UNIVERSITY OF CALIFORNIA, SAN DIEGO

Functional analysis of Sall4 in modulating embryonic stem cell fate

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Molecular Pathology

by

Pei Jen A. Lee

Committee in charge:

Professor Steven Briggs, Chair
Professor Geoff Rosenfeld, Co-Chair
Professor Alexander Hoffmann
Professor Randall Johnson
Professor Mark Mercola

2009

The dissertation of Pei Jen A. Lee is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2009

Dedicated to

my parents, my brother ,and my husband

for their love and support

Table of Contents

Signature Page……iii

Dedication…..……..iv

Table of Contents…….v

List of Figures……...vi

List of Tables……..……...ix

Curriculum vitae……x

Acknowledgement…….……..……...xi

Abstract……..……....xiii

Chapter 1

Introduction ..……1

Chapter 2

Materials and Methods……..…12

Chapter 3

Proteomic profiling and dissecting the role of Sall4 in Embryonic Stem cell fate determination……..…42

Chapter 4
A novel function of Sall4 in modulation of microRNA processing……..…76

References……..91

List of Figures

Fig. 2-1 Human ES cell-differentiated XEN cells under phase contrast……..20

Fig. 2-2 Method for isolating XEN cells from hESC culture by Collagenase..21

Fig. 2-3 Morphology change of XEN showed when they have been cultured on different martrix……22

Fig. 2-4 iTRAQ labeling efficiency……..…25

Fig. 2-5 Experimental scheme for enrichment of Sall4 interacting proteins in mES cells……...37

Fig. 3-1 H9 hESCs grow on mouse feeder (MEFs)……49

Fig. 3-2 Mouse ESCs (mESCs) grow on feeder-free condition……..……....50

Fig. 3-3 Subcellular localization distribution of whole H9/Embroid body proteome. …….……....52

Fig. 3-4 Biological processes distribution of whole H9/Embryoid body proteome. ……....53

Fig. 3-5 Subcellular localization distribution of whole H9/XEN cell proteome. ….....54

Fig. 3-6 Biological process distribution of whole H9/XEN cell proteome…...55

Fig. 3-7 The *Sall4*-mutant ESCs that have been used in this study……..….57

Fig. 3-8 *Sall4*-KO ESCs show proliferation defect compared to Sall4-het ESCs by cell viability assay....…….....…58

Fig. 3-9 Overexpression of Sall4 can rescue the *Sall4*-KO ESCs to form compact colonies……..……..…….…….....59

Fig. 3-10 Knockdown of exon1 and 4 of Sall4 does not change the phenotype of Sall4-KO ESCs…….....59

Fig. 3-11 Flow cytometry revealed 4.63% of Sall4 het ESCs and 4.84% of Sall-KO ESCs are apoptotic cells……..60

Fig. 3-12 Biological process distribution of whole *Sall4* mutant ES cell proteome. …….62

Fig. 3-13 Selected signaling pathways represented from whole Sall4 mutant ES cell proteome……..……..63

Fig. 3-14 Sall4 co-IP with and without enrichment……..……....67

Fig. 3-15 Confirmed interactions of selected Sall4 interacting proteins……..68

Fig. 3-16 Interacting domain mapping…….....69

Fig. 3-17 Selected signaling pathways represented by Sall4-interacting proteins……..70

Fig. 3-18 Sall4-Interacting network generated by DAVID…….....71

Fig. 3-19 A new Sall4 interacting network.……..…...……....72

Fig. 4-1 Confirmation of the interaction between Sall4 and Lin28 *in vitro* and *in vivo*…….....……...80

Fig. 4-2 Sall4 directly interacts with Lin28 and binds to pri-let-7 microRNA...81

Fig. 4-3 pri-let-7 specifically interferes with the Sall4-Lin28 binding affinity...81

Fig. 4-4 Dynamics of endogenous pri-let miRNAs in *Sall4* mutant ES cells upon differentiation……...83

Fig. 4-5 Quantitative PCR showing changes in levels of endogenous pri-let-7 upon sh-Lin28 induction in mouse EC cell line (P19)…….84

Fig. 4-6 Quantitative PCR showing changes in levels of endogenous pri-let-7 and Lin28 upon sh-Sall4 induction in two lines of mouse ECs……....84

Fig. 4-7 Levels of the pluripotency markers Nanog and Oct-4 in P19 cells infected with either shSall4 or shLacZ. ...……..…85

Fig. 4-8 Quantitative PCR showing changes in levels of endogenous pri-let-7 and Lin28 upon sh-Sall4 induction in mouse ESCs.……..…….…….....85

Fig. 4-9 NB analysis of the ectopically expressed pri-let-7g, Lin28, and Sall4 in HEK293T cells. Lane 2 and 3 represent different clones of Lin28…….87

Fig. 4-10 Signaling NB analysis of the ectopically expressed pri-let-7g, Lin28, and Sall4 in HEK293T cells……...….88

Fig. 4-11 Co-immunoprecipitation showed that Sall4 is an integral component of the Drosha protein complex....…….....…88

List of Tables

Table 3-1 Total proteins and phosphoproteins that were detected by our proteome and phosphoproteome. ……..….51

Table 3-2 Phosphorylation sites of Sall4 that have been detected from our phosphoproteome from different cell line.…… ……...56

Table 3-3 Selected Proteins detected from selected Signaling Pathways….63

Table 3-4 Selected proteins that decreased in expression in *Sall4*-KO ESCs compared to *Sall4*-het ESCs……..……64

Table 3-5 Selected proteins that increased in expression in *Sall4*-KO ESCs compared to *Sall4*-het ESCs…….……65

Table 3-6 Proteins detected from selected signaling pathways……..…….70

Curriculum vitae

Education

2004-2009 Ph.D., Molecular Pathology, University of California, San Diego

1996-2000 B.S., Zoology, National Taiwan University

Publication

Lee JP, Jeyakumar M, Gonzalez R, Takahashi H, Lee PJ, Baek RC, Clark D, Rose H, Fu G, Clarke J, McKercher S, Snyder EY "Stem cells act through multiple mechanisms to benefit mice with neurodegenerative metabolic disease." ***Nature Medicine*** 13: 439-47, 2007

Eminy H. Y. Lee, W. L. Hsu, Y. L. Ma, P. J. Lee and C.C. Chao. "Enrichment enhances the expression of *SGK*, a glucocorticoid-induced gene, and facilitates spatial learning through glutamate AMPA receptor mediation." ***Eur J. of Neuroscience*** 18: 2842-52, 2003

Acknowledgement

I am most grateful to my advisor, Professor Steven P. Briggs, for his mentorship, support, and inspiration throughout the course of my research. He was not just a mentor to me scientifically, but also a good friend who always helped me through numerous personal hardships and cared about me. His wisdom and patience guided me not just to be a mature scientist, but also a leader of a group. I feel lucky to have been working with him for these past three years.

I would like to acknowledge the members of my committee who have been keeping me grounded and providing me critical suggestions. From them, I see the model of great scientists. I would also like to thank Professor Evan Snyder, my previous advisor, for the almost 3 years of training in his lab.

I really appreciate all the members of the Briggs lab for their help in various aspects of my research. In particular, I would like to thank Dr. Kiyoshi Tachikawa and Dr. Zhouxin Shen for their experimental assistance and helpful discussions. I would also like to thank Darwin Yee for his excellent technical assistance. My best friend, Tenai Eguen, provided invaluable writing, editorial guidance, encouragement, and backing. She made my research life with lots of fun.

A few of scientists deserve specific acknowledgement. Eric Yu played a central role in initiating some projects. He helped me with lots of experiments. Stephen Soonthornvacharin helped me with the proofreading of my dissertation. In the end of each chapter, I listed all of the people who are involved in the projects respectively.

I would also like to give special thanks to family, my parents, and brother for their love and support. I also want thank my husband Meng-Ping Ben Kao for his forbearance and support during my studies. Finally, I would also love to thank all my friends who has helped me throughout these years and other colleagues who are too numerous to name in full.

Chapter 2 and 3, in full, are being prepared for publication. I have obtained the permissions from co-authors, Steve Briggs, Zhouxin Shen, Kiyoshi Tachikawa, and Darwin Yee to reproduce material in this dissertation. The dissertation author will be the primary author of the papers.

Pei Jen Angel Lee
Nov 2009

ABSTRACT OF THE DISSERTATION

Functional analysis of Sall4 in modulating embryonic stem cell fate

by

Pei Jen A. Lee

Doctor of Philosophy in Molecular Pathology

University of California, San Diego, 2009

Professor Steven Briggs, Chair

Professor Geoff Rosenfeld, Co-chair

Cell fate decisions of embryonic stem (ES) cells are dictated by repression and activation of lineage-specific genes. How are those regulators important in activating a stem cell to begin differentiating into a specific cell type regulated by the cellular and molecular signals? Understanding how these lineages arise during development will illuminate efforts to understand the establishment and

maintenance of the stem cell state and the mechanisms that restrict stem cell potency. Sall4 has been reported to be part of the core transcriptional network in embryonic stem cells; however, how Sall4 regulates ES cell fate is still unknown. I utilized mass spectrometry to generate whole ES cell proteomes and phosphoproteomes of human and mouse ES cells. The whole cell proteomic and phosphoproteomic analysis provides us a large repertoire of protein lists to unmask the regulatory elements in ES cells. Moreover, I generated a novel list of Sall4-binding proteins by modifying a previously established enrichment protocol. Our findings demonstrate the effectiveness of using a synthetic crosslinker, DSP, to enrich for weak and transient Sall4 interacting protein complexes. Working with molecular weight cut-off columns, allows us to identify new Sall4 binding partners, leading to a better understanding of the regulatory machinery behind the self-renewal and pluripotency of ES cells. Pathway analysis indicates that Sall4-binding proteins are involved in Wnt signaling pathway, FGF signaling pathway, and p53 pathway. Our data indicates that the ES cell proliferation defect may be regulated by Integrin, TGF-beta, Cadherin, and PI3K signaling pathways. Furthermore, the interactome has led me to explore a possible novel function of Sall4 in microRNA biogenesis. By understanding how Sall4 regulates cell fate determination can substantially push ESC research forward.

Chapter 1

Introduction

Cell fate determination

The fate of a cell describes what it will become in the course of normal development. The determinant of different cell types (cell fates) involves progressive restrictions in their developmental potentials. Determination implies a stable change - the fate of determined cells does not change. Differentiation follows determination, as the cell elaborates a cell-specific developmental program. Differentiation results in the formation of cell types such as muscle cells, nerve cells, and skin cells. The building of an organism from a single cell to a multicellular structure of characteristic shape and size is the result of coordinated gene action that directs the developmental fate of individual cells. The acquisition of different cell fates orchestrates an intricate interplay of cell proliferation, quiescence, migration, growth, self-renewal and differentiation, and death, elaborating and bringing together cellular ensembles in a precise manner. How intrinsic and extrinsic factors are integrated in ontogeny to specify cell fates defines the central question of developmental biology (4).

Why stem cells?

Embryogenesis is a complex developmental process involving coordinated cues that direct cell behavior. It is the maturation of an inner cell mass into tissues and then organ systems that comprise a living organism. A stem cell is a special kind of cell that has a unique capacity to renew itself and to give rise to specialized cell types. Although most cells of the body, such as heart or skin cells, are committed to conduct a specific function, a stem cell is

uncommitted and remains uncommitted, until it receives a signal to develop into a specialized cell. ES cells are derived from the inner cell mass (ICM) of the mammalian blastocyst-stage embryos. Their importance to modern biology and medicine derives from two unique characteristics that distinguish them from all other organ-specific stem cells identified to date. First, they are able to undergo cell division under specific cell culture conditions for extended periods (20). Unlike transformed tumor cell lines, ES cells can retain normal karyotypes following extensive passaging in culture. Second, they are pluripotent; possessing the capacity to give rise to all of the various cell types that make up the body, except for extra-embryonic tissues such as the amnion, chorion, and other components of the placenta (33). Therefore, this class of ES cells holds the promise to repairing or replacing cells or tissues damaged or destroyed by diseases and disabilities. Three kinds of stem cells have been derived: embryonic stem (ES) cells, trophoblast stem (TS) cells, and extra-embryonic endoderm (XEN) cells. These stem cells appear to be derived from three distinct tissue lineages within the blastocyst: the epiblast (EPI), the trophectoderm (TE), and the primitive endoderm (PE), respectively. A new pluripotent cell line was "invented" recently. Several groups reported that somatic cells could be reprogrammed into induced pluripotent stem cells (iPSCs), which functionally and phenotypically resemble ESCs (52, 67-69, 88). Although iPSCs meet the defining criteria for pluripotent stem cells, it is not known if iPSCs and embryonic stem cells differ in clinically significant ways.

Development of the blastocyst lineages

Following fertilization of the oocyte, there are three rounds of cell division leading to formation of the morula. During the morula stage, cells compact against each other by the formation of intercellular junctions. Two consecutive differentiative steps lead to blastocyst formation: TE and PE formation. The first recognized differentiation event in mouse development occurs at the late morula stage when the outer layer of the compacted ball of cells adopts an epithelial structure. This structure, the nascent TE, seals off the inside of the ball, and pump-like proteins in the TE facilitate the formation and expansion of an inner cavity or blastocoel. The embryo is now called a blastocyst and contains two cell types: TE and ICM. The TE will go on to produce trophoblast, the precursor to the non-maternal components of the placenta, while the inner cell mass will undergo a second differentiative step (85-86).

The second differentiative step occurs when the ICM differentiates into EPI and primitive endoderm. The PE forms as an epithelium covering the ICM and facing the blastocoel. The EPI, sometimes referred to as PE, develops between PE and TE and will give rise to the embryo itself. By the late blastocyst stage, three cell types, the EPI, TE, and PE, are committed to their different later lineages and are no longer totipotent. EPI gives rise to the entire fetus as well as extra-embryonic mesoderm. TE gives rise to all the trophoblast cell types that make up the majority of the fetal part of the placenta, and the PE forms the XEN layers of the visceral and parietal yolk sacs. The TE and PE extra-embryonic

lineages are required to support the growth of the mammalian fetus in the uterine environment and are sources of signals to the EPI to initiate axial patterning (3, 6, 58).

Factors to control EPI/ PE cell fate determination

For over 20 years, scientists have attempted to understand the intrinsic controls that keep stem cells from differentiation. A unique network of transcription factors and signaling molecules has shown to be essential for maintaining ES cell properties. Several key regulators identified are both essential for formation of the ICM during mouse preimplantation development and self-renewal of pluripotent ES cells (5, 18, 47). These core factors contribute hallmark characteristics of ES cells by: (1) activation of target genes that encode pluripotency and self-renewal mechanisms, and (2) repression of signaling pathways that promote differentiation (51).

One of the hallmarks of an undifferentiated pluripotent cell is the expression of the *Pou5f1* gene, which encodes the transcription factor Oct4. During mouse preimplantation development, zygotic *Pou5f1* expression is activated at the four-cell stage and is later restricted to the pluripotent cells of the ICM and EPI (50). During PE differentiation, Oct4 expression is downregulated in the presumptive PE, as is another gene called *Nanog*. *Nanog* encodes a transcription factor that, like Oct4, is expressed in the ICM/ EPI lineage(18, 47). Several observations suggest that Nanog acts as a selector of EPI fate over PE

fate. In the absence of a functional *Nanog*, embryos form TE and PE, but do not form EPI. Secondly, loss of *Nanog* from ES cells causes them to assume XEN character (47). Although Oct4 and Nanog are co-expressed, they appear to participate in parallel pathways. As loss of *Nanog* does not affect *Oct4* expression, nor does loss of *Oct4* affect *Nanog* expression in the blastocyst(17). Nanog has therefore been proposed to act as a selector gene during the EPI /PE lineage decision.

Genes required within the PE lineage have also been identified.

Gata6 encodes a transcription factor expressed in the PE at the blastocyst stage and is also expressed in XEN cells (42). Recent genetic evidence has implicated *Nanog* and *Gata* family transcriptional factors in specifying epiblast versus PE fate. Nanog could substitute for LIF to promote self-renewal of ES cells (18). Disruption of this gene leads to failure to maintain pluripotency in ES cells *in vitro* and peri-implantation lethality with failure of EPI formation *in vivo*. In a complementary manner, overexpression of *Gata6* is sufficient to transform ES cells to PE lineages and ES cells lacking *Gata6* cannot form a functional visceral endoderm layer in *in vitro* embryoid body cultures (26).

Molecular mechanism of EPI/ PE formation and segregation

The EPI- and PE-specific transcription factors, Nanog and Gata6, were expressed in the ICM in a random mutually exclusive “salt and pepper” pattern. Lineage tracing showed predominant lineage restriction of single ICM cells at

E3.5. Chazaud et al. proposed a model in which the ICM develops as a mosaic of EPI and PE progenitors at E3.5, dependent on Grb2-Ras-MAP kinase signaling, followed by later segregation of progenitors into appropriate cell layers (19). *Gata6* forced expression in ES cells causes differentiation into XEN, suggesting Gata6 functions early in the hierarchy regulating endoderm development. The mechanism that normally blocks these transcription factors in ES cells is unknown, but may involve Nanog and the LIF/Stat3 signaling pathway rather than Oct3/4 (17). However, *Gata6* mutants do not exhibit XEN defects until several days after blastocyst formation (26, 40) and may indicate other factors initially specify PE in the blastocyst. Although no known mutation in individual transcription factors leads to complete absence of PE formation, there is evidence that receptor tyrosine kinase signaling is required for PE development. *Grb2* encodes an adaptor protein for receptor tyrosine kinases such as the FGF receptor. In embryos lacking *Grb2* where no PE forms, *Gata6* expression was lost and all ICM cells were Nanog positive. The identity of receptor tyrosine kinase(s) involved has not been determined (19). It is interesting to observe that culturing ES cells *in vitro* seems to have similar cell segregation.

How is Sall4 important for cell fate decision?

The *Sall4* gene encodes a zinc finger protein and contains highly conserved C2H2 zinc finger domains and a C2HC motif at the N terminal. It belongs to *SALL* gene family and is human homolog of Drosophila *spalt* (*spl*) (2,

37, 41). Mutations in human *SALL4* result in an autosomal-dominant syndrome called Okihiro syndrome, characterized by limb deformity, eye movement deficits, and less commonly, anorectal, ear, heart, and kidney anomalies(2, 35) also known as Duane-radial ray syndrome, DRRS, Duane. A syndrome of Duane's anomaly associated with cervical spine and radial ray abnormalities and deafness (9-10, 36).

Sall4 is essential for embryonic development during peri-implantation period. *Sall4*-null mice did not survive beyond embryonic day (E) 6.5. Independent of TE, *Sall4* is essential for ICM outgrowth. No extra-embryonic endodermal (XEN) cell lines could be established from *Sall4*-KO blastocysts suggesting Sall4 is cell-autonomously required for the development of PE from the ICM (23). Sall4 is also involved in forelimb and ear development in mice (39). *SALL4* is also an oncogene and constitutively expresses in human acute leukemia (AML) (44). Sall4 also plays an important role in limb and organ development by partnering with Tbx5 to form a transcription factor complex (38). Disruption of *Sall4 and* Tbx5 causes limb mutant phenotype suggesting binding to each other and acting synergistically to regulate *FGF10* gene expression. Tbx5 appears to activate Sall4 gene expression in these developing tissues.

Zhang *et al.* showed by overexpression and knockdown of *Sall4* by RNA interference (RNAi), Sall4 regulates transcription of *Pou5f1* by binding to the highly conserved regulatory region of the distal enhancer and activates *Pou5f1*

expression *in vivo* and *in vitro.* They suggest a critical role of Sall4 by maintenance of ES pluripotency through modulating Oct-4 expression (89) and binding to Nanog, a key regulator of ES cell self-renewal. Sakaki-Yumoto et al. successfully isolated *Sall4*-null ES cells from *Sall4*-Knockout blastocysts and showed no decreased *Pou5f1* or Oct-4 expression. This result argues against the conclusion from Zhang *et al*. It is possible that there was off-target effects by RNAi knockdown of *Sall4 in vitro*. *Sall4*-null ES cells showed decreased proliferation compared to wild-type ES cell, suggesting Sall4 is essential for ES cell proliferation, but not pluripotency. Interestingly, *Nanog* was slightly decreased in Sall4-deficient cells (60). *Nanog* heterozygous cells show no reduction in proliferation, and *Nanog*-deficient cells differentiate into PE (47). Nanog and Sall4 co-occupied *Nanog* enhancer region in living ES cells. Sall4 is an important component of the transcription regulatory networks in ES cells by cooperating with Nanog (83). It raises the question whether Sall4 can regulate the PEI/PE fate determination by interfering with *Nanog* transcriptional regulation.

The biggest concern is whether these cells are representative of primitive endoderm of the blastocyst. Mouse blastocyst-derived XEN cell lines have been shown to represent extra-embryonic endoderm providing a new cell culture model of early mammalian lineage (42). It is important to identify markers on both mouse blastocyst- and ES cell-derived XEN cell lines. I will attempt to profile whole cell proteome of human and mouse ES cell- derived XEN cells. By comparison of published gene expression analysis of mouse blastocyst-derived

XEN cell lines, unique proteins expressed in XEN cell will be identified for future studies.

Perhaps the confounding factor to this field is due to variability in culture conditions between different laboratories. The signaling complexity reflects self renewal in culture; a true signaling pathway balancing act. There are several ways of balancing the self-renewal equation, but ultimately culture conditions must suppress differentiation and promote proliferation. In hESCs, by using defined media with purified recombinant growth factors may circumvent the problems. XEN cells may provide unique in vitro tools to study the interactions between these lineages during development. Furthermore, these studies would provide a foundation for efforts to guide differentiation of ES cells along selected developmental pathways for potential therapeutic use (56).

Significance

Regenerative medicine attempts to manipulate the developmental process within an engineered scaffold, but with varying degrees of success (57). *In vitro* stem cells provides a tool to mimic the *in vivo* model; however, one main problem that may limit medical applications is that ESCs have been observed to display high genomic instability and non-predictable differentiation after long-term growth (63-64). Although several genes essential for maintaining pluripotency have been identified in ESCs, little is known about why some cells are pluripotent, others are restricted in developmental potential, and how

undifferentiated ESCs commit to undergo which cell fates. The goal of my dissertation is to attempt to answer the following open questions: What are the intrinsic factors that direct ESCs along a particular differentiation pathway? How are those regulators regulated by the cellular and molecular signals that are important in activating a stem cell to begin differentiating into a specific cell type? Understanding how these lineages arise during development will illuminate efforts to understand the establishment and maintenance of the stem cell state and the mechanisms that restrict stem cell potency. By taking advantage of spontaneous differentiating ES cell culture *in vitro*, I expect to elucidate the functions of *Sall4* in ES cell fate determination. This *in vitro* study may not only provide a foundation for efforts to guide differentiation of ES cells along selected developmental pathways but the enormous potential implications for medical research and therapeutic uses. Through proteomic approaches to different experimental systems will help to address many of the interesting biological questions that I have raised here.

http://stemcells.nih.gov/info/scireport/execSum.asp

Chapter 2

Materials and Methods

1. Materials

Sall4 mutant ES cell lines

From Dr. Ryuichi Nishinakamura group

Drosha, DGCR8 expression vectors

From Nary Kim group

Human ESCs (H9)

Purchased from WiCell

Mouse ESCs (129/svev) and Mouse Embryonic Fibroblasts (CF-1)

Purchased from Millipore corp.

Mouse Embryonal Carcinomas (P19 and F9)

Purchased from ATCC

Lentiviral vectors

From Biosettia

Buffer A

20mM HEPES 200mM NaCl, pH 7.5

Buffer D

200mM KCl, 0.2mM EDTA, 20mM Tris-HCl, pH 8.0

HBS washing Buffer

10mM HEPES, 150mM NaCl, pH 7.4

6x loading dye for EMSA

0.1% bromophenol blue, 0.1% xylene cyanol, and 50% glycerol

2X HBS for Calcium Phosphate Tranfection (500 mL):

8.0 g NaCl.

0.37 g KCl

106.5 mg Na2HPO4 (anhydrous, 201.1 mg if 7X H2O)

1.0 g dextrose

5.0 g Hepes

Add 450 ml ddH2O, adjust pH to 7.05 with NaOH and bring final volume to 500 ml. Sterile filter through 0.45 M filter, discarding first few mls through and store in cold room.

Lysis buffer (50mL)

0.1% SDS, 1% Triton X-100, 1% Decycholate, 5mM EDTA, 1x PBS, 1 tablet of complete (EDTA free) protease inhibitor (Roche)

ECL Reagents

Stocks:

250mM Luminol in DMSO (keep at -20˚C)

For 10ml 250mM luminol: 0.44g luminol-free acid (Sigma A8511) in 10ml DMSO

90mM p-Coumaric acid in DMSO (keep at -20˚C)

For 10ml 90mM p-Coumaric acid: 0.147 g p-Coumaric acid (Sigma C9008) in 10ml DMSO

1M Tris-HCl, pH 8.5

30% Hydrogen Peroxide

To make working solution: mix 20mL Tris with 180mL ddH_2O. Add 0.44ml of the coumaric acid and 1ml of luminol. This can be kept for months at 4˚C. This, plus the stocks should be kept in brown bottles.

To do ECL: Mix 10ml working solution with 3µl H_2O_2. Incubate blot for 1 minute.

2. Methods

2.1 Cell Culture

Protocols for mouse, human ESCs, and MEFs were modified and available from the National Stem Cell Bank website. All cells indicated in this study were cultured and incubated in 100% humidity, 37°C incubator supplied with 5% CO_2.

Every two months, mycoplasma test was performed by PCR to ensure the quality and condition of the cells.

Splitting

Except for human ESCs, I used the same protocol for splitting the HEK293T, MEFs, mESCs, and mECs. When the cells reached 90% confluency, they were at the point to be split and passaged. Cells were first washed with 1xPBS twice and then incubated with 0.05% Trypsin at room temperature for one to two minutes. 3 volumes of basal media (described at 2.1.1) were then added to neutralize the protease. The cells were dissociated by pipetting up and down several times. The cells were then centrifuged at 1000g for 1 minute and the supernatant was removed. New media was then added and mixed with the cells to ensure homogeneity. Cells were then plated out at the desired ratio depending on the purpose (1:5 to 1:10). Media was changed every other day.

2.1.1 HEK293T cells

HEK293T cells were culture in Basal media that contained DMEM, 10% FBS, 100U of Penicillin, 100ug of Streptomycin, and 1x non-essential amino acids. If cells reached over confluency (> 100%), the plates were discarded and a new vial of cells was thawed. Media were changed every other day.

Calcium Phosphate Transfection

One day prior to transfection, the cells were plated such that they are exponentially growing by the day of transfection (i.e. 50-60% confluent at the time of transfection). Immediately prior to transfection, medium was changed with 8ml of fresh media for a 10-cm dish. DNA was added (usually 10ug of DNA was added to ddH20 to reach a volume of 438µl in total, then 62µl of 2M CaCl2 was added). Then 500µl 2X HBS pH 7.05 was added dropwise while vortexing, or bubbling. This solution was added directly to the cells, dropwise through the medium, scattering the drops uniformly and slowly over the surface of the plate. The cells were returned to the incubator for 6 hours and medium was changed. Transient assays for gene expression in transfected cells are performed 48-72hrs post transfection.

2.1.2 Mouse Embryonic Fibroblasts

Mouse Embryonic Fibroblasts (MEFs) served as feeder cells that supported the growth of undifferentiated mouse or human embryonic stem cells. MEFs were freshly isolated from mouse embryos and could be used immediately. We purchased CF-1 MEFs at passage 3 from ATCC and then expanded them for future experiments. The culture conditions were identical to HEK293T cells as described in 2.2.1. Before use, MEF cells were treated by mitomycin-C to stop their proliferation. MEFs were seeded on 10-cm dishes to 95% confluency. Mitomycin C was applied to the culture media to a final concentration of 10ug/mL at 37°C for at least 3hrs. To ensure sufficient removal of Mitomycin C, MEFs

were trypsinized and washed at least three times with 1x PBS before reseeding on gelatin-coated cell culture dishes at a density of 50,000cells/cm^2.

2.1.3 Mouse Embryonal Carcinomas

In this study, I cultured two different cell lines of mouse embryonal carcinomas, P19 and F9. The culture conditions were identical to HEK293T cells as described in 2.2.1.

2.1.4 Mouse Embryonic Stem Cells

Mouse embryonic stem cells (ESCs) were propagated in an undifferentiated state on gelatinized flasks. Culture media consisted of Knockout-DMEM (Gibco) supplemented with 15% ES-screened FBS (Hyclone), 1x non-essential amino acids, 100U of Penicillin, 100ug of Streptomycin, and 0.1mM beta-mercaptoethanol. Leukemia inhibitory factor (LIF) was added to culture media at a final concentration of 0.2% during undifferentiated expansion. The LIF was harvested from the conditioned media of CHO cell-secreting human LIF. Cell culture media was changed daily, and the cells were passaged every 2days (approximately 90% confluency).

2.1.5 Human Embryonic Stem Cells

The initial hES cell line H9 was maintained on feeders in human ESC medium, which contained 80% KNOCKOUT−Dulbecco's modified Eagle's medium (KO-DMEM), 20% KNOCKOUT serum replacement, 1mM L-glutamine,

0.1mM beta-mercaptoethanol, 1% nonessential amino acids, 100U/mL penicillin/ streptomycin, and 8ng/ml human basic fibroblast growth factor (bFGF). For feeder-free culture, cells were passaged until 90% confluency was reached by incubation in 4mg/mL collagenase IV, dissociated, and then seeded onto Matrigel-coated plates in Condition Media (CM) prepared from MEFs as follows.

Condition Media from MEFs

MEFs were seeded at ~50,000 cells/cm^2 in MEF medium. The medium was exchanged with hESC medium every day. CM was collected daily and supplemented with an additional 4ng/ml of bFGF before feeding the hES cells. MEF cells were again fed with human ESC medium daily and used for 12 days for CM collection (84). CM could be used after filteration with 0.22μm filter unit.

Splitting Human ESCs

To passage human ESCs, Collagenase type I (Millipore) was used at a 4mg/mL working solution. To prepare the working solution, 400mg of collagenase type I powder was dissolved in 100mL Knockout-DMEM (Gibco) and filter-sterilized with a 0.22μm filter unit; the solution was then aliquoted and stored at -20°C until needed. Cells were overlaid with 0.1mL/cm^2 of the collagenase type I working solution and placed at 37°C in the incubator. After 10 minutes, the cultures were checked for signs of curling up at the edges of colonies. If not, cells were then put back at 37 °C for another 10 minutes before proceeding.

2.1.6 Derive extra-embryonic endoderm (XEN) cells from human ES cells *in vitro*

In order to avoid the contamination of mouse feeders, the human ESCs (H9) were maintained in medium conditioned by mitotically inactivated mouse embryonic fibroblast supplemented with 8 ng/ml basic Fibroblast Growth Factor on matrigel-coated plates (Fig.2-1). Media was changed every other day and continually cultured until the ES cells reached to about 90% confluency. Cells were washed with 1xPBS, and then incubated with 1mg/mL Collagenase IV at room temperature for 10 minutes. Based on the differences in adherence to the plate, the differentiated PE-like cells detached first. The loosen cells could be washed away by 1xPBS (Fig.2-2). Supernatant containing the cells was centrifuged at 1000*g* for 5 minutes and the cell pellet was re-suspended in 10% FBS media and transferred to fresh gelatin or matrigel-coated wells. The cells were always passed before they reached 80% confluency (Fig.2-3). We termed this primitive endoderm like cells as ES cell-derived XEN cells.

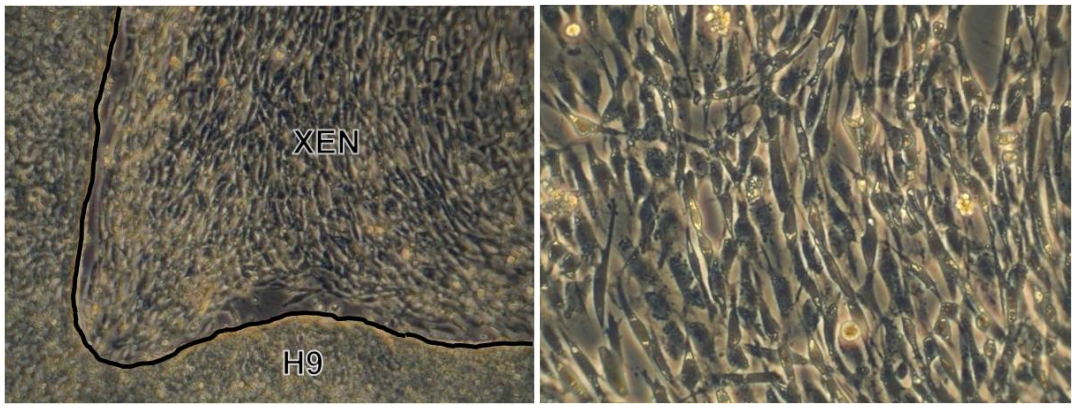


Fig. 2-1 Human ES cell-differentiated XEN cells under phase contrast. The higher magnification shows the fibroblast-like morphology of XEN cells.

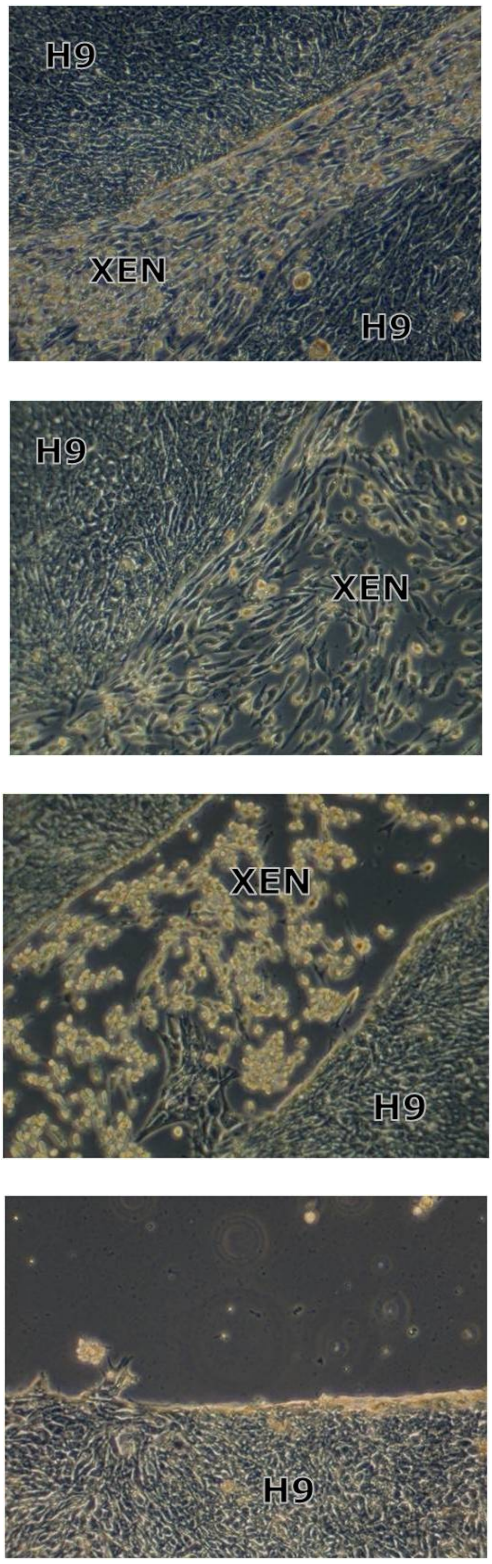


Fig. 2-2 Method for isolating XEN cells from hESC culture by Collagenase.

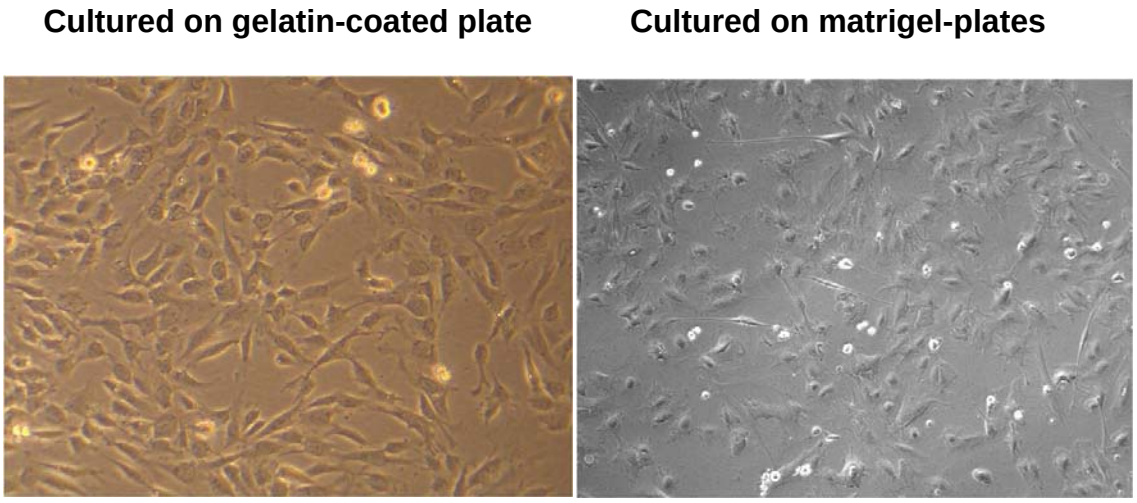


Fig. 2-3 Morphology change of XEN showed when they have been cultured on different martrix.

2.1.7 Cross-Linking

Synthetic cross linker DSP (Pierce # 22586) was non-sulfonated; it was first dissolved in an organic solvent (such as DMSO) and added to the aqueous reaction mixture. Before addition of the crosslinker solution, ESCs were washed by 1x PBS twice and changed with new media in order to remove dead cells. The samples were incubated on a shaker at gentle speed at room temperature for 30mins. The final concentration of the cross linkers was 2mM. Stop solution (1M Tris, pH 7.5) was added to a final concentration of 20mM and incubated at room temperature for 15mins, The samples were then washed twice with HBS buffer (10mM HEPES, 150mM NaCl, pH 7.4) for co-IP.

2.1.8 Cellular fractionation

Separation of Embryonic Stem Cell nucleus and cytoplasm was performed with NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific Cat#78833) according to the manufacturer's instruction. Appropriate volume of protease inhibitor was added (Roche cOmplet, EDTA-free Protease Inhibitor Cocktail Tablets Cat # 1187358001). After centrifuging, the cell pellets were freshly harvested and separation was performed immediately at 4^{o}C. The nuclear and cytoplasm fractions were directly used in co-immunoprecipitation or Western blotting analyses.

2.2 Proteomic Analysis

2.2.1 Protein extraction and digestion

50uL cell pellets were lysed in 100uL lysis buffer (2% (w/v) RapiGest (Waters, 186002122), 1mM EDTA, and 50mM Hepes buffer (pH 7.2)). Cysteines were reduced and alkylated using 1mM Tris(2-carboxyethyl) phosphine (TCEP, Fisher, AC36383)) at 95°C for 5 minutes followed by 2.5mM iodoacetamide (Fisher, AC12227) at 37°C in the dark for 15 minutes. Protein concentrations were measured using Bradford assay (Pierce). Proteins were digested with trypsin (Roche, 03 708 969 001) at the enzyme-to-substrate ratio (w:w) = 1:50 overnight.

Our standard cell lysis buffer contains 2% RapiGest (Waters) in 10mM HEPES buffer. Benzonase is added to degrade DNA and RNA to obtain a clear

solution. The extracted proteins are treated with 0.5M TCEP (Tris(2-carboxyethyl)phosphine) to a final concentration of 2mM TCEP at 37^{o}C for 30 minutes to reduce all of the disulfide bonds. Then 0.5M iodoacetamide (IAA) is added to a final concentration of 5mM IAA and the sample is incubated in the dark at 37^{o}C for 30 minutes to alkylate all of the sulfhydryl groups. We then dilute 5-fold and measure pH (should be ~8) and protein concentration (should be 1-2.5 mg/ml) with the Bradford assay. The proteins are digested by addition of trypsin (1:50) and shaken at 37^{o}C overnight. The completion of digestion is validated by silver staining. The digested peptides are acidified by adding Trifluoroacetic acid (TFA) to a final concentration of 0.5% (v/v, pH=1.5) to break down RapiGest. Samples are incubated at 4^{o}C overnight and then centrifuged at 16,100g for 15 minutes. Supernatant is taken for iTRAQ labeling.

2.2.2 iTRAQ labeling

For iTRAQ (Applied Biosystems, Foster City, CA) derivatization, an aliquot of each digested sample (100 g of total protein) was treated with one tube of each of the iTRAQ reagents in 70% isopropanol at pH 7.2 for 2 hours at room temp. Labeled samples were dried down in a vacuum concentrator. 100uL of water was added to each tube to dissolve the peptides. Samples tagged with 4 different iTRAQ reagents were pooled together. 1% trifluoroacetic acid (TFA), pH 1.4 was added to precipitate RapiGest. Samples were incubated at 4°C overnight and then centrifuged at 16,100g for 15 minutes. Supernatant was collected and centrifuged through a 0.22 micrometer filter and used for LC-

MS/MS analysis. iTRAQ labeling efficiency was calculated by searching through the MS/MS data for samples that contain the 4 specific iTRAQ modifications: 1) fully labeled; 2) n-terminus-labeled only; 3) lysine-labeled only; and 4) non-labeled. Using the above protocol we obtained higher than 90% iTRAQ labeling efficiency for all datasets (Fig. 2-4)

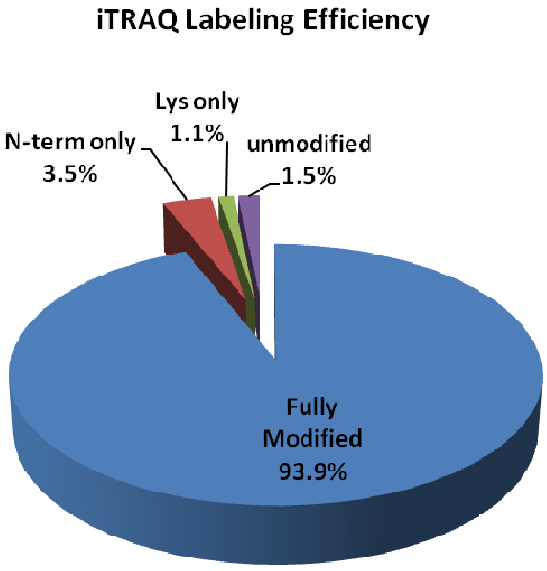


Fig.2-4. iTRAQ labeling efficiency. MS/MS spectra analyzed for the 4 possible iTRAQ modifications: 1) fully labeled; 2) n-terminus-labeled only; 3) lysine-labeled only; and 4) non-labeled. Only 1.5% identified MS/MS spectra were non-labeled, while 93.9% identified spectra were fully labeled.

2.2.3 Chromatography

We used automated 2D nanoflow LC-MS/MS to produce the data described in Preliminary Results. An Agilent 1100 HPLC system (Agilent Technologies, Santa Clara, CA) delivered a flow rate of 300nL per minute to a 3-phase capillary chromatography column through a splitter. Using a custom pressure cell, 5μm Zorbax SB-C18 (Agilent) was packed into fused silica capillary tubing (200μm ID, 360μm OD, 20cm long) to form the first reverse phase column (RP1). A 5cm long strong cation exchange (SCX) column packed with 5μm PolySulfoethyl (PolyLC, Inc.,) was connected to RP1 using a zero dead volume 1μm filter (Upchurch, M548) attached to the exit of the RP1 column. A fused silica capillary (100μm ID, 360μm OD, 20cm long) packed with 5μm Zorbax SB-C18 (Agilent) was connected to SCX as the analytical column (the second reverse phase column). The electro-spray tip of the fused silica tubing was pulled to a sharp tip with the inner diameter smaller than 1 μm using a laser puller (Sutter P-2000). The peptide mixtures were loaded onto the RP1 using the custom pressure cell. Columns were not re-used. Peptides were first eluted from the RP1 to the SCX column using a 0 to 80% acetonitrile gradient for 150 minutes. The peptides were fractionated by the SCX column using a series of salt gradients (from 10mM to 1M ammonium acetate for 20 minutes), followed by high resolution reverse phase separation using an acetonitrile gradient of 0 to 80% for 120 minutes. Typically it takes 4 days (38 salt fractions) for each full proteome analysis.

2.2.4 Mass spectrometry

All of our analyses were performed using LTQ linear ion trap tandem mass spectrometers (Thermo Electron Corporation, San Jose, CA) employing automated, data-dependent acquisition. The mass spectrometer was operated in positive ion mode with a source temperature of 150°C. As a final fractionation step, gas phase separation in the ion trap was employed to separate the peptides into 3 mass classes prior to scanning; the full MS scan range of 400-2000m/z was divided into 3 smaller scan ranges (400-800, 800-1050, and 1050-2000Da) to improve the dynamic range. Each MS scan was followed by 4 MS/MS scans of the most intense ions from the parent MS scan. A dynamic exclusion of 1 minute was used to improve the duty cycle of MS/MS scans. We will test other parameters for gas phase separation, including eliminating it, to optimize the trade off between dynamic range and observation of rapidly eluting peptides. We will use an empirical approach to determine the best length of time for dynamic exclusion, which in theory should equal the median elution time for all peptides.

PQD (Pulsed-Q Dissociation) is a new fragmentation method offered on the Thermo Finnigan LTQ 2.2 control software. PQD eliminates the "1/3 Rule" or “Low Mass Cut-Off” for MS/MS data by decoupling the ion activation and the fragmentation processes, providing the ability to scan product ions down to 50m/z. PQD enables the detection of MS/MS reporter ions (m/z=114, 115, 116, and 117) from iTRAQ, which are normally not detectable on ion trap mass

spectrometers. For our iTRAQ labeled samples, an extra PQD scan is added after each CID (Collision Induced Dissociation) MS/MS scan of the same precursor ion. The iTRAQ reporter ion (114-117) intensities are used for relative quantitation while both CID and PQD fragmentation peaks are used for peptide identification (7, 16, 70).

2.2.5 Protein Identification

Raw data was extracted and analyzed using SpectrumMill (Agilent, version A.03.03). MS/MS spectra with a sequence tag length of 1 or less were considered to be poor spectra and discarded. The rest of the MS/MS spectra were compared against the IPI (International Protein Index) database for human or mouse proteins. The enzyme parameter was limited to full tryptic peptides with a maximum miscleavage of 1. All other search parameters were set to SpectrumMill's default settings (carbamidomethylation of cysteines, +/- 2.5Da for precursor ions, +/- 0.7Da for fragment ions, and a minimum matched peak intensity of 50%).

Search results for individual spectra were automatically validated using filtering criteria from in-situ FDR (False Discovery Rate) calculation. The empirical FDR for such criteria was calculated by searching the data against a concatenated forward-reverse database. The FDR of our filtering criteria is <0.1% spectra, and <1% protein. Most of the matches from the reverse database are 1 or 2 peptide hits. Therefore, only proteins with at least 3 or more

unique peptides are generally selected for quantitative analysis between different samples.

2.2.6 Protein Quantitation

iTRAQ utilizes the MS/MS intensities of the reporter ions for the relative quantitation of up to 4 samples in a run. Spectrum Mill has built-in support of iTRAQ quantitation. The reporter ion intensities are extracted and exported in both spectrum and protein levels. Protein iTRAQ intensities were calculated by summing the peptide iTRAQ intensities from each protein group. Peptides shared among different protein groups were removed before quantitation. A minimal total iTRAQ intensity of 100 was used to filter out low intensity spectra. Isotope impurities of iTRAQ reagents were corrected using correction factors provided by the manufacturer (Applied Biosystems). Median normalization was performed to normalize the protein iTRAQ reporter intensities in which the log ratios between different iTRAQ tags (115/114, 116/114, 117/114) are adjusted globally such that the median log ratio is zero.

Quantitative analysis was performed on the normalized protein iTRAQ intensities. Protein ratios between undifferentiated and differentiated cells were calculated by taking the ratios of the total iTRAQ intensities from the corresponding iTRAQ reporters. T-test (two tailed, paired) was used to calculate the p-values. Proteins with more than 50% change and p-values less than 0.05 were considered significantly changed after differentiation.

2.2.7 Protein function analysis

IDs for identified proteins were converted to gene symbols and Unigene IDs using the IPI gene cross reference table. Protein function analysis is executed through the NCI DAVID (Database for Annotation Visualization and Integrated Discovery) bioinformatics resources. Proteins were categorized according to their cellular distribution, with p-values less than 0.05. Various annotation outputs from DAVID including gene ontology, pathway, protein domain, protein interaction, and disease can be used to cluster and simplify the up- and down-regulated protein lists.

To compare protein profiles to mRNA profiles, we measure mRNA using the Affymetrix GeneChip® Human Exon 1.0 ST Array from the Affymetrix Human Exon 1.0 chip. For comparison, the IPI human protein database v3.16 April 4, 2006 contains 20,227 unique Unigene IDs; 62,322 entries; compiled from a total of 217,390 entries coming from UniProtKB, RefSeq, Ensembl, H-InvDE, and Vega. The Affymetrix GeneChip® Human Exon 1.0 ST Array contains 23,167 unique Unigene IDs; 1.4 million probe sets based on exon clusters, FLmRNA, Ensembl transcripts, ESTs, mouse and rat mRNA, and gene predictions. We have matched these databases and found that there are 18,324 genes shared between the two sets, which will be sufficient for our purposes.

2.2.8 Phosphopeptide enrichment

Phosphopeptide enrichment was performed using home-made TiO_2 columns. 0.2g TiO_2 powder (Sigma# 224227) was weighted and added to an empty 0.22uM filter (Fisher# 07-200-386). Beads were washed with 500uL water twice by centrifuging at 5,000g for 5 minutes followed by 0.5% TFA. Digested peptides (pH 1.4) or iTRAQ labeled peptide mixtures are added to the column and incubated at room temperature for 15 minutes. Samples were centrifuged at 1,000g for 30 minutes until the column was dry. Flow through was discarded. Metal oxide column was washed by 500uL 1% TFA solution (pH 1.5) twice by centrifuging at 5,000g for 5 minutes. Enriched phosphopeptides were eluted by 200uL 100mM $(NH_4)_2HPO_4$ by centrifuging at 1,000g for 30 minutes until the column was dry. Eluted peptides were acidified by adding 3% (v/v) formic acid to a final pH of 3 and concentration of 2%.

Anti-phosphotyrosine antibody is used to enrich the phosphotyrosine peptides due to their low abundance compared to serine and threonine phosphopeptides. Digested peptides are incubated with anti-phosphotyrosine antibody coupled to agarose beads (Sigma) at 4^{o}C overnight. Beads are washed 3 times by Tris buffer (10mM). Phosphotyrosine peptides are eluted by phosphotyrosine solution (1mg/ml, pH 3) and analyzed by 2D-LC-MS/MS.

2.3 Flow cytometry

Cells were harvested with 0.25% trypsin at 37^{o}C for 5-10 minutes to collect single cell suspension and then washed with PBS in the presence of 1% FBS.

The cells were fixed with 70% ethanol at -20^{o}C for 2 hours and stained with propidium iodide solution containing 1mM Tris/HCl pH 8.0, 0.1mM EDTA, 0.1% Triton X-100, 0.1% Na-Citrate, 50mg/ml propidium iodide, 50 mg/ml RNAse A on ice for 1 hour. The apoptotic cells were determined by FACScan and analyzed by FlowJo software.

2.4 Cell Viability Assay

The assay was performed with CellTiter-Glo Luminescent Cell Viability Assay according to the manufacturer's protocol (Promega Cat# G7572). The rationale of this kit is to determine the number of viable cells in a culture by quantification of ATP. There is a direct relationship between the number of viable cells and the amount of ATP present. Cells were seeded in a 96-well plate with 5000 cells per well. Each sample was prepared with three replicates on the same plate, and one plate was analyzed each day. 100µl of reagent was added to 100µl of cell culture media in each well and the plate was mixed rigorously followed by 10 minutes incubation at room temperature. Fluorescent values were normalized with the negative control which contained media only, and the values from three or four consecutive days were plotted with Standard Errors.

2.5 Lentiviral shRNA and miRNA Expression

Molecular Cloning

The mouse and human Sall4 shRNAs were designed and cloned into the lentiviral vector pLV-H1-EF1α-puro using the single oligonucleotide RNAi technology developed by Biosettia (San Diego, CA). The shRNA targeting an irrelevant gene β–galactosidase was used as a negative control. The miRNA (has-let-7a and has-let-7g) precursors and approximately 100-bp upstream and downstream flanking genomic sequences were PCR amplified and cloned into the XhoI site of the human EF1α intron on the lenti-miRNA expression vector pLV-miRNA (Biosettia). The pLV-miRNA vector without insertion was employed as negative control in this study.

Lenti-viral infection

The lentiviral sh-Sall4 and has-let-7 constructs were confirmed by DNA sequencing and co-transfected with packaging mix into 293T cells according to the manufacturer's protocol to generate lentiviral stock. Forty-eight hours post transfection, the supernatant was collected and filtered through a 0.45um filter. One day before transduction, mouse embryonic carcinoma cell lines p19 and F9 were seeded on 6-well plates. The cells were transduced with lentivirus-containing supernatant supplemented with 8μg/ml polybrene and centrifuged at 1,000g for 60 minutes, the medium was replaced immediately after spin transduction. 24 hours post-infection, different antibiotics were added to select for infected cells. 48 hours post-selection, cells were harvested. Target message RNA levels were detected by real-time PCR. Actin mRNA was used as an internal control.

2.6 Protein expression and purification

GST-lin28 and HA-Sall4-His were produced with *E. coli* strain BL21 (DE3). Cells were grown until OD600 was about 0.6, then expression was induced with 0.2mM IPTG at 15°C. 20 hours after induction, cells were harvested and stored at -20°C. For GST-lin28 purification, the cell pellets were suspended in PBS and sonicated on ice for 4 minutes of total stroke time. Cell debris was then spun down at 40,000 x g for 45 minutes. Supernatant was collected and added to 2ml of GST sepharose pre-equlibrated with PBS. GST sepharose beads were then washed with 50ml of PBS. GST-lin28 was then eluted by 5ml of PBS with 20mM glutathione. More PBS was added to the eluted protein to dilute the concentration of glutathione. GST-lin28 was then concentrated with a 10kDa cut-off Centricon device to ~5 mg/ml. For HA-Sall4-His purification, the cell pellets were suspended in Buffer A and sonicated on ice for 4 minutes of total stroke time. Cell debris was then spun down at 40,000 x g for 45 minutes. Supernatant was collected and added to 2ml of NiNTA sepharose pre-eqlibrated with buffer A. NiNTA beads were then washed with 20ml of buffer A plus 25mM imidazole followed by 20ml of buffer A plus 25mM imidazole and 500mM NaCl. Next, NiNTA beads were equilibrated with 20ml of buffer A plus 25mM imidazole. HA-Sall4-His was then eluted by 5ml of buffer A plus 250mM imidazole. The protein was then concentrated and further purified by gel flitration chromatography with a Superdex 200 column, which was pre-equlibrated with

PBS. Fractions containing HA-Sall4-His were collected and then concentrated with a 50kDa cut-off Centricon device to ~5mg/ml.

2.7 In vitro binding assay

1μg of GST or GST-lin28, and 2μg of HA-Sall4-His were mixed. PBS was added to keep the total volume of the mixture at 50μl. After 10 minutes at room temperature, 10μl of GST sepharose was added to the mixture and incubated for another 10 minutes. GST beads were then spun down and the supernatant was discarded. The beads were washed with 500μl of PBS with 0.1% Triton X-100 and 20mM imidazole three times. GST beads were then mixed with SDS-PAGE loading buffer and subjected to SDS-PAGE/immunoblot analysis.

2.8 Western Blot

Samples with SDS loading dye were heated at 94^{o}C for 10 mins. 20μg of total proteins were loaded into 8% SDS-PAGE electrophoresis (100V for 20 minutes, followed by 160V for 80 minutes). Proteins were transferred to 0.2um PDVF membrane by semidry blotting (Bio-Rad) with Transfer buffer (25mM Tris-Base, 200mM Glycine, 20% Methanol) for 1 hour per membrane. Blocking was performed with 5% milk in 0.1% PBS-Tween20 for 1 hour and the membrane was incubated overnight 4^{o}C with appropriate primary antibodies (1:1500 in 5% BSA). The membrane was rinsed with 0.1% PBST three times (15 minutes each, room temperature) and incubated with second antibodies that were raised

against mouse or rabbit (1:10000 in 5% milk w/ 0.1% PBST) for 30 minutes at RT. The membrane was rinsed again with 0.1% PBST three times (10 minutes each, room temperature) and visualized with ECL detection reagent.

2.9 Immunoprecipitation (IP)

Co-IP with Molecular Weight Cut-Off column (MWCO)

Cells were lysed with Lysis buffer. After resuspension, the samples were sonicated for 10 minutes until solutions were clear. Each sample was separated into two tubes at equal volumes. Protein G beads (Santa Cruz Biotechnology #SC2002) were washed 3 times with Lysis buffer. Lysis buffer of the same volume was added to beads to create a 1:1 Protein G solution; and 40µl of Protein G solution was added to each tube in conjunction with 5µg of Rabbit control IgG (Santa Cruz Biotechnology # SC2027) in order to preclear the samples. After two hours of incubation at 4°C, one tube from each sample was taken out and beads were removed through centrifuging at 3000 x g for 1 minute. Another 40µl of 1:1 Protein G solution was then added together with 5µg of Rabbit Sall4 Antibody (Abcam ab31968) into the tubes and incubated overnight at 4°C. The samples were then spun down at 3000 x g for 1 minute and the supernatant was removed. The samples were then washed two times with 400µl lysis buffer and HBS buffer. 2% RapiGest in HEPES was added enough to completely cover the beads and 5M NaCl to a final concentration of 500mM NaCl. Samples were mixed well and incubate at 65°C for 30 minutes

and 94°C for 15 minutes. The mixture was spun through a filter at maximum speed at 4°C, then the volume of the eluent was adjusted to a final volume of 250µl with HBS buffer. Following the manufacturer's recommendation, samples were passed through a 200kDa MWCO column (Advantec Ultra filter USY-20 70517U12). The flow-through was discarded and the filter was washed with 100µl 2% RapiGest in HEPES. The samples were collected and analyzed through LC MS/MS mass spectrometry and Western blotting (Fig.2-5).

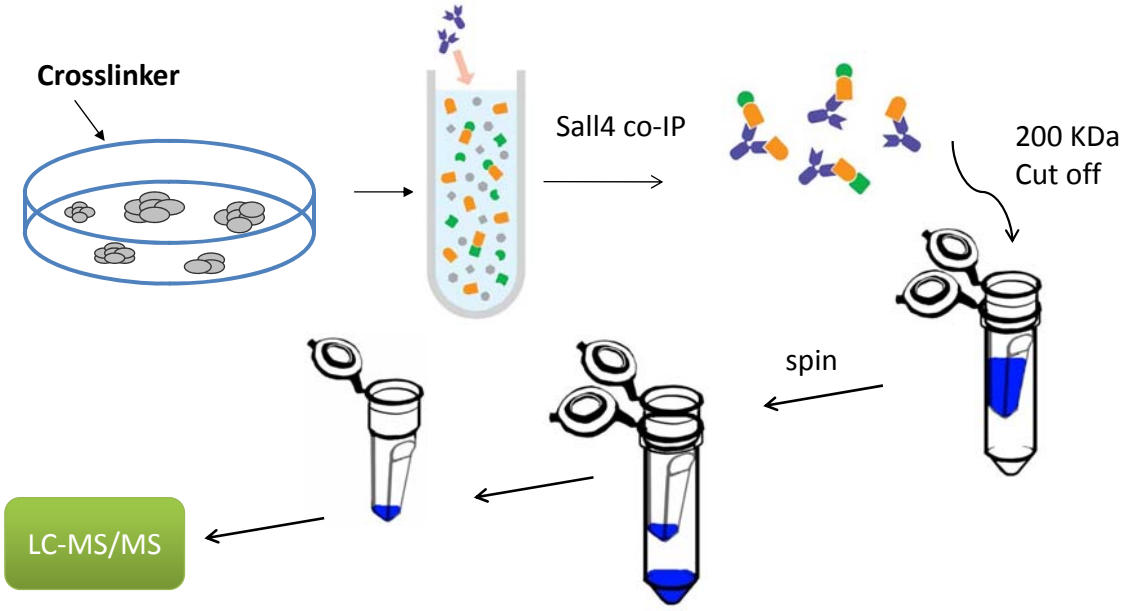


Fig. 2-5 Experimental scheme for enrichment of Sall4 interacting proteins in mES cells.

Flag IP

Transfected HEK293 cells were lysed with four times cell pellet volume of Buffer D. The samples were sonicated for 15 minutes at 4^{o}C and spun down for 20 minutes (16100g, 4^{o}C, Eppendorf 5415R). The pellets were discarded and protein concentration was measured. V5-Beads or Flag-Beads were washed with Buffer D three times and 30ul (1:1, Buffer D: Beads) were added to samples accordingly. The samples were incubated on rotor at 4^{o}C for 1 hour and supernatants were removed through centrifugation (3000g, 4^{o}C, and 1 minute). The beads were washed three times with Buffer D and 20ul of 2x SDS Loading dye (w/b-mercaptoethanol) for Western blot analysis.

2.10 Reverse Transcription

RNA isolation

RNA was extracted with Trizol Reagent (Invitrogen Cat# 15596-018) following the manufacturer's recommendation. 500µl of Trizol Reagent was added to 100µl of cell pellets. The pellets were resuspended with no washing and incubated at room temperature (22^{o}C) for 5 minutes. 100µl of chloroform was added and mixed rigorously for 15 seconds and incubated at room temperature for 3 minutes. The samples were centrifuged at 12000g for 15 minutes at 4^{o}C and upper aqueous layers were saved. 5µl of 2mg/ml glycogen were added as a carrier along with 250µl of isopropanol, and the samples were mixed and incubated at room temperature for 10 minutes. The tubes were

centrifuged at 12000g for 10 minutes at 4^{o}C and the pellets were washed with 600µl of 75% ethanol in DEPC water. The samples were centrifuged at 7500g for 5 minutes at 4^{o}C. Supernatant was removed and the pellets were air-dried for 15 minutes. RNA pellets were resuspended with DEPC water.

DNase treatment

10µg of extracted RNA was taken for DNase Treatment. DNase Treatment was performed with Ambion TURBO DNase Treatment and Removal Reagents (Cat# AM1907) according to the manufacturer's recommendation.

Reverse Transcription

2.5µg of DNase treated RNA was used for cDNA synthesis. cDNA synthesis was performed using Invitrogen SuperScript II First Strand method. 100ng of Pd(N)6 Random Hexamer and 10nM of dNTP were added to 2.5µg of RNA and incubated for 5 minutes at 70^{o}C followed by 1 minute at 4^{o}C. 4µl of 5x First Strand Buffer, 2µl of 0.1M DTT, 1µl of SuperScript II Reverse Transcriptase (Invitrogen Cat# 18064-014), and 1µl of RNase OUT (Invitrogen Cat#10777-019) were added to the mixture and incubated at 25^{o}C for 10 minutes, 42^{o}C for 50 minutes, followed by 70^{o}C for 15 minutes and 4^{o}C for 1 minute. 1µl of RNase H were added (Invitrogen Cat#18021-014) and the samples were incubated at 37^{o}C for 20 minutes.

2.11 Quantitative PCR (qPCR)

qPCR was performed using Stratagene Mx3000P QPCR System (Stratagene Cat#401403). 200-600ng of cDNA was mixed with Stratagene Brilliant SYBR Green QPCR Master Mix (Stratagene Cat#600548) according to the manufacturer's recommendation. Each data set contains three biological and individual replicates. Data Analysis was performed using the comparative Ct method, with the assumption that primer efficiency is 100%. Ct values were converted to fold differences in expression according to the equation $2^{[Ct(reference) - Ct(reference\ \beta\text{-}Actin)] - [Ct(targeted) - Ct(targeted\ \beta\text{-}Actin)]}$.

2.12 PAGE Northern Blot for Small RNAs

Polyacrylamide gel electrophoresis (PAGE) Northern methods were used to detect small RNAs (<200 nt)(53). The 5′-radiolabeled DNA oligo probes were hybridized at 50°C. The control 5S probe was generated by Prime-It II Random Primer Labeling Kit (Stratagene). Storage phosphor screens exposed to Northern blots were scanned on a Typhoon Trio PhosphorImager (GE Healthcare) and band signals were quantified using ImageQuant software. The mean and SEM were calculated from three or more independent experiments where the control sample was set to 1. A two-tailed Student's *t*-test was used to determine if relative mRNA levels were significantly different between conditions (29).

2.13 In-vitro Transcription

In-Vitro Transcription was performed with Riboprobe *in vitro* Transcription Systems (Promega Cat#P1440) following the manufacturer's protocol. 1µl of template DNA was mixed with 4µl of 5x Transcription Optimized Buffer, 2µl of 100mM DTT, 0.6µl of 40U Recombinant RNasin Ribonuclease Inhibitor, 6µl of rNTP (2.5mM each) mix, and 1µl of 20U T7 RNA Polymerase in a 20µl reaction. The mixture was incubated at 37^{o}C for 1hr, and 1µl of RQ1 RNase-Free DNase was added. The sample was incubated at 37^{o}C for 15 minutes, and Phenol-Chloroform extraction was performed to purify the RNA.

2.14 EMSA (Electrophoretic Mobility Shift Assay)

The 5' end 32P-labeled synthetic pre-miRNA (10fmol, 0.5 µL) of 1x10^{4} - 1x10^{5} cpm is mixed with 2µL H_2O and 2.5µL 6x loading dye. It is mixed with another mixture composed of 1µL 10x EMSA buffer, 0.5µL 100mM DTT, 0.2µL RNase inhibitor (40U/µL, Takara), 1µL rSall4, and 2.8µL H2O. After 10 minutes of incubation at room temperature, the 11µL solution is loaded into native gel. The gel is run at 150V for 3 hours at 4^{o}C. The gel is directly exposed to Phosphor Imaging Plate (Fujifilm) and is read with the BAS-2500 system (Fujifilm) for quantification. The 10x EMSA buffer is composed of 150mM Hepes (pH 7.4), 1M KOAc, 20mM Mg(OAc)2, and 2% BSA. The native gel is composed of 5% polyacrylamide (40:1), 40% glycerol, and 1x TBE. The pre-miRNAs were labeled at the 5' end with T4 polynucleotide kinase (Takara) and [γ-32P] ATP (28).

4.5 Acknowledgements

3. Acknowledgements

Chapter 2, in full, is being prepared for publication. I have obtained the permissions from co-authors, Steve Briggs, Zhouxin Shen, and Darwin Yee to reproduce material in this dissertation. The dissertation author will be the primary author of the papers.

Chapter 3

Proteomic profiling and dissecting the role of Sall4 in Embryonic Stem cell fate determination

3.1 Abstract

Sall4 has been reported to be part of the core transcriptional network in embryonic stem cells; however, how Sall4 regulates ES cell fate is still unknown. In this chapter, I took advantage of mass spectrometry to generate whole ESC proteome and phosphoproteome of human and mouse ESCs. Moreover, I generated a novel list of Sall4-binding proteins by modifying a previously established enrichment protocol. Our findings demonstrate the effectiveness of using a synthetic crosslinker, DSP, to enrich for weak and transient Sall4 interacting protein complexes. Working with MWCO, allows us to identify new Sall4 binding partners, leading to a better understanding of the regulatory machinery behind the self-renewal and pluripotency in ESCs. Pathway analysis indicates that Sall4-binding proteins are involved in Wnt signaling pathway, FGF signaling pathway, and p53 pathway. Our data indicates that the *Sall4* mutant ES cell proliferation defect may be regulated by Integrin, TGF-beta, Cadherin, and PI3K signaling pathways. Furthermore, the interactome has led me to explore a possible novel function of Sall4 in microRNA biogenesis. By understanding how Sall4 regulates cell fate determination can substantially push ESC research forward.

3.2 Introduction

Signal transduction processes are of key importance for the biological function of eukaryotic cells. Protein phosphorylation, an important kind of post-translational modifications, plays a central role in signal transduction. Many biological events involved in cellular response are mediated by the dynamics of phosphorylation and dephosphorylation, such as regulation of protein conformational change, enzymatic activity, subcellular localization, protein-protein interaction, and protein ubiquitination. Protein phosphorylation can control various biological functions such as cell proliferation, differentiation, metabolism, survival, and gene transcription (12-13, 61). In the past decade, quantitative proteomics has emerged as a powerful tool to study signaling on a genome-wide scale (21, 54).

Mass spectrometry-based large-scale proteomics is currently one of the most powerful tools used to dissect the molecular mechanisms underlying embryonic stem cell properties. Many groups have successfully established and improved protocols that enable them to use this tool to answer very important biological questions. Nagano *et al* discovered 1790 proteins, including 365 nuclear, and 260 membrane proteins in mouse ES cells with the help of multidimensional-LC Tandem Mass Spectrometry (MS/MS) (48). With nano-LC FT-ICR MS/MS, Van Hoof *et al* identified 1775 proteins in human ES cells, 1532 in differentiated human ES cells, 1871 in mouse ES cells and 1552 in differentiated mouse ES cells. Furthermore, they also suggested new candidates

for ES cell markers from the many uncharacterized proteins that were identified (76). Swaney *et al* studied the hES cell phosphoproteome using electron transfer dissociation (ETD) MS/MS and discovered 10844 phosphorylation sites (66). Some researchers combine conventional proteomic methods with mass spectrometry to analyze the protein-protein interactome. MS/MS was employed to analyze the Nanog co-immunoprecipitated samples in an attempt to reveal the Nanog interactome in ES cells (78). A more advanced work was performed to obtain a large-scale hES cell protein-protein interactome generated by using LC-ESI-MS/MS analysis. More than 6400 protein-protein interactions between 2235 distinct proteins were identified (24).

The whole cell proteomic and phosphoproteomic analysis provides us large repertoire of protein lists that can potentially unmask the regulatory elements in ESCs. What do these protein lists tell us? How do these proteins work together in a complex biological system? How do they communicate within a cell and signal among different cells? In order to answer these questions and further study the biological meanings, the raw data has to be analyzed and displayed systematically by using both theoretical and computational approaches. By taking advantages of the system biological tools available online, such as PANTHER and DAVID, proteins can be sub-grouped according to their functions. I will briefly introduce these two tools and present the graphs and figures in the Results section.

PANTHER (Protein Analysis Through Evolutionary Relationships) Pathway

PANTHER is a free online software system that organizes genes and proteins by evolution, specific functions, and biological roles (45-46). It was designed to model evolutionary sequence-function relationships on a large scale. Input protein list could be divided into groups based on PANTHER classification (either molecular functions or biological process pathway) (72-73).

DAVID (Database for Annotation, Visualization and Integration Discovery)

DAVID is a high-throughput functional annotation that systematically maps a large number of interesting genes in a list to associated Gene Ontology (GO) terms, and then statistically highlights the enriched GO terms (22, 30). This tool utilizes a novel algorithm to measure relationships among the annotation terms based upon the degrees of co-association of the genes in effort to group annotation contents from the same or different resources into additional annotation groups.

http://david.abcc.ncifcrf.gov/content.jsp?file=functional_annotation.html#DATA_INPUT

Sall4-enriched mass spectroscopic analysis

The *Sall4*-null ESCs lacks exon 2 and 3. In order to make these mutant cells a better control for our co-IP experiments, I compared and tested several commercially available antibodies. Eventually, I decided to use the rabbit Sall4 polyclonal antibody (abcam# 29112), which recognizes the epitope within

residues 1000 to the C-terminus of mouse Sall4 to pull-down Sall4 binding partners. There are always non-specific binding proteins when we perform co-immunoprecipitation (co-IP). The non-specifically binding proteins are usually abundant in the cells and causes difficulty in detection of low abundant proteins by mass spectrometry (LC ESI MS/MS). In order to reduce this problem, I have developed an enriched co-IP technique to identify new Sall4 binding partners. While the interpretation of this large data set is ongoing, I have found that these enriched proteins are tightly associated with post-translational modification and cell growth.

Choosing a Suitable Cross-Linker

In our quantitative proteomic data, it showed that Sall4 is highly expressed in undifferentiated ESCs. It has been reported that Sall4 is a zinc finger protein, and it binds to DNA (83, 89). I decided not to use formaldehyde, a cross-linker commonly used in chromatin immunoprecipitation, for the following reasons: 1) I was not solely interested in transcriptional regulation of Sall4. Moreover, I would like to decipher the mystery of Sall4 and study how it regulates ES cell fate determination. For this reason, I tried to avoid the protein-protein interaction through the conjugation of chromatin by formaldehyde. 2) It has been shown that by using formaldehyde, the Oct4-Nanog protein complex could not be preserved (90).

In this study, I chose Dithiobis (succinimidyl propionate) (DSP) to crosslink Sall4 interacting proteins. DSP is a homobifunctional, thiol-cleavable cross-linker. It contains an amine-reactive*N*-hydroxysuccinimide (NHS) ester at both ends of a cleavable, 8-atom (12Å) spacer arm. NHS esters react with primary amines at pH 7-9 to form stable amide bonds, along with release of the *N*-hydroxy-succinimide leaving group. Proteins, including antibodies, generally have several primary amines in the side chain of lysine (K) residues and the N-terminus of each polypeptide that are available as targets for NHS-ester cross linking reagents. Because DSP does not possess a charged group, it is lipophilic and membrane-permeable and is useful for intracellular and intramembrane conjugation.
(http://www.piercenet.com/Objects/View.cfm?type=File&ID=0544)

Molecular Weight Cut-Off column (MWCO)

In our quantitative proteomic data, it shows that Sall4 is highly expressed in undifferentiated ESCs. After crosslinking the proteins, there were still many free Sall4 proteins in the cell lysates. This interfered with the results from the LC-MS/MS. In addition to the cross-linker, I also applied a Molecular Weight Cut-Off column (MWCO) to enrich Sall4-binding proteins. MWCO is a column with ultra filtration membranes with different size designations. Usually the membrane retains 90% of the protein elution after filtration based on their molecular weight cut off values. The molecular weight of mouse Sall4 is about 140KDa. I

assumed the size of Sall4-binding complexes would be greater than 140KDa. Since the 150KDa MWCO is not commercially available, I chose 200KDa instead.

3.3 Results

Derive extra-embryonic endoderm (XEN) cells from human ESCs

Human and mouse ES cells can spontaneously differentiate extra-embryonic endoderm cells when cultured on cell culture dishes (Fig3.1 and Fig. 3.2). In order to study how ES cells decide their fate to differentiation, I developed a special method to harvest and maintain these cells for future proteomic analysis. The method was detailed in chapter 2.

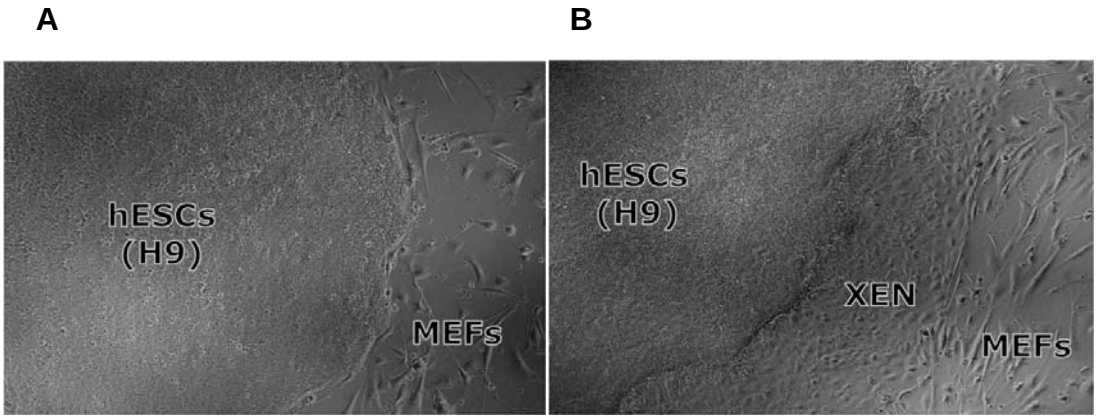


Fig.3-1 H9 hESCs grow on mouse feeder (MEFs). (A) The undifferentiated hESCs usually grow as compact colonies and are difficult to distinguish individual cells. The colonies have defined edges and with high nuclear to cytoplasmic ratio. Under phase contrast microscopy, mouse feeders have distinguishing spindle-shape morphology. (B) H9 cells can spontaneously differentiate to XEN cells under the same culture conditions. H9, XEN cells, and MEFs can be easily discriminated by their morphologies.

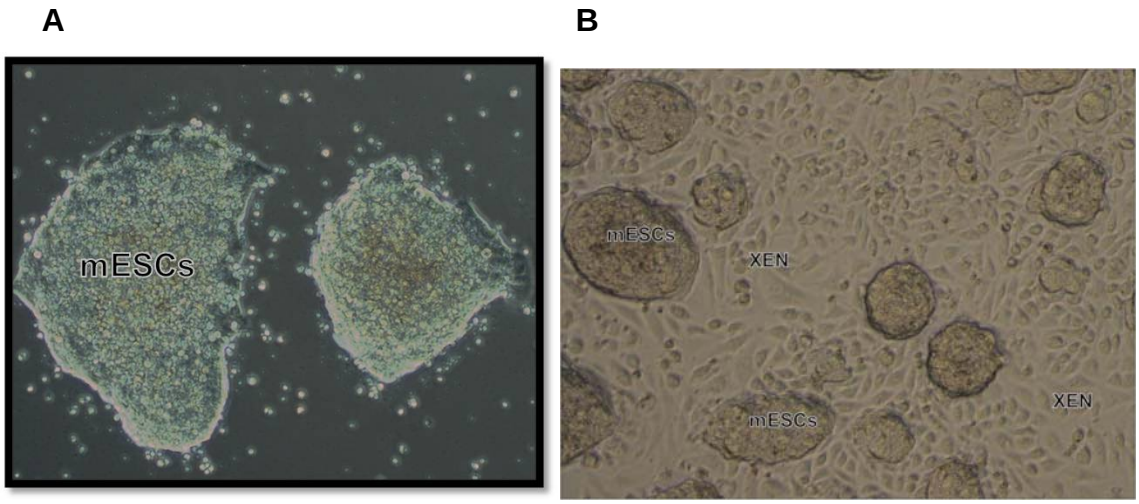


Fig. 3-2 Mouse ESCs (mESCs) grow on feeder-free condition. (A) Two undifferentiated colonies of mESCs under phase contrast. (B) Mouse XEN cells can also be observed under the culture conditions.

Proteomic and Phosphoproteomic analysis of hESCs

We have previously established the quantitative phosphoproteomic profiling of mouse embryonal carcinoma (mECs) and the p19 cell line by using liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS-MS). Undifferentiated mEC cells and retinoic acid (RA)-treated differentiated mEC cells were harvested and the proteins were extracted and digested. Phosphopeptides were enriched by using TiO_2 affinity chromatography. We identified 1270 phosphoproteins in undifferentiated mECs; however, the phosphorylations are not present at detectable levels in RA-treated mECs. We repeated the experiment on undifferentiated mESCs (129/svev) and hESCs (H9)

and differentiated embryoid bodies (EB) up to day 8. We not only confirmed that phosphorylations are conserved in many stemness proteins in undifferentiated hESCs, but also observed more phosphoproteins in ESCs than in differentiated ESCs (Table 3-1).

A total of 4182 proteins in H9/EBs were identified, 2038 were phosphorylated with 1.9% False Discovery Rate (FDR). All proteins identified were also classified by subcellular localization, biological process and molecular function by DAVID (Fig. 3-3 and 3-4). The experiments were repeated on H9/XEN cells (Fig. 3-5 and 3-6) in order to profile differences between the biological processes of XEN cells versus heterogeneously differentiating ESCs in EBs. As expected, more proteins were involved in diverse biological processes in the differentiating ESCs, presumably reflecting the “cell fate switches” involved in differentiation.

Table 3-1 Total proteins and phosphoproteins that were detected by our proteome and phosphoproteome.

Species	Cell types	Proteins	Phospho-proteins
Human	H9/ XEN	4092 (1.6% FDR)	2038 (1.9 % FDR)
Human	H9/ EB	4182 (4.1% FDR)	2038 (1.9 % FDR)
Mouse	129 ESC/ EB	4672 (1.4% FDR)	1957 (1.1 % FDR)
Mouse	p19 EC/ EB	4036 (2.3% FDR)	1470 (2.7 % FDR)

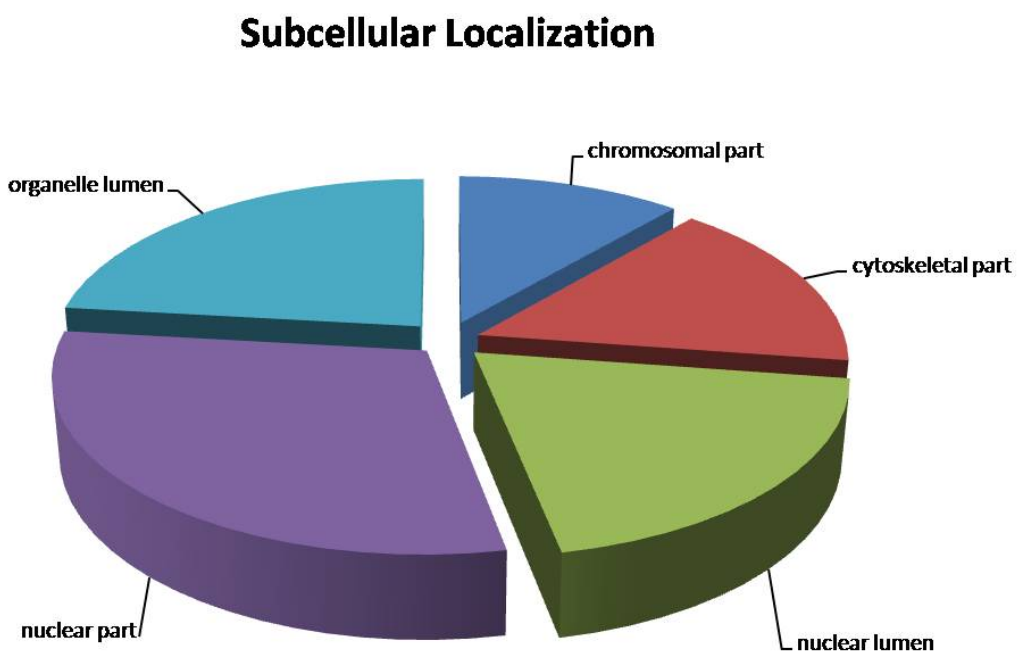


Fig. 3-3 Subcellular localization distribution of whole H9/Embroid body proteome.

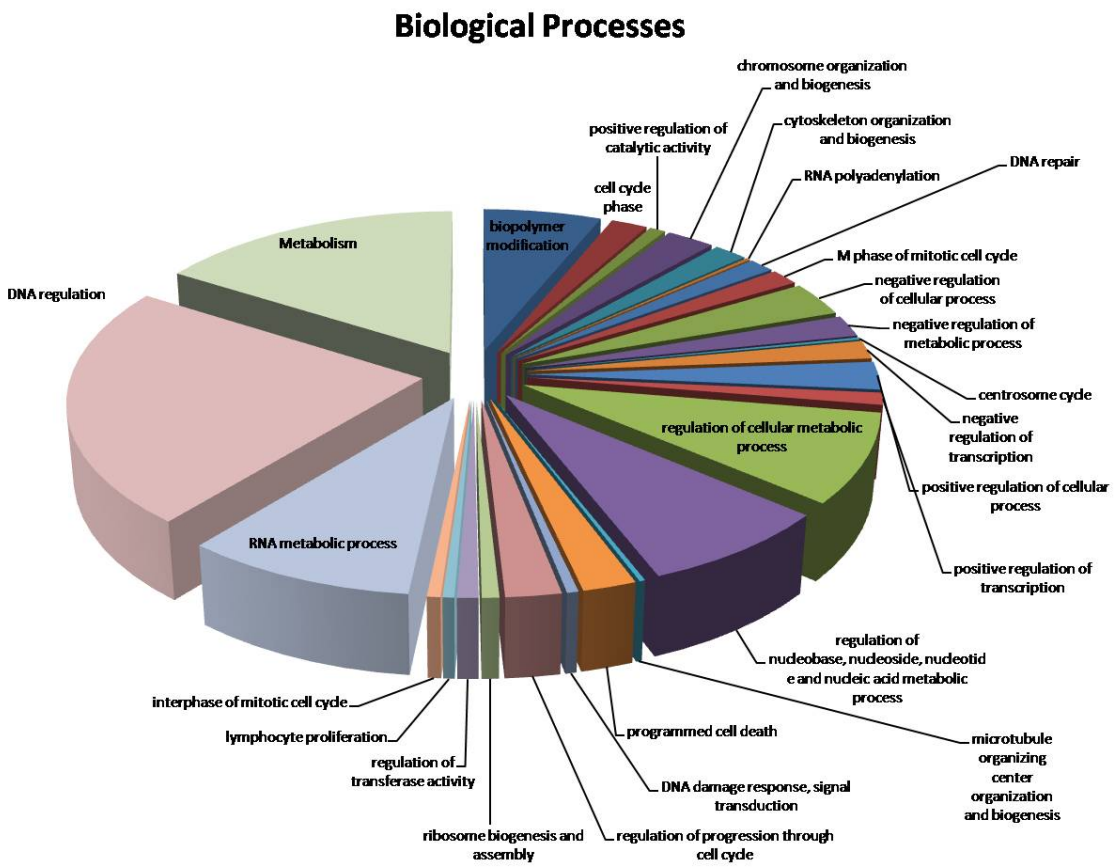


Fig. 3-4 Biological processes distribution of whole H9/Embryoid body proteome.

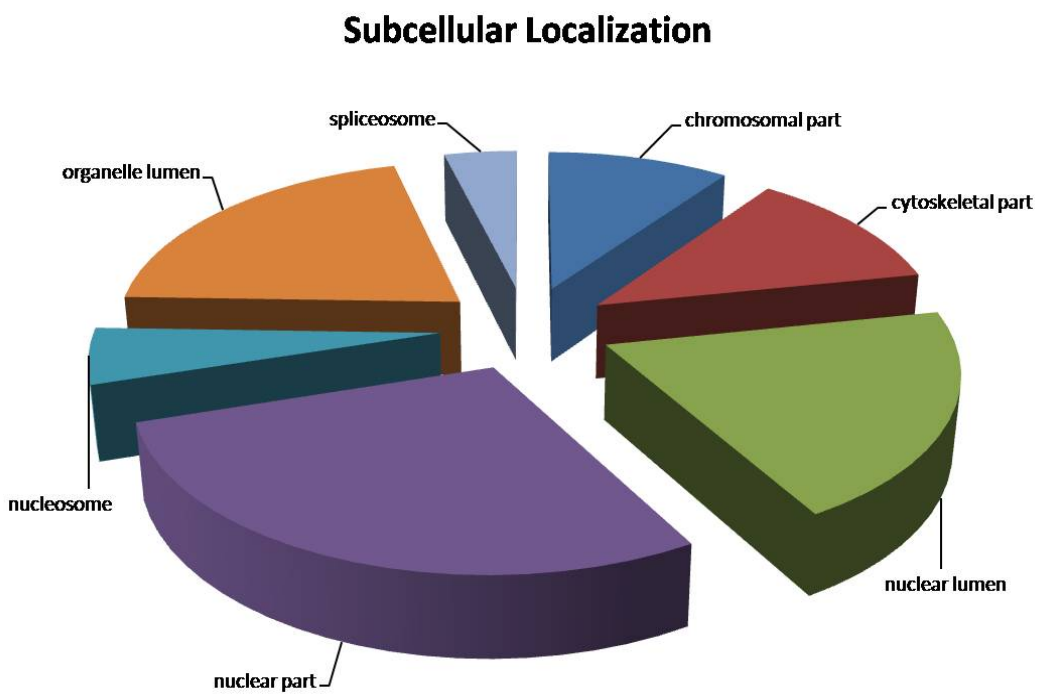


Fig. 3-5 Subcellular localization distribution of whole H9/XEN cell proteome.

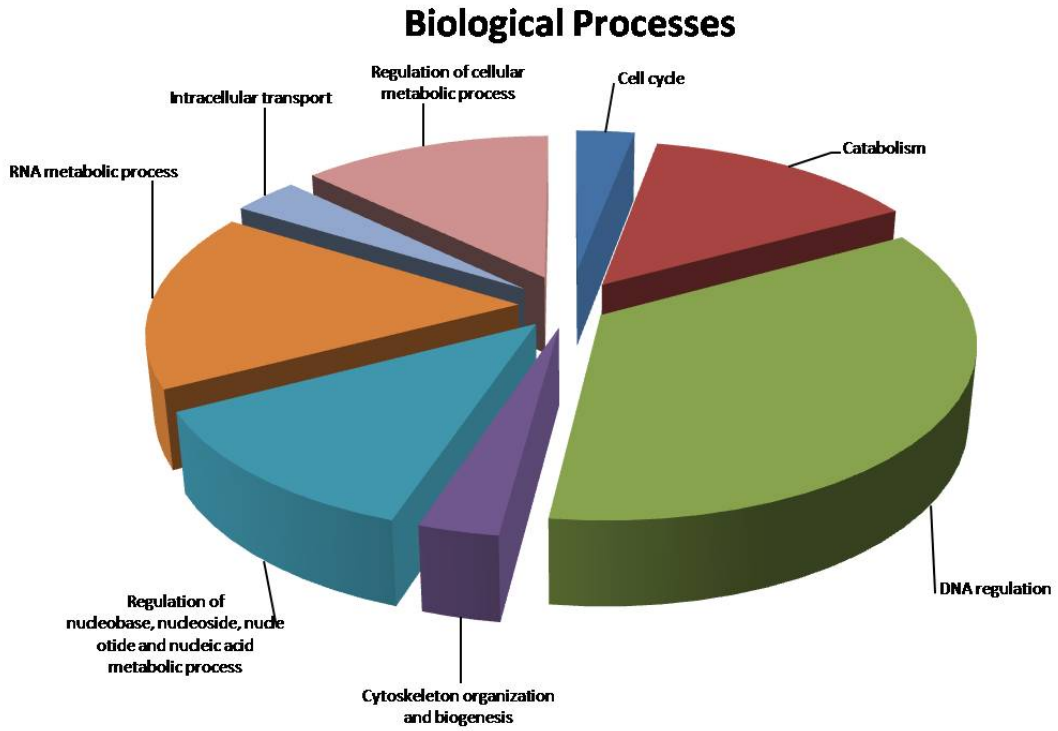


Fig. 3-6 Biological process distribution of whole H9/XEN cell proteome.

Dissecting the functions of Sall4 in cell fate determination

Phosphorylation of Sall4 has not yet been reported. We observed that amino acids T93, T789, T791, S785 are phosphorylated in mES cells, but not in EC cells differentiated into neurons following RA treatment (Table 3-2). Interestingly, all of the phosphorylation sites are located in exon 2. These results raise the question whether phosphorylation of Sall4 is responsible for the ES cell pluripotency.

Table 3-2 Phosphorylation sites of Sall4 that have been detected from our phosphoproteome from different cell lines.

Cell types	Phosphorylation sites
Human ESCs	S112, S116, S776, S789
Mouse ESCs	T93, S122, S785, T791
Mouse ECs	T93, S122,

I next characterized the phenotype of mouse *Sall4*-mutant ESCs that were obtained from the Nishinakamura group. They removed exon 2 and 3 that contains all zinc finger domains (60) (Fig. 3-7). We observed that the compact colony phase of ES cell growth was significantly delayed in *Sall4*-KO ESCs and a reduced proliferation rate of 50% compared to *Sall4*-het ESCs (Fig. 3-8). Our discovery suggests that *Sall4* is required for ES cell to cell interaction including compact colony morphology, and for proliferative or apoptotic signals that are generated within compact colonies. In order to answer the question whether exon 2 and 3 of Sall4 are essential for maintenance of this morphology, I overexpress Sall4 in S*all4*-KO ESCs. The compact morphology could be recovered by Sall4 (Fig. 3-9). What does exon 1 or 4 do to *Sall4*-KO ESCs? Since we detected two unique peptides of exon 4 in the *Sall4*-KO ESCS from our mass spectrometry, I designed shRNA to target 3' UTR of *Sall4* to knockdown

exon 1 and 4. The result showed that exon 1 and 4 were not essential for the compact morphology (Fig. 3-10)

I next ask what causes proliferation defect in *Sall4*-KO ESCs. I hypothesized that the reduced cell proliferation of *Sall4*-KO ESCs is due to apoptosis. In order to detect apoptosis, the ESCs were stained with Annexin V-FITC conjugate and Propidium Iodide. The cells were sorted by FACS. Flow cytometry revealed that 4.84% of Sall4-KO ESCs are apoptotic (Fig. 3-11). The rate is close to the WT ES cells (4.63%). My data suggests that the reduced proliferation rate may not be due to apoptosis.

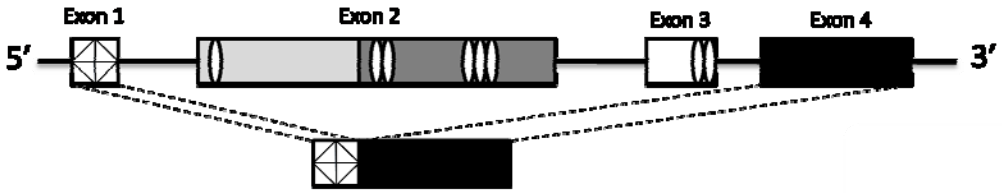


Fig. 3-7 The *Sall4*-mutant ESCs that have been used in this study. The exon 2 and 3 has been removed, but exon 1 and 4 are still present in the cells. I also called them *Sall4*-KO ESCs.

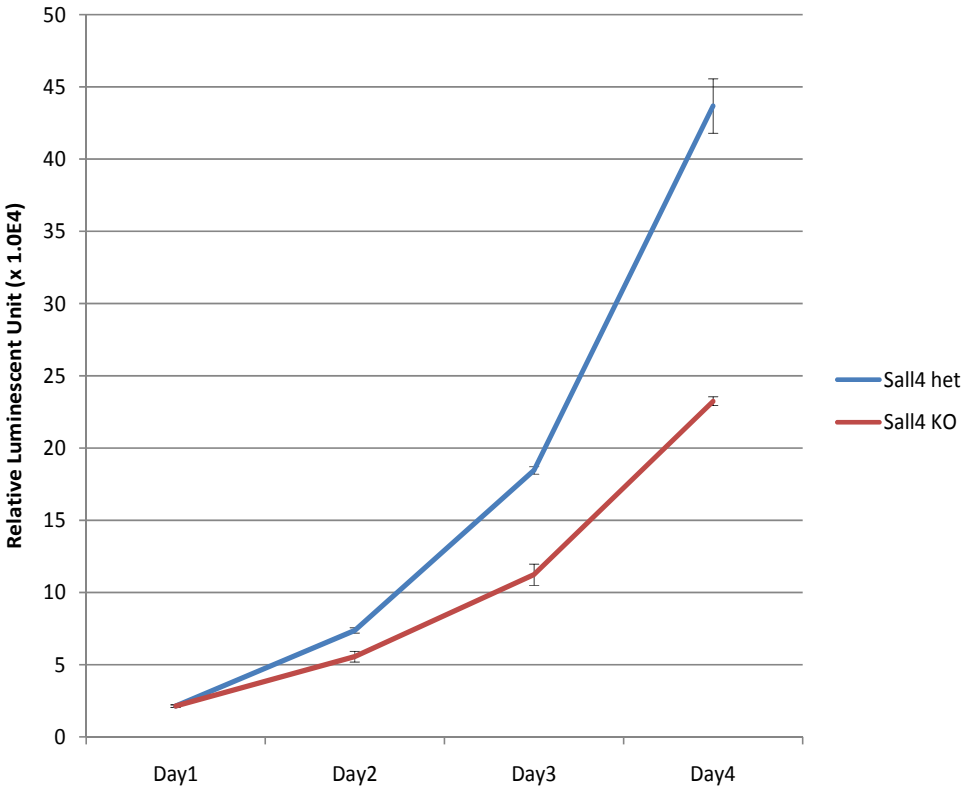
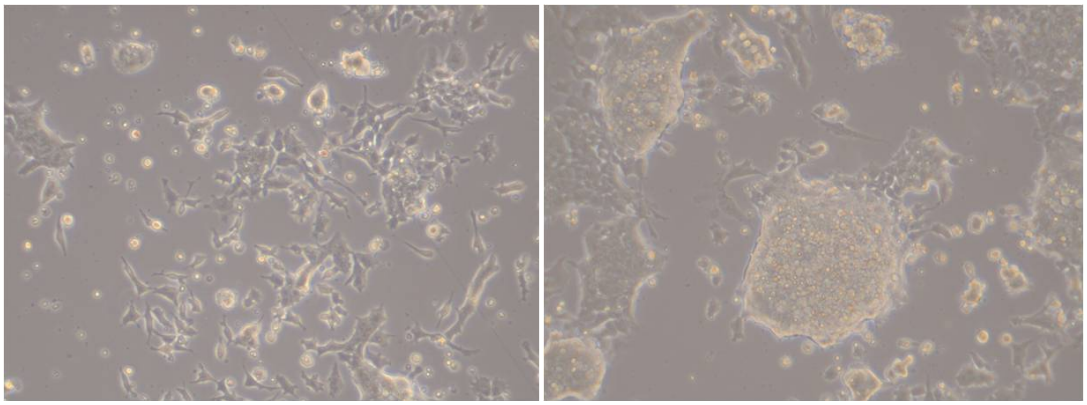
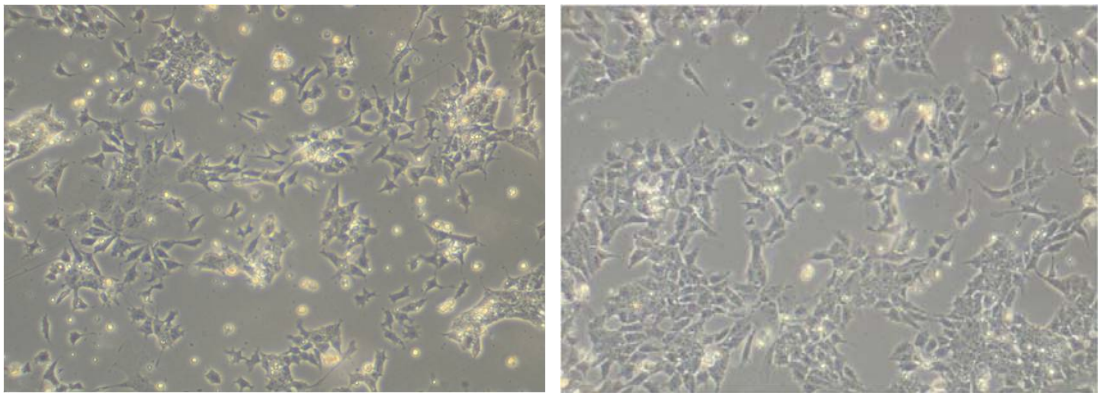


Fig. 3-8 *Sall4*-KO ESCs show proliferation defect compared to Sall4-het ESCs by cell viability assay. Cell number correlates with luminescent output. Values are normalized by background signal resulting from serum-supplemented medium without cells and represent the mean ± S.D. of 3 replicates for each cell number.



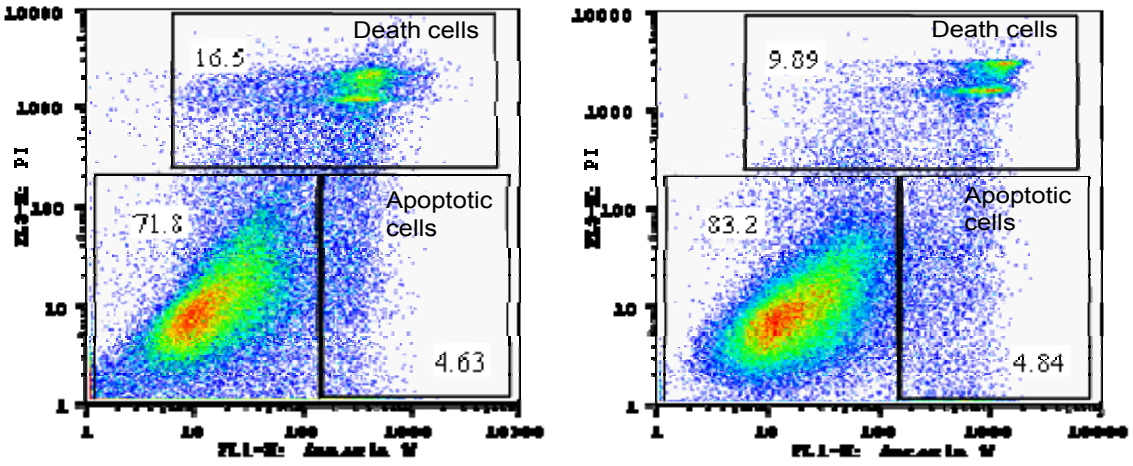
Sall4-KO ESCs + pcDNA **Sall4-KO ESCs + Sall4-pcDNA**

Fig. 3-9 Overexpression of Sall4 can rescue the *Sall4*-KO ESCs to form compact colonies.



Sall4-KO ESCs + shLac **Sall4-KO ESCs + shSall4**

Fig. 3-10 Knockdown of exon1 and 4 of Sall4 does not change the phenotype of Sall4-KO ESCs



Sall4- het ESCs (Sall4 flox/- ESCs)

Sall4- null ESCs (Sall4-KO ESCs)

Fig. 3-11 Flow cytometry revealed 4.63% of Sall4 het ESCs and 4.84% of Sall-KO ESCs are apoptotic cells. The ESCs were stained with Annexin V-FITC conjugate and Propidium Iodide.

Sall4 whole cell proteomic profiling

In order to study the mechanism of Sall4 in regulating ES cell proliferation, I obtained the *Sall4* het ESCs and *Sall4*-KO ECSs proteome and tried to identify the possible pathways that may control the phenotype. I have identified 4000 proteins by iTRAQ, and I am interested in studying those proteins that undergo at least 3-fold change (increase or decrease) between Sall4 het and KO ESCs. Taking advantage of the online bioinfomatics resource DAVID, the list of proteins was inputted into the database for pathway analysis and gene ontology (Fig. 3-12 and 3-13). The functional categories of the Sall4 binding proteins in mouse ES cells revealed a wide variety of cellular processes. I summarized the highlighted

pathways and listed detected proteins on Table 3-3. Surprisingly, I observed several interesting pathways that maybe regulated or influenced by Sall4 downregulation, such as Integrin signaling, TGF-beta signaling, Cadherin signaling, PI3K signaling, Wnt signaling, FGF signaling, and p53 signaling pathways.

We developed a method for identifying differentially expressed proteins using the relative abundance of iTRAQ reporter ions acquired in mass spectra. This process involves several scoring steps for both within and between MS/MS experiments. For a given MS/MS experiment, we first calibrate the intensity ratio of two reporter ions of interest by their intensity dependent variation. The intensity ratio of a specific peptide is then defined as the averaged reporter ratio among the spectra, which have been recognized as from this peptide. Similar to the peptide ratios, the fold change of a protein between the two conditions of interest is the averaged ratio of all mapped peptides. The protein ratios from individual experiments are then integrated into a single score to assess the significance and robustness of the observed changes. To estimate the p-values of the observed changes, the protein scores are compared to the random scores sampled from a permutation test of 300 random trials across multiple MS/MS experiments. For each MS/MS experiment in a random trial, we randomly assign the intensity ratios of two reporter ions to the spectra and then perform the same scoring procedure as that of real data. Such permutation disrupts the correlation between spectra, which are generated

from the same peptide. The score of each protein is indexed on the null distribution of all random scores. The protein names were listed on Table 3-4 and 3-5 according to their expression changes.

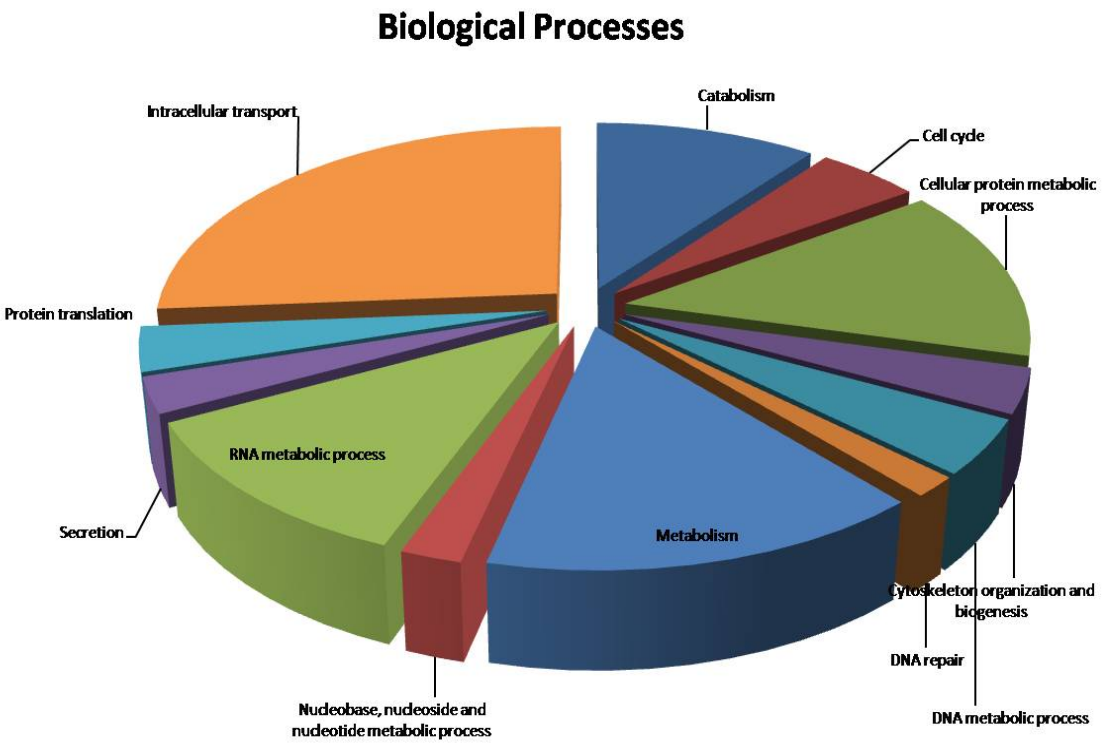


Fig. 3-12 Biological process distribution of whole *Sall4* mutant ES cell proteome.

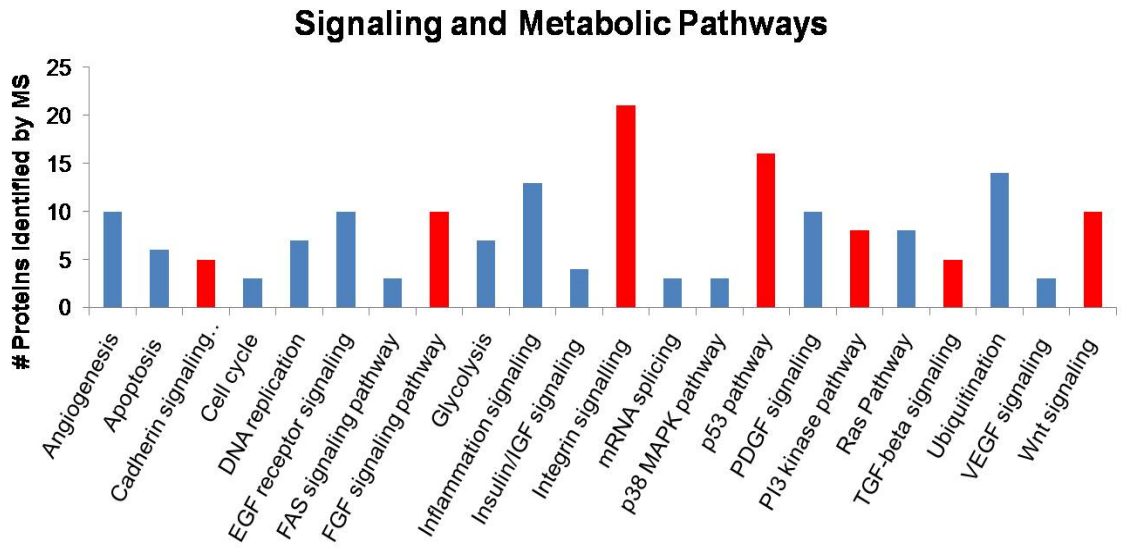


Fig. 3-13 Selected signaling pathways represented from whole Sall4 mutant ES cell proteome. The Panther Pathway Mapping system was used to map 845 proteins into a total of 75 pathways. Interesting ES cell regulatory pathways with relative high coverage are highlighted as red.

Table 3-3 Proteins detected from selected Signaling Pathways.

Pathway	Protein detected
FGF signaling	Pebp1, Pik3c3, Ppp2r1b, Ppp2r5d, Rac1, Rras, Rras2, Sfn, Ywhab, Ywhae
p53 pathway	Cdc2a, Grse1, Pdk1, Pik3c3, Sfn, Ubtf, Ywhab, Ywhae, Zmat2, Ihbp1, Ctnnb1, Rras, Rras2, Taldo1
Integrin signaling	Actn1, Arf6, Arfgap1, Arl1, Arpc5l, Cdc42, Flna, Flnb, Flnc, Fn1, Itgav, Lama1, Lama5, Lamb1-1, Pik3c3, Rac1, Rras, Rras2, Vcl
TGF-beta signaling	Cdc42, Fkbp1a, Rab10, Rras, Rras2
Cadherin signaling	Ctnnb1, Ctnnd1, Gsk3b, Ptpn2, Taldo1
PI3K signaling	Gnai2, Gsk3b, Pdk1, Rras, Rras2, Sfn, Ywhab, Ywhae
Wnt signaling	Adss, Arid1a, Csnk1a1, Ctnnb1, Gsk3b, Ppp2r5d, Smarca4, Smarcb1, Smarce1, Taldo1

Table 3-4 Selected proteins that decreased in expression in *Sall4*-KO ESCs compared to *Sall4*-het ESCs.

Gene Symbol	Entrez ID	Het/null Fold Change	Consistent between protein and gene	Protein Name
Blvrb	233016	2.04	N/A	Flavin reductase
Ckb	12709	1.67	Yes	Creatine kinase B-type
Eef1d	66656	1.07	Yes	eukaryotic translation elongation factor 1 delta isoform a
Eif2s1	13665	1.13	No	Eukaryotic translation initiation factor 2 subunit 1
Ftsj3	56095	1.57	No	Putative rRNA methyltransferase 3
Gfpt2	14584	1.74	Yes	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] 2
Gstm1	14862	1.73	Yes	Glutathione S-transferase Mu 1
Hspd1	15510	1.29	Yes	60 kDa heat shock protein, mitochondrial precursor
Mdn1	100019	1.15	Yes	similar to Midasin (MIDAS-containing protein) isoform 1
Mtap	66902	1.96	No	S-methyl-5-thioadenosine phosphorylase
Rad23b	19359	1.24	N/A	RAD23 homolog B
Rps4x	20102	1.39	N/A	40S ribosomal protein S4, X isoform
Ruvbl2	20174	1.15	No	RuvB-like 2
Sall4	99377	6.50	Yes	Isoform 1 of Sal-like protein 4
Sfxn1	14057	1.65	Yes	Sideroflexin-1
Srm	20810	1.14	Yes	Spermidine synthase
Tubb6	67951	1.84	No	Tubulin beta-6 chain
Ubap2	68926	1.42	No	Ubiquitin-associated protein 2

Table 3-5 Selected proteins that increased in expression in *Sall4*-KO ESCs compared to *Sall4*-het ESCs.

Gene Symbol	Entrez ID	Het/null Fold Change	Consistent between protein and gene	Protein Name
Asns	27053	0.68	No	Asparagine synthetase
Atp5f1	11950	0.79	N/A	ATP synthase B chain, mitochondrial precursor
Canx	12330	0.72	No	Calnexin precursor
Cdc42	12540	0.42	Yes	Cell division control protein 42 homolog precursor
Ckap4	216197	0.47	Yes	Cytoskeleton-associated protein 4
Dpysl3	22240	0.42	Yes	Dihydropyrimidinase-related protein 3
Etfa	110842	0.64	Yes	Electron transfer flavoprotein subunit alpha, mitochondrial precursor
Hadh	15107	0.45	No	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial precursor
Hist1h1b	56702	0.57	Yes	Histone H1.5
Hmga2	15364	0.44	Yes	High mobility group protein HMGI-C
Hmgb2	97165	0.62	Yes	High mobility group protein B2
Hnrnpa0	77134	0.81	No	heterogeneous nuclear ribonucleoprotein A0
Igf2bp3	140488	0.47	Yes	insulin-like growth factor 2, binding protein 3
Nucks1	98415	0.60	Yes	Nuclear ubiquitous casein and cyclin-dependent kinases substrate
P4hb	18453	0.73	Yes	Protein disulfide-isomerase precursor
Pdia3	14827	0.89	Yes	protein disulfide isomerase associated 3
Pgk1	18655	0.73	N/A	phosphoglycerate kinase 1
Psmc3	19182	0.74	Yes	26S protease regulatory subunit 6A
Raver1	71766	0.56	No	ribonucleoprotein, PTB-binding 1
Slc2a1	20525	0.73	Yes	Solute carrier family 2, facilitated glucose transporter member 1
Syncrip	56403	0.67	No	Heterogeneous nuclear ribonucleoprotein Q
Tfrc	22042	0.72	Yes	Transferrin receptor protein 1
Tpi1	21991	0.71	No	Triosephosphate isomerase
Tyms	22171	0.64	No	Thymidylate synthase
Vim	22352	0.67	Yes	Vimentin
Ywhae	22627	0.79	Yes	14-3-3 protein epsilon

Sall4 protein-protein interactome

Our Sall4-enriched mass spectroscopic analysis was detailed in the Chapter 2. Six individual biological replicates of WT and Sall4 het ESCs treated with crosslinker and MWCO for LC-MS/MS. For negative controls, I used *Sall4*-KO ESCs and normal rabbit IgG. As a result, a total of 621 proteins were detected in our six runs of the samples. 334 proteins remained after removal of proteins overlapping with *Sall4*-KO ESCs and normal rabbit IgG crosslinker + MWCO treated samples (Fig. 3-14). For *Sall4* binding confirmation, co-IP and Western blotting was performed. 334 proteins were from different biological processes as annotated by DAVID and PANTHER. Interesting biological pathways are shown on (Fig. 3-17). Several interesting mechanisms are highlighted such as cell cycle, FGF signaling, p53 signaling, and Wnt signaling. Entire protein lists are shown on Table 3-6. It has been shown that overexpression of the novel oncogene SALL4 and activation of the Wnt/beta-catenin pathway in myelodysplastic syndrome (62). SALL4 is also directly activated by TCF/LEF in the canonical Wnt signaling pathway (8). Our data supports these findings. It would be promising to hypothesize that Sall4 maybe also involved in p53 and FGF signaling in regulating of proliferation.

The interaction network generated from DAVID indicates the function of Sall4 in DNA methylation, cell cycle, RNA binding, and embryonic development (Fig. 3-18). Interestingly, when I looked into the individual proteins and searched each protein from GeneCards, I could draw a totally different interacting network

(Fig. 3-19). This interactome implied several possible mechanisms such as microRNA biogenesis and reprogramming that Sall4 maybe involved with and allows us to look at Sall4 not just only as a transcription factor.

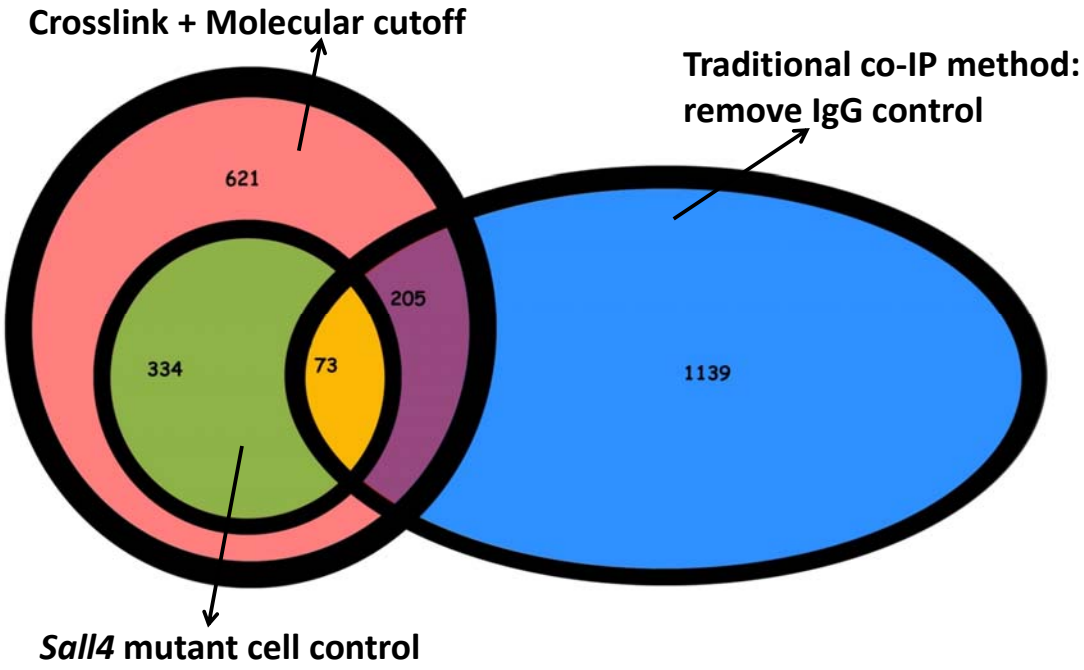


Fig. 3-14 Sall4 co-IP with and without enrichment. Venn diagram comparing the number of proteins detected in samples with and without enrichment is shown. 205 proteins were overlapping. Removal of proteins detected in *Sall4-null* sample revealed 73 proteins that were in both enriched and unenriched samples.

Confirmation of *Sall4* Interacting Proteins

Lin28, Dnmt3l, and Ranbp1 were selected for reciprocal co-IP verification. They were involved in mRNA metabolic processes, embryonic development/DNA

metabolism, and nuclear transport respectively. Reciprocal co-IP was performed with either *Sall4* het or WT mES cells with α-Lin28 (Abcam ab46020), α-Dnmt3l (Santa Cruz sc20705), α-Ranbp1 (Bethyl A300-512A), and rabbit α-Sall4 (Abcam ab29112) was used for Western blotting. The binding between *Sall4* and the three proteins was confirmed (Fig. 3-15). *Sall4*-KO reciprocal co-IP for these proteins was performed in parallel, and no *Sall4* was detected. Based on the result from the Western blot, I summarized the possible interacting domains of Sall4 (Fig. 3-16).

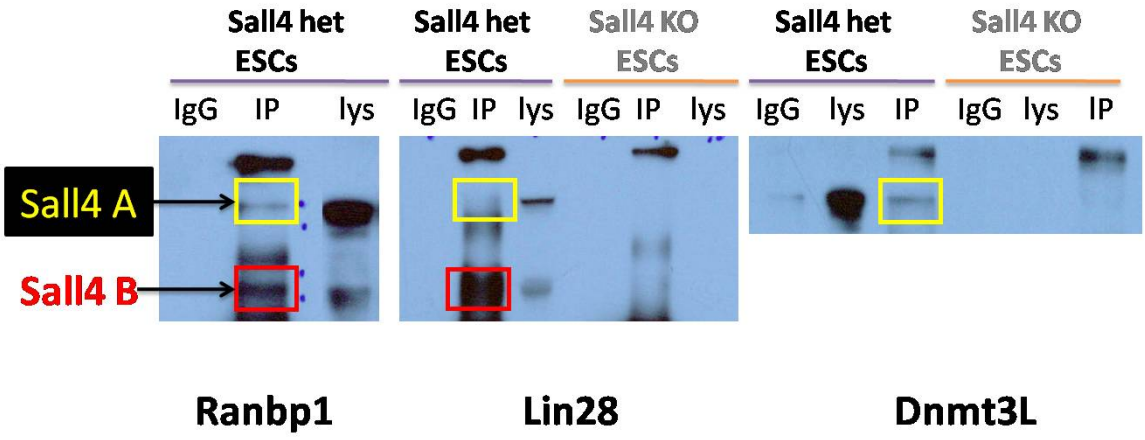


Fig. 3-15 Confirmed interactions of selected Sall4 interacting proteins. Ranbp1, Lin28, and Dnmt3L antibodies were used for co-immunoprecipitation in mESCs. *Sall4*-KO cells are the negative controls. Each experiment was repeated 3 times. The input represents cell lysates.

Protein name	Interact with Sall4 isoform	Possible binding domain
Ranbp1	A and B	1st half of exon 2 or exon 3
Lin28	A and B	1st half of exon 2 or exon 3
Dnmt3L	A	2nd half of exon 2

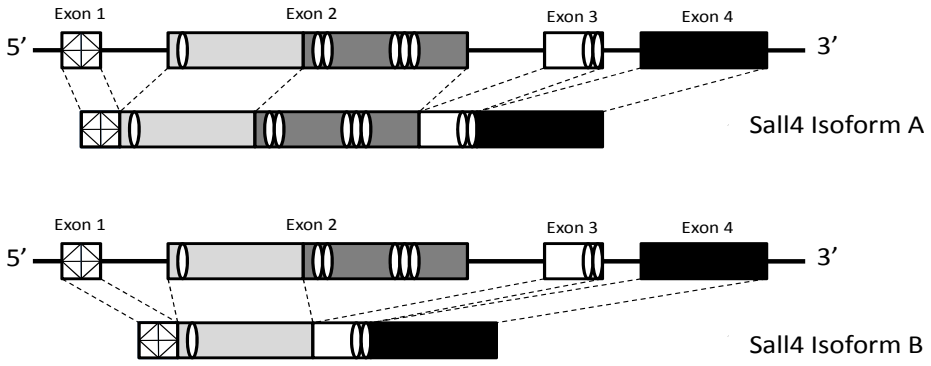


Fig. 3-16 Interacting domain mapping. Based on the Western blot result, the possible domains of Sall4 that may interact with the three tested proteins are shown. This figure was generated by Darwin Yee.

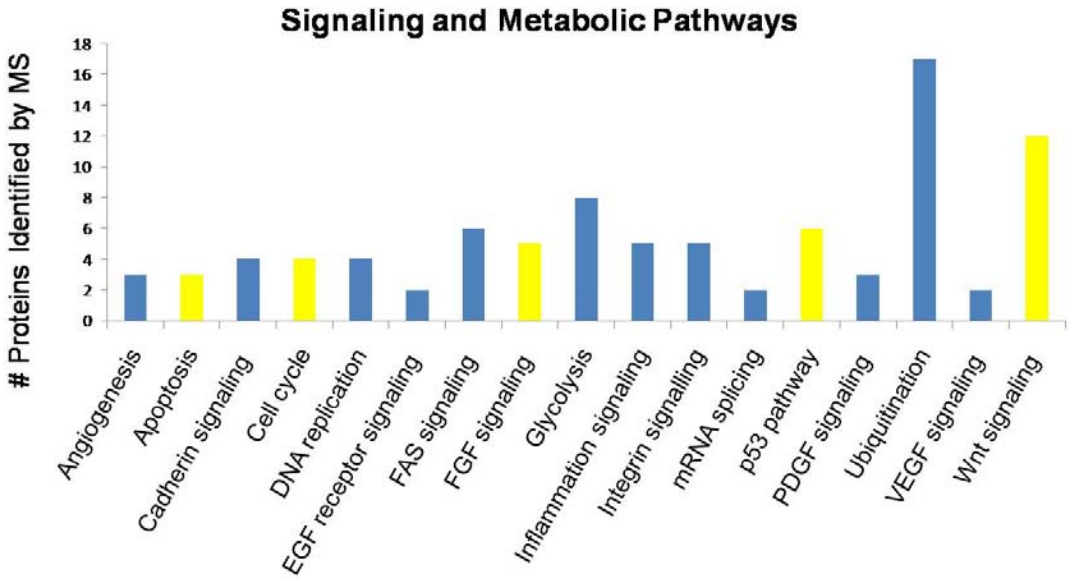


Fig. 3-17 Selected signaling pathways represented by Sall4-interacting proteins. The Panther Pathway Mapping system was used to map 621 Sall4-interacting proteins into total 42 pathways. Interesting ES cell regulatory pathways with relative high coverage are highlighted as yellow.

Table 3-6 Proteins detected from selected signaling pathways.

Pathway	Protein detected
Cell cycle	Psmd3, Psmd11, Psmd12, Psmd13
FGF signaling	Plcg2, Ppp2r1a, Ppp2r1b, Prkci, Ptpn6
p53 pathway	Aifm2, Cdc2a, Gtse1, Hdac1, Mta2, Sin3a
Wnt signaling	Csnk2a1, Ctbp2, Ctnna1, Fat3, Hdac1, Hells, Prkci, Smad2, Smarca4, Smarca5, Smarcb1, Smarcd1

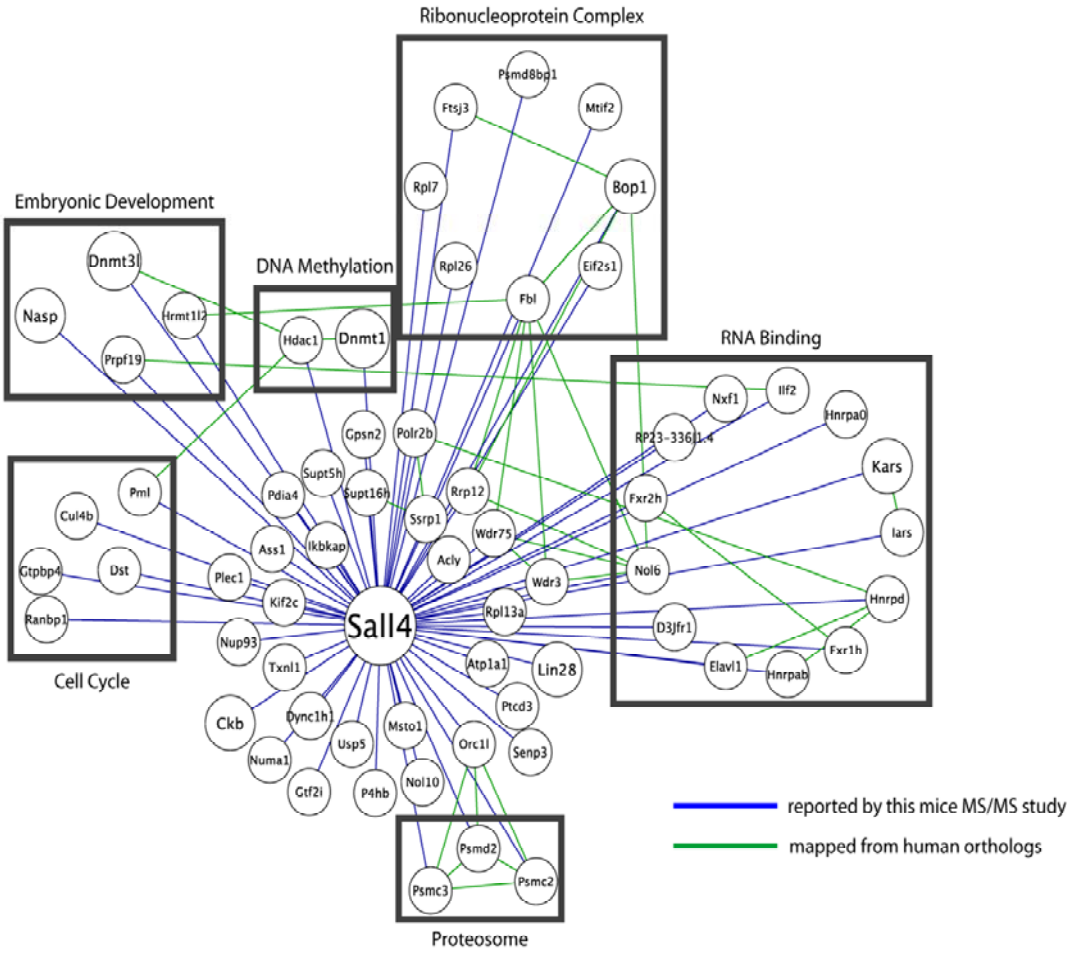


Fig. 3-18 Sall4-Interacting network generated by DAVID.

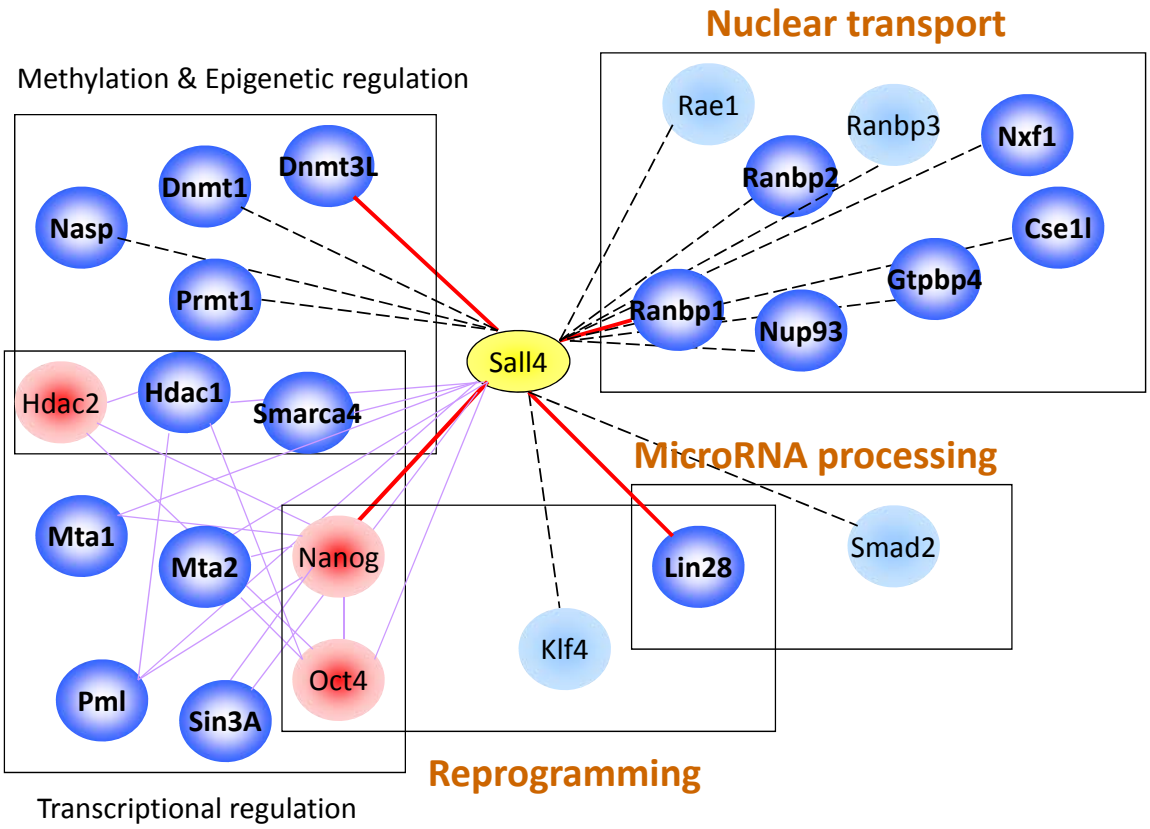


Fig. 3-19 A new Sall4 interacting network. The light and dark blue colored circles represent the proteins that are detected from my Sall4 interactome. Red circles means not. The solid line indicates that the interactions have been confirmed by me or the other published results. The dash line indicates the interactions still remain to be confirmed. Here I highlight 3 possible novel functions of Sall4 including nuclear transport, microRNA processing, and reprogramming.

3.4 Discussion

Stem cell scientists including myself have provided a rich resource of information for future investigation on the function of new proteins that may influence human ES cell fate determination. Recently, two groups have also profiled phosphoproteome of hESCs by using different approaches to stimulate differentiation (14, 31, 75). Linking whole genomic, transcriptomic, proteomic, and phosphoproteomic approaches will facilitate our understanding of the mystery of stem cell and developmental biology. There are two very important issues that we should keep in mind: 1) It has been proven difficult to directly differentiate either mouse or human ESCs to any pure population of cells. The most cheap and common way to differentiate ESCs is to use Retinoic Acid (RA) to derive cells into a neural ectodermal lineage. However, when I repeated these experiments in my laboratory, I observed more than 5 groups of cells distinct by morphology. The RA-treated cells also expressed the other lineage markers such as CD105, a mesodermal precursor marker. We should pay attention to the implications of multiple signaling cascades that direct differentiation along these diverse paths. 2) Phosphorylation can occur in seconds. In our protocol, we cultured cells for days and obtained the undifferentiated and differentiated cell lysates afterward. We were not able to observe the dynamics of the phosphorylation at any particular time point. It is possible that we have already missed the peak of the phosphorylation, so we did not detect other phosphopeptides. Therefore, we have to further evaluate this data to more carefully make hypotheses.

Dnmt3l (DNA (cytosine-5-)-methyltransferase 3-like; ecat7)

CpG methylation is an epigenetic modification that is important for embryogenesis (87). DNA methylation has been demonstrated to be required for mouse development (80-81). In our database, novel phosphorylation of both Dnmt3b and Dnmt3l has been observed. *Dnmt3l* encodes a nuclear protein but not function as a DNA methyltransferase because it does not contain the amino acid residues necessary for the activity. It directly activates the *de novo* DNA methylase activity of both Dnmt3a and Dnmt3b (65). There is a requirement for Dnmt3l in the *de novo* methylation of dispersed repeated sequences in male germ cells prior to meiosis, and a lack of Dnmt3l causes male sterility (11, 79). Dnmt3l has also been reported that mediates transcriptional repression by binding HDAC1 (1). Given the broad impact that Dnmt3l activity can have, it is difficult to predict the phenotype that will be caused by reduction of Dnmt3l activity in ESCs. We observed that amino acids S61, S63, S64 of Dnmt3l are phosphorylated in mESCs cells, but not in EC cells differentiated into neurons following RA treatment. We also confirmed that Dnmt3l and Sall4 formed a protein complex. Phosphorylation of Dnmt3l may determine its level of activity either directly or by affecting its interactions with Dnmt3a, Dnmt3b, HDAC1, or Sall4. It would also be interesting to study whether these phosphorylations can secondarily mediate the epigenetic modification through binding affinity of Dnmt3a and Dnmt3b upon ES cell differentiation.

3.5 Acknowledgements

I would like to thank Dr. Ryuichi Nishinakamura for the gift of the *Sall4* mutant ES cell lines; Darwin Yee for helping me to repeat co-IP experiments and analyze the data; Han-Yu Chuang, Chi-Chiang Tu, and Shann-Ching Chen for computational analysis; Chiung-fang Chang and Chih-Wen Shu for the FACS experiments. The proteome and phosphoproteome profiling data sets were generated and collaborated with Dr. Zhouxin Shen. I would also like to thank Han-Yu Chuang, Chi-Chiang Tu, and Shann-Ching Chen for computational analysis.

Chapter 3, in full, is being prepared for publication. I have obtained the permissions from co-authors, Steve Briggs, Zhouxin Shen, Kiyoshi Tachikawa, and Darwin Yee to reproduce material in this dissertation. The dissertation author will be the primary author of the papers.

Chapter 4

A novel function of Sall4 in modulation of microRNA processing

4.1 Abstract

MicroRNAs (miRNAs) play important roles in development, cell proliferation, and have been implied in human cancer. The biogenesis of mature miRNA is a multi-step process and each of these steps provides opportunities for the regulation of miRNA expression; however, the mechanism is still unclear. Here we report Sall4, a zinc finger protein that has been reported to be required for early embryonic development and embryonic stem cell pluripotency, as a member in the miRNA-processing pathway. Sall4 directly interacts with Lin28 and binds to pri-let-7 miRNAs. Co-immunoprecipitation showed Sall4 serves as a component of Drosha. Knockdown of Sall4 results in a decrease of pri-let-7 transcript in mouse ES cells. Our study reveals an unexpected function that links Sall4 to the machinery of miRNAs maturation; this may underlie key aspects of stem cell research and cancer biology.

4.2 Introduction

Embryonic stem cells are pluripotent cells that can self-renew or differentiate into specific lineages. In order to undergo self-renewal, ESCs require mircoRNA silencing of certain genes (32, 71). MicroRNAs (miRNAs) are small single-stranded noncoding RNA molecules that repress target mRNAs through an antisense mechanism (25). In mammalian system, miRNA biogenesis starts when transcription is initiated by RNA polymerase II to form long primary transcripts (pri-miRNAs). The nuclear RNase III Drosha digests the pri-miRNAs into hairpin-shaped precursor miRNAs (pre-miRNAs) (43). The pre-

miRNAs are then exported out to the cytoplasm and turned into ~22nt miRNA duplexes by Dicer, another RNase III ribonuclease (34). These small RNAs are incorporated into the RNAi effector complex RISC and direct gene silencing of targeted mRNAs. However, the regulatory mechanism of miRNA biogenesis remains largely unknown.

It has been shown that the pluripotency promoting factor Lin28 can regulate let-7 accumulation and processing in a negatively regulated manner (15). From the previously generated Sall4 interactome, we first reported Lin28 forms a protein complex with Sall4 in ES cells. This information initiated us to address the question whether Sall4 is also involved in the microRNA biogenesis. In this chapter, I report that Sall4 regulates let-7 miRNA processing in a Lin28-independent manner. Sall4 not only interacts with Lin28, but also forms a complex with Drosha, leading us to hypothesize that Sall4 may be also involved in the post-transcriptional regulation of miRNA.

4.3 Results

Sall4 directly interacts with Lin28 and binds to pri- let-7 miRNA

We further analyzed the immunopurified Sall4-binding complex from mouse ESCs by mass spectrometry and identified a set of proteins that have been reported to be involved in regulating miRNA mechanism. Surprisingly, Lin28 stood out on the list because Sall4 is a nuclear protein; however Lin28 is mainly localized in the cytoplasm of ESCs. We launched this project from

confirmation of the interactions by co-immunoprecipitation in mouse ESCs (data shown in chapter 4). Overexpression experiments were further performed with Lin28 and Sall4. HA-tagged Sall4 and myc-tagged Lin28 vectors were transfected into HEK293T cells for *in vivo* binding analysis. Transfected HEK293T cells were subjected to anti-HA antibody (Sigma H6908) co-IP and Western blotting with anti-myc antibody (Santa Cruz sc-40). By co-IP RNase A treated HEK293T, we further confirmed that the interaction was not caused by RNA bridging in ESCs (Fig. 4-1A). Recombinant Lin28 and Sall4 proteins purified from *E. coli* were applied to *in vitro* binding assays to show the direct interaction (Fig. 4-1B). It has been shown that Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing (49, 55). Using electrophoretic mobility shift assay (EMSA), we demonstrated that recombinant Sall4 (rSall4) directly interacted with pri-let-7a-1 (Fig. 4-2). The interaction between them *in vitro* was disrupted by pri- let-7 miRNAs, but not pri-mir-16 miRNAs (Fig. 4-3).

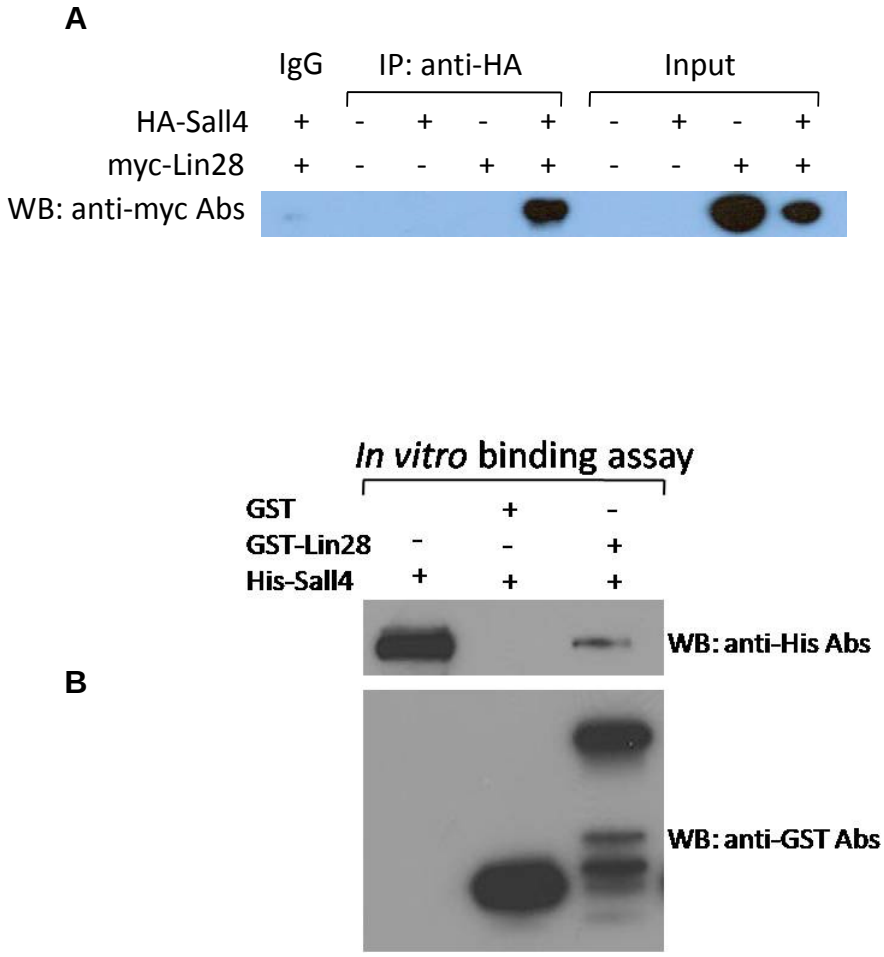


Fig.4-1 Confirmation of the interaction between Sall4 and Lin28 *in vitro* and *in vivo*. (A) Co-IP of HA-Sall4 and Myc-Lin28 in HEK293T. (B) *In vitro* binding assay was carried out to confirm the direct interaction of recombinant Lin28 and Sall4.

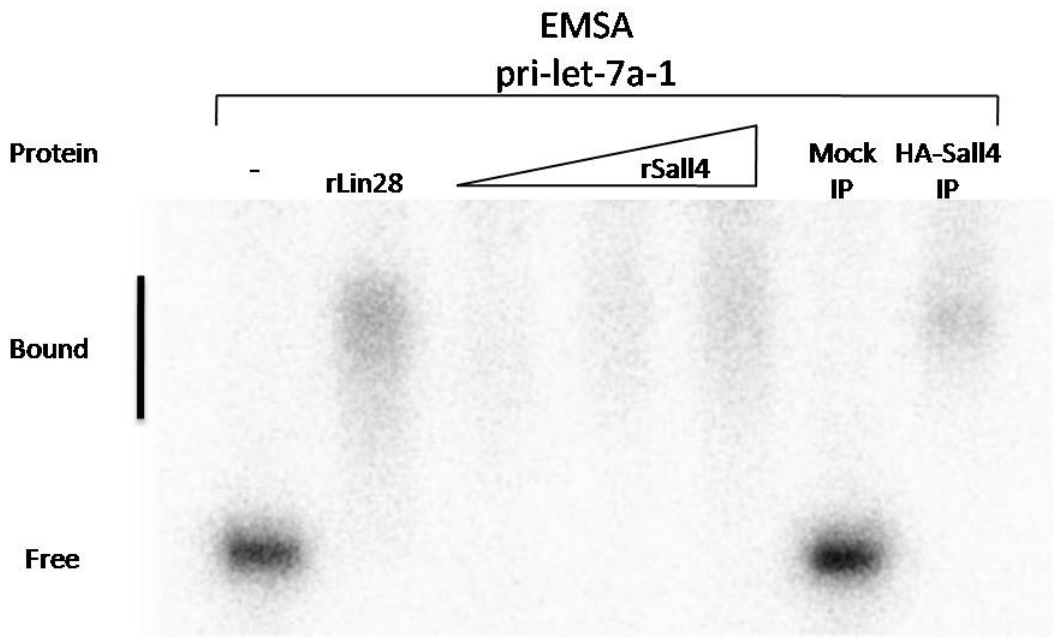


Fig.4-2 Sall4 directly interacts with Lin28 and binds to pri-let-7 microRNA. Electrophoretic Mobility Shift Assay (EMSA) was carried out with 5′ end-labeled synthetic pre-let-7a-1. “Free” represents RNA free of proteins and “bound” represents RNA bound to proteins. The concentrations of rLin28b used are 300nM. The concentrations of rSall4 used are 100nM, 1uM, and 10uM. Co-IP was carried out with HA-Sall4-overexpressed HEK293T cell extracts.

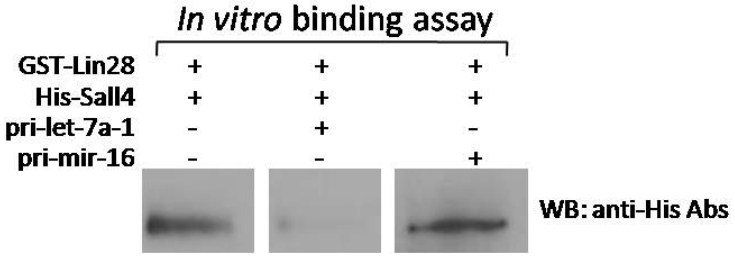


Fig. 4-3 pri-let-7 specifically interferes with the Sall4-Lin28 binding affinity. *In vitro* binding assays were performed as described in methods. Cold RNA was mixed with recombinant proteins and Western blotting was performed.

Sall4 regulates miRNA biogenesis in embryonic cells

We sought to gain insight into the molecular mechanism of Sall4-mediated miRNA biogenesis. Since Sall4 and Lin28 form a protein complex and the majority of Sall4 proteins localizes in the nucleus, we started to examine the mechanism before nuclear export of miRNA. We observed that the endogenous pri-let-7g transcript is downregulated in undifferentiated *Sall4*-null ESCs compared to *Sall4* flox/- ESCs (Fig. 4-4). To test whether Sall4 regulates miRNA processing in a Lin28-dependent manner, we introduced lentiviral shRNA to knockdown Lin28 and Sall4 in mouse ECs. The level of pri-let-7g upon knockdown of Lin28 in P19 cells remained unaltered (Fig. 4-5). This result is consistent with the previous report (77). When we knockdown Sall4, we observed an increased level of pri-let-7g by 30% in P19 cells and 70% in another mouse EC cell line, F9 cells (Fig. 4-6). We observed no decrease in levels of the pluripotency markers Nanog and Oct-4 upon knockdown of Sall4 over the time course (Fig. 4-7). Interestingly, the level of pri-let-7 decreased up to 40% in mESCs (Fig. 4-8). The discrepancy observed here could be due to cell line differences. This could imply that Sall4 may also play different roles in ESCs and carcinomas. Our data suggests that Sall4 may regulate miRNA processing independent from Lin28 and play multiple roles of miRNA biogenesis in different tissues.

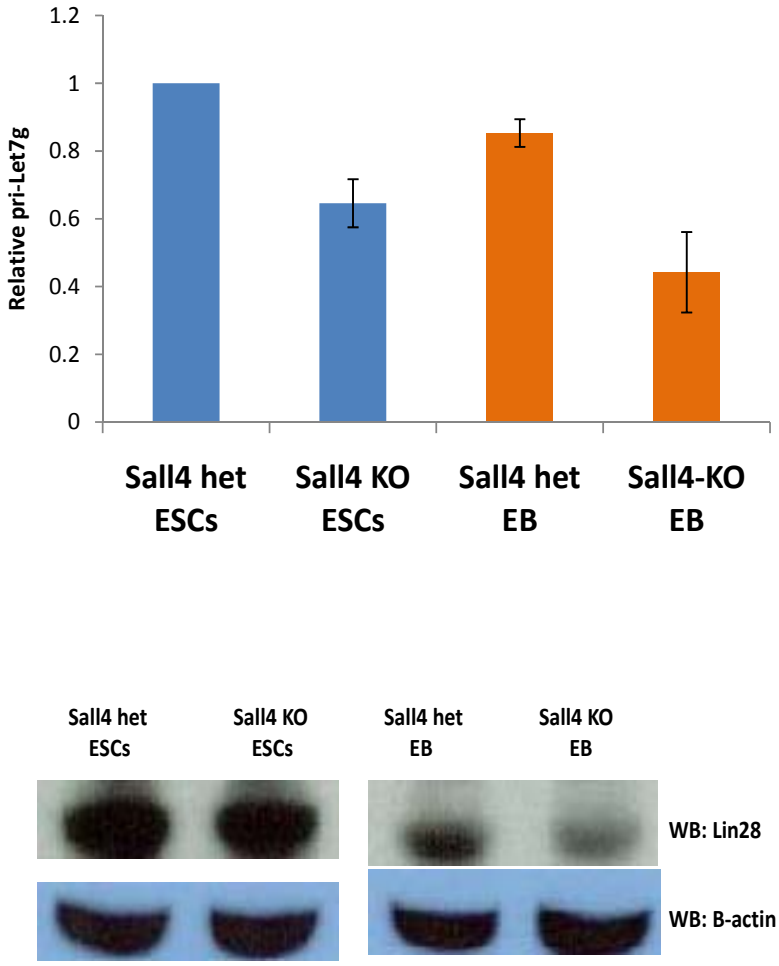


Fig. 4-4 Dynamics of endogenous pri-let miRNAs in *Sall4* mutant ES cells upon differentiation. The level of pri-let-7g is downregulated in *Sall4*-KO ESCs compared to *Sall4* het ESCs, while the protein level of Lin28 remained unchanged. When ESCs grew in suspension and differentiated to Embryoid Bodies (EBs) for 8 days, pri-let-7g levels dropped more than 50% in *Sall4*-KO EBs. Upon differentiation, the pri-let-7g transcript slightly decreased in *Sall4* het EBs. The pri-let-7g transcripts were measured by Reserve Transcription qPCR. Actin served as a control.

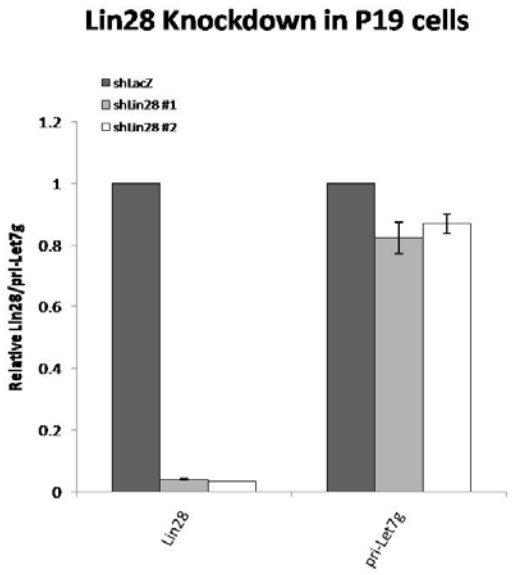


Fig. 4-5 Quantitative PCR showing changes in levels of endogenous pri-let-7 upon sh-Lin28 induction in mouse EC cell line (P19). Error bars represent SEM with *N*=3.

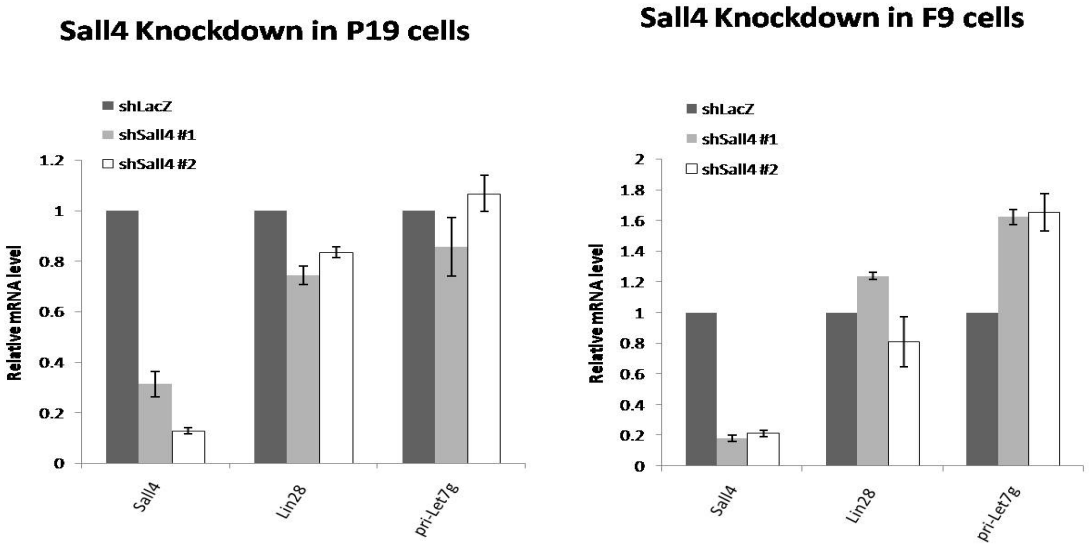


Fig. 4-6 Quantitative PCR showing changes in levels of endogenous pri-let-7 and Lin28 upon sh-Sall4 induction in two lines of mouse ECs. Error bars represent SEM with *N*=3.

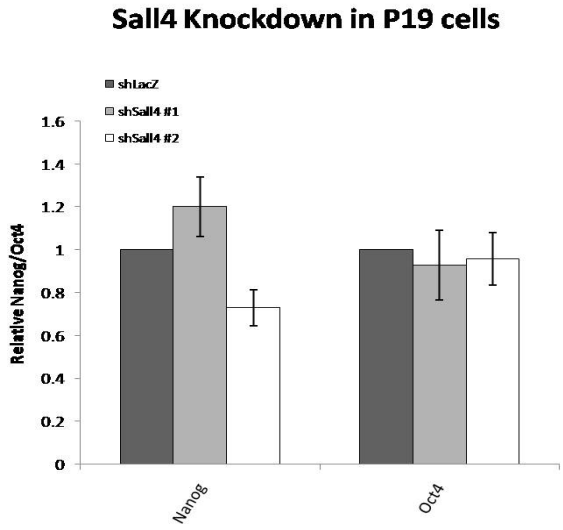


Fig. 4-7 Levels of the pluripotency markers Nanog and Oct-4 in P19 cells infected with either shSall4 or shLacZ. Error bars represent SEM with *N*=3.

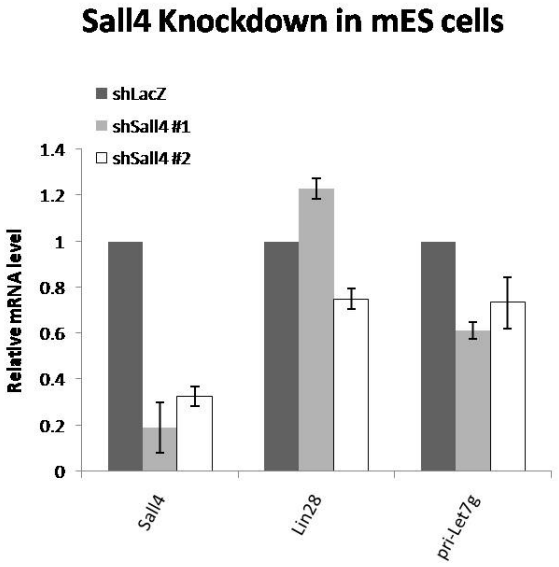


Fig. 4-8 Quantitative PCR showing changes in levels of endogenous pri-let-7 and Lin28 upon sh-Sall4 induction in mouse ESCs. Error bars represent SEM with *N*=3.

To investigate whether Sall4 is also involved in post-transcriptional regulation of miRNA, we quantified the endogenous pre-let-7 and mature let-7 miRNAs in *Sall4* mutant cells. The levels of pre-let-7a-1 miRNAs were unchanged when ESCs were differentiated into embryoid bodies for 8 days. We observed an induced expression of mature let-7 in undifferentiated *Sall4*-null ESCs. The accumulation of mature let-7 dramatically increased upon differentiation (Fig. 4-9). We then ectopically cotransfected Sall4 and pri-let-7a-1 into HEK293T cells, which contains very low levels of endogenous Sall4 protein and let-7 RNA. Northern blot analysis showed increased mature let-7g (Fig. 4-10). Since knockdown of Sall4 causes increased pri-let-7, but not accumulation of precursor or mature let-7 miRNAs in p19 cells, it indicates a loss of function of microprocessor in the nucleus. Co-immunoprecipitation reveals that Sall4 is an integral component of the Drosha protein complex (Fig. 4-11). The detail of mechanism of how Sall4 regulates Drosha processing still remains to be studied. In summary, Sall4 binds specifically to let-7 miRNA precursors and promotes their processing by favoring the association of Drosha.

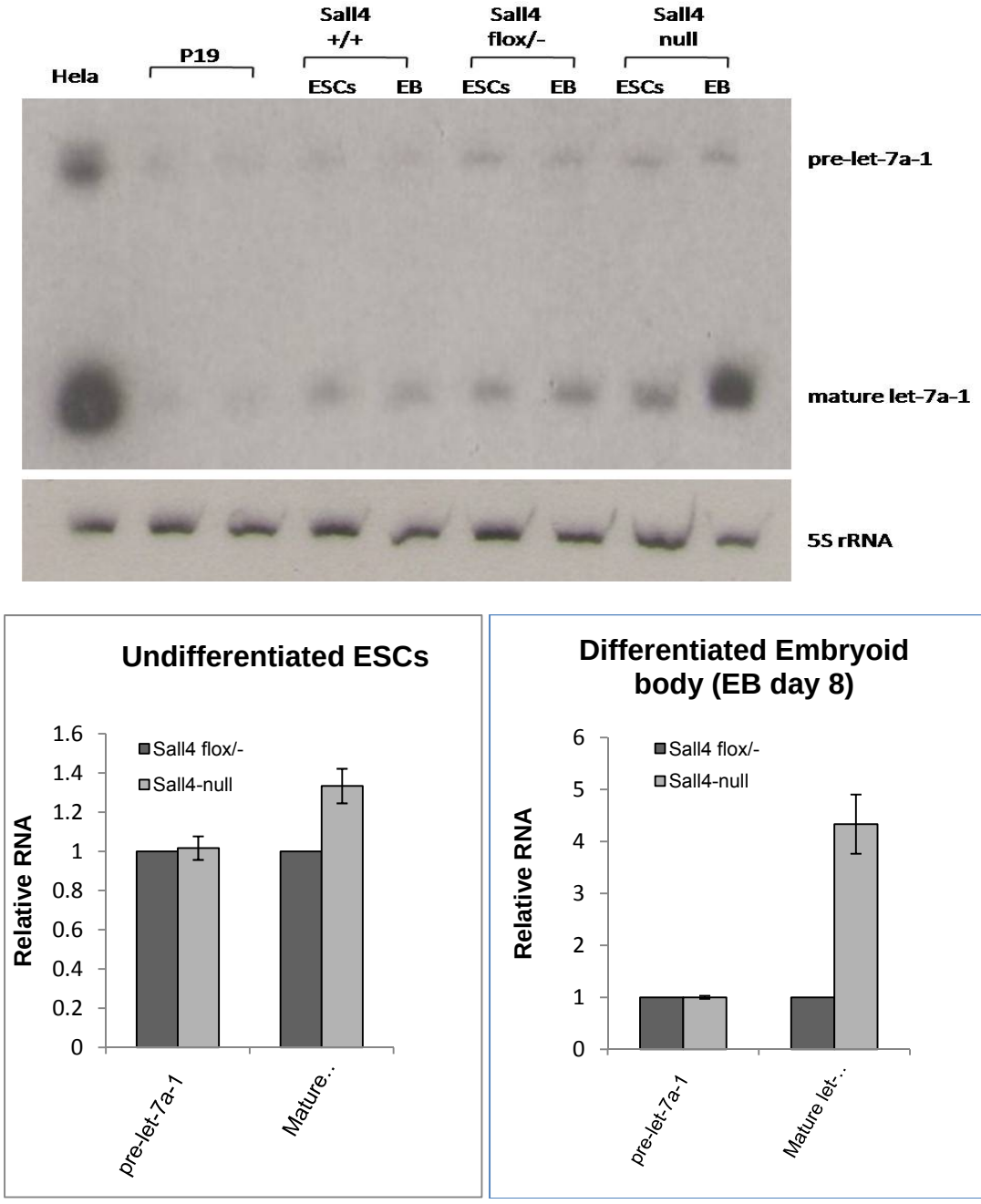


Fig. 4-9 Sall4 is involved in let-7 microRNA biogenesis. Northern blot (NB) showing posttranscriptional induction of mature let-7a-1 during embryoid body (EB) differentiation; mouse ESCs were differentiated to EB for 8 days. Storage phosphor screens exposed to Northern blots were scanned and quantified using ImageQuant software. The endogenous pre- and mature let-7 in undifferentiated ESCs and differentiated EBs were normalized to the level of 5S RNA. The level of pre-let-7 and mature let-7 was quantified based on the signal intensity from NB. Error bars represent SEM with *N*=3.

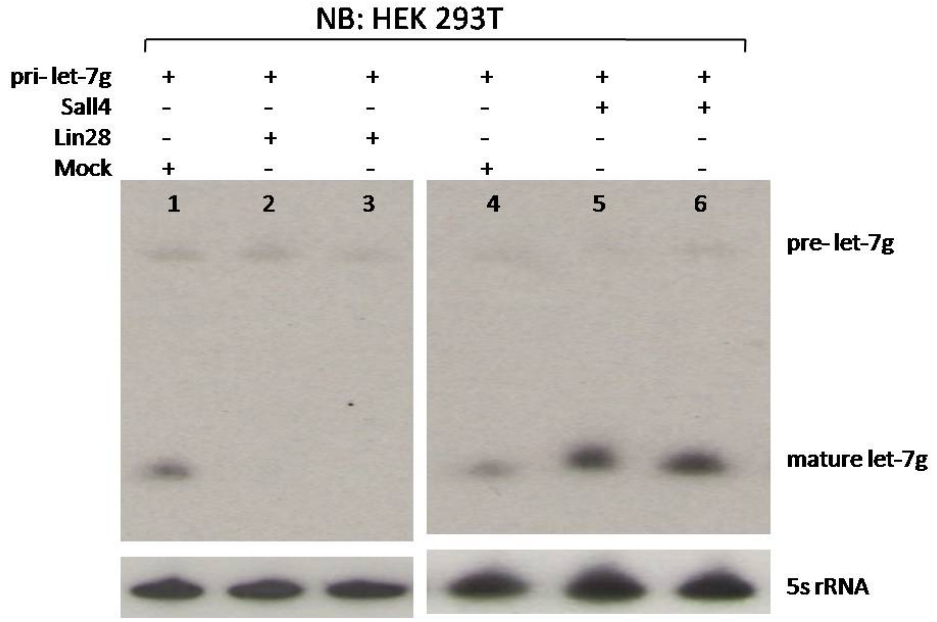


Fig. 4-10 NB analysis of the ectopically expressed pri-let-7g, Lin28, and Sall4 in HEK293T cells. Lane 2 and 3 represent different clones of Lin28. Lane 5 and 6 represent different clones of Sall4.

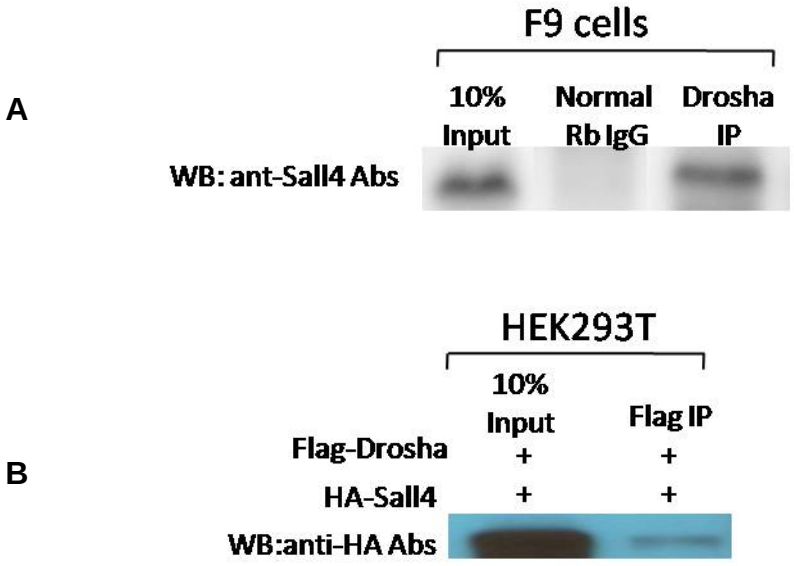


Fig. 4-11 Co-immunoprecipitation showed that Sall4 is an integral component of the Drosha protein complex. (A) Co-IP of endogenous Sall4 and Drosha in F9 cell extracts. (B) Co-overexpression of HA-Sall4 and Flag-Drosha in HEK293T.

4.4 Discussion

Recently, it has been reported that a combination of Lin-28, Oct-3/4, Sox2, and Nanog are sufficient to reprogram human somatic cells into pluripotent stem cells (iPSCs) (88). Two groups have reported that Lin28 interferes with Drosha processing of pri-let-7 in embryonic cells (49, 77). The other two groups showed maturation of let-7 is blocked and indicated Lin28 is involved in Dicer processing (27, 59). It suggests that the restoration of the let-7 processing block is an essential step for iPSCs. Here we introduce Sall4 as a new regulator of let-7 miRNA biogenesis. Sall4 has also been used as part of an enhancer for somatic cell and nuclear reprogramming (68, 82). It would be interesting to study if depletion of let-7 through antisense knockdown can substitute for the requirement for Lin28. While the mechanism is still unclear, by demonstration that KSRP interacts with Drosha and Dicer complexes, indicates KSRP is essential for the efficient recruitment of miRNA precursors to these proteins (74). It implies that protein-protein interactions with other components of the biogenesis machinery may be required.

The exact mechanism of how Lin-28 inhibits Drosha processing is unknown. In this study, we also showed Sall4 is associated with Drosha complex. We also have shown that Sall4 regulates let-7 miRNAs and directly interacts with Lin28; however, how Sall4 coordinates with Lin28 or Drosha to mediate miRNA biogenesis remains to be further studied. One possible mechanism is that Sall4 sequesters Lin28 from the cytoplasm to the nucleus. The high binding affinity to

let-7 may simply block access of Drosha complex to the pri-miRNAs. Lin28/let-7 complex is then exported out of the nucleus and shuttled to P-bodies for degradation. However, we cannot rule out the possibility that Sall4, alone or in concert with the other factors, such as KRPS or p53, may block other pri-miRNAs in different physiological contexts. Altogether, our findings uncovered a novel regulator, Sall4, in miRNA processing. Future investigations of Sall4 may show promise to reveal how a great portion of pri-miRNAs are regulated and how miRNA expression patterns are important determinants of stem cell fate and oncogenesis.

4.5 Acknowledgements

I thank the members in Dr. Narry Kim's and Dr. Amy Pasquinelli's laboratory for sharing reagents, technical support, and helpful discussion. I also thank Dr. Yang Xu for his critical comments on this study and his advice to launch this project. And finally, I would like to thank Dr. Wen-Yuan Hu for helping me with the experiment with lenti-viral infection and teaching me how to design shRNA to knockdown gene expression.

References

1. **Aapola U, Liiv I, and Peterson P**. Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity. *Nucl Acids Res* 30: 3602-3608, 2002.

2. **Al-Baradie R, Yamada K, St Hilaire C, Chan WM, Andrews C, McIntosh N, Nakano M, Martonyi EJ, Raymond WR, Okumura S, Okihiro MM, and Engle EC**. Duane radial ray syndrome (Okihiro syndrome) maps to 20q13 and results from mutations in SALL4, a new member of the SAL family. *Am J Hum Genet* 71: 1195-1199, 2002.

3. **Ang LP, and Tan DT**. Ocular surface stem cells and disease: current concepts and clinical applications. *Ann Acad Med Singapore* 33: 576-580, 2004.

4. **Artavanis-Tsakonas S, Rand MD, and Lake RJ**. Notch signaling: cell fate control and signal integration in development. *Science* 284: 770-776, 1999.

5. **Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, and Lovell-Badge R**. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 17: 126-140, 2003.

6. **Beddington RS, and Robertson EJ**. Anterior patterning in mouse. *Trends Genet* 14: 277-284, 1998.

7. **Bischof LJ, Kao CY, Los FC, Gonzalez MR, Shen Z, Briggs SP, van der Goot FG, and Aroian RV**. Activation of the unfolded protein response is required for defenses against bacterial pore-forming toxin in vivo. *PLoS Pathog* 4: e1000176, 2008.

8. **Bohm J, Sustmann C, Wilhelm C, and Kohlhase J**. SALL4 is directly activated by TCF/LEF in the canonical Wnt signaling pathway. *Biochem Biophys Res Commun* 348: 898-907, 2006.

9. **Borozdin W, Boehm D, Leipoldt M, Wilhelm C, Reardon W, Clayton-Smith J, Becker K, Muhlendyck H, Winter R, Giray O, Silan F, and Kohlhase J**. SALL4 deletions are a common cause of Okihiro and acro-renal-ocular syndromes and confirm haploinsufficiency as the pathogenic mechanism. *J Med Genet* 41: e113, 2004.

10. **Borozdin W, Wright MJ, Hennekam RC, Hannibal MC, Crow YJ, Neumann TE, and Kohlhase J**. Novel mutations in the gene SALL4 provide further evidence for acro-renal-ocular and Okihiro syndromes being allelic entities, and extend the phenotypic spectrum. *J Med Genet* 41: e102, 2004.

11. **Bourc'his D, and Bestor TH**. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431: 96-99, 2004.

12. **Brennan C, Momota H, Hambardzumyan D, Ozawa T, Tandon A, Pedraza A, and Holland E**. Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations. *PLoS One* 4: e7752, 2009.

13. **Brennan SC, and Conigrave AD**. Regulation of cellular signal transduction pathways by the extracellular calcium-sensing receptor. *Curr Pharm Biotechnol* 10: 270-281, 2009.

14. **Brill LM, Xiong W, Lee KB, Ficarro SB, Crain A, Xu Y, Terskikh A, Snyder EY, and Ding S**. Phosphoproteomic analysis of human embryonic stem cells. *Cell Stem Cell* 5: 204-213, 2009.

15. **Bussing I, Slack FJ, and Grosshans H**. let-7 microRNAs in development, stem cells and cancer. *Trends Mol Med* 14: 400-409, 2008.

16. **Castellana NE, Payne SH, Shen Z, Stanke M, Bafna V, and Briggs SP**. Discovery and revision of Arabidopsis genes by proteogenomics. *Proc Natl Acad Sci U S A* 105: 21034-21038, 2008.

17. **Chambers I**. The molecular basis of pluripotency in mouse embryonic stem cells. *Cloning Stem Cells* 6: 386-391, 2004.

18. **Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, and Smith A**. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113: 643-655, 2003.

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

20. **Chew JL, Loh YH, Zhang W, Chen X, Tam WL, Yeap LS, Li P, Ang YS, Lim B, Robson P, and Ng HH**. Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* 25: 6031-6046, 2005.

21. **de la Fuente van Bentem S, Mentzen WI, de la Fuente A, and Hirt H**. Towards functional phosphoproteomics by mapping differential phosphorylation events in signaling networks. *Proteomics* 8: 4453-4465, 2008.

22. **Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, and Lempicki RA**. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4: P3, 2003.

23. **Elling U, Klasen C, Eisenberger T, Anlag K, and Treier M**. Murine inner cell mass-derived lineages depend on Sall4 function. *Proc Natl Acad Sci U S A* 103: 16319-16324, 2006.

24. **Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y, Vasilescu J, Abu-Farha M, Lambert JP, Duewel HS, Stewart, II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, and Figeys D**. Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* 3: 89, 2007.

25. **Filipowicz W, Bhattacharyya SN, and Sonenberg N**. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 9: 102-114, 2008.

26. **Fujikura J, Yamato E, Yonemura S, Hosoda K, Masui S, Nakao K, Miyazaki Ji J, and Niwa H**. Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev* 16: 784-789, 2002.

27. **Heo I, Joo C, Cho J, Ha M, Han J, and Kim VN**. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* 32: 276-284, 2008.

28. **Heo I, Joo C, Kim YK, Ha M, Yoon MJ, Cho J, Yeom KH, Han J, and Kim VN**. TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* 138: 696-708, 2009.

29. **Holtz J, and Pasquinelli AE**. Uncoupling of lin-14 mRNA and protein repression by nutrient deprivation in Caenorhabditis elegans. *RNA* 15: 400-405, 2009.

30. **Hosack DA, Dennis G, Jr., Sherman BT, Lane HC, and Lempicki RA**. Identifying biological themes within lists of genes with EASE. *Genome Biol* 4: R70, 2003.

31. **Hutchins AP, and Robson P**. Unraveling the human embryonic stem cell phosphoproteome. *Cell Stem Cell* 5: 126-128, 2009.

32. **Ivey KN, Muth A, Arnold J, King FW, Yeh RF, Fish JE, Hsiao EC, Schwartz RJ, Conklin BR, Bernstein HS, and Srivastava D**. MicroRNA regulation of cell lineages in mouse and human embryonic stem cells. *Cell Stem Cell* 2: 219-229, 2008.

33. **Keller G**. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* 19: 1129-1155, 2005.

34. **Kim VN**. MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* 6: 376-385, 2005.

35. **Kohlhase J, Heinrich M, Schubert L, Liebers M, Kispert A, Laccone F, Turnpenny P, Winter RM, and Reardon W**. Okihiro syndrome is caused by SALL4 mutations. *Hum Mol Genet* 11: 2979-2987, 2002.

36. **Kohlhase J, Schubert L, Liebers M, Rauch A, Becker K, Mohammed SN, Newbury-Ecob R, and Reardon W**. Mutations at the SALL4 locus on chromosome 20 result in a range of clinically overlapping phenotypes, including Okihiro syndrome, Holt-Oram syndrome, acro-renal-ocular syndrome, and patients previously reported to represent thalidomide embryopathy. *J Med Genet* 40: 473-478, 2003.

37. **Kohlhase J, Schuh R, Dowe G, Kuhnlein RP, Jackle H, Schroeder B, Schulz-Schaeffer W, Kretzschmar HA, Kohler A, Muller U, Raab-Vetter M, Burkhardt E, Engel W, and Stick R**. Isolation, characterization, and organ-specific expression of two novel human zinc finger genes related to the Drosophila gene spalt. *Genomics* 38: 291-298, 1996.

38. **Koshiba-Takeuchi K, Takeuchi JK, Arruda EP, Kathiriya IS, Mo R, Hui C-c, Srivastava D, and Bruneau BG**. Cooperative and antagonistic interactions between Sall4 and Tbx5 pattern the mouse limb and heart. *Nat Genet* 38: 175-183, 2006.

39. **Koshiba-Takeuchi K, Takeuchi JK, Arruda EP, Kathiriya IS, Mo R, Hui CC, Srivastava D, and Bruneau BG**. Cooperative and antagonistic interactions between Sall4 and Tbx5 pattern the mouse limb and heart. *Nat Genet* 38: 175-183, 2006.

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

41. **Kuhnlein RP, Frommer G, Friedrich M, Gonzalez-Gaitan M, Weber A, Wagner-Bernholz JF, Gehring WJ, Jackle H, and Schuh R**. spalt encodes an evolutionarily conserved zinc finger protein of novel structure which provides homeotic gene function in the head and tail region of the Drosophila embryo. *EMBO J* 13: 168-179, 1994.

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

43. **Lee Y, Jeon K, Lee JT, Kim S, and Kim VN**. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* 21: 4663-4670, 2002.

44. **Ma Y, Cui W, Yang J, Qu J, Di C, Amin HM, Lai R, Ritz J, Krause DS, and Chai L**. SALL4, a novel oncogene, is constitutively expressed in human acute myeloid leukemia (AML) and induces AML in transgenic mice. *Blood* 108: 2726-2735, 2006.

45. **Mi H, Guo N, Kejariwal A, and Thomas PD**. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res* 35: D247-252, 2007.

46. **Mi H, and Thomas P**. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol* 563: 123-140, 2009.

47. **Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, and Yamanaka S**. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 113: 631-642, 2003.

48. **Nagano K, Taoka M, Yamauchi Y, Itagaki C, Shinkawa T, Nunomura K, Okamura N, Takahashi N, Izumi T, and Isobe T**. Large-scale identification of proteins expressed in mouse embryonic stem cells. *Proteomics* 5: 1346-1361, 2005.

49. **Newman MA, Thomson JM, and Hammond SM**. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* 14: 1539-1549, 2008.

50. **Niwa H, Miyazaki J, and Smith AG**. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 24: 372-376, 2000.

51. **Orkin SH**. Chipping away at the embryonic stem cell network. *Cell* 122: 828-830, 2005.

52. **Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, and Daley GQ**. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 451: 141-146, 2008.

53. **Pasquinelli AE, McCoy A, Jimenez E, Salo E, Ruvkun G, Martindale MQ, and Baguna J**. Expression of the 22 nucleotide let-7 heterochronic RNA throughout the Metazoa: a role in life history evolution? *Evol Dev* 5: 372-378, 2003.

54. **Pieroni E, de la Fuente van Bentem S, Mancosu G, Capobianco E, Hirt H, and de la Fuente A**. Protein networking: insights into global functional organization of proteomes. *Proteomics* 8: 799-816, 2008.

55. **Piskounova E, Viswanathan SR, Janas M, LaPierre RJ, Daley GQ, Sliz P, and Gregory RI**. Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J Biol Chem* 283: 21310-21314, 2008.

56. **Ralston A, and Rossant J**. Genetic regulation of stem cell origins in the mouse embryo. *Clin Genet* 68: 106-112, 2005.

57. **Rehfeldt F, Engler AJ, Eckhardt A, Ahmed F, and Discher DE**. Cell responses to the mechanochemical microenvironment--implications for regenerative medicine and drug delivery. *Adv Drug Deliv Rev* 59: 1329-1339, 2007.

58. **Rossant J**. Lineage development and polar asymmetries in the peri-implantation mouse blastocyst. *Semin Cell Dev Biol* 15: 573-581, 2004.

59. **Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, and Wulczyn FG**. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol* 10: 987-993, 2008.

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

61. **Sawyer TK, Shakespeare WC, Wang Y, Sundaramoorthi R, Huang WS, Metcalf CA, 3rd, Thomas M, Lawrence BM, Rozamus L, Noehre J, Zhu X, Narula S, Bohacek RS, Weigele M, and Dalgarno DC**. Protein phosphorylation and signal transduction modulation: chemistry perspectives for small-molecule drug discovery. *Med Chem* 1: 293-319, 2005.

62. **Shuai X, Zhou D, Shen T, Wu Y, Zhang J, Wang X, and Li Q**. Overexpression of the novel oncogene SALL4 and activation of the Wnt/beta-catenin pathway in myelodysplastic syndromes. *Cancer Genet Cytogenet* 194: 119-124, 2009.

63. **Stojkovic M, Lako M, Stojkovic P, Stewart R, Przyborski S, Armstrong L, Evans J, Herbert M, Hyslop L, Ahmad S, Murdoch A, and Strachan T**. Derivation of human embryonic stem cells from day-8 blastocysts recovered after three-step in vitro culture. *Stem Cells* 22: 790-797, 2004.

64. **Stojkovic M, Lako M, Strachan T, and Murdoch A**. Derivation, growth and applications of human embryonic stem cells. *Reproduction* 128: 259-267, 2004.

65. **Suetake I, Shinozaki F, Miyagawa J, Takeshima H, and Tajima S**. DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. *J Biol Chem* 279: 27816-27823, 2004.

66. **Swaney DL, Wenger CD, Thomson JA, and Coon JJ**. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* 106: 995-1000, 2009.

67. **Takahashi K, Okita K, Nakagawa M, and Yamanaka S**. Induction of pluripotent stem cells from fibroblast cultures. *Nat Protoc* 2: 3081-3089, 2007.

68. **Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, and Yamanaka S**. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131: 861-872, 2007.

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

70. **Tanner S, Shen Z, Ng J, Florea L, Guigo R, Briggs SP, and Bafna V**. Improving gene annotation using peptide mass spectrometry. *Genome Res* 17: 231-239, 2007.

71. **Tay Y, Zhang J, Thomson AM, Lim B, and Rigoutsos I**. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455: 1124-1128, 2008.

72. **Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, and Narechania A**. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 13: 2129-2141, 2003.

73. **Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, Ladunga I, Ulitsky-Lazareva B, Muruganujan A, Rabkin S, Vandergriff JA, and Doremieux O**. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res* 31: 334-341, 2003.

74. **Trabucchi M, Briata P, Garcia-Mayoral M, Haase AD, Filipowicz W, Ramos A, Gherzi R, and Rosenfeld MG**. The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* 459: 1010-1014, 2009.

75. **Van Hoof D, Munoz J, Braam SR, Pinkse MW, Linding R, Heck AJ, Mummery CL, and Krijgsveld J**. Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell* 5: 214-226, 2009.

76. **Van Hoof D, Passier R, Ward-Van Oostwaard D, Pinkse MW, Heck AJ, Mummery CL, and Krijgsveld J**. A quest for human and mouse embryonic stem cell-specific proteins. *Mol Cell Proteomics* 5: 1261-1273, 2006.

77. **Viswanathan SR, Daley GQ, and Gregory RI**. Selective blockade of microRNA processing by Lin28. *Science* 320: 97-100, 2008.

78. **Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW, and Orkin SH**. A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444: 364-368, 2006.

79. **Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, Craig J, Harrison DK, Aung H, Phutikanit N, Lyle R, Meachem SJ, Antonarakis SE, de Kretser DM, Hedger MP, Peterson P, Carroll BJ, and Scott HS**. Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc Natl Acad Sci U S A* 102: 4068-4073, 2005.

80. **Weng A, Engler P, and Storb U**. The bulk chromatin structure of a murine transgene does not vary with its transcriptional or DNA methylation status. *Mol Cell Biol* 15: 572-579, 1995.

81. **Weng A, Magnuson T, and Storb U**. Strain-specific transgene methylation occurs early in mouse development and can be recapitulated in embryonic stem cells. *Development* 121: 2853-2859, 1995.

82. **Wong CC, Gaspar-Maia A, Ramalho-Santos M, and Reijo Pera RA**. High-efficiency stem cell fusion-mediated assay reveals Sall4 as an enhancer of reprogramming. *PLoS One* 3: e1955, 2008.

83. **Wu Q, Chen X, Zhang J, Loh YH, Low TY, Zhang W, Sze SK, Lim B, and Ng HH**. Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem* 281: 24090-24094, 2006.

84. **Xu C, Inokuma MS, Denham J, Golds K, Kundu P, Gold JD, and Carpenter MK**. Feeder-free growth of undifferentiated human embryonic stem cells. *Nat Biotechnol* 19: 971-974, 2001.

85. **Yamanaka S**. [Molecular mechanisms underlying pluripotency of embryonic stem cells]. *Seikagaku* 78: 27-33, 2006.

86. **Yamanaka Y, Ralston A, Stephenson RO, and Rossant J**. Cell and molecular regulation of the mouse blastocyst. *Dev Dyn* 235: 2301-2314, 2006.

87. **Young LE, and Beaujean N**. DNA methylation in the preimplantation embryo: the differing stories of the mouse and sheep. *Anim Reprod Sci* 82-83: 61-78, 2004.

88. **Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin, II, and Thomson JA**. Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318: 1917-1920, 2007.

89. **Zhang J, Tam WL, Tong GQ, Wu Q, Chan HY, Soh BS, Lou Y, Yang J, Ma Y, Chai L, Ng HH, Lufkin T, Robson P, and Lim B**. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol* 8: 1114-1123, 2006.

90. **Zhang L, Rayner S, Katoku-Kikyo N, Romanova L, and Kikyo N**. Successful co-immunoprecipitation of Oct4 and Nanog using cross-linking. *Biochem Biophys Res Commun* 361: 611-614, 2007.

http://www.piercenet.com/Objects/View.cfm?type=File&ID=0544

http://david.abcc.ncifcrf.gov/content.jsp?file=functional_annotation.html#DATA_INPUT

http://stemcells.nih.gov/info/scireport/execSum.asp