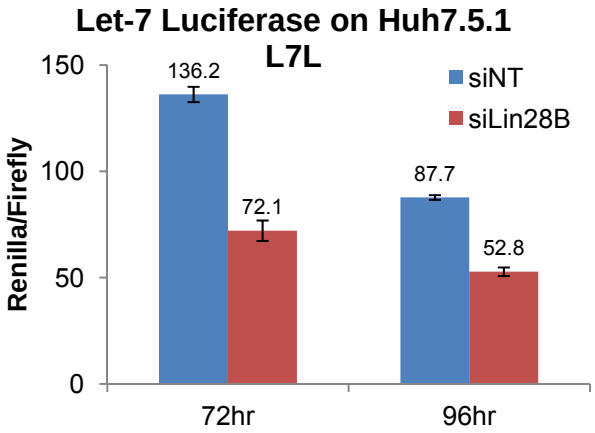
(C) Normalized Renilla/Firefly value in untrasnfected Huh7.5.1 and Huh7.5.1 let-7 luciferase reporter line at the two plating concentration


Relative Lin28B mRNA expression (normalized to GAPDH)
1.2
1
0.8
0.6
0.4
0.2
0
Expt1
Expt2
siNT
siLin28B


siNT
siLin28B


Relative miRNA expression (normalized to u6b)
3.5
3
2.5
2
1.5
1
0.5
0
siNT
siLin28B
mir-15a
Let7a
Let7b
Let7c
Let7d
Let7e
Let7f
Let7g
Let7i

D


Let-7 Luciferase on Huh7.5.1 L7L
Renilla/Firefly
150
100
50
0
136.2
72.1
87.7
52.8
72hr
96hr
siNT
siLin28B

E

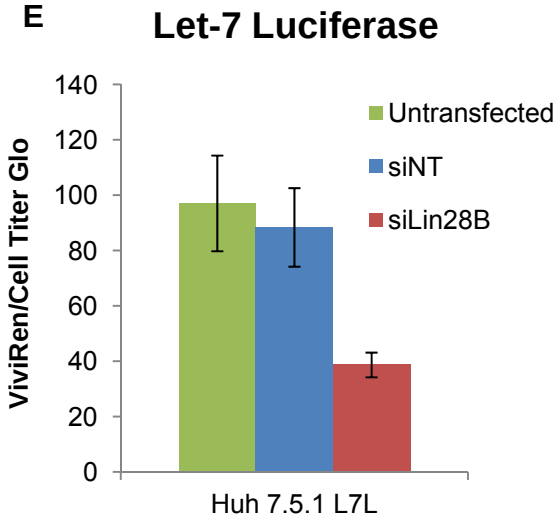

Let-7 Luciferase
ViviRen/Cell Titer Glo
140
120
100
80
60
40
20
0
Huh 7.5.1 L7L
Untransfected
siNT
siLin28B

**Figure 3-3. Effect of *LIN28B* knock down on *LIN28B* expression and let-7 activity in Huh7.5.1 L7L.**

(A) Huh7.5.1 L7L was trasnfected with either non-targeting siRNA (siNT) or siRNA against *LIN28B*. Lysis was collected 48hrs after transfection, mRNA extraction and reverse-transcription PCR was carried out to assay the relative amount of *Lin28B* mRNA.

(B) Huh7.5.1 L7L was trasnfected with either non-targeting siRNA (siNT) or siRNA against *LIN28B*. Transfected cells were replated on coverslips and fixed 72hrs after transfection. Immunofluorescence staining of LIN28B (red) and DNA (blue) was carried out to assay the amount of LIN28B protein.

(C) Huh7.5.1 L7L was trasnfected with either non-targeting siRNA (siNT) or siRNA against LIN28B. Lysis was collected 48hrs after transfection, total RNA extraction and reverse-transcription PCR was carried out to assay the relative amount of mature let-7 miRNA. U6b and mir-15a serve as internal controls.

(D) Huh7.5.1 L7L was trasnfected with either non-targeting siRNA (siNT) or siRNA against *LIN28B*. Transfected cells were replated in triplicates and lysed 72 or 96hrs after transfection. Dual-glo luciferase was carried out to assay for relative let-7 activity.

(E) Huh7.5.1 L7L was trasnfected with either non-targeting siRNA (siNT) or siRNA against *LIN28B*. Untransfected and transfected cells were replated in 20 replicates and assayed 72hrs after transfection. Viviren life cell substrate was used to assay let-7 activity whereas Cell Titer Glo was used to assay for cell titer.

***Construction of a Huh7 cell line stably expressing the let-7 activity reporter***

To facilitate reproducible results in screening, we created a cell line with a built-in luciferase reporter activity. We cloned a Neomycin resistance cassette into the PSI-Check2 let-7 luciferase. We stably introduced the let-7 luciferase plasmid in the Huh7.5.1 cell line and selected with G418 for 3 weeks. The stable cell line was subjected to dual-glo luciferase assay, where it displayed a stable luciferase unit per cell in both Firefly and Renilla (Fig 3-2).

To demonstrate the dynamic range of detection in let-7 activity, we transfected this Huh7.5.1 let-7 luciferase reporter line (termed Huh7.5.1 L7L hereafter) with siRNA against *LIN28B*, as well as let-7 mimics. siRNA effectively knocked down *LIN28B* expression by at least 90% (Fig 3-3A and 3-3B). In response to the downregulation of *LIN28B*, mature microRNA levels raised about 2 to 3 fold for all let-7 family members (Fig 3-3C). As a result, the let-7 activity was reduced by 25-50%, as assayed by dual-glo luciferase (Fig 3-3D and 3-3E).

***High Throughput Screening of Small Molecules***

With the help of the UCLA MSSR facility staff, we have screened through 15489 small molecules as described in Materials and Methods. Out of these small molecules, 682 have luciferase readings downregulated 1.5 to 2 fold, and 163 downregulated over 2 fold, indicating potential let-7 stimulators. Out of these small molecules, 87 have luciferase readings upregulated 1.5 to 2 fold, and 15 upregulated over 2 fold, indicating potential let-7 inhibitors (Table 3-1). We proceeded to validate the top 80 potential let-7 stimulators and top 28 potential let-7 inhibitors. Using small aliquots provided by the UCLA MSSR facility, we carried out primary validation on Huh7.5.1 L7L in a 48-well or 96-well format using dual-glo luciferase assay. After eliminating false positive hits, 17 potential let-7 stimulators and 23 potential let-7 inhibitors remained (Table 3-2 and Table S3-1 (in orange)).

***Secondary Validation of presumptive let-7 activity regulators***

| | | | let7 activity stimulators | | let7 activity inhibitors | |
|---|---|---|---|---|---|---|
| Library | Number of Plates | Number of Small Molecules | Down > 2 fold | Down 1.5 to 2 fold | Up 1.5 to 2 fold | Up > 2 fold |
| Biomol | 2 | 505 | 7 | 16 | 5 | 1 |
| Prestwick | 4 | 1120 | 5 | 37 | 12 | 3 |
| Emerald | 6 | 1800 | 56 | 166 | 3 | 0 |
| Microsource | 7 | 1999 | 63 | 107 | 30 | 8 |
| Target | 27 | 9545 | 11 | 298 | 36 | 3 |
| NIH | 2 | 520 | 21 | 58 | 1 | 0 |
| Total | 48 | 15489 | 163 | 682 | 87 | 15 |

**Table 3-1. Summary of High Throughput Screening results.**

The table shows the number of plates and small molecules screened in each library. Normallized let-7 luciferase readings for each small molecule was compared to the average of DMSO controls on each plate, and fold change was calculated. Potential let-7 activity stimulators would have decreased luciferase readings, as demonstrated in the $4^{th}$ and $5^{th}$ columns. Potential let-7 activity inhibitors would have increased let-7 luciferase readings, as demonstrated in the $6^{th}$ and $7^{th}$ columns. The last row showed the tally of the number of plates, small molecules and potential hits from the small molecule screening.

| | All hits from HTS Screen | Top hits from HTS Screen | Primary Validation | Secondary Validation |
|---|---|---|---|---|
| let-7 stimulators | 845 | 80 | 17 | 5 |
| let-7 inhibitors | 102 | 28 | 23 | 9 |

**Table 3-2. Summary of primary and secondary validation from HTS screen results.**

The table shows the number of small molecules identified as potential let-7 stimulators or inhibitors (column 1); the number of top hits chosen from the HTS for further validation (column 2); the number of hits validated in primary validation (column 3) and the number of hits validated in secondary validation (column 4)

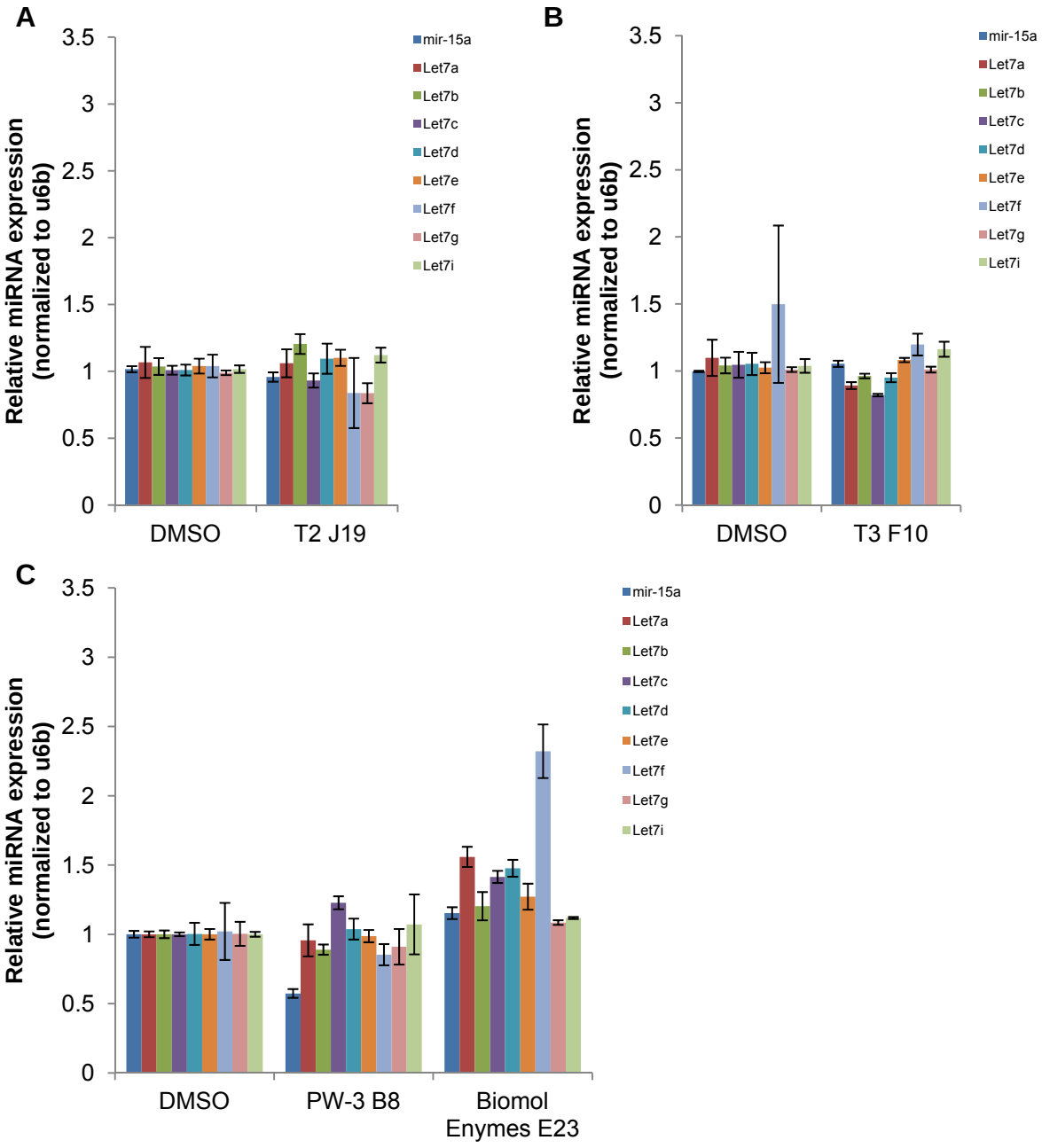



**Figure 3-4. let-7 miRNA expression analysis of Huh7.5.1 L7L treated with small molecules.**

Huh7.5.1 L7L was treated with small molecules or vehicle (DMSO). Lysis was collected 36 or 48hrs after treatment, total RNA extraction and reverse-transcription PCR was carried out to assay the relative amount of mature let-7 miRNA. U6b and mir-15a serve as internal controls.

(A) Huh7.5.1 L7L was treat with DMSO, PW-3 B8 (candidate let-7 inhibitor) or Biomol Enzymes E23 (candidate let-7 stimulator) for 48hrs.

(B) Huh7.5.1 L7L was treat with DMSO or T3F 10 (candidate let-7 inhibitor) for 36hrs.

(C) Huh7.5.1 L7L was treat with DMSO or T2 J19 (candidate let-7 inhibitor) for 48hrs.

*(1) Dual-glo Luciferase Assay on Huh7.5.1 L7L treated with small molecules*

After acquiring the potential small molecule modulators of let-7 activity (from Sigma Aldrich, Asinex or Molport), we treated Huh7.5.1 L7L with 3uM, 10uM, 30uM, 100uM of each of the small molecules at time points varying from 24, 36, 48, and 72hrs. We monitored cell health as well as the modulation of let-7 activity to choose the best window to screen for let-7 modulation. Out of 17 potential let-7 stimulators, 5 were validated to have more than 1.5-fold changes over DMSO control. Out of 23 potential let-7 inhibitors, 9 were validated to have more than 1.5-fold changes over DMSO control (Table 3-2 and Table S3-1 (in dark orange)).

*(2) Let-7 microRNA expression analysis*

For the small molecules that survive the secondary luciferase validation, the amount of mature let-7 microRNA was measured using reverse transcription-PCR specific to let-7 microRNA, according to Materials and Methods. Compared to induction of let-7 microRNA by RNAi against *LIN28B* (Figure 3-3C), the changes brought about by the let-7 inhibitor and stimulator candidates are modest. The most significant changes were induced by the Biomol Enymes E23, a candidate let-7 stimulator, which stimulated the expression of mature let-7 microRNA by 1.5 to 2 fold (Figure 3-4). We are continuing to screen for more active compounds in additional libraries.

### 3.5 Discussion

We have generated a cell-based model suitable for high throughput small molecule screening for let-7 activity modulators. Theoretically, because no transfection is required, this model is also suitable for high throughput screening involving cDNA and siRNA libraries.

We have screened through 6 libraries, including Biomol, Prestwick, Emeral, Microsource, Target, and NIH. Out of about 15,500 small molecules screened, we have found one small molecule, Biomol Enymes E23, which modestly increased let-7 miRNA level and was shown to increase

let-7 activity by let-7-luciferase assay. 15,500 small molecules only constitute a small portion of the small molecules available at the UCLA MSSR facility. Further efforts should be put towards screening through more of these available compounds. It may also be advisable to consider using affinity models to predict the pools of small molecules likely to bind to LIN28, small RNAs or cell machinery pertaining to the LIN28-let7 pathway.

We have adopted a stable let-7-luciferase reporter line (Huh7.5.1 L7L), which expresses far less luciferase mRNAs (and proteins) than transiently transfected cell lines. Since this luciferase system reports for let-7 mediated degradation of the Renilla luciferase mRNA, it allows a higher sensitivity for any reagents that can modestly change the let-7 activity. For example, we can detect the drop in let-7 activity from knocking down *LIN28B* in the Huh7.5.1 L7L (Figure 3-3D,E) but not in Huh7.5.1 transiently transfected with let-7 luciferase (data not shown). However, using a highly sensitive system may also result in high numbers of false positives (and potentially false negatives that were lost). This should be addressed by repeating the screening with the Huh7.5.1 cell line transiently transfected with the let-7 luciferase.

We also encountered the problem of non-specific effects of small molecules on luciferase readings during the high through-put screening, primary and secondary luciferase validations. This should be addressed by comparing the psiCHECK2- let-7 8× luciferase reporter and the psiCHECK2 control luciferase reporter during the early screening process to weed out false positive hits and prevent the loss of false negatives. Alternatively, a fluorescence-based reporter on let-7 activity should also be considered.

If small molecules are discovered that significantly enhance let-7 activity, the functional role of such small molecules should be tested on cancer cell lines and pluripotent stem cell derivatives. Increased let-7 activity could lead to reduced expression of *LIN28* as well as oncogenes, and thereby could possibly reduce the malignancy of cancer cells. Cancer cells have been show to exhibit reduced malignancy and motility when *LIN28* is suppressed and let-7 activity is elevated [25-27]. In addition, it is possible that the developmental maturity of pluripotent stem cell

derivatives is controlled by the LIN28/let-7 pathway. Pluripotent-stem cell derivatives, which have a higher level of *LIN28* and lower level of Let-7 than somatic counterparts, are thought to exhibit an early fetal-like signature.

Although this screening for let-7 modulators is still work in progress, it has laid down a foundation for the first human cell-based screening to modulate the LIN28/let-7 pathway. Any small molecules found to alter this pathway will have great potential in facilitating the study of the maturation in pluripotent stem cell derivatives as well as human development. Moreover, any discovery of low-toxicity, highly bioavailable small molecules can have great value in therapeutics to battle or prevent malignant cancers.

### 3.6 Bibliography


1. Patterson, M., et al., *Defining the nature of human pluripotent stem cell progeny.* Cell Res, 2012. **22**(1): p. 178-93.
2. Nam, Y., et al., *Molecular basis for interaction of let-7 microRNAs with Lin28.* Cell. **147**(5): p. 1080-91.
3. Heo, I., et al., *Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA.* Mol Cell, 2008. **32**(2): p. 276-84.
4. Rybak, A., et al., *A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment.* Nat Cell Biol, 2008. **10**(8): p. 987-93.
5. Newman, M.A., J.M. Thomson, and S.M. Hammond, *Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing.* RNA, 2008. **14**(8): p. 1539-49.
6. Hagan, J.P., E. Piskounova, and R.I. Gregory, *Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells.* Nat Struct Mol Biol, 2009. **16**(10): p. 1021-5.

7. Piskounova, E., et al., *Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms.* Cell. **147**(5): p. 1066-79.

8. Moss, E.G., R.C. Lee, and V. Ambros, *The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA.* Cell, 1997. **88**(5): p. 637-46.

9. Bussing, I., F.J. Slack, and H. Grosshans, *let-7 microRNAs in development, stem cells and cancer.* Trends Mol Med, 2008. **14**(9): p. 400-9.

10. Viswanathan, S.R. and G.Q. Daley, *Lin28: A microRNA regulator with a macro role.* Cell. **140**(4): p. 445-9.

11. Zhu, H., et al., *Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies.* Nat Genet. **42**(7): p. 626-30.

12. Lettre, G., et al., *Identification of ten loci associated with height highlights new biological pathways in human growth.* Nat Genet, 2008. **40**(5): p. 584-91.

13. Lu, L., et al., *Pluripotent factor lin-28 and its homologue lin-28b in epithelial ovarian cancer and their associations with disease outcomes and expression of let-7a and IGF-II.* Eur J Cancer, 2009. **45**(12): p. 2212-8.

14. Ong, K.K., et al., *Genetic variation in LIN28B is associated with the timing of puberty.* Nat Genet, 2009. **41**(6): p. 729-33.

15. Perry, J.R., et al., *Meta-analysis of genome-wide association data identifies two loci influencing age at menarche.* Nat Genet, 2009. **41**(6): p. 648-50.

16. Sulem, P., et al., *Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche.* Nat Genet, 2009. **41**(6): p. 734-8.

17. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.

18. Melton, C., R.L. Judson, and R. Blelloch, *Opposing microRNA families regulate self-renewal in mouse embryonic stem cells.* Nature. **463**(7281): p. 621-6.

19. Martinez, N.J. and R.I. Gregory, *MicroRNA gene regulatory pathways in the establishment and maintenance of ESC identity.* Cell Stem Cell. **7**(1): p. 31-5.

20. Yuan, J., et al., *Lin28b reprograms adult bone marrow hematopoietic progenitors to mediate fetal-like lymphopoiesis.* Science. **335**(6073): p. 1195-200.

21. Ichida, J.K., et al., *A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog.* Cell Stem Cell, 2009. **5**(5): p. 491-503.

22. Shi, Y., et al., *A combined chemical and genetic approach for the generation of induced pluripotent stem cells.* Cell Stem Cell, 2008. **2**(6): p. 525-8.

23. Viswanathan, S.R., et al., *Lin28 promotes transformation and is associated with advanced human malignancies.* Nat Genet, 2009. **41**(7): p. 843-8.

24. Ben-Porath, I., et al., *An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors.* Nat Genet, 2008. **40**(5): p. 499-507.

25. Chang, T.C., et al., *Lin-28B transactivation is necessary for Myc-mediated let-7 repression and proliferation.* Proc Natl Acad Sci U S A, 2009. **106**(9): p. 3384-9.

26. Iliopoulos, D., H.A. Hirsch, and K. Struhl, *An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation.* Cell, 2009. **139**(4): p. 693-706.

27. Yu, F., et al., *let-7 regulates self renewal and tumorigenicity of breast cancer cells.* Cell, 2007. **131**(6): p. 1109-23.

28. Zhong, J., et al., *Robust hepatitis C virus infection in vitro.* Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9294-9.

29. Iwasaki, S., T. Kawamata, and Y. Tomari, *Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression.* Mol Cell, 2009. **34**(1): p. 58-67.

| | | | | Normalized Luc Reading Fold Change | | |
|---|---|---|---|---|---|---|
| Compound | Screen | Library | let7 S or I | HTS | Primary | Secondary |
| Biomol Enymes | 1 | Biomol | S | 0.42355 | NA | |
| Biomol Enymes | 1 | Biomol | S | 0.42081 | NA | |
| Biomol Enymes | 1 | Biomol | S | 0.341751 | NA | |
| Biomol Enymes | 1 | Biomol | S | 0.404188 | NA | 0.3495255 |
| PW-3 H16 | 1 | Prestwick | S | 0.4105 | NA | |
| PW-3 B8 | 1 | Prestwick | I | 3.92465 | NA | 1.0356113 |
| ES-1 A3 | 2 | Emerald | S | 0.111302 | 0.605668268 | |
| ES-1 A4 | 2 | Emerald | S | 0.209725 | 0.976675337 | |
| ES-1 A10 | 2 | Emerald | S | 0.187512 | 0.408063809 | |
| ES-1 B7 | 2 | Emerald | S | 0.150984 | 0.820338759 | |
| ES-1 G3 | 2 | Emerald | S | 0.14938 | 0.694183653 | |
| ES-1 G5 | 2 | Emerald | S | 0.133012 | 0.926747423 | |
| ES-2 A4 | 2 | Emerald | S | 0.258038 | 0.753406693 | |
| ES-2 E6 | 2 | Emerald | S | 0.21048 | 0.750478761 | |
| ES-2 E11 | 2 | Emerald | S | 0.261894 | 0.823273204 | |
| ES-2 G5 | 2 | Emerald | S | 0.250003 | 0.513413988 | |
| ES-2 K9 | 2 | Emerald | S | 0.178292 | 1.087874866 | |
| ES-3 B5 | 2 | Emerald | S | 0.18279 | 0.692356545 | |
| ES-3 G5 | 2 | Emerald | S | 0.144469 | 0.213963749 | 0.3508794 |
| ES-6 A6 | 2 | Emerald | S | 0.100679 | 0.597920159 | |
| ES-6 B12 | 2 | Emerald | S | 0.182803 | 0.579509037 | |
| MS1 A10 | 2 | MicroSource | S | 0.141593 | 1.46043221 | |
| MS1 E5 | 2 | MicroSource | S | 0.251005 | 1.402731418 | |
| MS1 H16 | 2 | MicroSource | S | 0.188442 | 1.515397414 | |
| MS3 N10 | 2 | MicroSource | S | 0.218643 | 0.826942987 | |
| MS4 A4 | 2 | MicroSource | S | 0.219427 | 0.941250666 | |
| MS4 O20 | 2 | MicroSource | I | 2.470343 | 1.403804124 | |
| MS5 G4 | 2 | MicroSource | S | 0.230021 | 0.644410672 | |
| MS6 A10 | 2 | MicroSource | S | 0.13632 | 1.285639709 | |
| MS7 E9 | 2 | MicroSource | S | 0.10836 | 1.363228105 | |
| MS7 G7 | 2 | MicroSource | S | 0.095189 | 1.541101083 | |
| MS7 G8 | 2 | MicroSource | S | 0.083928 | 1.185002199 | |
| MS7 G10 | 2 | MicroSource | S | 0.221355 | 1.405023111 | |
| T2 J14 | 3 | Target | S | 0.574362 | 1.287913196 | |
| T2 J19 | 3 | Target | S | 0.625234 | 0.931411128 | 0.594872 |
| T2 H13 | 3 | Target | S | 0.639935 | 0.636767698 | |
| T2 K14 | 3 | Target | S | 0.648141 | 1.338086174 | |
| T3 F10 | 3 | Target | S | 0.473944 | 0.428711026 | 0.4615037 |
| T3 E10 | 3 | Target | S | 0.52699 | 0.490944092 | 0.9455441 |
| T4 E6 | 3 | Target | S | 0.588347 | 1.073185684 | |
| T4 G20 | 3 | Target | S | 0.630835 | 0.646726335 | 1.037208 |
| T4 K11 | 3 | Target | S | 0.529368 | 0.90349783 | |
| T4 L11 | 3 | Target | S | 0.618393 | 0.897537273 | |
| T4 J13 | 3 | Target | S | 0.627584 | 0.916953376 | |
| T4 I15 | 3 | Target | S | 0.614613 | 0.597899415 | |
| T5 G8 | 3 | Target | S | 0.5838 | 0.928529276 | |
| T5 I8 | 3 | Target | S | 0.62829 | 1.420084322 | |
| T5 K20 | 3 | Target | S | 0.65062 | 0.739242687 | 1.0292849 |
| T5 K22 | 3 | Target | S | 0.546481 | 0.702649533 | |
| T7 J16 | 3 | Target | S | 0.509086 | 1.255159375 | |
| T7 K11 | 3 | Target | S | 0.617315 | 1.071733928 | |
| T8 I21 | 3 | Target | S | 0.555563 | 1.04139279 | |
| T9 G12 | 3 | Target | S | 0.509461 | 0.40441402 | 0.4829768 |
| T9 F12 | 3 | Target | S | 0.537888 | 0.377035234 | 0.3485532 |
| T10 H18 | 3 | Target | S | 0.572443 | 0.591959928 | 0.6814287 |
| T10 K14 | 3 | Target | S | 0.575935 | 1.027967782 | |
| T10 K19 | 3 | Target | S | 0.561492 | 0.860416795 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| T11 M20 | 3 | Target | S | 0.076997 | 2.740305922 | |
| T13 K17 | 3 | Target | S | 0.087182 | 1.014324782 | |
| T14 G7 | 3 | Target | S | 0.397198 | 1.287913196 | |
| T15 A19 | 4 | Target | I | 1.774087 | 1.682403514 | |
| T15 D12 | 4 | Target | I | 1.888458 | 4.785506617 | 5.8858113 |
| T16 B21 | 4 | Target | I | 1.888945 | 3.439045699 | |
| T16 G13 | 4 | Target | S | 0.53698 | 1.193630661 | |
| T16 H13 | 4 | Target | S | 0.600404 | 1.390331984 | |
| T16 I22 | 4 | Target | S | 0.54006 | 7.722280357 | |
| T16 J22 | 4 | Target | S | 0.362427 | 4.985294631 | |
| T16 L13 | 4 | Target | S | 0.613352 | 1.090180601 | |
| T17 A3 | 4 | Target | I | 1.657296 | 1.296039325 | |
| T17 A6 | 4 | Target | I | 1.722083 | 4.79657573 | |
| T18 D4 | 4 | Target | I | 2.239457 | 5.547081094 | |
| T18 F4 | 4 | Target | I | 1.575902 | 5.092850254 | |
| T18 F18 | 4 | Target | S | 0.469765 | 5.068767566 | |
| T19 B6 | 4 | Target | I | 1.502801 | 2.708616179 | |
| T19 D6 | 4 | Target | I | 1.596709 | 5.840984213 | |
| T19 F18 | 4 | Target | I | 1.636427 | 5.884691322 | |
| T19 H22 | 4 | Target | I | 1.511302 | 4.289863165 | |
| T19 J16 | 4 | Target | S | 0.463676 | 0.869496198 | |
| T20 P21 | 4 | Target | I | 1.283425 | 2.931407494 | 2.5354664 |
| T20 P22 | 4 | Target | I | 1.422606 | 1.919422157 | 2.5242011 |
| T21 C6 | 4 | Target | S | 0.522479 | 0.791953061 | |
| T21 C10 | 4 | Target | I | 1.273727 | 1.212830369 | |
| T21 D10 | 4 | Target | I | 1.438278 | 2.713757221 | 5.244546 |
| T21 G13 | 4 | Target | S | 0.609766 | 1.029449127 | |
| T23 A22 | 4 | Target | I | 1.416737 | 2.187617317 | |
| T23 G4 | 4 | Target | I | 1.304861 | 3.209667449 | 4.2781564 |
| T23 L3 | 4 | Target | I | 1.326835 | 2.546823879 | 2.9739543 |
| T24 A8 | 4 | Target | S | 0.35983 | 2.617202195 | |
| T24 G9 | 4 | Target | I | 1.444915 | 2.487427226 | 3.6162729 |
| T26 C14 | 4 | Target | I | 1.502048 | 3.725143657 | 5.1277389 |
| T26 G20 | 4 | Target | I | 1.322043 | 1.770370429 | 2.4429103 |
| T26 J16 | 4 | Target | S | 0.518096 | 0.3908476 | |
| T26 L13 | 4 | Target | S | 0.525294 | 0.963531851 | |
| T27 J3 | 4 | Target | I | 1.81576 | 3.554779285 | 4.7593282 |
| T27 L12 | 4 | Target | I | 1.701977 | 1.800060411 | |
| N1 H13 | 4 | NIH 1 | S | 0.19669 | 0.9069203 | |
| N1 J5 | 4 | NIH 1 | S | 0.407115 | 1.145427413 | |
| N1 K14 | 4 | NIH 2 | S | 0.49762 | 1.143779202 | |
| N1 P20 | 4 | NIH 2 | I | 1.656408 | 0.763434129 | |
| N2 K14 | 4 | NIH 2 | S | 0.433098 | 1.46287812 | |
| N2 L15 | 4 | NIH 2 | S | 0.35436 | 1.23417151 | |
| N2 M13 | 4 | NIH 2 | S | 0.44297 | 1.295489404 | |
| N2 O4 | 4 | NIH 2 | I | 1.188587 | 1.348043112 | |
| N2 P8 | 4 | NIH 2 | S | 0.254058 | 0.889160744 | |
| N2 P19 | 4 | NIH 2 | S | 0.231497 | 0.9818218 | |
| N2 C22 | 4 | NIH 2 | S | 0.490316 | 1.005780483 | |
| N2 F21 | 4 | NIH 2 | S | 0.510524 | 0.946871343 | |

**Table S3-1. List of small molecules in primary and secondary luciferase validation.**

The table listed all potential let-7 activity stimulators or inhibitors (indicated in 4$^{th}$ column). The normalized luciferase readings display the fold change difference from vehicle control (DMSO). A lower normalized luciferase reading indicates more let-7 activity, vice versa. The small molecules colored in both light and dark orange are validated through the primary luciferase validation to have significant

induction or inhibition on let-7 activity. The small molecules colored in dark orange are also validated through the secondary luciferase validation to have significant induction or inhibition on let-7 activity.

# CHAPTER 4

## Modeling the first human developmental EMT *in vitro*

## 4.1 Abstract

Epithelial to mesenchymal transitions (EMTs) are thought to be essential to generate diversity of tissues during early fetal development, but these events are essentially impossible to study at the molecular level *in vivo* in humans. The first EMT event that has been described morphologically in human development occurs just prior to generation of the primitive streak. Because human embryonic stem cells (hESCs) and induced pluripotent stem cells (hiPSCs) are thought to most closely resemble cells found in epiblast stage embryos prior to formation of the primitive streak, we sought to determine whether this first human EMT could be modeled *in vitro* with pluripotent stem cells. The data presented here suggest that generating embryoid bodies from hESCs or iPSCs drives the procession of EMT events that can be observed within 24-48 hours after EB generation. These structures possess the typical hallmarks of developmental EMTs, and portions also display evidence of primitive streak and mesendoderm. In addition, we identify PTK7 as an especially useful marker of this EMT population, which can also be used to purify these cells for subsequent analyses and identification of novel markers of human development. Finally, we also describe the generation of a product of mesendoderm, the definitive endoderm, from hESCs, which is not only distinguishable from the EMT population but also serve to elucidate the next developmental cornerstone.

### 4.2 Introduction

Human pluripotent stem cells are widely regarded to resemble the epiblast of a developing embryo, in what is known as a primed state, as opposed to murine ESCs which more closely resemble an earlier state, the ground state [1]. It has been demonstrated that embryoid bodies can be routinely generated from both hESCs and hiPSCs [2-6]. Embryoid bodies (EBs) are cell aggregates generated from pluripotent stem cells, which spontaneously differentiate and have the potential to form all three embryonic germ layers [7]. EBs have been adopted as one of the standards to assess the pluripotency and differentiation potential of pluripotent stem cells [2, 3]. Understanding how human pluripotent stem cells (hPSCs) spontaneously undergo differentiation is important for: (1) understanding how to maintain pluripotency in culture; (2) improving protocols to induce different progeny cell types of significance; (3) establishing an "*ex vivo*" model of human development.

Because of its three-dimensional organization and cell of origin, it has been suggested that embryoid bodies can be used as a model to study early human development [7, 8]. The early events of development have been well studied in model organisms such as sea urchin and mouse [9], but the process in human is still elusive at the molecular level. Although the mouse model has generally been accepted as a close alternative to studying human development, there are well-established differences between murine and human development. First and foremost, mouse embryonic development occurs much faster than human development. Secondly, the molecular events underlining development are not identical [10]. Understanding exactly how human development occurs will be of value to modeling congenital diseases and other developmental defects, particularly with the advent of human induced pluripotent stem cells (hiPSCs), where disease states can be captured *in vitro* [11, 12]. Here we demonstrate that embryoid bodies from either hESCs or hiPSCs can be used to demonstrate early human developmental events such as epithelial-to-mesenchymal transition (EMT).

Epithelial-to-mesenchymal transition is probably best studied in the context of carcinoma progression, but is clearly critical for the development of many cell types during normal development of all species [13]. Mesenchymal-to-epithelial transition, a process reverse to EMT, is shown to initiate and facilitate somatic reprogramming [14]. We have shown that a spontaneous EMT event occurs in hPSC culture as well as in EBs. In addition to a three-dimensional model for development, the use of EBs (with serial sections) enables the identification of multiple markers in the same population that underwent EMT. Meanwhile, EMT also underlies several important developmental events, such as gastrulation and neural crest migration [13, 15, 16]. The first wave of primary EMT after implantation occurs when mesendodermal progenitors are formed from the developing epiblast, prior to the formation of the primitive streak [13]. Since pluripotent stem cells closely resemble cells from the developing epiblast [1], we speculate that the EMT we observed in hPSCs and EBs could correlate with this step of gastrulation in human development.

Protein tyrosine kinase 7 (PTK7) is a transmembrane receptor initially identified in colon carcinoma as colon carcinoma kinase-4, CCK4 [17]. Orthologs of PTK7 have been identified in Hydra (*Lemon*), *Drosophila* (*offtrack, otk*), chicken (*kinase-like gene, KLG*) and mouse (*Ptk7*) [18]. PTK7 has been shown to play critical roles in developmental processes, most notably the regulation of gastrulation and neural tube closure [14, 19]. Both of these events entail convergent extension, which involves the regulation of planar cell polarity [20, 21], in which PTK7 is a crucial regulator [19, 22, 23]. In addition, PTK7 is also involved in other developmental events such as axon guidance, neural crest migration and cardiac morphogenesis [24-28].

Although PTK7 has a conserved tyrosine kinase domain, there is no evidence that kinase activity is required for its function [29, 30]. Nevertheless, considering its essential role in embryonic development, the signaling mechanism of PTK7 is of considerable importance. Recent functional analysis suggested that PTK7 acts as a Wnt co-receptor that activates the

planar cell polarity pathway (non-canonical Wnt signaling) while inhibiting canonical Wnt signaling [31]. Additionally, PTK7/Otk has been shown to complex with PlexinA1 and VEGFR1 to confer specificity to downstream signaling [26, 32-34]. This suggested the role of PTK7 as a versatile co-receptor for modulating downstream signaling responses or cell-cell contacts [35]. Because of its important role in early development, we investigated the expression of PTK7 in embryoid body development. We found PTK7 to be a reliable marker for EMT in EBs. We further demonstrated the utility of PTK7 as a marker to capture and isolate the cells that underwent developmental EMT to produce cells that appeared to reflect a mesendodermal state. To further define these cells, we compared them to definitive endoderm derived from hPSCs. The characterization of these three stages of human development (epiblast, mesendoderm and endoderm) can potentially unveil novel markers that are important for human development, and provide a platform for the study of early embryonic developmental disorders.

### 4.3 Materials And Methods

*Culture of Pluripotent Stem Cells*

Pluripotent stem cells were cultured as described [6, 36] according to UCLA Escro.

*Generation of Embryoid Bodies*

Embryoid Bodies were generated using established methods [6]. Briefly, embryonic stem cell colonies are dissociated from the plate mechanically or by collagenase. The cell colonies are centrifuged and then pipeted into ultra low attachment plates (Corning).

*Serial sections of Embryoid Bodies*

Embryoid bodies were fresh embedded in O.C.T. Compound (Sakura), rapidly frozen with dry ice, and then stored at -80 for at least 30minutes. The frozen embryoid bodies were sectioned on a CM3050S cryostat (Leica) at 6-10uM thickness.

*Immunofluorescence*

Immunostaining was performed as described [36]. Embryoid body sections or cells on coverslips are fixed in 4% PFA, blocked for 5 to 30min, stained with primary antibody overnight at 4°C, washed 3X with PBST, incubated with secondary antibody at RT for 1hr, washed 3X with PBST, mounted with Prolong Gold (Invitrogen) and imaged. All imaging was performed on Zeiss Axio Imager A1. Antibodies used include the following: chicken anti-Vimentin (Covance), 1:1000, mouse anti-SOX17, goat anti-BRACHYURY (R&D), both at 1:50, mouse anti-PTK7 (Sigma-Aldrich), mouse anti-EPCAM, rabbit anti-cleaved CASPASE3, rabbit anti-ECADHERIN, rabbit anti-LIN28A, rabbit anti-LIN28B, rabbit anti-NANOG, rabbit anti-OCT4, rabbit anti-SOX2, rabbit anti-SLUG (Cell Signaling Technology), rabbit anti-NCADHERIN (Santa Cruz), rabbit anti-EOMES, rabbit anti-Ki67 (Abcam) and goat anti-FOXA2 (Santa Cruz Biotech), all at 1:100.

*FAC-Sorting of Embryoid Bodies and Embryonic Stem Cells*

Embryoid bodies were dissociated with TrypLE and then filtered through a 40um cell strainer (BD Biosciences). Embryonic stem cells were dissociated with PBS and filtered through a 40um cell strainer. Dissociated cells were blocked with 5% FBS (FACS buffer), stained with primary antibody for 30min, washed 2X with FACS buffer, incubated with secondary antibody for 30min, washed 3X with FACS buffer, stained with 7AAD for 10min, then filtered through 40um cell strainer before FACS. FAC-sorting was performed at the UCLA BSCRC FACS Core with FACSAria II (BD Biosciences).

*Expression Analysis*

RNA isolation was performed with either Stratagene Absolutely RNA miniprep kit or Qiagen RNEasy micro kit; reverse transcription, and real-time PCR were performed as described [6, 36]. Microarray profiling was performed with Affymetrix Human HG-U133 2.0 Plus arrays as

described [6, 36]. Data were normalized with Robust Multichip Algorithm in Genespring. Probe sets that were not expressed at a raw value of > 50 in at least 10% of samples were eliminated from further analysis.

## 4.4 Results

### *Establishing embryoid body as a model system for studying early differentiation*

To demonstrate the utility of embryoid bodies as model system, we investigated embryoid bodies that were differentiated for 5 days. Immunofluorescence of serial EB cyro-sections revealed that E-CADHERIN and N-CADHERIN, markers for epithelial and mesenchymal cells respectively, were mutually exclusive (Fig 4-1). EBs at earlier stages of differentiation also displayed the same mutual exclusivity between E-CAD and N-CAD (Fig 4-2 and data not shown). Since undifferentiated hESCs/hiPSCs express E-CAD but not N-CAD, this suggested that the N-CAD positive cells represent a switch from an epithelial to a mesenchymal state. In mammalian development, the first EMT occurs during gastrulation, therefore we examined whether these N-CAD positive cells co-expressed other markers relevant to gastrulation. PTK7 is a transmembrane protein that is critical for normal gastrulation in mouse and zebrafish [14, 19]. In serial sections of EBs, PTK7 was shown to be co-expressed in N-CAD positive and E-CAD negative cells. In addition, PTK7 positive cells also displayed an upregulation of VIMENTIN, another marker for mesenchymal cells (Fig 4-2).

### *PTK7 is expressed in populations that have undergone epithelial-to-mesenchymal transition*

EBs were generated from various hESC and hiPSC lines (with or without feeders) under standard culture conditions. EBs were embedded and cryo-sectioned before immunofluorescence staining. Serial sections were stained and analyzed together to display expression of various pluripotency and EMT markers. We observed that most PTK7 positive

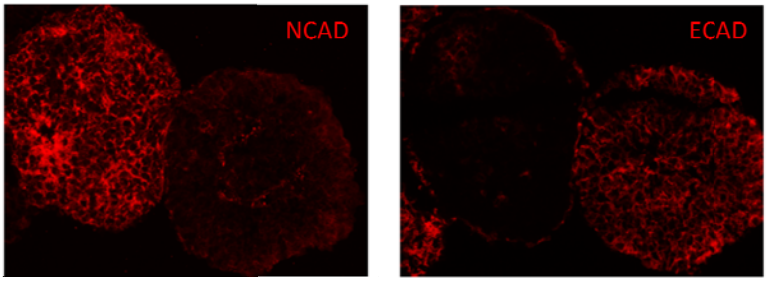



**Figure 4-1. Epithelial and mesenchymal marker analysis in hEBs.**

Immunofluorescence staining of H9 hEB serial sections. Serial sections of H9 Day5 hEB are stained with mesenchymal markers (N-CADHERIN) and epithelial markers (E-CADHERIN)

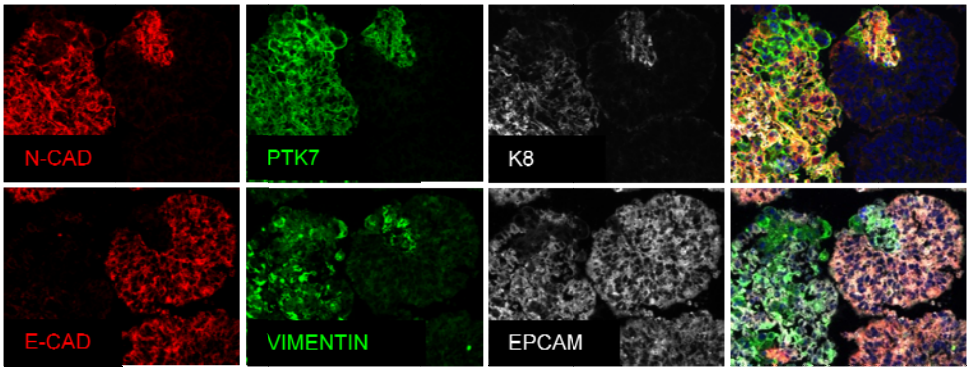



**Figure 4-2. EMT marker analysis of PTK7 population in hEBs.**

Immunofluorescence staining of H9 hEB serial sections. Serial sections of H9 24hr hEB are co-stained with PTK7 (green) and epithelial markers (E-CADHERIN, red; EPCAM, white) or mesenchymal markers (N-CADHERIN, red; VIMENTIN, green). The rightmost column shows the merged images from three fluorescent channels and DAPI.

cells in EBs displayed a loss of pluripotency markers such as OCT4 and NANOG and a loss of epithelial markers such as E-CAD and EPCAM (Fig 4-3 A,C,D,I and Fig S4-1 A,C,D,I). On the other hand, PTK7 positive populations strongly expressed mesenchymal markers such as N-CAD, NCAM and VIMENTIN (Fig 4-3 all, Fig S4-1 all and data not shown). Unexpectedly, OCT4 and SOX2 are still expressed in some PTK-7 positive clusters (Fig 4-3 A,B and Fig S4-1B), indicating that this EMT is likely a very early differentiation event. To assay for proliferation and viability, the PTK7-positive populations were also co-stained with Ki67 and cleaved CASPASE3. PTK7-positive populations were largely proliferating and few cells underwent apoptosis. We also assayed for markers for developmental potential such as LIN28A, LIN28B, and HMGA2. While HMGA2 appeared unchanged (Fig 4-3I and Fig S4-1I), there appeared to be a subtle switch from LIN28A to LIN28B when some cells started expressing PTK7 and underwent EMT differentiation (Fig S4-1 G,H), which is consistent with reported differentiation progression of EBs. To further understand the fate decision of this population, we characterized mesendodermal markers such as SNAIL2 and EOMES. In all stainings we have analyzed, SNAIL2 and EOMES-expressing cells are strictly within PTK7-positive populations (Fig 4-3 J,K, and Fig S4-1 J,K) and data not shown). Of note, SNAIL2 and EOMES are expressed in a mutually exclusive fashion, suggesting possibly different mesendodermal outcomes in individual clusters of the PTK7 positive population.

***PTK7 expression preceded the epithelial-to-mesenchymal transition***

While PTK7 expression largely corresponded with EMT differentiation, we wondered if this expression preceded such an event, of if it was simply a marker of EMT. To visualize a bigger sampling of PTK7 population, we allowed adherent hESC and hiPSC to undergo differentiation *in situ* over different time courses and performed immunofluorescence staining on fixed hESC/hiPSC colonies. Undifferentiated hESC/hiPSC colonies expressed E-CAD and had no PTK7 staining (Fig 4-4 A). Although the majority of PTK-expressing cells in differentiating

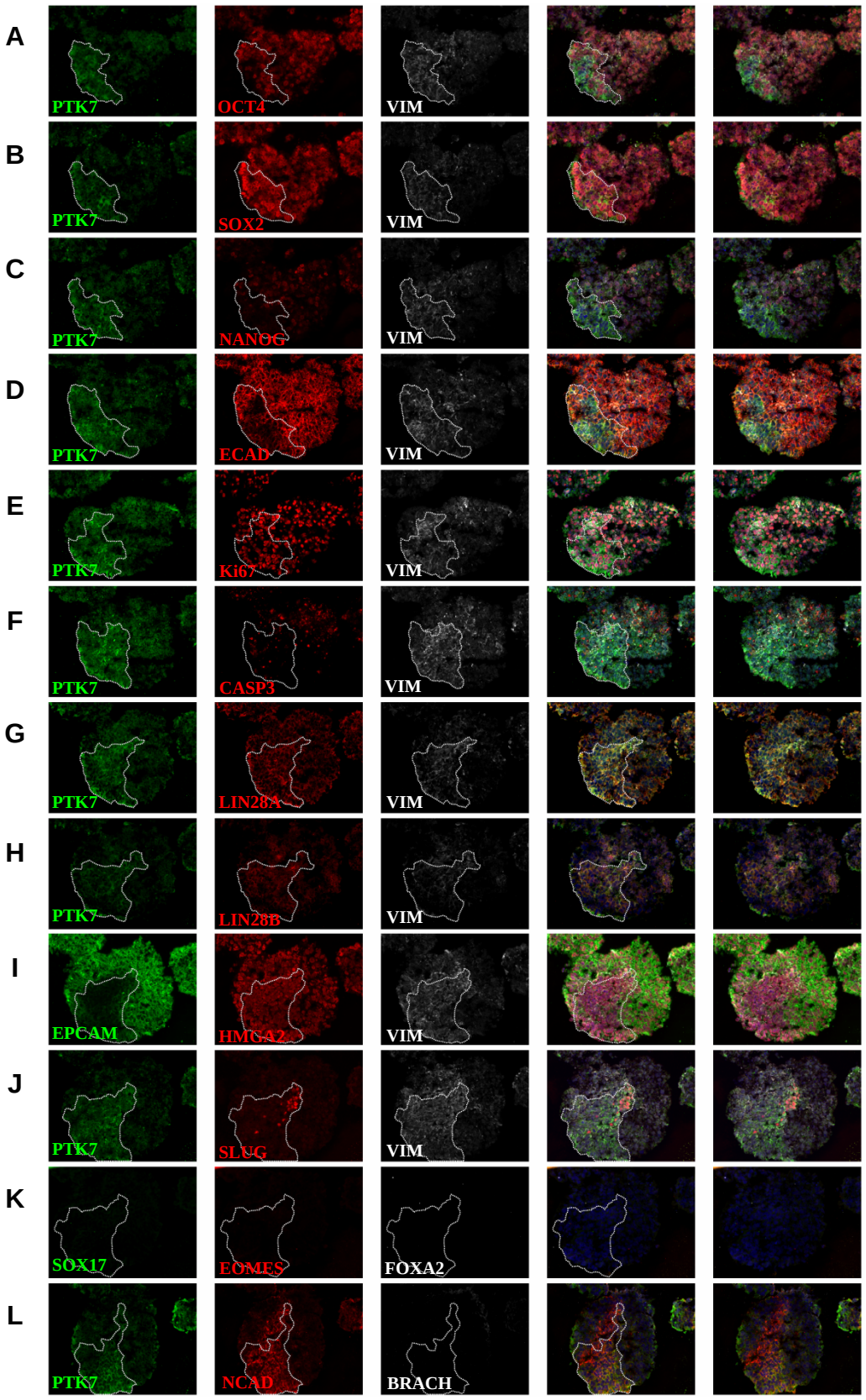

A
PTK7
OCT4
VIM
B
PTK7
SOX2
VIM
C
PTK7
NANOG
VIM
D
PTK7
ECAD
VIM
E
PTK7
Ki67
VIM
F
PTK7
CASP3
VIM
G
PTK7
LIN28A
VIM
H
PTK7
LIN28B
VIM
I
EPCAM
HMGA2
VIM
J
PTK7
SLUG
VIM
K
SOX17
EOMES
FOXA2
L
PTK7
NCAD
BRACH

**Figure 4-3. Pluripotency, lineage, and viability marker analysis of PTK7 population in hEBs.**

Immunofluorescence staining of XFiPSC2 hEB serial sections. White boundary indicates the PTK7 population. XFiPSC2 hEB were cultured for 24hrs before cryosectioned. Serial sections of XFiPSC2 24hr hEB are co-stained with PTK7, (green) epithelial markers (E-CADHERIN, red; EPCAM, green), mesenchymal markers (N-CADHERIN, red; VIMENTIN, white), pluripotency markers (OCT4, SOX2, NANOG, red), developmental genes (LIN28A, LIN28B, HMGA2, red), primitive streak markers (SLUG, red; Brachyury, white) and endodermal markers (SOX17, green; EOMES, red; FOXA2, white). The 4$^{th}$ and 5$^{th}$ columns show the merged images from three fluorescent channels and DAPI.

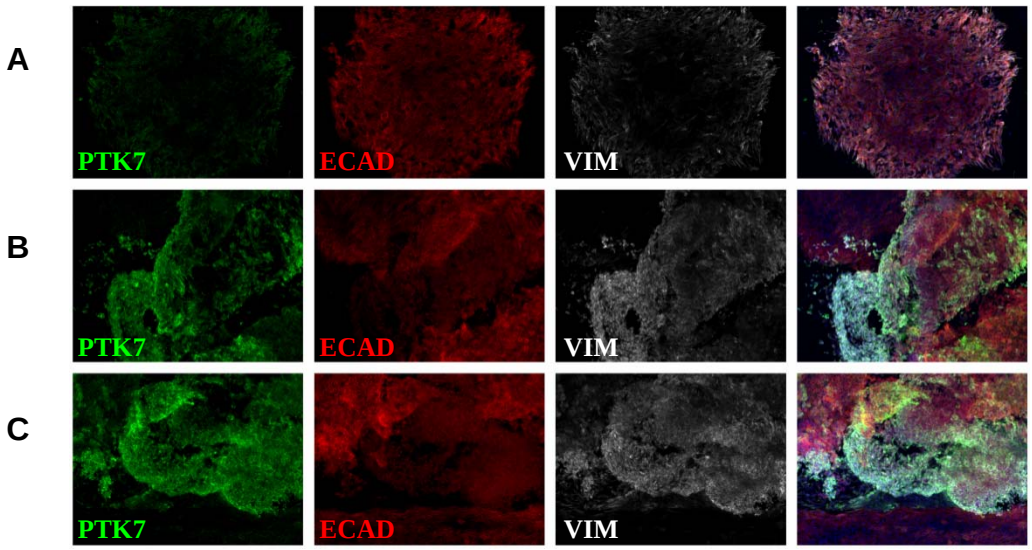



**Figure 4-4. PTK7 and EMT marker analysis in adherent differentiation of hPSCs.**

Immunofluorescence staining of adherent XFiPSC2. XFiPSC2 were plated on coverslips and fixed before immunostaining with PTK7, epithelial maker (E-CADHERIN) and mesenchymal marker (VIMENTIN). (A): undifferentiated XFiPSC2. (B) and (C): XFiPSC2 differentiated in low FGF conditions for 2 days. The 4$^{th}$ column shows the merged images from three fluorescent channels and DAPI.

cultures were depleted of E-CAD and other pluripotent markers, a small population of which still expressed E-CAD (Fig 4-4 B,C). This suggested that PTK7 expression preceded loss of E-CAD and that the full EMT event might require the physical reorganization imparted by EB formation.

In an attempt to understand the mechanism behind this EMT event, the activity of various signaling pathways was assessed by immunostaining for appropriate markers. Assaying for the activity of TGF, BMP, or MAPK pathways did not uncover any trends that would correlate with EMT in this setting (Fig S4-2 and data not shown), suggesting that none of these pathways plays a role in this cell fate decision. This could suggest that more direct, cell-cell interactions might drive the EMT event. It is tempting to speculate that simply the rearrangement of cells from 2-D cell culture into a sphere and the subsequent shape and cell contact changes induced in such an even might drive cell fate decisions in this case, as has been proposed in other settings. To further define the molecular basis for this EMT event and to better understand the consequences, we isolated both populations and performed gene expression profiling.

***Capture and Transcriptional Profile of PTK7+ population***

The antibody used to identify PTK7 expression also appeared suitable to isolate cells by FACS. Therefore, we sorted PTK7 positive, PTK7 negative and undifferentiated hESCs to perform transcriptome profiling. We first validated the sorting strategy by RT-PCR as shown in Fig 4-5.

From microarray analysis, PTK7 positive population displayed upregulation of markers of EMT and gastrulation, such as *TWIST, SNAIL2, ZEB1, MMP2, MMP10, FOXC2* and *VIMENTIN* (Fig 4-6G). In addition, PTK7+ population also expressed increased levels of TGF-beta and targets such as *TGFβ1, TGFβ2, TGFβ-I* and *CDKN2b* (Fig 4-6F). Fibronectin (*FBN1*) and fibronectin-interacting proteins, including *DECORIN, LUMICAN, TENASIN-C* and a panel of collagens (*COL1A1, COL2A1, COL3A1, COL4A1, COL4A2, COL5A1, COL6A1, COL6A2, COL6A3,*

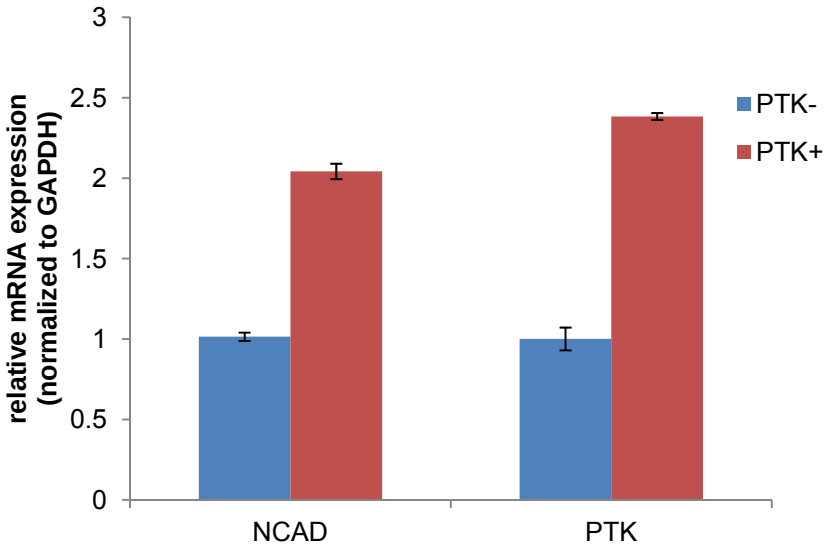
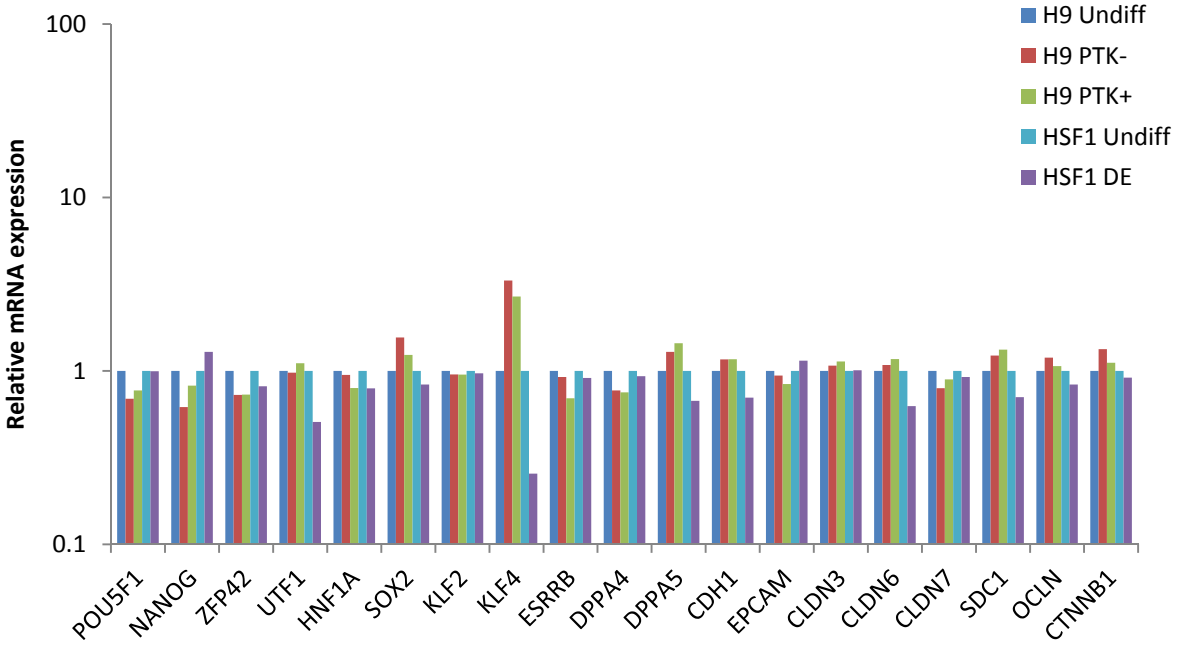


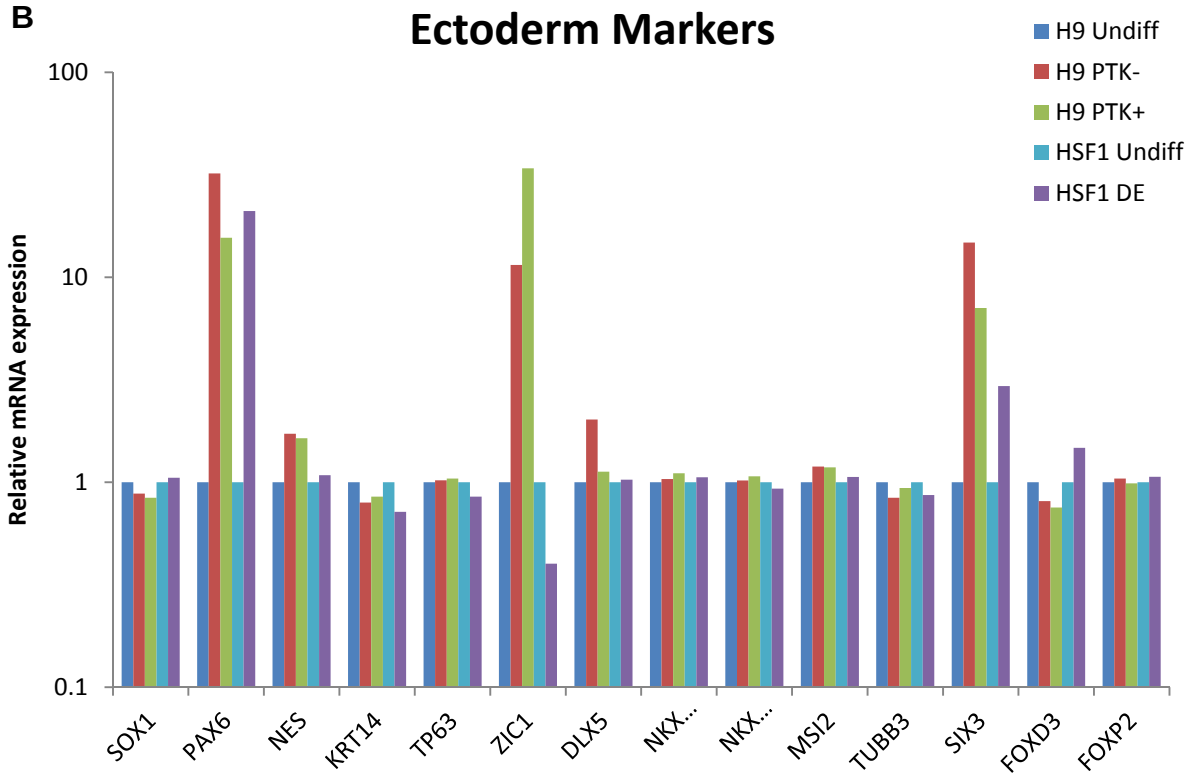

**Figure 4-5. mRNA expression analysis on PTK7-sorted hEBs.**

XFiPSC2 EBs were sorted into PTK+ and PTK- population. Reverse-transcription PCR was performed. mRNA expression was normalized to GAPDH. Error bars represent the standard error over four technical replicates.

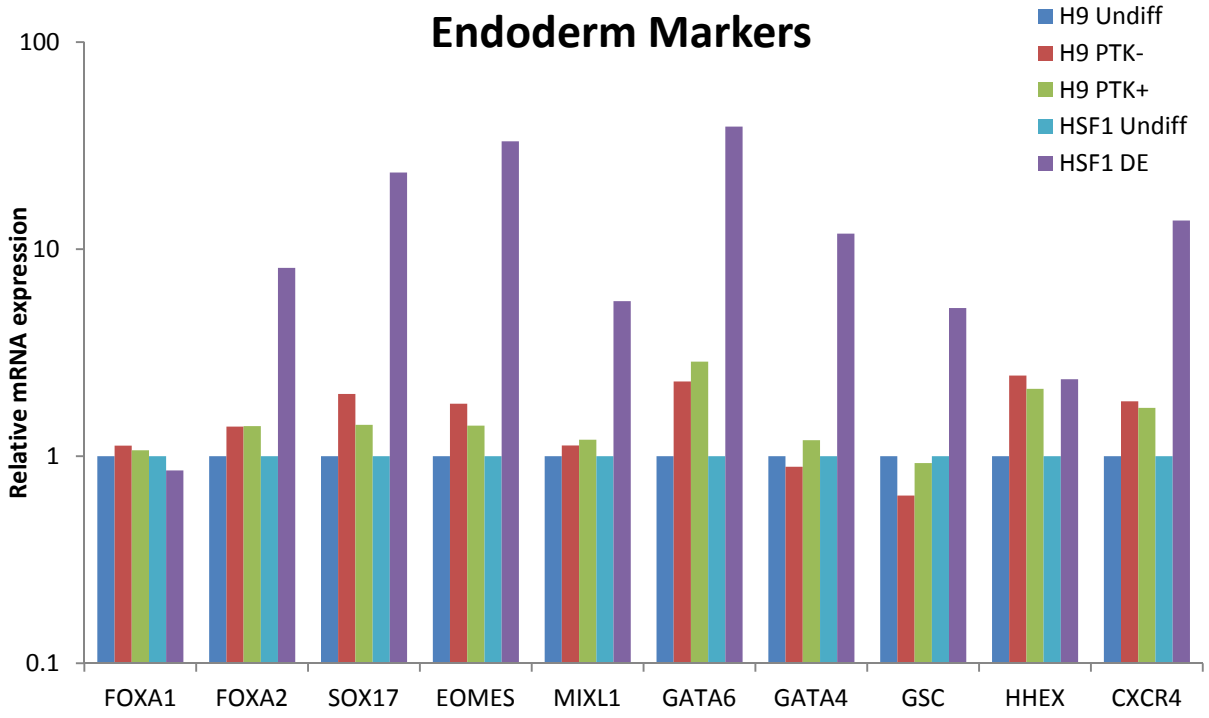
A


Pluripotency Markers
Relative mRNA expression
100
10
1
0.1
H9 Undiff
H9 PTK-
H9 PTK+
HSF1 Undiff
HSF1 DE
POU5F1
NANOG
ZFP42
UTF1
HNF1A
SOX2
KLF2
KLF4
ESRRB
DPPA4
DPPA5
CDH1
EPCAM
CLDN3
CLDN6
CLDN7
SDC1
OCLN
CTNNB1

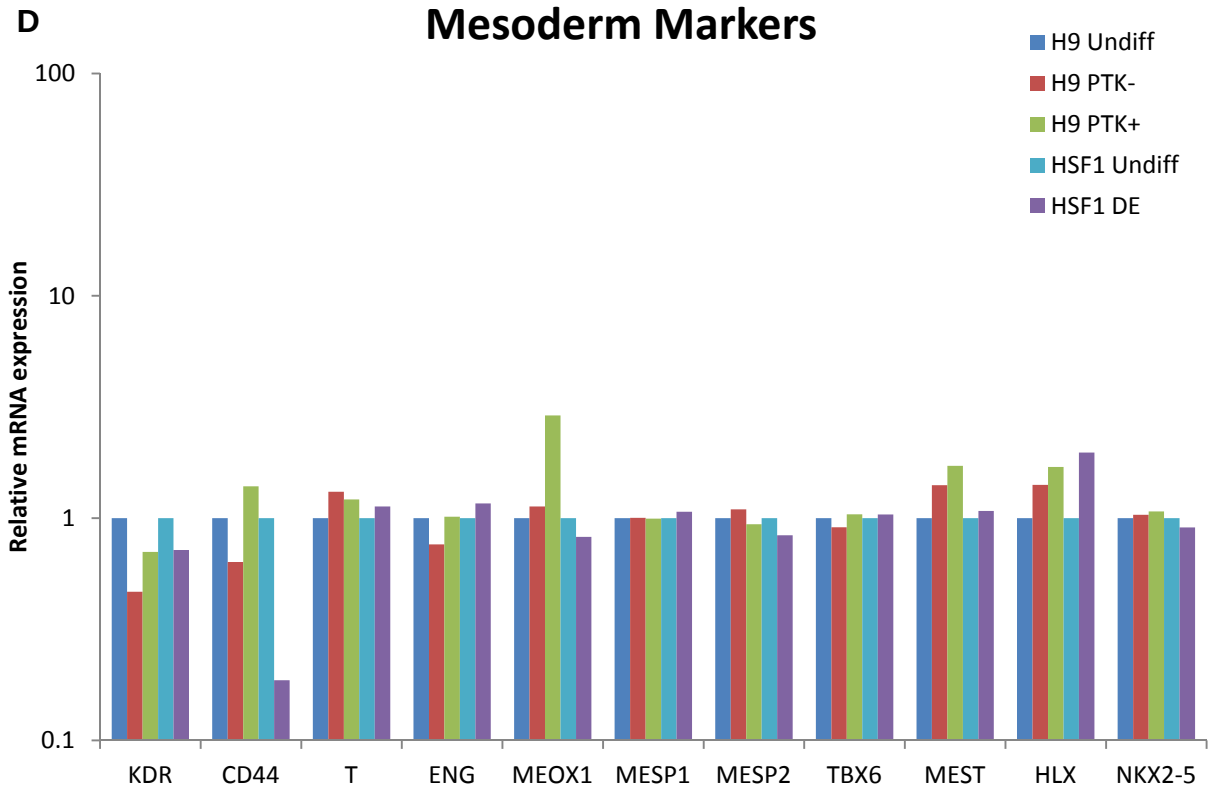
B

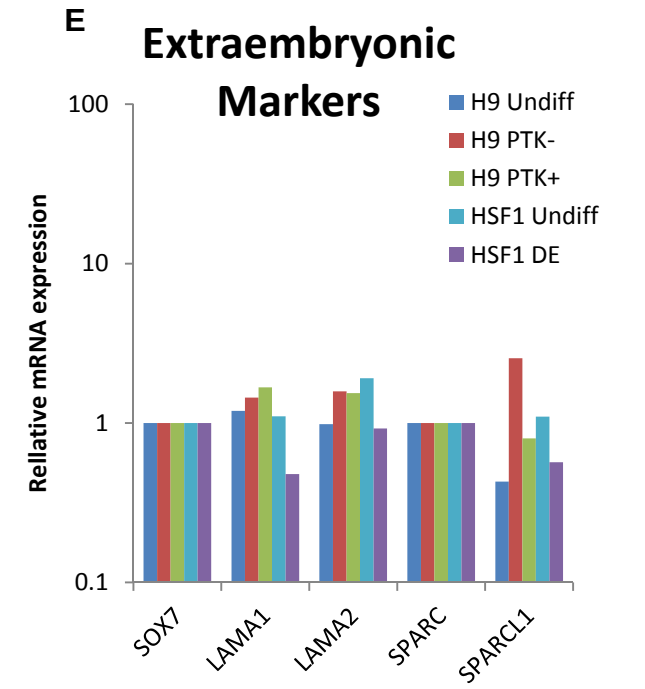

Ectoderm Markers
Relative mRNA expression
100
10
1
0.1
H9 Undiff
H9 PTK-
H9 PTK+
HSF1 Undiff
HSF1 DE
SOX1
PAX6
NES
KRT14
TP63
ZIC1
DLX5
NKX...
NKX...
MSI2
TUBB3
SIX3
FOXD3
FOXP2

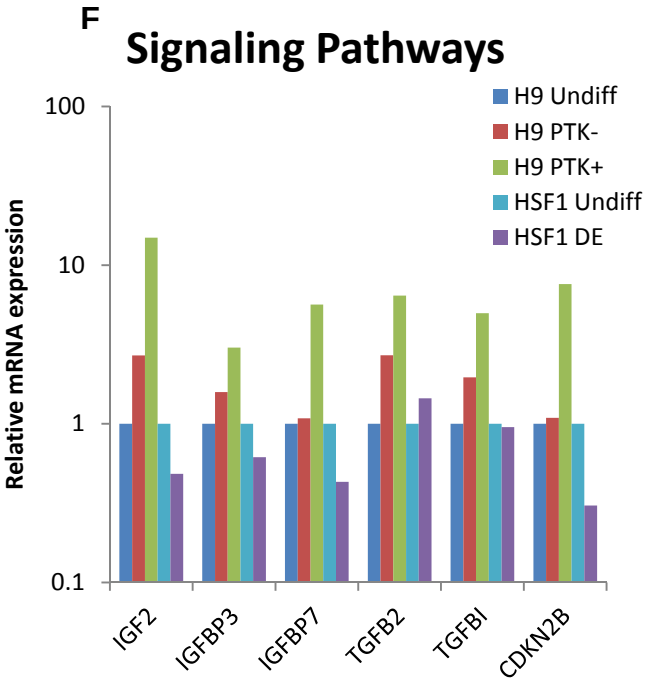
C

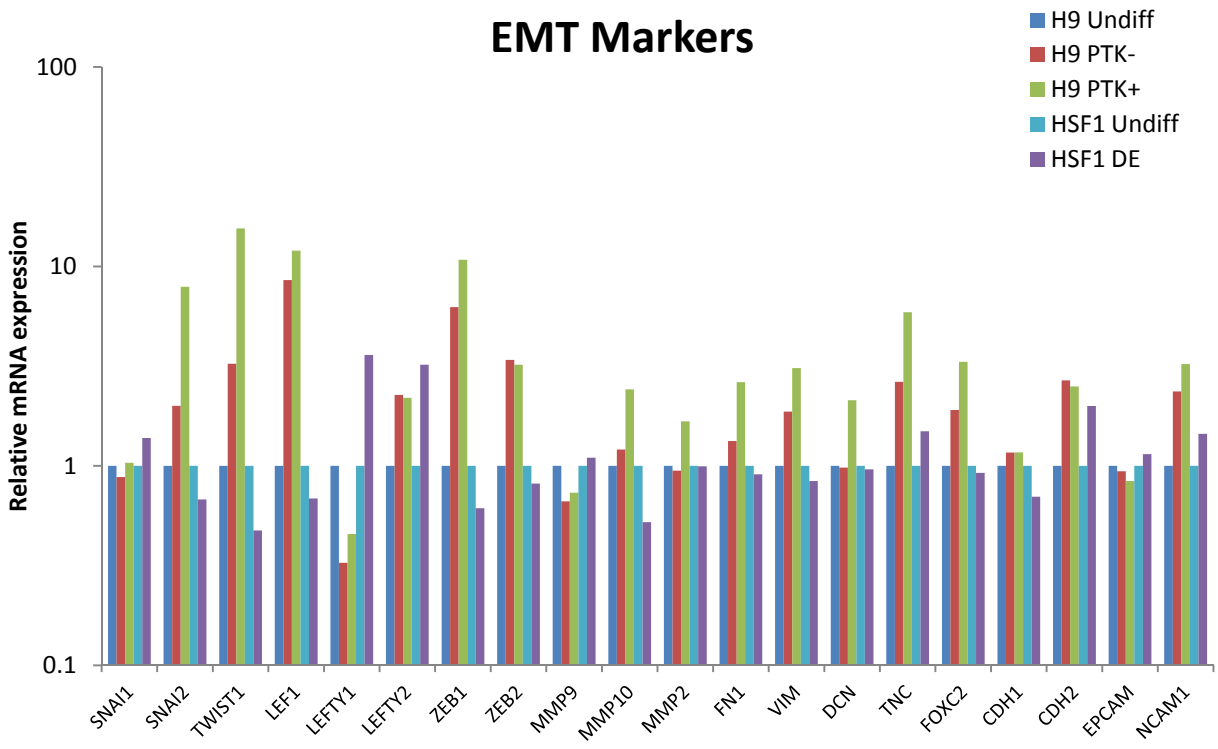

Endoderm Markers
H9 Undiff
H9 PTK-
H9 PTK+
HSF1 Undiff
HSF1 DE
Relative mRNA expression
100
10
1
0.1
FOXA1
FOXA2
SOX17
EOMES
MIXL1
GATA6
GATA4
GSC
HHEX
CXCR4

D


Mesoderm Markers
H9 Undiff
H9 PTK-
H9 PTK+
HSF1 Undiff
HSF1 DE
Relative mRNA expression
100
10
1
0.1
KDR
CD44
T
ENG
MEOX1
MESP1
MESP2
TBX6
MEST
HLX
NKX2-5


E
Extraembryonic Markers
H9 Undiff
H9 PTK-
H9 PTK+
HSF1 Undiff
HSF1 DE
Rellative mRNA expression
100
10
1
0.1
SOX7
LAMA1
LAMA2
SPARC
SPARCL1


F
Signaling Pathways
H9 Undiff
H9 PTK-
H9 PTK+
HSF1 Undiff
HSF1 DE
Relative mRNA expression
100
10
1
0.1
IGF2
IGFBP3
IGFBP7
TGFB2
TGFBI
CDKN2B


EMT Markers
H9 Undiff
H9 PTK-
H9 PTK+
HSF1 Undiff
HSF1 DE
Relative mRNA expression
100
10
1
0.1
SNAI1
SNAI2
TWIST1
LEF1
LEFTY1
LEFTY2
ZEB1
ZEB2
MMP9
MMP10
MMP2
FN1
VIM
DCN
TNC
FOXC2
CDH1
CDH2
EPCAM
NCAM1

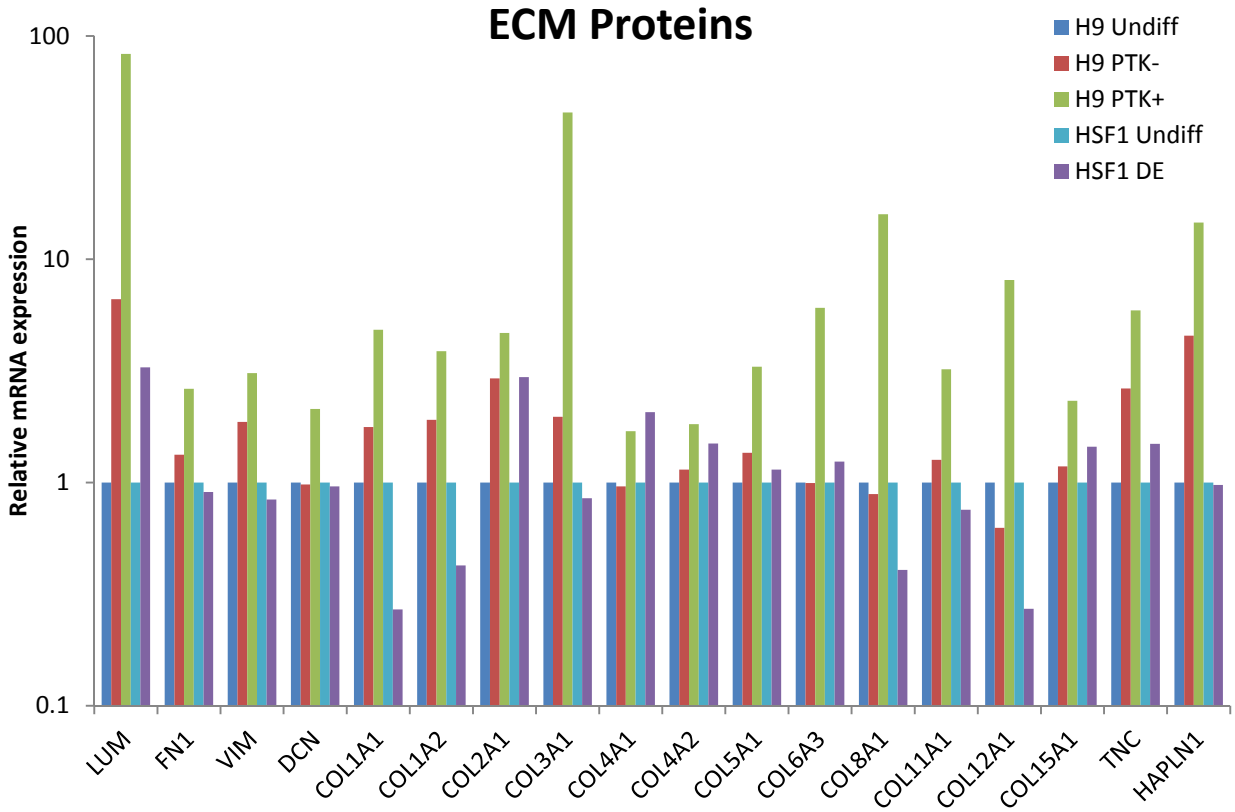

ECM Proteins
H9 Undiff
H9 PTK-
H9 PTK+
HSF1 Undiff
HSF1 DE
Relative mRNA expression
100
10
1
0.1
LUM
FN1
VIM
DCN
COL1A1
COL1A2
COL2A1
COL3A1
COL4A1
COL4A2
COL5A1
COL6A3
COL8A1
COL11A1
COL12A1
COL15A1
TNC
HAPLN1

**Figure 4-6. Lineage marker analysis on undifferentiated hPSC, PTK-, PTK7+ and DE populations.**

Feeder-free H9 were sorted into PTK+ and PTK- population. Definitive endoderm was induced by treating HSF1 with Activin A for 5 days (HSF1 DE). Expression level was determined by microarray analysis as described in materials and methods. All H9 samples were normalized to undifferentiated H9 (H9 Undiff). All HSF1 samples were normalized to undifferentiated HSF1 (HSF1 Undiff). The relative expression of all samples were analyzed for (A) pluripotency markers; (B) ectoderm markers, (C) endoderm markers, (D) mesoderm markers, (E) extraembryonic markers, (F) IGF and TGFβ signaling pathways, (G) EMT markers, and (H) ECM proteins.

*COL8A1, COL11A1, COL12A1, COL15A1*), were also significantly induced (Fig 4-6H). It has been shown that TGFβ signaling is important for gastrulation and EMT both *in vivo* and *in vitro* [13]. In addition, an intact fibronectin matrix complex is also essential for developmental patterning and gastrulation [37], while DECORIN has been shown to play a role in convergent extension [38]. These findings support the notion that the PTK7 population is representative for human cells fated to undergo developmental EMT or even gastrulation.

***Transcriptional profiling of definitive endoderm***

Because the PTK7 positive population appeared to transcriptionally resemble mesendodermal progenitors, we sought to isolate and profile a subsequent developmental stage, the definitive endoderm. In this case, undifferentiated hESCs were exposed to activin A, a TGFβ agonist (as described previously) [36, 39, 40], and at least 90% of the cells in culture expressed SOX17 and FOXA2 after 5 days of specification (termed DE hereafter). We normalized these transcriptional profiles and studied the changes in pluripotency, extra-embryonic, ectoderm, endoderm, and mesoderm markers.

Both PTK7 positive and DE cells expressed similar or slightly lower levels of pluripotency markers as undifferentiated ES cells, with the exception of *KLF4*. Upon differentiation to DE, *KLF4* was downregulated by 75%, whereas in PTK7 positive cells, it was increased by 2.7-fold (Fig 4-6A). Both PTK7 positive and DE cells did not display appreciable inductions of extra-embryonic markers (Fig 4-6E). Among typical ectodermal markers, DE had increased levels of *PAX6* (20-fold). In comparison, PTK7 positive cells also overexpressed *PAX6* (15-fold) and *SIX3* (7-fold), but it should be noted that PTK7 negative cells expressed even higher levels of *PAX6* (32-fold) and *SIX3* (15-fold). Except for *PAX6*, *SIX3* and *ZIC1*, all other ectodermal markers remained unchanged in DE and PTK7 positive cells (Fig 4-6B). This suggested that these ectodermal markers may upregulate upon spontaneous differentiation, but PTK7 negative population appears to be somewhat enriched for them. Among endodermal markers, DE has

significant upregulation of most (*FOXA2, SOX17, EOMES, MIXL1, GATA4, GATA6, GSC* and *CXCR4*) as expected (Fig 4-6C). The PTK7 positive population only displayed slight upregulation of *GATA6* and *HHEX* (2.9-fold and 2.1-fold). Compared to populations shown to be committed to the mesoderm [41], PTK7+ cells only displayed modest upregulation in a handful of these markers (*CD44* at 1.3-fold, *MEOX1* at 2.9-fold, *MEST1* at 1.7-fold and *HLX* at 1.7-fold), suggesting that this population is likely not committed to mesoderm (Fig 4-6D). It is worth noting, however, that PTK7+ did appear to have higher levels of certain mesodermal markers than PTK7- (*CD44* at 2.2-fold, *MEOX1* at 2.6-fold over PTK7 negative). DE in general did not express increased levels of mesodermal markers.

## 4.5 Discussion

The development of human embryonic stem cells, and more recently, human induced pluripotent stem cells, has brought about a glimpse of the future in human embryology studies. Embryoid bodies have long been fascinating research tools since the discovery of teratocarcinoma cell lines. *A priori*, we presumed that EBs followed a spontaneous random differentiation into three specific germ layers. Here we demonstrated that PTK7, a planar cell polarity regulator, is expressed in a population of cells that have undergone epithelial-to-mesenchymal transition. Although PTK7 was somewhat induced before the loss of epithelial nature, the EMT only appeared to complete when a three-dimension structure formed on the plate, or when embryoid bodies were aggregated. While most human studies currently focus on differentiation in terms of growth factor induction and gene expression initiation, regulation of polarity and polarity genes has been shown to play important roles in mouse embryonic development.

### *Uncoupling between regulation at RNA level and protein level*

While we observed decreases in epithelial and pluripotency markers such as E-CADHERIN, EPCAM, NANOG and OCT4 in PTK7 positive population on the protein level, we did not capture the same trend on the RNA level, suggesting uncoupled regulation between the two. Such regulation of E-CADHERIN have been reported previously in mouse [42-44]. While SNAIL can downregulate *E-Cadherin* [45], we observed downregulation of E-CADHERIN protein prior to induction of *SNAIL* family. These data raise the intriguing notion that post-translational regulation of genes may underlie the onset of human developmental EMT.

***Upregulation of ZIC1 in PTK7+ population suggests embryonic mesoderm formation***

Among various established ectoderm markers, we observed a marked upregulation of *ZIC1* (but not *ZIC2*, *ZIC3*) in the PTK7 positive population. While *ZIC1* has long been regarded as a classic ectoderm gene, its expression is restricted to the neuroectoderm only after the neuralation stage in mouse and *Drosophila* [46-48]. However, during early primitive streak stage, *ZIC* genes are upregulated in the embryonic mesoderm. While *ZIC2* and *ZIC3* are also detected in the early ectoderm, the highest level of *ZIC1* is restricted to the embryonic mesoderm at this stage [49]. The upregulation of *ZIC1* transcript corroborates with the assessment that this PTK7 positive population is a product of an early EMT event. Further analysis needs to be carried out to determine whether *ZIC1* is a reliable determinant of early mesodermal differentiation in human pluripotent stem cells.

***Upregulation of PAX6 and SIX3 in PTK- population suggests anterior ectoderm formation***

*PAX6* and *SIX3* are ectodermal markers that were upregulated in PTK7 positive population compared to undifferentiataed ES cells. However, the upregulation of these two genes is much more dramatic in the PTK7 negative cells (Fig 4-6B). It has been established that *PAX6* and *SIX3* are some of the earliest determinants for the establishment of the forebrain and an eye field in the anterior neuroectoderm at the late gastrula/early neurala stage [50-59]. This is

suggestive of a more ectodermal fate of the PTK7 negative population. The rapid development of an eye field, and eventually optic cups from embryoid bodies have been reported [60], and it would be intriguing to use lineage tracing with a PTK7 reporter to see if only PTK7 negative populations can mature to generate a functional optic cup. In addition, it is unknown whether genes that are highly upregulated in the PTK7 negative population are indicative of early human ectoderm differentiation. These candidates include *LIX1*, a gene suggested to be involved in neurogenesis and limb development [61, 62].

***PTK7+ population shows upregulation of EMT markers, but not mesoderm or definitive endoderm markers***

As we have demonstrated, PTK7 positive cells displayed an upregulation of a panel of EMT/primitive streak markers such as *TWIST1*, *SNAIL2*, *VIMENTIN*, *ZEB1*, *FOXC2* and *MMP2*. It is plausible that this EMT event is still at an early stage based on two observations: (1) some of the classic EMT markers, such as *SNAIL1*, are not turned on; (2) in some of the PTK7 positive cells in adherent culture differentiation, E-CADHERIN is not completely turned off, both on the RNA and protein level. It is likely that the incomplete silencing of *E-CAD* transcription stems from the lack of induction of *SNAIL1*, a well-established inhibitor of *E-CAD* transcription. It is possible that primitive streak formation/gastrulation is initiated by the induction of the genes associated with EMT that we uncovered in the PTK7 positive population, whereas the induction of *SNAIL1* (by growth factors we did not provide in our spontaneous differentiation) completes the transcriptional silencing of *E-Cadherin* and drives full epithelial-to-mesenchymal transition. The most upregulated genes in the PTK7+ population includes LUMICAN, DECORIN, FBN1, *TENASIN-C* as well as a vast collection of *COLLAGEN* isotypes, indicating that early remodeling of extra-cellular matrix may pave the way of large-scale tissue patterning and cell migration in human embryonic EMT.

Also corroborative with the notion that the PTK7 positive population resembles early developmental EMT is the lack of appreciable induction in mesdoermal and endodermal markers, as compared to committed mesoderm and definitive endoderm respectively (Fig 4-6 B,C) and [41]. It is likely that the PTK7 positive population is not yet committed to mesodermal or endodermal lineages, as we have seen expression of mesodermal markers (*SNAIL2*) and endodermal markers (*EOMES*) in PTK7 positive population, albeit in a mutually exclusive fashion (Fig S4-1 J,K). The propensity of PTK7 positive and PTK7 negative populations to form ectodermal derivatives would be a topic of interest to study if cell fates are restricted as early as the gastrula stage in human development.

***Impact of findings on human embryology***

Our findings in studying the earliest human pluripotent stem cell has answered some questions on the earliest event of EMT while opening new ventures for future discoveries. It also suggests that some well established markers, such as *ZIC1*, may have different expression patterns and functions under different contexts. The FACS isolation of PTK7 positive and negative cells allows us to separate two populations with distinct developmental characteristics. The analysis of the transcriptional profile reveals some expected markers as well as a panel of new markers to be explored. While the PTK7 positive population appears to be a mesendoderm progenitor, there may be more specific markers and determinants for driving mesoderm and endoderm formation within this population, which should be further studied. Our approach has demonstrated, in principle, the potential of pluripotent stem cells and embryoid bodies in modeling the very early stage of human development.

### 4.6 Bibliography


1. Nichols, J. and A. Smith, *Naive and primed pluripotent states.* Cell Stem Cell, 2009. **4**(6): p. 487-92.
2. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.
3. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
4. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
5. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.
6. Lowry, W.E., et al., *Generation of human induced pluripotent stem cells from dermal fibroblasts.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 2883-8.
7. Itskovitz-Eldor, J., et al., *Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers.* Mol Med, 2000. **6**(2): p. 88-95.
8. Sharon, N., et al., *Molecular and functional characterizations of gastrula organizer cells derived from human embryonic stem cells.* Stem Cells. **29**(4): p. 600-8.
9. Oliveri, P., Q. Tu, and E.H. Davidson, *Global regulatory logic for specification of an embryonic cell lineage.* Proc Natl Acad Sci U S A, 2008. **105**(16): p. 5955-62.
10. Winston, N.J. and M.H. Johnson, *Can the mouse embryo provide a good model for the study of abnormal cellular development seen in human embryos?* Hum Reprod, 1992. **7**(9): p. 1291-6.
11. Hanna, J., et al., *Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin.* Science, 2007. **318**(5858): p. 1920-3.

12. Kazuki, Y., et al., *Complete genetic correction of ips cells from Duchenne muscular dystrophy.* Mol Ther. **18**(2): p. 386-93.

13. Thiery, J.P., et al., *Epithelial-mesenchymal transitions in development and disease.* Cell, 2009. **139**(5): p. 871-90.

14. Yen, W.W., et al., *PTK7 is essential for polarized cell motility and convergent extension during mouse gastrulation.* Development, 2009. **136**(12): p. 2039-48.

15. Heisenberg, C.P. and L. Solnica-Krezel, *Back and forth between cell fate specification and movement during vertebrate gastrulation.* Curr Opin Genet Dev, 2008. **18**(4): p. 311-6.

16. Sauka-Spengler, T. and M. Bronner-Fraser, *A gene regulatory network orchestrates neural crest formation.* Nat Rev Mol Cell Biol, 2008. **9**(7): p. 557-68.

17. Mossie, K., et al., *Colon carcinoma kinase-4 defines a new subclass of the receptor tyrosine kinase family.* Oncogene, 1995. **11**(10): p. 2179-84.

18. Grassot, J., et al., *Origin and molecular evolution of receptor tyrosine kinases with immunoglobulin-like domains.* Mol Biol Evol, 2006. **23**(6): p. 1232-41.

19. Lu, X., et al., *PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates.* Nature, 2004. **430**(6995): p. 93-8.

20. Roszko, I., A. Sawada, and L. Solnica-Krezel, *Regulation of convergence and extension movements during vertebrate gastrulation by the Wnt/PCP pathway.* Semin Cell Dev Biol, 2009. **20**(8): p. 986-97.

21. Wallingford, J.B., S.E. Fraser, and R.M. Harland, *Convergent extension: the molecular control of polarized cell movement during embryonic development.* Dev Cell, 2002. **2**(6): p. 695-706.

22. Golubkov, V.S., et al., *The Wnt/planar cell polarity protein-tyrosine kinase-7 (PTK7) is a highly efficient proteolytic target of membrane type-1 matrix metalloproteinase: implications in cancer and embryogenesis.* J Biol Chem. **285**(46): p. 35740-9.

23. Wehner, P., et al., *RACK1 is a novel interaction partner of PTK7 that is required for neural tube closure.* Development. **138**(7): p. 1321-7.

24. Cafferty, P., L. Yu, and Y. Rao, *The receptor tyrosine kinase Off-track is required for layer-specific neuronal connectivity in Drosophila.* Development, 2004. **131**(21): p. 5287-95.

25. Pulido, D., et al., *Dtrk, a Drosophila gene related to the trk family of neurotrophin receptors, encodes a novel class of neural cell adhesion molecule.* EMBO J, 1992. **11**(2): p. 391-404.

26. Winberg, M.L., et al., *The transmembrane protein Off-track associates with Plexins and functions downstream of Semaphorin signaling during axon guidance.* Neuron, 2001. **32**(1): p. 53-62.

27. Shnitsar, I. and A. Borchers, *PTK7 recruits dsh to regulate neural crest migration.* Development, 2008. **135**(24): p. 4015-24.

28. Toyofuku, T., et al., *Dual roles of Sema6D in cardiac morphogenesis through region-specific association of its receptor, Plexin-A1, with off-track and vascular endothelial growth factor receptor type 2.* Genes Dev, 2004. **18**(4): p. 435-47.

29. Kroiher, M., M.A. Miller, and R.E. Steele, *Deceiving appearances: signaling by "dead" and "fractured" receptor protein-tyrosine kinases.* Bioessays, 2001. **23**(1): p. 69-76.

30. Miller, M.A. and R.E. Steele, *Lemon encodes an unusual receptor protein-tyrosine kinase expressed during gametogenesis in Hydra.* Dev Biol, 2000. **224**(2): p. 286-98.

31. Peradziryi, H., et al., *PTK7/Otk interacts with Wnts and inhibits canonical Wnt signalling.* EMBO J. **30**(18): p. 3729-40.

32. Wagner, G., et al., *PlexinA1 interacts with PTK7 and is required for neural crest migration.* Biochem Biophys Res Commun. **402**(2): p. 402-7.

33. Shin, W.S., et al., *Soluble PTK7 inhibits tube formation, migration, and invasion of endothelial cells and angiogenesis.* Biochem Biophys Res Commun, 2008. **371**(4): p. 793-8.

34. Lee, H.K., et al., *Flt-1 regulates vascular endothelial cell migration via a protein tyrosine kinase-7-dependent pathway.* Blood. **117**(21): p. 5762-71.

35. Peradziryi, H., N.S. Tolwinski, and A. Borchers, *The many roles of PTK7: A versatile regulator of cell-cell communication.* Arch Biochem Biophys. **524**(1): p. 71-6.

36. Patterson, M., et al., *Defining the nature of human pluripotent stem cell progeny.* Cell Res. **22**(1): p. 178-93.

37. Darribere, T. and J.E. Schwarzbauer, *Fibronectin matrix composition and organization can regulate cell migration during amphibian development.* Mech Dev, 2000. **92**(2): p. 239-50.

38. Zoeller, J.J., et al., *A central role for decorin during vertebrate convergent extension.* J Biol Chem, 2009. **284**(17): p. 11728-37.

39. Song, Z., et al., *Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells.* Cell Res, 2009. **19**(11): p. 1233-42.

40. Si-Tayeb, K., et al., *Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells.* Hepatology. **51**(1): p. 297-305.

41. Evseenko, D., et al., *Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells.* Proc Natl Acad Sci U S A. **107**(31): p. 13742-7.

42. Lee, J.D., et al., *The FERM protein Epb4.1l5 is required for organization of the neural plate and for the epithelial-mesenchymal transition at the primitive streak of the mouse embryo.* Development, 2007. **134**(11): p. 2007-16.

43. Hirano, M., et al., *EPB41L5 functions to post-transcriptionally regulate cadherin and integrin during epithelial-mesenchymal transition.* J Cell Biol, 2008. **182**(6): p. 1217-30.

44. Zohn, I.E., et al., *p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation.* Cell, 2006. **125**(5): p. 957-69.

45. Oda, H., S. Tsukita, and M. Takeichi, *Dynamic behavior of the cadherin-based cell-cell adhesion system during Drosophila gastrulation.* Dev Biol, 1998. **203**(2): p. 435-50.

46. Aruga, J., et al., *The mouse zic gene family. Homologues of the Drosophila pair-rule gene odd-paired.* J Biol Chem, 1996. **271**(2): p. 1043-7.

47. Aruga, J., et al., *A novel zinc finger protein, zic, is involved in neurogenesis, especially in the cell lineage of cerebellar granule cells.* J Neurochem, 1994. **63**(5): p. 1880-90.

48. Benedyk, M.J., J.R. Mullen, and S. DiNardo, *odd-paired: a zinc finger pair-rule protein required for the timely activation of engrailed and wingless in Drosophila embryos.* Genes Dev, 1994. **8**(1): p. 105-17.

49. Nagai, T., et al., *The expression of the mouse Zic1, Zic2, and Zic3 gene suggests an essential role for Zic genes in body pattern formation.* Dev Biol, 1997. **182**(2): p. 299-313.

50. Seo, H.C., et al., *Expression of two zebrafish homologues of the murine Six3 gene demarcates the initial eye primordia.* Mech Dev, 1998. **73**(1): p. 45-57.

51. Kobayashi, M., et al., *Overexpression of the forebrain-specific homeobox gene six3 induces rostral forebrain enlargement in zebrafish.* Development, 1998. **125**(15): p. 2973-82.

52. Loosli, F., et al., *Six3, a medaka homologue of the Drosophila homeobox gene sine oculis is expressed in the anterior embryonic shield and the developing eye.* Mech Dev, 1998. **74**(1-2): p. 159-64.

53. Oliver, G., et al., *Six3, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development.* Development, 1995. **121**(12): p. 4045-55.

54. Jean, D., K. Ewan, and P. Gruss, *Molecular regulators involved in vertebrate eye development.* Mech Dev, 1998. **76**(1-2): p. 3-18.

55. Winkler, S., et al., *The conditional medaka mutation eyeless uncouples patterning and morphogenesis of the eye.* Development, 2000. **127**(9): p. 1911-9.

56. Bovolenta, P., et al., *Expression pattern of cSix3, a member of the Six/sine oculis family of transcription factors.* Mech Dev, 1998. **70**(1-2): p. 201-3.

57. Chow, R.L., et al., *Pax6 induces ectopic eyes in a vertebrate.* Development, 1999. **126**(19): p. 4213-22.

58. Hill, R.E., et al., *Mouse small eye results from mutations in a paired-like homeobox-containing gene.* Nature, 1991. **354**(6354): p. 522-5.

59. Hogan, B.L., et al., *Small eyes (Sey): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse.* J Embryol Exp Morphol, 1986. **97**: p. 95-110.

60. Eiraku, M., et al., *Self-organizing optic-cup morphogenesis in three-dimensional culture.* Nature. **472**(7341): p. 51-6.

61. Moeller, C., et al., *Murine Lix1, a novel marker for substantia nigra, cortical layer 5, and hindbrain structures.* Brain Res Gene Expr Patterns, 2002. **1**(3-4): p. 199-203.

62. Swindell, E.C., et al., *Cloning and expression analysis of chicken Lix1, a founding member of a novel gene family.* Mech Dev, 2001. **109**(2): p. 405-8.

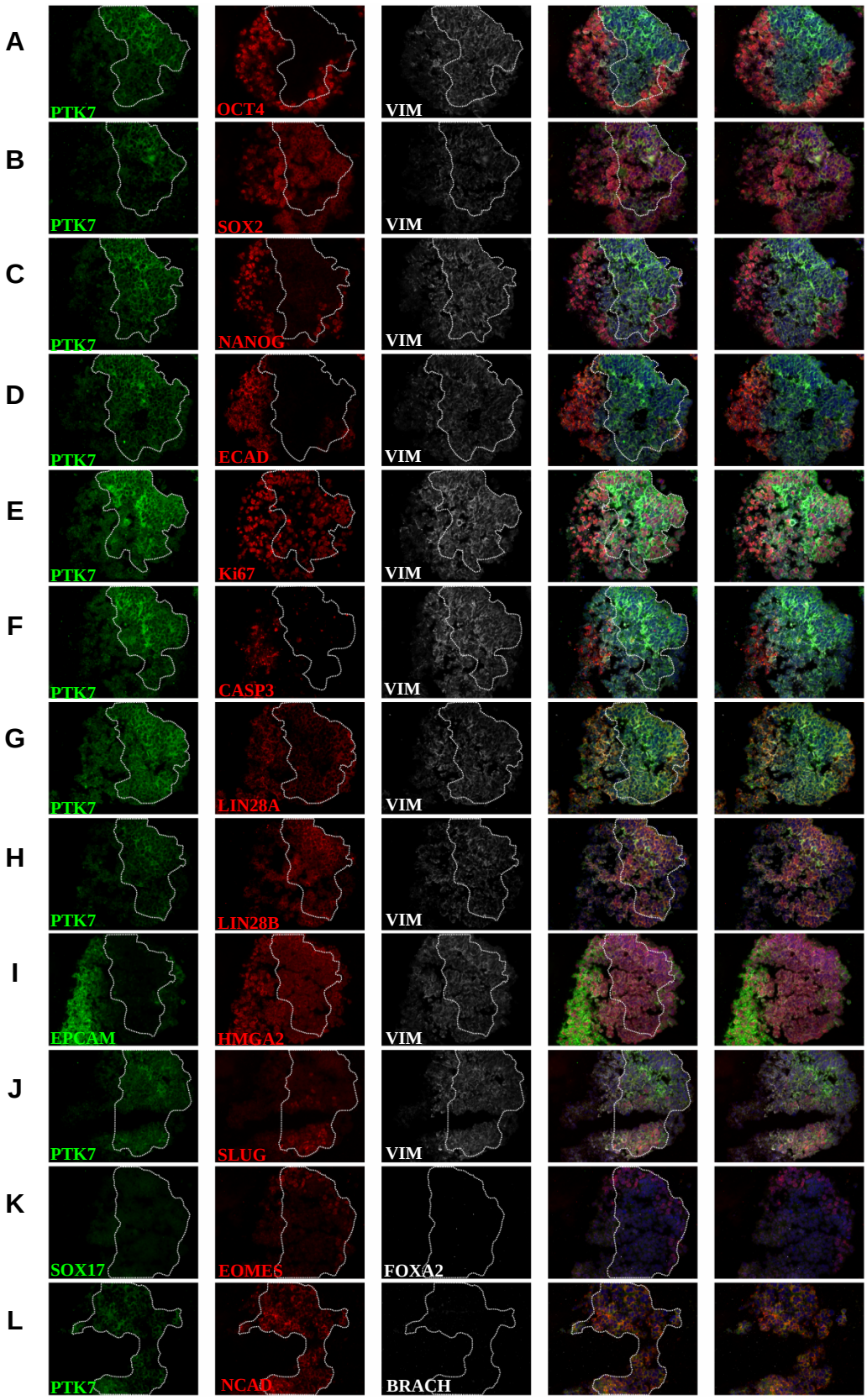

A
PTK7
OCT4
VIM
B
PTK7
SOX2
VIM
C
PTK7
NANOG
VIM
D
PTK7
ECAD
VIM
E
PTK7
Ki67
VIM
F
PTK7
CASP3
VIM
G
PTK7
LIN28A
VIM
H
PTK7
LIN28B
VIM
I
EPCAM
HMGA2
VIM
J
PTK7
SLUG
VIM
K
SOX17
EOMES
FOXA2
L
PTK7
NCAD
BRACH

**Figure S4-1. Pluripotency, lineage, and viability marker analysis of PTK7 population in hEBs.**

Immunofluorescence staining of XFiPSC2 hEB serial sections. White boundary indicates the PTK7 population. XFiPSC2 hEB were cultured for 24hrs before cryosectioned. Serial sections of XFiPSC2 24hr hEB are co-stained with PTK7, (green) epithelial markers (E-CADHERIN, red; EPCAM, green), mesenchymal markers (N-CADHERIN, red; VIMENTIN, white), pluripotency markers (OCT4, SOX2, NANOG, red), developmental genes (LIN28A, LIN28B, HMGA2, red), primitive streak markers (SLUG, red; Brachyury, white) and endodermal markers (SOX17, green; EOMES, red; FOXA2, white). The 4$^{th}$ and 5$^{th}$ columns show the merged images from three fluorescent channels and DAPI.

**Figure S4-2. Signaling pathway analysis of PTK7 population in hEBs.**

Immunofluorescence staining of XFiPSC2 hEB serial sections. XFiPSC2 hEB were cultured for 24hrs before cryosectioned. Serial sections of XFiPSC2 24hr hEB are co-stained with PTK7, (green) epithelial markers (E-CADHERIN, red; EPCAM, green), pluripotency markers (NANOG, red), TGFβ signaling output (pSmad2, red) and BMP signaling output (pSmad1,5,8, red). The rightmost column shows the merged images from three fluorescent channels and DAPI.

# CHAPTER 5

## Conclusions and Perspectives

***Transcriptional differences between PSC-derivatives and natural counterparts***

While embryonic and induced pluripotent stem cells display somewhat distinct transcriptional profiles [1-3], we have shown by driving them down any of the three germ layer lineages that these differences dissipate upon differentiation [4]. However, pluripotent stem cell derivatives are transcriptionally discernible from somatic counterparts despite displaying similar functional capabilities. These differences also coincide with the fact that in vitro derivatives from pluripotent stem cells display markedly fetal characteristics and markers. It is our hypothesis that some of these differentially expressed genes may hold the key to the regulation of developmental maturity [4].

***Regulation of LIN28/let-7 pathway***

Within the differentially expressed genes between PSC-derivatives and somatic counterparts, let-7 targets including *LIN28A* and *Lin28B* are over-represented, suggesting their possible roles in determining developmental maturity in these derivatives [4]. Mammalian development corresponds with decreasing Lin28 expression and increasing let7 activity [5-7]. In addition, many aggressive cancers have elevated Lin28 levels and low let-7 activity, which presumably contributed to an undifferentiated tumor with poor prognosis [6, 8, 9]. Our current understanding of the regulation of Lin28/let-7 pathway is still not complete. While it has been elucidated how LIN28 inhibit the maturation of let-7 and how let-7 targets *Lin28* mRNA [10-15], there are still areas poorly studied. Some of these questions include how the transcription of *LIN28* and let7 pri-miRNA is regulated. The exploration of these mechanisms would allow us to better understand how the Lin28/let-7 pathway is initiated and maintained in different contexts, and would possibly aid in the efforts of manipulating this pathway.

***Efforts in manipulating LIN28/let-7 pathway***

We have focused our efforts in on manipulating the LIN28/let-7 pathway to demonstrate whether this would affect developmental maturity in hPSC derivatives because of its known role in regulating developmental timing [6, 7, 16, 17]. Our transcriptome analysis also shows a disproportionate misexpression of LIN28 and let-7 targets in our PSC-derivatives [4]. Indeed it has been demonstrated that by overexpression Lin28, adult HSCs could be reprogrammed to exhibit fetal-like lymphopoiesis [18].

While it is possible to transiently modulate the amounts of LIN28 or let-7 by transfection in some of the differentiating cultures (e.g. PSC-NPCs), it is not conclusive whether it is possible to fixate a determined amount of LIN28/let-7 activity permanently during differentiation. In some three-dimensional differentiation cultures where transfection is difficult (e.g. PSC-hepatocytes), we have been unable to alter the LIN28/let-7 activity. All these call for a more consistent and penetrant way to modulate the LIN28/let-7 pathway. One possible method would be to stably introduce in undifferentiated hPSC drug inducible overexpression of, or shRNA against *LIN28* and let-7 respectively. Upon differentiation, the LIN28/let-7 activity can then be titrated by using various dosages of drugs. Such an approach could, however, be hampered by leakiness of the stably introduced construct, silencing of the transgene, and off-target effects, among other concerns.

We have proposed to screen for small molecules that modulate the LIN28/let-7 pathway. The discovery of such small molecules can lead to fast and effective way of controlling LIN28/let-7 pathway not only in *in vitro* differentiation, but probably also in a clinical setting to reduce the LIN28 load in tumors. We have built a stable let-7 luciferase reporter cell line for high throughput small molecule screening. Other than small molecule screening, our human cell line model is also suitable for siRNA and cDNA library screening to uncover novel regulators of the LIN28/let-7 pathway. Nevertheless, we have not yet found a small molecule that can significantly alter let-7 activity. It is possible that only a specific class of chemical that can interact with the LIN28 protein, pre-miRNA, and/or other let-7 processing machinery, will have

an effect on overall let-7 activity. Further efforts should be focused on either rapid screening of more libraries of small compounds, or the use of computer models to predict classes of chemicals likely to bind to players in the LIN28/let-7 pathway. The discovery of classes of small molecules that can affect LIN28/let-7 activity will have revolutionary impact on understanding the LIN28/let-7 pathway, on developmental maturity of hPSC-derivatives and on the importance of this pathway in cancer therapeutics.

***Using in vivo models to study the maturation of hPSC-derivatives***

There have been examples where fetal cells can be induced to mature by injecting them into an adult organism. For example, human fetal hepatocytes and reprogrammed hepatocytes can be matured by injecting into the livers of fumarylacetoacetate hydrolase (*Fah*)-deficient ($Fah^{-/-}$) recipient mice [19, 20]. While it is improbable to study all the maturity-inducing signals from the adult microenvironment, one can study the progress of maturation by isolating and profiling these cells. Such studies may uncover important regulators for developmental maturity, and in our case, reveal the role of Lin28/let-7 pathway and downstream targets on developmental maturity.

***Uncovering the early determinants in spontaneous EMT from hPSCs***

Using embryoid body differentiation, we intended to probe for an early EMT event from a pluripotent state. We focused on PTK7, a planar polarity gene known to be essential in development events such as gastrulation and neural tube folding [21, 22]. In vertebrates, the first post-implantation developmental event occurs during gastrulation, when part of the pluripotent embryonic ectoderm undergoes an EMT event to form the mesendoderm, which internalizes, converges to the midline, and extends along the AP axis [23, 24]. Since hPSCs resemble the pluripotent embryonic ectoderm [25], we are interested in modeling this early EMT event in human development with hPSCs. Using cryosections of embryoid bodies, we have

demonstrated that PTK7, a transmembrane receptor tyrosine kinase with null kinase activity, is expressed in cells having undergone EMT. We used PTK7 as a marker to FACS-purify this population for transcriptional profiling. The analysis revealed the upregulation of most of the known EMT markers in PTK7 positive population, as predicted. Nevertheless, the PTK7 positive population did not display significant inductions of ectodermal, endodermal, mesodermal, or extraembryonic markers yet still expressed a high level of pluripotency genes, indicating an early differentiation event. We propose hPSCs and embryoid bodies as feasible and ethically acceptable models to study early human development.

***Challenges in human embryology—understanding early cell fate determination***

While much research has been focused on differentiating a specific cell type from hPSCs, less emphasis has been put on studying early cell fate decision. Our study of PTK7 in hPSC differentiation has uncovered several interesting trends while raising more questions at the same time.

In the PTK7 positive population, the upregulation of markers like Twist, Snail2, and Vimentin reconfirmed the identity of a group of cells undergoing EMT. However, the lack of transcriptional induction in *SNAIL1* (and subsequently downregulation of *E-cadherin* mRNA) showed that this EMT has not been driven to completion. It is unknown whether we are capturing a group of cells early in EMT that display a basket of genes needed for EMT initiation; or whether our spontaneous differentiation conditions prohibit this EMT from being driven to completion.

Another interesting observation in the PTK7 population is the mutually exclusive expression of SNAIL2 and EOMES. This indicated that this PTK7 population likely represents a mesendoderm progenitor, with the ability to differentiate into different subpopulations towards the in mesodermal or endodermal lineages. PTK7-inducible lineage tracing may be a good way to track the cell fate decision, and to even track if PTK7 positive cells can shuffle back to a

pluripotent state through an MET event. Perhaps by exploring the protein expression pattern corresponding to the upregulated genes in the PTK7 positive population, we could further elucidate the decision to commit for these two germ layers, and find more specific markers/regulators for early germ layer commitment.

In the PTK7 negative population, the higher upregulation of *PAX6* and *SIX3*, two genes known to play a role in mammalian anterior neuroectoderm determination [26-35], suggests a potential ectodermal fate of the PTK7 negative population. In addition, most of the other ectodermal markers we probed for remained unchanged, indicating that either (1) *PAX6* and *SIX3* are some of the earliest determinants for ectoderm specification, or (2) under our spontaneous differentiation conditions, we have enriched for this *PAX6+/SIX3+* population of ectodermal differentiation. Again, PTK7-inducible lineage tracing may answer if only the PTK7 negative cells are fated to become ectoderm. Alternatively, a PTK7-inducible cell ablation (e.g. with diphtheria toxin) can be employed to observe fate determination.

### *Conclusions*

Our work in using human pluripotent stem cells to model development has yielded interesting discoveries as well as opening the door to more pressing questions on human embryonic development. We have shown that in addition to therapeutic promise, perhaps the power of hPSCs to model human development is more impressive and merits significant attention and effort.

### *Bibliography*

1. Chin, M.H., et al., *Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures.* Cell Stem Cell, 2009. **5**(1): p. 111-23.

2. Ghosh, Z., et al., *Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells.* PLoS One. **5**(2): p. e8975.

3. Marchetto, M.C., et al., *Transcriptional signature and memory retention of human-induced pluripotent stem cells.* PLoS One, 2009. **4**(9): p. e7076.

4. Patterson, M., et al., *Defining the nature of human pluripotent stem cell progeny.* Cell Res. **22**(1): p. 178-93.

5. Bussing, I., F.J. Slack, and H. Grosshans, *let-7 microRNAs in development, stem cells and cancer.* Trends Mol Med, 2008. **14**(9): p. 400-9.

6. Viswanathan, S.R. and G.Q. Daley, *Lin28: A microRNA regulator with a macro role.* Cell. **140**(4): p. 445-9.

7. Zhu, H., et al., *Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies.* Nat Genet. **42**(7): p. 626-30.

8. Viswanathan, S.R., et al., *Lin28 promotes transformation and is associated with advanced human malignancies.* Nat Genet, 2009. **41**(7): p. 843-8.

9. Ben-Porath, I., et al., *An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors.* Nat Genet, 2008. **40**(5): p. 499-507.

10. Nam, Y., et al., *Molecular basis for interaction of let-7 microRNAs with Lin28.* Cell. **147**(5): p. 1080-91.

11. Heo, I., et al., *Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA.* Mol Cell, 2008. **32**(2): p. 276-84.

12. Rybak, A., et al., *A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment.* Nat Cell Biol, 2008. **10**(8): p. 987-93.

13. Newman, M.A., J.M. Thomson, and S.M. Hammond, *Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing.* RNA, 2008. **14**(8): p. 1539-49.

14. Hagan, J.P., E. Piskounova, and R.I. Gregory, *Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells.* Nat Struct Mol Biol, 2009. **16**(10): p. 1021-5.

15. Piskounova, E., et al., *Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms.* Cell. **147**(5): p. 1066-79.

16. Moss, E.G., R.C. Lee, and V. Ambros, *The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA.* Cell, 1997. **88**(5): p. 637-46.

17. Lettre, G., et al., *Identification of ten loci associated with height highlights new biological pathways in human growth.* Nat Genet, 2008. **40**(5): p. 584-91.

18. Yuan, J., et al., *Lin28b reprograms adult bone marrow hematopoietic progenitors to mediate fetal-like lymphopoiesis.* Science. **335**(6073): p. 1195-200.

19. He, Z., et al., *Liver xeno-repopulation with human hepatocytes in Fah-/-Rag2-/- mice after pharmacological immunosuppression.* Am J Pathol. **177**(3): p. 1311-9.

20. Azuma, H., et al., *Robust expansion of human hepatocytes in Fah-/-/Rag2-/-/Il2rg-/- mice.* Nat Biotechnol, 2007. **25**(8): p. 903-10.

21. Yen, W.W., et al., *PTK7 is essential for polarized cell motility and convergent extension during mouse gastrulation.* Development, 2009. **136**(12): p. 2039-48.

22. Lu, X., et al., *PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates.* Nature, 2004. **430**(6995): p. 93-8.

23. Thiery, J.P., et al., *Epithelial-mesenchymal transitions in development and disease.* Cell, 2009. **139**(5): p. 871-90.

24. Heisenberg, C.P. and L. Solnica-Krezel, *Back and forth between cell fate specification and movement during vertebrate gastrulation.* Curr Opin Genet Dev, 2008. **18**(4): p. 311-6.

25. Nichols, J. and A. Smith, *Naive and primed pluripotent states.* Cell Stem Cell, 2009. **4**(6): p. 487-92.

26. Seo, H.C., et al., *Expression of two zebrafish homologues of the murine Six3 gene demarcates the initial eye primordia.* Mech Dev, 1998. **73**(1): p. 45-57.

27. Kobayashi, M., et al., *Overexpression of the forebrain-specific homeobox gene six3 induces rostral forebrain enlargement in zebrafish.* Development, 1998. **125**(15): p. 2973-82.

28. Loosli, F., et al., *Six3, a medaka homologue of the Drosophila homeobox gene sine oculis is expressed in the anterior embryonic shield and the developing eye.* Mech Dev, 1998. **74**(1-2): p. 159-64.

29. Oliver, G., et al., *Six3, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development.* Development, 1995. **121**(12): p. 4045-55.

30. Jean, D., K. Ewan, and P. Gruss, *Molecular regulators involved in vertebrate eye development.* Mech Dev, 1998. **76**(1-2): p. 3-18.

31. Winkler, S., et al., *The conditional medaka mutation eyeless uncouples patterning and morphogenesis of the eye.* Development, 2000. **127**(9): p. 1911-9.

32. Bovolenta, P., et al., *Expression pattern of cSix3, a member of the Six/sine oculis family of transcription factors.* Mech Dev, 1998. **70**(1-2): p. 201-3.

33. Chow, R.L., et al., *Pax6 induces ectopic eyes in a vertebrate.* Development, 1999. **126**(19): p. 4213-22.

34. Hill, R.E., et al., *Mouse small eye results from mutations in a paired-like homeobox-containing gene.* Nature, 1991. **354**(6354): p. 522-5.

35. Hogan, B.L., et al., *Small eyes (Sey): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse.* J Embryol Exp Morphol, 1986. **97**: p. 95-110.