โปรตีนที่มีปฏิสัมพันธ์กับออคโฟร์ในเซลล์ต้นกำเนิดตัวอ่อนมนุษย์ และการผลิตเบสิคไฟโบรบลาสต์โกรธแฟคเตอร์ของมนุษย์

นายสุเมธ อิ่มสุนทรรักษา

มหาวิทยาลัยเทคโนโลยีสุรนารี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีชีวภาพ
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2558

IDENTIFICATION OF OCT4 ASSOCIATED PROTEINS IN HUMAN EMBRYONIC STEM CELLS AND PRODUCTION OF HUMAN BASIC FIBROBLAST GROWTH FACTOR

Sumeth Imsoonthornruksa

มหาวิทยาลัยเทคโนโลยีสุรนารี

A Thesis Submitted in Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Biotechnology

Suranaree University of Technology

Academic Year 2015

IDENTIFICATION OF OCT4 ASSOCIATED PROTEINS IN HUMAN EMBRYONIC STEM CELLS AND PRODUCTION OF HUMAN BASIC FIBROBLAST GROWTH FACTOR

Suranaree University of Technology has approved this thesis submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

_____
(Assoc. Prof. Dr. Rangsun Parnpai)
Chairperson

_____
(Assoc. Prof. Dr. Mariena Ketudat-Cairns)
Member (Thesis Advisor)

_____
(Dr. Ruttachuk Rungsiwiwut)
Member

_____
(Dr. Chanchao Lorthongpanich)
Member

_____
(Dr. Chuti Laowtammathron)
Member

_____
(Dr. Weerapong Prasongchean)
Member

_____
(Prof. Dr. Sukit Limpijumnong)
Vice Rector for Academic Affairs and Innovation

_____
(Prof. Dr. Neung Teaumroong)
Dean of Institute of Agricultural Technology

มหาวิทยาลัยเทคโนโลยีสุรนารี

สุเมธ อิ่มสุนทรรักษา : โปรตีนที่มีปฏิสัมพันธ์กับออคท์โฟร์ในเซลล์ต้นกำเนิดตัวอ่อนมนุษย์และการผลิตเบสิคไฟโบรบลาสต์โกรธแฟคเตอร์ของมนุษย์ (IDENTIFICATION OF OCT4 ASSOCIATED PROTEINS IN HUMAN EMBRYONIC STEM CELLS AND PRODUCTION OF HUMAN BASIC FIBROBLAST GROWTH FACTOR) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.มารินา เกตุทัต-คาร์นส์, 156 หน้า.

เทคโนโลยีชีวภาพสามารถนำมาเป็นเครื่องมือในการศึกษาชีววิทยาเซลล์ต้นกำเนิด การศึกษานี้ได้ทำการสำรวจโปรตีนที่มีปฏิสัมพันธ์กับออคท์โฟร์ (OCT4) ในเซลล์ต้นกำเนิดตัวอ่อนมนุษย์ (hESCs) โดยใช้เทคโนโลยีทาเลนต์ (TALENs) เพื่อสร้าง hESCs ที่สามารถสร้างโปรตีน OCT4 ที่เชื่อมต่อกับโปรตีน FLAG และไบโอติน (FLAG-Biotin-tagged OCT4) จากนั้นใช้เทคนิคการทำให้บริสุทธิ์แบบสัมพรรคภาพ (Affinity purification) ร่วมกับแมสสเปกโตรเมทรี (mass spectrometry) เพื่อระบุโปรตีนที่มีปฏิสัมพันธ์กับ OCT4 จากผลการทดลองพบว่า OCT4 มีปฏิสัมพันธ์กับโปรตีนที่เกี่ยวข้องกับกระบวนการทรานสคริปชั่น ทรานสเลชั่น การเปลี่ยนแปลงโครมาติน การซ่อมแซมดีเอ็นเอ การตัดแต่งอาร์เอ็นเอ การย้ายหมู่กลูโคส (ไกลโคซิเลส) และโปรตีนที่ควบคุมกระบวนการเอพิเจเนติก เช่น NuRD, SWI/SNF และ FACT complexes นอกจากนี้ได้ทำการยับยั้งการแสดงออกของโปรตีนบางตัว (BEND3, FUS, RBM25, WDR82 และ SFPQ) เพื่อทดสอบว่าโปรตีนดังกล่าวมีผลต่อ hESCs หรือไม่ จากการทดลองพบว่าหลังจากการยับยั้งการแสดงออกของโปรตีน SFPQ ส่งผลให้เซลล์ตาย โดยอาจจะเป็นไปได้ว่าปฏิสัมพันธ์ระหว่าง SFPQ กับ OCT4 มีความเกี่ยวพันธ์กับกลไกที่ควบคุมกระบวนการทรานสคริปชั่น, การซ่อมแซมดีเอ็นเอ และการบำรุงรักษาโครโมโซมซึ่งมีบทบาทสำคัญต่อความเป็น hESCs การทดลองนี้เผยให้เห็นถึงโปรตีนต่างๆที่มีปฏิสัมพันธ์กับ OCT4 ซึ่งเป็นองค์ความรู้พื้นฐานที่จะใช้ในการศึกษากลไกการทำงานของ OCT4 เพื่อให้เข้าใจเครื่อข่ายที่เกี่ยวข้องกับคุณสมบัติพลูริโพเทนท์ของ hESCs โดยในอนาคตต้องมีการศึกษาว่าปฏิสัมพันธ์ของโปรตีนต่างๆ กับ OCT4 มีบทบาท และหน้าที่อย่างไร

การศึกษานี้ยังได้แสดงวิธีการที่ง่ายและมีประสิทธิภาพในการผลิตเบสิคไฟโบรบลาสต์โกรธแฟคเตอร์ของมนุษย์ (hbFGF) โดยได้มีการทดสอบปัจจัยต่างๆ ที่มีอิทธิพลต่อการเพิ่มประสิทธิภาพในการผลิต hbFGF เช่น ฟิวชั่นโปรตีนที่ติดแท็ก ชนิดของเซลล์เจ้าบ้าน (BL21(DE3) หรือ ArcticExpress®RIL) ความเข้มข้นของ IPTG ระยะเวลาที่ใช้กระตุ้นให้เกิดการสร้าง และ วิธีการทำให้

บริสุทธิ์ พบว่าภายใต้สภาวะที่ดีที่สุดสามารถผลิต hbFGF ได้ 60-80 มิลลิกรัมโปรตีน ต่ออาหารเลี้ยงเชื้อ 1 ลิตร และ hbFGF บริสุทธิ์ที่ได้มีคุณสมบัติทางชีวภาพหลังจากทดสอบกับเซลล์ NIH3T3 hESCs และเซลล์ต้นกำเนิดจากการเหนี่ยวนำของมนุษย์ (iPSCs) อีกทั้งยังพบว่า ฟิวชั่นโปรตีนที่ติดแท็กอยู่กับ hbFGF (6xHis-hbFGF และ Trx-6xHis-hbFGF) ไม่มีผลต่อคุณสมบัติทางชีวภาพ และ 6xHis-hbFGF และ Trx-6xHis-hbFGF บริสุทธิ์ที่ได้สามารถนำมาใช้เลี้ยงเซลล์ hESCs และ iPSCs โดยเซลล์ดังกล่าวหลังจากการเลี้ยงหลายรุ่น (passage) ยังคงรูปร่าง คุณสมบัติพลูริโพเทนท์ และการแสดงออกของยีนบางยีนที่ศึกษา ไม่ต่างจากเซลล์ที่เลี้ยงด้วย hbFGF ที่นำเข้าจากต่างประเทศ จากการนำ hbFGF ที่ผลิตด้วยวิธีการผลิตที่ง่าย และมีประสิทธิภาพนี้มาใช้ในการเลี้ยงเซลล์ต้นกำเนิดมนุษย์สามารถลดต้นทุนในการเตรียมอาหารเลี้ยงเซลล์ได้ประมาณ 30-60 เปอร์เซ็นต์ จึงมีความเป็นได้ที่จะสามารถนำกระบวนการนี้ไปประยุกต์ใช้ในการผลิตโปรตีนไซโตไคน์ และโกรธแฟคเตอร์ตัวอื่นๆ เพื่อใช้ในงานวิจัยเซลล์ต้นกำเนิด

มหาวิทยาลัยเทคโนโลยีสุรนารี

สาขาวิชาเทคโนโลยีชีวภาพ ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2558 ลายมือชื่ออาจารย์ที่ปรึกษา_____

SUMETH IMSOONTHORNRUKSA : IDENTIFICATION OF OCT4 ASSOCIATED PROTEINS IN HUMAN EMBRYONIC STEM CELLS AND PRODUCTION OF HUMAN BASIC FIBROBLAST GROWTH FACTOR. THESIS ADVISOR : ASSOC. PROF. MARIENA KETUDAT-CAIRNS, Ph.D., 156 PP.

PROTEIN-PROTEIN INTERACTION/OCT4 INTERACTOME/HUMAN STEM CELL/RECOMBINANT PROTEIN/BASIC FIBROBLAST GROWTH FACTOR

Biotechnology provides powerful and useful tools to study stem cell biology. Here, OCT4 associated proteins in human embryonic stem cells (hESCs) were surveyed. TALENs technology was used to engineer hESCs to express FLAG-Biotin-tagged OCT4. Then, affinity purification techniques in combination with mass spectrometry were used to purify and identify OCT4 associated proteins. OCT4 was shown to associate with proteins related to transcription, translation, chromatin remodeling, DNA repair, RNA processing, glycosylation as well as epigenetic regulation machineries including NuRD, SWI/SNF and FACT complexes. Validation of some interactors (BEND3, FUS, RBM25, WDR82 and SFPQ) revealed that knockdown expression of SFPQ protein led to hESC dysfunction and death. This novel result demonstrated that SFPQ-OCT4 interaction may provide a novel protein-protein interaction complex that plays crucial function for maintaining hESC pluripotency. Here, some OCT4 interacting proteins were uncovered. This knowledge will be a resource to help with the investigation of OCT4 function. However, further analysis of the listed candidates at the functional level will be needed to assist in the understanding of the hESC pluripotency network.

An easy and efficient expression system for in-house human basic fibroblast growth factor (hbFGF) production was also performed. Several parameters for optimizing the expression of soluble hbFGF, including fusion tagged strategy (6xHis-hbFGF and Trx-6xHis-hbFGF), *E. coli* expression host strains (BL21(DE3) or ArcticExpress®RIL), concentration of IPTG used for induction, duration of induction and affinity purification procedures were investigated. Under the optimal condition, approximately 60-80 mg of hbFGF fusion proteins was obtained from 1 liter culture broth. The biological activity of purified hbFGF was assessed using NIH3T3 cells and undifferentiated hESCs and induced pluripotent stem cells (iPSCs). The results demonstrated that the N-terminal fusion tags of 6xHis-hbFGF and Trx-6xHis-hbFGF fusion proteins did not interfere with their biological activities. Human ESCs and iPSCs supplemented with the fusion proteins could maintain undifferentiated stage for several passages. No significant differences in the stem cells morphology and gene expression were observed when supplement with our purified hbFGF fusion proteins or commercialized hbFGF. The use of the fusion proteins in our laboratory was found to significantly reduce the cost of media preparation, approximately 30-60%. We have demonstrated an easy and efficient expression system for in-house hbFGF production, and this system may not only facilitate further production of recombinant hbFGF, but also allow possible large-scale production of other biologically active cytokines and growth factors used for stem cell research.

School of Biotechnology Student's Signature_____

Academic Year 2015 Advisor's Signature_____

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Mariena Ketudat-Cairns (MKC) for her continuous support of my study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all of the research and writing of this thesis. I could not have imagined having a better advisor and mentor for my study. I would like to thank her very much for her professional guidance, valuable support and understanding over these past years.

Besides my advisor, I would like to thank the rest of my thesis committee: Assoc. Prof. Dr. Rangsun Parnpai (RP), Dr. Ruttachuk Rungsiwiwut, Dr. Chanchao Lorthongpanich, Dr. Chuti Laowtammathron and Dr. Weerapong Prasongchean for their insightful comments and encouragements and also for the hard questions which incented me to widen my research from various perspectives.

My sincere thanks also go to Prof. Dr. Rudolf Jaenisch and Dr. Thorold W. Theunissen who provided me an opportunity to join their team at the Department of Biology, Massachusetts Institute of Technology and Whitehead Institute for Biomedical Research (WIBR), Cambridge, MA, USA. Without they precious generosity it would not be possible to conduct the study of chapter III.

I would like to offer my special thanks to my colleagues from Human Embryonic Stem Cell Research Center at Chulalongkorn University, Black Family Stem Cell Institute, Mount Sinai School of Medicine in New York and WIBR in Boston who provided insight and expertise that greatly assisted the research.

I express my warm thanks to the present and passed members of Molecular

Biology and Protein Engineering Laboratory (MKC team), Embryo Technology and Stem Cell Research Center (RP team) and School of Biotechnology at SUT for helpful discussions and for all the fun we have had in the years together. Getting through my dissertation required more than academic support, and I have many, many people to thank for supporting me throughout the times, having to tolerate me over the past years.

Most importantly, none of this could have happened without my family. They were always supporting me and encouraging me with their best wishes. This dissertation stands as a testament to their unconditional love and encouragement. I would also like to express my definitely appreciation to Natteewan Udomsil for her love, encouragement and support in every way during my study. She was always there cheering me up and stood by me through the good and bad times.

Last but not the least, I am grateful to God for the good health and wellbeing that were necessary to complete this dissertation. Again, I would like to thank all those living organisms including *E. coli*, mice, and cell lines which made this thesis possible and an unforgettable experience for me.

My study was funded by the Suranaree University of Technology, SUT-PhD Scholarship Program (SUT-PhD/07/2554) under the National Research University (NRU) project of Thailand, Office of the Higher Education Commission and other funds from MKC group.

Sumeth Imsoonthornruksa

CONTENTS

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES (Continued)

Figure **Page**

LIST OF ABBREVIATIONS

AP-MS	=	affinity purification-mass spectrometry
BirA	=	*Escherichia coli* biotin holoenzyme synthetase
Co-IP	=	co-immunoprecipitation
DSBs	=	DNA double strand breaks
E. coli	=	*Escherichia coli*
ESCs	=	embryonic stem cells
FBS	=	fetal bovine serum
GO	=	gene ontology
6xHis	=	hexa-histidine fusion protein
hbFGF	=	human basic fibroblast growth factor
hESCs	=	human embryonic stem cells
HDR	=	homology-directed repair
IMAC	=	immobilized metal affinity chromatography
iPSCs	=	induced pluripotent stem cells
IPTG	=	isopropyl-β-D-thiogalactoside
LC-MS/MS	=	liquid chromatography-tandem mass spectrometry
MTT	=	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NHEJ	=	nonhomologous end-joining
NuRD	=	nucleosome remodeling histone deacetylase
OCT4	=	octamer binding transcription factor 4
PBS	=	phosphate buffered saline

LIST OF ABBREVIATIONS (Continued)

PPIs	=	protein-protein interactions
SA	=	streptavidin bead
SDS-PAGE	=	sodium dodecyl sulfate polyacrylamide gel electrophoresis
shRNA	=	short hairpin RNA
SILAC	=	stable isotope labeling by amino acids in cell culture
TALEN	=	transcription activator-like effector nuclease
Trx	=	thioredoxin fusion protein

มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER I

INTRODUCTION

Regenerative medicine is a broad field of biomedical research especially in tissue engineering and therapeutic stem cells. The hope is to regenerate the tissues or organs to establish and restore normal function. In mammals, four types of pluripotent stem cells have been identified, isolated and grown in culture including embryonic stem cells (ESCs), embryonic germ cells (EGCs), embryonal carcinoma cells (ECCs) and induced pluripotent stem cells (iPSCs). ESCs can *in vitro* differentiate into all kinds of cells and provides a powerful model system for cellular and molecular mechanism studies of mammalian developmental biology. These features provide the cells the potential to be alternative tools for the treatment of human diseases, since pluripotent stem cells differentiation could be used to regrow, repair or replace damaged or diseased tissues or organs via transplantation. However, research and development of methods for *in vitro* regulated, controlled and reproducible differentiation of the stem cells are required before the applications of the cells differentiation capability can be fully realized. An understanding of pluripotency at the molecular level should provide signs into ways to maintain as well as directly differentiate the stem cells which specific lineages or somatic cell reprograming to iPSCs. The analysis of protein complex and physical protein-protein interaction networks of pluripotency factors should provide valuable information lead to a better understanding of pluripotency mechanisms that involve self-renew and maintain pluripotency of the stem cells. The proteome in human ESCs has not been extensively study in comparison to other cellular systems. One reason for this is that the culture of

human ESCs is difficult to establish and manipulate. A lot of money on expensive stem cell culture media and growth factors are also needed.

In this study, we examine the pluripotency interactome to try to understand the intricate protein interaction networks and protein complexes surrounding the OCT4 transcription factor. To recover the transcription factors with its associated proteins, a biotinylation proteomic approach was used. The transcription factors complementary DNA bearing a carboxyl-terminal triple (3x)-FLAG epitopes followed by a short peptide tag that severs as a substrate for *in vivo* biotinylation were expressed in human ESCs that were previously engineered to express *Escherichia coli* biotin holoenzyme synthetase (BirA). The OCT4 tagged transcription factors and its associates were recovered from nuclear extracts with streptavidin beads or immunoprecipitation with anti-FLAG and anti-OCT4 antibodies. Protein complexes recovered from the purification were fractionated on SDS-PAGE followed by in-gel digestion then subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS) to identify component of the protein complexes. Here, we report OCT4 interactomes and demonstrate that the interactomes were connected with the pluripotency machinery in human ESCs that are important for human stem cells to maintain pluripotency and somatic cell reprograming.

The studies of protein-protein interaction in human ESCs need large-scale purification of multi-protein complexes of approximately 20 to 40 15-cm dishes of human ESCs to provide enough cells to make 50 to 100 mg of nuclear extract. To maintain self-renew human ESCs in culture, human basic fibroblast growth factor (hbFGF) is a crucial growth factor needed for cell culture. To decrease the cost of the growth factor in human ESCs research, here, we describe a simple procedure for production of recombinant hbFGF. This study demonstrated an uncomplicated, fast and low-cost preparation of homemade soluble biologically active hbFGF fusion proteins

expressed in *E. coli*. The purified hbFGF proteins, both with and without the fusion tags, were biologically active. Moreover, the fusion proteins have the ability to support several culture passages of undifferentiated human ESCs and iPSCs. This work provided a simple tool for production of the hbFGF that is friendly and easy to apply in stem cell research.

CHAPTER II

LITERATURE REVIEWS

2.1 Stem cells

Stem cells are undifferentiated cells that contained the capacity to differentiate into different cell types and self-renew with the ability to extensively proliferate without senescence. Stem cells can be classified according to their potency as totipotent, pluripotent, multipotent, oligopotent and unipotent (Table 2.1). They can also be defined by the origin as embryonic stem cells, adult stem cells or induced pluripotent stem cells.

2.1.1 Embryonic stem cells (ESCs)

ESCs originally derived from inner cell mass (ICM) of the blastocyst of a pre-implantation embryo (Evans and Kaufman, 1981). These cells exhibit an unlimited proliferated capacity in culture and maintain their pluripotent potential to differentiate into tissue of the three embryonic germ layers (Yao et al., 2006). Human ESCs have been derived from ICM of day5/day6 fresh as well as frozen blastocyst. They are co-cultured on fibroblast feeder cells of either mouse or human origin, combined with fetal bovine serum or more human serum or serum replacement with or without bovine or human serum albumin and supplemented with one or more cytokines, usually basic fibroblast growth factor (bFGF). The bFGF has been determined to be the key factor to maintain human ESC pluripotency. While leukemia inhibitory factor (LIF) activates the LIF/Stat pathway to sustain pluripotency of mouse ESCs. LIF does not have the same effect in human ESCs. There have been reports that bFGF support human ESC self-renewal via

TGFβ/Activin A signaling pathway (Thomson et al., 1998; Amit et al., 2003; Richards et al., 2003; Koivisto et al., 2004; Daheron et al., 2004; Park et al., 2011). In culture, human ESC colonies are tightly packed with clear borders of small round cells. Individual human stem cell has a high nucleus-cytoplasm ratio with distinctive nucleoli. Mouse ESCs grow and develop in three-dimensional rounded colonies with regular borders, while human ESCs proliferate forming bi-dimensional flat-shape colonies with clear border. ESCs are identified by the presence of high level of alkaline phosphatase activity and expression of ESC markers (Marti et al., 2013). Unlike mouse ESCs, human ESCs express transcription factors OCT4, SOX2 and NANOG as well as the stage-specific embryonic antigen SSEA-3, SSEA-4 but negative for SSEA-1 expression. Although, both human and mouse ESCs are derived from ICM of blastocyst, they have different biological properties with respect to under lying mechanism of self-renewal including gene expression profile, growth factor/signaling confidence and colony morphology. Molecular analysis has shown that human ESCs retain their pluripotency similar to mouse-derived epiblast stem cells (Ginis et al., 2004). The pluripotent of ESC properties include not only maintenance the capable of self-renewal indefinitely in undifferentiated stage but also demonstrated clonally derived cell capable of differentiation into all three embryonic germ layers *in vitro* and maintaining normal karyotype during growth.

2.1.2 Adult stem cells

Adult stem cells are derived from adult tissue and characterized by their capacity for self-renewal and differentiation into tissue-specific cell types which they committed to differentiate into. These cells could be obtained from all tissues of the three germ layers as well as placenta including skin, muscle, intestine and bone marrow (Artlell et al., 1998; Matsuoka et al., 2001; Coulombe et al., 2004; Humphries et al.,

2011). Different cell types of adult multipotent cells are capable of differentiating into various cell lineages (Table 2.2) (Slack, 2000; Blau, 2001).

Table 2.1 Differentiation potential classification of self-renewal cells.

Term	Definition
Totipotent	The most undifferentiated cells found in early development, zygote and the cells of the first two divisions. They can differentiate into embryonic as well as extra-embryonic tissues.
Pluripotent	Cells that have the potential to differentiate into derivatives of all three germ layers, ectoderm, endoderm and mesoderm. Usually are embryonic stem cells (ESCs), which were first derived from ICM of the blastocyst. Recently, reprogramming of somatic cells also generated pluripotent cells, named iPSCs which share similar characteristic with ESCs.
Multipotent	Cells found in most tissues which differentiate into closely related family of cells. Mesenchymal stem cells are the most recognized cells derived from bone marrow, bone, adipose tissue, Wharton's jelly, umbilical cord blood and peripheral blood.
Oligopotent	Cells that are able to form two or more lineages within a specific tissue and self-renew. Example, hematopoietic stem cells that can differentiate into both myeloid and lymphoid lineages.
Unipotent	Precursor cells that are able to differentiate into only one specific cell type and have the property of self-renewal. Example, spermatogenic stem cell giving rise to sperm and not any other cells.

Table 2.2 Example of adult stem cells and their commitment of differentiation.

Cell type	Specific tissue source	Cells/Tissue destination
Hematopoietic stem cells	Bone marrow, peripheral blood	Bone marrow and blood lymphohematopoietic cells
Mesenchymal stem cells	Bone marrow, peripheral blood	Bone, cartilage, tendon, adipose tissue, muscle, marrow stroma, neural cells
Neural stem cells	Ependymal cells, astrocytes (subventricular zone) of the central nervous system	Neurons, astrocytes, oligodendrocytes
Hepatic stem cells	In or near the terminal bile ductules (canals of Hering)	Oval cells that subsequently generate hepatocytes and ductular cells
Pancreatic stem cells	Intraislet, nestin-positive cells, oval cells, duct cells	Beta cells
Skeletal-muscle stem cells or satellite cells	Muscle fibers	Skeletal muscle fibers
Stem cells of the skin (keratinocytes)	Basal layer of the epidermis, bulge zone of the hair follicles	Epidermis, hair follicles
Epithelial stem cells of the lung	Tracheal basal and mucussecreting cells, bronchiolar Clara cells, alveolar type II pneumocyte	Mucous and ciliated cells, type I and II pneumocytes
Stem cells of the intestinal epithelium	Epithelial cells located around the base of each crypt	Paneth's cells, brush-border enterocytes, mucussecreting goblet cells, enteroendocrine cells of the villi

(Source: Korbling and Estrov, 2003)

Hematopoietic stem cells are typical example of adult stem cells that can be found abundantly in the bone marrow and are tissue-specific and can only be used to reconstitute the hematopoietic system. Hematopoietic stem cells can be used after myeloablation to repopulate the bone marrow in patients with hematologic disorders, potentially curing the underlying disorder. Although, the molecular mechanisms of lineage switches within the hematopoietic system have been extensively studies (Graf, 2002), but the mechanisms that determine the transitions in the fate of adult stem cell remain poorly understood.

Mesenchymal stem cells are another type of adult stem cells that are able to differentiate into various mesodermal cell lineages including myocytes, osteoblasts, chondroblasts, fibroblast, adipocytes and other stromal element. Mesenchymal stem cells present in all organs and are organ specific. For therapeutic purposes, mesenchymal stem cells are mostly isolated from bone marrow and umbilical cord.

Adult stem cells have several advantages over ESCs, mainly they can be autologous cells which do not raise issues of rejection or ethical controversies (Korbling and Estrov, 2003; McCormick and Huso, 2010). Several studies have shown that adult stem cells transplantation restores damaged organ *in vivo*, such as bone tissue repair and revascularization of the ischemic cardiac tissue via stem cell differentiation and generation of new specialized cells (Obradovic etal., 2004; Chimutengwende-Gordon and Khan, 2012).

2.1.3 Induced pluripotent stem cells (iPSCs)

iPSCs are differentiated cell that have been reprogrammed to exhibit similar features to ESCs including cell morphology, growth properties, telomerase activity, cell surface marker, gene expression, epigenetic status of pluripotent cell-specific genes. They

could differentiate into all cell type of the 3 germ layer. First reported in 2006, Takahashi and Yamanaka demonstrated the generation of mouse iPSCs by retrovial transduction of Yamanaka's factors (OCT4, SOX2, KLF4 and c-MYC). A year later, in 2007, iPSCs can be made from both mouse and human with several cell types, as show in Table 2.3 (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007). Additional, scientists are still investigating new methods to generate safer iPSCs. A number of different reprogramming factors and methods have been established and the results demonstrated that various cell source, delivery methods and combination of transcription factors differ significantly in their efficiency (Table 2.4).

Table 2.3 Generation of iPSCs from different cell sources and different combinations of transcription factors (O-OCT4; S-SOX2; K-KLF4; M-cMYC; N-NANOG, L-LIN28).

Cell types	Transcription factors	References
Fibroblast	OSKM	Takahashi and Yamanaka, 2006
	OSLN	Yu et al., 2007
Keratinocytes	OSKM	Aasen et al., 2008
Cord blood endothelial cells	OSLN	Haase et al., 2009
Cord blood stem cells	OSKM	Ye et al., 2009
Neural stem cells	O	Kim et al., 2009a
Melanocytes	OSKM	Utikal et al., 2009
Amniotic cells	OSKM	Li et al., 2009
Adipose derived stem cells	OSKM	Sugii et al., 2010
Hepatocytes	OSKM	Liu et al., 2010
Circulating T cells	OSKM	Seki et al.,2010
Astrocytes	OSKM	Ruiz et al., 2010
Peripheral blood	OSKM	Kunisato et al., 2011
Kidney mesangial cells	OSKM	Song et al.,2011
Exfoliated renal epithelial cells	OS	Zhou et al., 2012

Table 2.4 The role of different factors in the generation of iPSCs (O-OCT4; S-SOX2; K-KLF4; M-cMYC; VPA-valproic acid; CHIR99021-inhibitor of Glycogen synthase kinase 3).

Transduction methods	Reprogramming Factors	Cell type	Efficiency %	References
Retroviral	OSKM	Mouse fibroblast	0.001-1	Takahashi and Yamanaka, 2006
	OSK + VPA	Neonatal	1	Huangfu et al., 2008a
Lentiviral	OSKM	Human fibroblast	0.1-1	Yu et al., 2007
Inducible lentiviral	OK + parnate + CHIR99021	Neonatal	0.02	Li et al., 2009
	OSKM	Human fibroblast	0.1-2	Maherali et al., 2008
Sendai virus	OSKM	Human fibroblast	~0.1	Fusaki et al., 2009
Adeno virus	OSKM	Mouse fibroblast	~0.001	Stadtfeld et al., 2008a
Plasmid	OSK	Fibroblast	<0.001	Okita et al., 2008
PiggyBAC	OSKM	Fibroblast	0.01	Woltjen et al., 2009
Polyarginine tagged	OSKM	Neonatal	<0.001	Kim et al., 2009b
lox P lentivirus	OSKM	Fibroblast	0.1-1	Somers et al., 2010
Synthetic mRNA	OSKM	Human fibroblast	4.40	Warren et al., 2010

Since the beginning, the use of viral vectors integration, such as lentiviral and retroviral vectors which possess comparatively high reprogramming efficiency, has been a classic method for the generation of iPSCs. However, these manipulations changed the structure of the genome by integrating other gene sequences, thus may lead to tumorogenesis (Okita et al., 2007), which would limit practical use. To date, several non-integrating approaches have been developed to generate human iPSCs, such as

adenovirus-mediated gene delivery (Zhou and Freed, 2009), non-integrative episomal vectors (Yu et al., 2009), piggyBac transposition (Woltjen et al., 2009), self-excisable vector (Somers et al., 2010), mRNAs (Warren et al., 2010), and by the delivery of reprogramming proteins (Zhou et al., 2009; Kim et al., 2009). Many of the non-integrative methods still have severe limitations, such as the challenges in generation and purification of proteins and sendai viruses (Fusaki et al., 2009, Zhou et al., 2009), the need for repeated administration of synthetic mRNA (Warren et al., 2010), and the low reprogramming efficiency of plasmid- and protein-based methods (Kim et al., 2009). Developments of efficient non-integrative methods for reprogramming are still required. iPSCs provide the opportunity to create an identical match to the cell donor (patients) and thus avoiding issues of rejection. Identical match donors are widely required in therapeutics for disease modeling, regenerative medicine, and drug discovery.

2.2 Methods used to study protein-protein interactions (PPIs)

In 1995, the word proteome was created to describe the protein complement of a genome while the word proteomic represents a systemic analysis of proteins composition of the genome (Wasinger et al., 1995). Proteomic can be divided into two fields. First, expression proteomic is the study of the global changes of protein expression and second, cell map proteomic which is the systematic study of protein-protein interactions (PPIs) through the isolation of protein complex (Blancksteck and Weir, 1999). The studies of PPIs have provided fundamental understanding of complex biology processes and pathways that control the growth and development including cell-to-cell interaction, metabolic and developmental control replication and selective elimination of cells. It is known that abnormal PPIs may lead to diseases and cancer. Assembly of transcriptional machinery also involves PPIs that provided an understanding of transcriptional regulatory

circuitry as response for pluripotency and self-renewal in every cells including human ESCs. Different methologies such as yeast two-hybrid, protein fragment complementary assay, and protein affinity purification have been developed for studying PPIs.

2.2.1 Yeast two-hybrid (Y2H)

Y2H method is an *in vivo* method developed by Fields and Song (1989) to study PPIs. In Y2H assay, both a DNA binging domain (DBD) that helps binding to DNA and an activation domain (AD) responsible for activating transcription of DNA are required for activating transcription of a reporter gene (Figure 2.1A). For example, *Saccharomyces cerevisiae* transcriptional activator GAL4, which is composed of two separable and functionally distinct domains, an N-terminal domain binds to specific DNA sequence and a C-terminal domain containing acidic regions, that necessary to activate transcription. A system of two hybrid protein was setup via the GAL4 DBD was fused to a protein X, termed the bait and a GAL4 AD fused to a protein Y, termed the prey. If protein X and protein Y are co-expressed and both proteins interact the complex of X and Y will bring the two domains of GAL4 into proximity, then a functional transcription factor is reconstituted at the GAL4 promoter site upstream of a reporter gene. Consequently, transcription of reporter gene is activated (Figure 2.1A). Typically, Y2H prey screening is replaced by a collection of unknown prey expression from cDNA or genomic libraries. Screening of the libraries against defined bait may lead to the discovery of novel interaction partners. Many studies demonstrated that Y2H can be easily automate for high-through put studies of PPIs on a genome wide scale including bacteriophage T7 (Bartel et al., 1996), *S. cerevisiae* (Uetz et al., 2000; Ito et al., 2001), *Drosophila melamogaster* (Formstecher et al., 2005), *Caenorhabditis elegans* (Obrdlik et al., 2004) and human (Stelzl et al., 2005; Rual et al., 2005). The reasons for the success of

Y2H system are an *in vivo* genetic screening system, easily scalable, no purification steps and robotic platforms are available. In addition, interactor partners could be identified by colony PCR followed DNA sequencing analysis which makes such screens less time consuming and less expensive. However, false positive and negative can occur and are considered to be the serious technical problem of the Y2H system. In the Y2H assay, PPIs must be localization to the nucleus, some of the proteins fused with DBD and AD cannot localize in the yeast nucleus, cannot fold properly and cannot function when expressed as a fusion protein. Some protein need appropriate post-translational modification to carry out their functions which are unlike to behave or interact normally in Y2H. Due to these limitations the Y2H is less popular now.

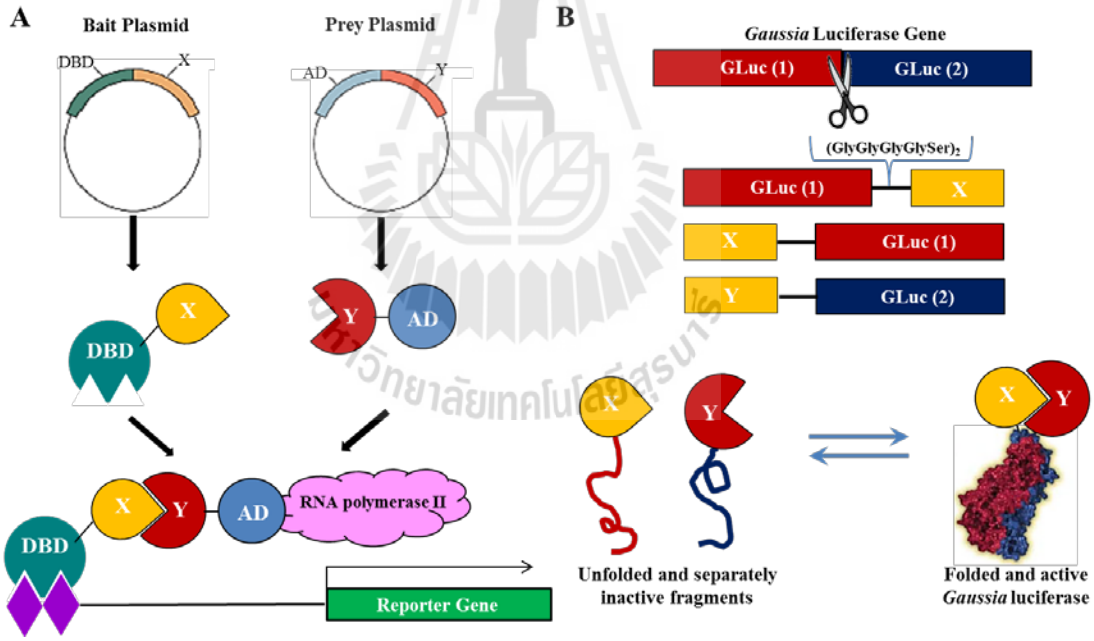

Figure 2.1 Techniques to study protein-protein interactions. (A) Yeast two-hybrid system and (B) Protein complementation assays with *Gaussia* luciferase reporter (Gluc). DBD: DNA binding domain, AD: activation domain, X, Y: gene encode for protein X and protein Y.

2.2.2 Protein fragment complementation assays (PCAs)

To observe PPIs directly in cells, PCAs can be use. This techniques are based on two proteins of interest (bait and prey) interact by fuse each of them with an inactive fragment of a reporter protein when the two come together the reporter protein is active. Each protein requires specific breaking points that allow the non-covalent protein reconstitution while minimizing the spontaneous folding. If the proteins under interact with each other, the reporter fragments are brought together allowing them to fold into native structure of the reporter and then reconstituting the activity of the reporter (Figure 2.1B). PCAs have been created with many different reporter proteins includes dihydrofolate reductase (DHFR) (Pelletier et al., 1998), β-lactamase (Galarneau et al., 2002), fluorescent proteins (Hu and Kerppola, 2003) and luciferase (Villalobos et al., 2007). The signal readout can be survival, fluorescence or colour. Originally, PCAs were developed with DHFR, widely used in libraries selection experiments but the availability of fluorescent or luminescent proteins has significant extended their application. To assure that PCA can occur, linker of about 10-15 amino acids flexible polypeptide that consist of $(Gly\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Ser)_n$ is inserted between the proteins of interest and inactive reporter fragments. The linker provides polarity and flexibility to allow the fragment to find each other and fold. The length of linker could be shorter if the protein of interests are short peptide to avoid entropic penalty upon binding (Remy and Michnick, 2001).The folding and activity of the reporter depends on the interaction between bait and prey. One critical point for PCAs is that the dissected inactive fragments of the reporter protein should not associate spontaneously in the absence of the binding proteins. If this happen, false positive will be observed. To avoid self-association, if possible the expression of fusion protein should be low levels as close to those of the endogenous counterparts and the use of endogenous promoter is advisable (Cabantous et

al., 2005). Like Y2H, PCAs permit *in vivo* genome wide screening for weak PPIs but the system still requires the localization in the nuclear of fusion proteins. In addition, Y2H assay are limited to the study of binary interaction whereas PCAs can be applied to study ternary protein complexes which the competition for the binding to specific protein or the subcellular localization of PPIs.

2.2.3 Protein affinity purification

In mammal cells, protein-protein interaction complexes have been studied by combining protein affinity purification with mass spectrometry and bioinformatics. Affinity purification refers to the capture of biological material by specific enrichment with ligand and coupled to a solid support (Roague and Lowe, 2008). Affinity purification can be classified into two main approaches according to the nature of the target molecule (Figure 2.2). When the target molecule is the native endogenous protein, then the antibody-based affinity purification, usually named co-immunoprecipitation is applied. However, if the target molecule is tagged with an epitope, then affinity purification will be based on the affinity tag. One more different affinity tags can be used for tandem affinity purification, which can increase specificity and reduce the background (Puig et al., 2001; Li, 2010).

2.2.3.1 Co-immunoprecipitation (Co-IP)

Co-IP, also known as antibodies pull down or antibody-based affinity purification was developed from the immunoprecipitation method. Immunoprecipitation involves the specific antigen-antibody reaction, then incubation of cellular extracts or protein mixture to form an immune complex with the target proteins. The know protein X is named the bait protein and it interaction protein Y is named prey. If protein X form a stable complex with protein Y in the cell. Immunoprecipitation of

protein X may be resulted in co-immunoprecipitation of protein Y (Figure 2.2). Co-IP is commonly used to investigate whether two proteins of interest are associated *in vivo* and it can be used to identify interacting partners of a target protein (Miernyk and Thelen, 2008). The interacting protein can be identified by Western blot if suitable antibodies are available or by mass spectrometry. The advantages of antibody-based affinity purification are the endogenous protein complexes can be purified, which signifies the biological and functional relevance of the identified interacting proteins and the experiment can be done in a speedy manner, which no transgenic lines need to be established. However, the disadvantages of this approach are most antibodies suffer from non-specific reactivity such that spurious protein complexes will co-purify with the *bona fide* protein complexes, often spend the time and money associated with the preparation of the specific antibody and some interactions may be weak and lost during purification procedure, due to the affinity between antibody and the target protein is low (Master, 2004; Berggard et al., 2007; Miernyk and Thelen, 2008).

2.2.3.2 Epitope-tag affinity purification

The epitope-tag affinity purification has been widely used in isolating protein complexes *in vivo*, after the proteomic study in yeast published in 1999 (Rigaut et al., 1999), In 2002, two studies utilized the epitope-tag affinity purification to isolate protein complex in yeast followed by an MS analysis, and identified hundreds of protein complexes *in vivo* (Gavin et al., 2002; Ho et al., 2002). After that several studies have followed this approach. This approach involves genetically engineering of an oligo peptide or an epitope-tag including hexahistidine (His6), hemagglutinin (HA), FLAG, GST, Myc, calmodulin-binding protein (CBP) or maltose-binding protein (MBP) (Stevens et al., 2000) on to the target protein of interested. The epitope tag may be fused

to either the N- or C-terminus of the full length coding region of the target protein in an expression vector for protein overexpression in the cells. More than two tags can be used to fuse together into either the N- or C-terminus of the target protein. This is called tandem affinity tags. The epitope tag can be translated directly into the affinity handle, or may be used for further modification, such as a biotinylatable sequence can be fused to a bait protein such that biotinylation can be occur *in vivo* following co-expression of a biotin ligase such as the bacterial BirA protein (Rodriguez et al., 2006). The tagged protein can be purified using an affinity matrix that specifically recognizes the epitope (Jarvik and Telmer, 1998; Brizzard, 2008). This technique provides many advantages, such as purity of protein complex purification can be achieved higher than conventional approaches. Rigaut et al. (1999) demonstrated that after tandem affinity tag purification, purified protein complexes are much cleaner than just by one-step purification. This also provides a uniform system for tagging and purifying different proteins for the comparison. Antibodies against the tags are also commercially available. In addition, an epitope-tagging experiment, important concerns relate to detrimental effects potentially mediated by the expression level of exogenously expressed recombinant proteins is needed to be concerned. Massive overexpressed or problematically tagged proteins may exhibit protein misfolding, mislocalization, or mis-regulation, leading to increases in both false-positives and false-negatives and increased background contaminants (Goelet al., 2000; Rumlova et al., 2001; Hofemeister et al., 2011).

2.3 An overview of mass spectrometry platforms for studying PPIs in ESCs

Several methodologies have been developed to study protein complexes in eukaryotes. The combination of affinity purification with mass spectrometry and

bioinformatics analysis has become the method of choice for the characterization of protein complexes in ESCs (Wang et al., 2006). This section describes the overview workflow of mass spectrometry based PPIs studies. The identification of interacting protein may help defining PPIs and proteins of unknown functions. Protein(s) can be engineered to be fused with a single tag or a more combination of various tags to allow the isolation of the bait and together with its associated proteins. To isolate protein complexes, single step or tandem affinity purification can be used. The purified protein complex can be separated by SDS-PAGE. Then, the fractionated proteins will be digested in the gel by trypsin enzyme and the peptide fragments can be analyzed using mass spectrometry. Protein identification is reached by the use of an algorithm against a sequence data base.

2.3.1 Engineering of interested protein

The selection of the proteins of interest, or baits is the first step for designing an affinity purification-mass spectrometry (AP-MS) experiment that will be used to characterize a network of PPIs. In ESCs, transcription factors OCT4, SOX2 and NANOG are core factors involved in maintaining pluripotency. The characteristic of the protein complexes and mapping the protein interaction networks of the pluripotency factors provide an understanding of pluripotency at the molecular level includes epigenetic modifications operating in a regulatory network. These methods have also uncovered new factors in self-renewal signaling pathways that will provide both direct differentiation with specific lineages and reprogram somatic cells to an ESCs. The experimental designs must balance optimizing bait expression to maintain relevance biological level. The expression of the bait protein should be comparable to the expression level of the endogenous counterparts (Al-Hakim et al., 2012). If the

expression level of the bait protein is too high, the protein complexes assembly may not occur under physiological conditions (Drewes and Bouwmeester, 2003). In ESCs, there are many techniques to generate *in vivo* affinity-tagged bait protein including episomal vectors transfection (for example, pEF1∝5His and pPyCAG), lentivirus transfection and recombinase-mediated cassette exchange (Wang et al., 2006; van den Berg et al., 2008; Pardo et al., 2010; Ding et al., 2012; Gagliardi et al., 2013). Programmable nuclease approaches, such as transcription activator-like effector nuclease (TALEN) and clustered regularly interspersed short palindromic repeats (CRISPR)/Cas9 systems, can also be considered to generate stable cell lines that express affinity-tagged proteins (Musunuru, 2013). These systems directly fuse affinity tags to bait proteins within their gene located in the genome. However, engineering stable expression cell lines may need to develop of multiple clones to select for optimal expression levels which require more money and time (Moorman et al., 2010).

2.3.2 *In vivo* biotinylation of bait proteins

Biotin is an occurring cofactor for metabolic enzymes, which is active only when covalently attached to the enzymes through the action of specific protein-biotin ligases. Biotin ligase is responsible for attaching biotin to a specific lysine at the active site of newly synthesized biotin enzymes (Chapman-Smith and Cronan, 1999). Biotinylation is an attractive approach for protein complex purification due to the very high affinity binding avidin and streptavidin to the biotinylated protein. Biotin/streptavidin binding is the strongest noncovalent interaction known in nature ($K_d \sim 10^{-15}$ M), allowing for more efficient and stable capture of the bait protein and associated factors without the use of antibodies. This offers advantages compared to the traditional FLAG and HA tags affinity purifications (Wang et al., 2006). In ESCs, there are a few

endogenously biotinylated proteins that have been well defined, thus reducing the chance for cross-reactivity when using biotinylation in protein purification (de Boer et al., 2003; Wang et al., 2006). The characterization of the minimal amino acid sequence requirements for *in vivo* biotinylated proteins has been developed for substrate peptide sequences. This 23 amino acids (MSGLNDIFEAQKIEWHEGAPSSR) known as biotinylation peptide that can be biotinylated in ESCs (Kim et al., 2009). Biotinylation can occur by co-expression of an exogenous biotin ligase. *Escherichia coli* biotin holoenzyme synthetase (BirA) is an enzyme that performs highly selective biotinylation to the biotinylation peptide sequence. Wang and his colleague established an *in vivo* biotinylation system in mouse ESCs and demonstrated that co-expressed BirA and transcription factors tagged by an N-terminal fusion of the biotinylation peptide can be done. The resulting stable line produced biotinylated transcription factors that can be can be efficiently purified by streptavidin affinity. This improves the AP-MS studies in ESCs (Wang et al., 2006; Kim et al., 2009; Ding et al., 2012). However, additional time is required to establish cell lines for *in vivo* biotinylation when compared to direct endogenous antibody immunoprecipitation, and like affinity tagging, cell lines expressing a controlled level of biotinylated proteins are necessary for analysis.

2.3.3 Affinity tags

In vivo affinity fusion-based protein purification has been implemented in AP-MS studies. It takes advantage of the selective binding of a genetically fused affinity tag. Typically, this method depends on the expression of a bait protein with an affinity tag. As shown in Table 2.7, many different types of affinity tags include c-Myc, FLAG, hemagglutinin (HA), biotin, poly histidine, Streptavidin-tag II, glutathione S-transferase (GST), green fluorescent protein (GFP), protein A, maltose-binding protein (MBP),

calmodulin-binding protein (CBP), FLAG-HA, FLAG-biotin and FLAG-CBP have been successfully applied to AP-MS studies. Two or more tags can be combined effectively to generate tandem affinity purification for investigating mammalian protein interaction. The tagged protein can be purified using an affinity matrix that specifically recognizes the tags. To increase affinity and specificity, the affinity tags can be multiplied such as 3xFLAG. The affinity tag can be fused single or tandem to the N- or C terminus of the bait protein. However, each tagging system has its optimal applicability to different number proteins with different purity, yield and cost.

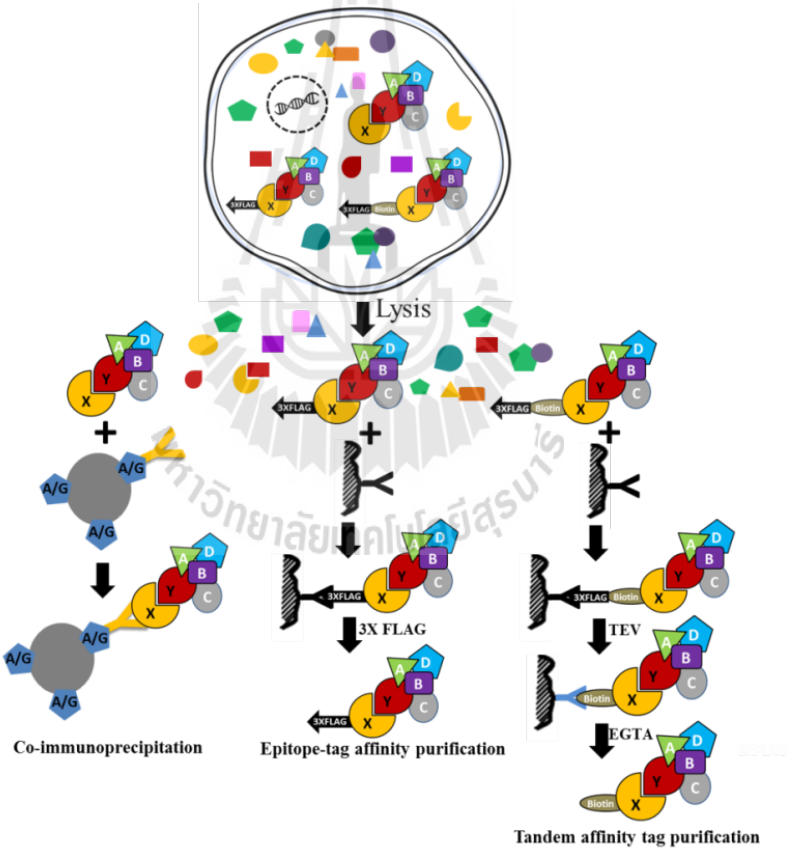

Figure 2.2 Affinity purification of protein complex. Co-immunoprecipitation performed with endogenous proteins and then the protein of interest is isolated with a specific antibody. The tagged protein complex is purified by either one-step or muti-steps affinity purification using affinity columns.

The affinity tags especially a big tag may affect the native structure of bait protein and its binding activity with other partners. Additional, it is important that overexpressed tagged bait proteins may exhibit protein misfolding or mislocalization leading to increases in both false-positives and -negatives and increased background contaminants (Dunham et al., 2012). However, every tag has its own specific background protein profile, such as STK38 and the associated CRAPome for FLAG-tagged baits or biotin-cofactor enzymes include some carboxylases and ribosomal proteins associated with the streptavidin tag system (Wang et al., 2006; Mellacheruvu et al., 2013). Recently, the 3xFLAG, *in vivo* biotinylation and endogenous antibody co-IP strategies have been the most widely used affinity tags for both immunodetection and affinity purification to studies the PPIs network in ESCs (Wang et al., 2006; van den Berg et al., 2008; Pardo et al., 2010; Ding et al., 2012; Gagliardi et al., 2013).

2.3.4 Affinity purification

Affinity tagged bait protein purification is an excellent method to purify and identify multiprotein complexes when compared with other methodologies. It allows rapid purification of protein complex under native conditions and the detection are well recognized. Purification of the tagged bait protein can be performed through either one- or multiple-step purification from the stable cell lines by affinity or immune-affinity columns. One-step affinity purification purifies bait protein with one tag fused to either the N- or C-terminus of the bait protein. With the affinity column, the tagged bait protein from the cell lysate can bind to the affinity column or antibody, and be eluted by competition tag peptide analog, changing salt concentrations, pH, ionic strength or detergents to displace the bait protein and bait-associated proteins (Table 2.5). The washing conditions determine purity and the composition of protein complex in the one-

step affinity purification. Using buffers containing high salt concentrations or detergents for stringent washes are used to reduce contaminating proteins. However, stringent washes may also reduce the binding of low-affinity targets and the weak interactions will also lose.

Table 2.5 Affinity tags used for purification of tagged bait protein in the AP-MS studies.

Tag	Sequence/size	Binding agent	Elution conditions
c-myc	EQKLISEEDL	Anti-c-myc	Low pH
FLAG	DYKDDDDK	Anti-FLAG	FLAG peptide/pH/EDTA
HA	YPYDVPDYA	Anti-HA	HA peptide
Biotin	MSGLNDIFEAQKIEWHEGAPSSR (biotinylation peptide)	Streptavidin	SDS
His-Tag	HHHHHH	Ni^{2+}/CO^{2+} NTA	Imidazole/low pH
Strep-tag II	WSHPQFEK	Streptavidin	Desthiobiotin
GST	26 kDa	Glutathione	Reduced glutathione
GFP	26.9 kD	Anti-GFP	pH
Protein A	45 kDa	IgG	Protein A/low pH
MBP	40 kDa	Amylose	Maltose
CBP	4 kDa	Calmodulin	EGTA

(Source: modified from Dunham et al., 2012)

In addition, a rapid strategy with high yield of protein complex under cellular native conditions of the one-step affinity is not only expected to produce more interaction candidates but also with additional false positives (Westermarck et al., 2013). For example, Wang and his colleague established an *in vivo* biotinylation system in mouse ESCs to perform one-step affinity purification of the biotin-streptavidin binding with successful construction of the NANOG interactome in mouse ESCs (Wang et al., 2006). In another example, van den Berg et al., (2008) and Pardo et al., (2010) utilized

3xFLAG tagging strategy to purify and identify OCT4 interaction network in mouse ESCs. The 3xFLAG tagged OCT4 and its associated protein complexes can be isolated with anti-FLAG and eluted with FLAG peptides by competition. Although, the FLAG strategy has been success in HeLa and HEK293 cells but its application in ESCs still suffers from non-specific binding. This may be due to the presence of many nuclear proteins in ESC extracts that react nonspecifically to the FLAG tag and the FLAG antibody (Wang, 2012). Therefore, to improve affinity purification strategy employing FLAG in tandem with a second tag is more often a preferred for the second step purification to achieve better discrimination of specific associations against unspecific protein background. Alternatively, *in vivo* biotinylation strategy has been also developed to complement and improve the affinity purification studies in ESCs (Kim et al., 2009).

2.3.5 Pre-Mass spectrometry (Peptide purification and digestion)

For identification of protein complexes, mass spectrometry is one of the best method that is superior in speed and sensitivity when compared to traditional protein sequencing by Edman degradation or other methods. Typically, the analysis starts from 1- or 2-D gels that are used to separate the components of a protein and stained with Coomassie or silver staining. After the gel is stained, whole lane or bands can be excised and processed generally include destaining, reduced disulfide bonds, alkylation of free cysteine residues and in-gel digestion. In-gel digestion approach has been developed for the enzymatic fragmentation of proteins embedded in gel pieces and the extraction of the resulting peptides. Peptides are normally generated by trypsin proteolytic digestion. Trypsin cleaves peptide bonds specifically C-terminal to arginine and lysine residues leads to peptides in the preferred mass range required for analysis by mass spectrometry (Olsen et al., 2004; Medzihradszky, 2005). After digestion, peptide mixtures are

concentrated and desalted with reversed-phase micro-columns to remove salts and other digestion impurities. The peptide fragments are then analyzed using mass spectrometry strategies.

2.3.6 Mass spectrometry

The mass spectrometer (MS) is used to determine the accurate molecular mass of proteins and the derived peptides, and tandem mass spectrometry (MS/MS) facilitates amino acid sequencing and mapping of post-translational modifications (Aebersold and Mann, 2003; Seet et al., 2006). MS measure the mass to charge (m/z) ratio of charged molecules in gas phase, allowing for the determination of their molecular masses after charge state resolution. Commonly, the MS platform for AP-MS analyses is implemented by liquid chromatography (LC), coupled to electrospray ionization (ESI) tandem mass spectrometers, as named LC-MS/MS systems. Briefly, high-performance liquid chromatography (HPLC), using nanoscale reversed-phase columns (50-100 microns in diameter), is used for separation peptides. Charged peptides are introduced into the MS by ESI, allows the transfer of analyses from a liquid phase to the gas phase at atmospheric pressure. The mixture of peptides is separated according to their m/z by the first MS and a list of the most intense peptide peaks is established. Then individually selected m/z species, presumably a unique peptide ion, is allowed to enter the second MS. These ions are directed into a collision cell and their kinetic energy is increased lead to break peptide bonds. The set of peptide fragments is then recorded and generate the MS/MS spectrum for that peptide. The m/z of the intact peptides and their corresponding fragment ions (MS/MS) are recorded. The MS/MS spectra contain fragmentation pattern of the amino acids in the peptide were then searched against a protein sequence database to find the peptides that best match the spectra. Matching raw MS data to protein

sequences can be performed using search engines such as Mascot, Sequest, X!tandem and Protein Prospector (Eng et al., 1994; Perkins et al., 1999; Craig and Beavis, 2004; Chalkley et al., 2008). However, the identity of the peptide is determined by comparison of the mass spectral data with an *in silico* protein data bank. The quality of mass spectra is critically dependent on the amount of sample introduced into the mass spectrometer and on sample purity.

2.4 Protein interaction networks in ESCs

Protein plays essential roles in all biological processes of cells for example; metabolism, signaling pathways, DNA replication, transcription, translation, protein degradation, and cell cycle control. Proteins do not function in isolation but they interact with one another and also with other molecules that mediate essential biological processes. All biological processes are regulated through association and dissociation of protein molecules. These protein complexes are important cellular machineries that work hand in hand to maintain normal cell homeostasis. Protein interactions control the mechanisms leading to healthy and diseased states in organisms. It is known that impaired or deregulated protein-protein interactions can lead to human disease and cancer (Charbonnier et al., 2008; Bader et al., 2008; Kar et al., 2009). A pluripotent state of ESCs and iPSCs is maintained through the combinatorial activity of core transcriptional factors in conjunction with many other factors. Transcription factors play key roles in undifferentiated state of ESC pluripotency because they control gene expression. OCT4, SOX2 and NANOG are the main molecules of the transcriptional regulatory hierarchy that specifies ESC pluripotency. These factors are known to play roles in stem cell biology as self-renewal regulators and key reprogramming factors. Understanding the molecular mechanisms of pluripotency will provide valuable information on how the

pluripotent state is established and maintained and enable scientist to utilize stem cells more effectively, particularly in the field of regenerative medicine. The strong evidence for the involvement of OCT4, SOX2 and NANOG in pluripotency makes them good starting points to study the protein interaction network of pluripotency. Recently, datasets that are used to build the mouse ESCs protein interaction network have been generated from affinity purification-mass spectrometry experiments. However, no protein interaction networks base on affinity purification-mass spectrometry approaches have been generated from human ESCs.

2.4.1 OCT4 interactome

OCT4 is a POU domain-containing transcription factors encoded by Pou5f1 and known to be central player in ESC self-renewal and early embryonic development. A decrease or increase in OCT4 levels by 50% induces differentiation into trophectoderm or endoderm and mesoderm lineages, respectively (Niwa et al., 2000). The loss of OCT4 in the mouse embryo causes failure in the development of inner cell mass of the zygote (Nichols et al., 1998). There have been reports suggested that OCT4 is essential for integrating the epigenetic machinery into the pluripotency network. Recently, three independent studies have investigated the OCT4 protein interaction network in mouse ESCs. The van den Berg study established ZHBTc4 ESCs line that expresses N-terminal triple FLAG-tag-OCT4 in the absence of endogenous OCT4 (van den Berg et al., 2010). The Pardo study introduced the 3xFLAG epitope and a calmodulin binding peptide (CBP) fused at the C-terminal of the OCT4 coding region by recombining into a BAC clone containing full-length OCT4. (Pardo et al., 2010). The Ding study established a transgenic mouse ESC line via lentivirus infection and Dox treatment, that expresses only biotinylated OCT4 replacing the doxycycline suppressible OCT4 in ZHBTc4 cells (Ding

et al., 2012). The current OCT4 network studies are shown in Figure 2.3. There are 54, 92 and 198 OCT4 interacting proteins identified by the van der Berg, Pardo and Ding studies, respectively. It is unclear at this point whether the *bona fide* OCT4 centered interactome should be constructed as a union or intersection of the data sets. However, when compared the dataset with the recently published OCT4 interactome studies (Figure 2.3) (Ding et al., 2012), 18 proteins were identified consistently observed in all three OCT4 interactomes experiments. They are either involved in the nucleosome remodeling histone deacetylase (NuRD) complex (CHD4, GATA2A, GATA2B, MTA2, MTA3, MBD3, HDAC1), SWI/SNF (also called BAF complex, BRG1, BAF155), or LSD1 (LSD1, RCOR2) complexes, or they are individual transcription factors (SALL1, SALL4, HCFC1, HELLS, etc.) playing important roles in ESCs. Additional, other proteins that have been used as baits include NANOG, SALL4, TCFCP2l1, DAX1, ESRRB, REX1, NAC1 and ZFP281, all of which also were shown to interact with OCT4. However, Ding et al., (2012) demonstrated that OCT4 links to multiple epigenetic regulatory pathways of the pluripotency network and the interactome components of regulatory protein complexes such as LSD1, FACT, COMPASS-like, MLL5-L, MutSalpha, ISWI, and PAF1 complexes (Figure 2.4). This has indicated that OCT4 either directly or indirectly, interact with an extensive set of proteins, including transcription factors, chromatin remodelers, and components of the basal transcriptional machinery. This could explain why many of the OCT4-interacting proteins are ubiquitously expressed in both differentiated and undifferentiated cells.

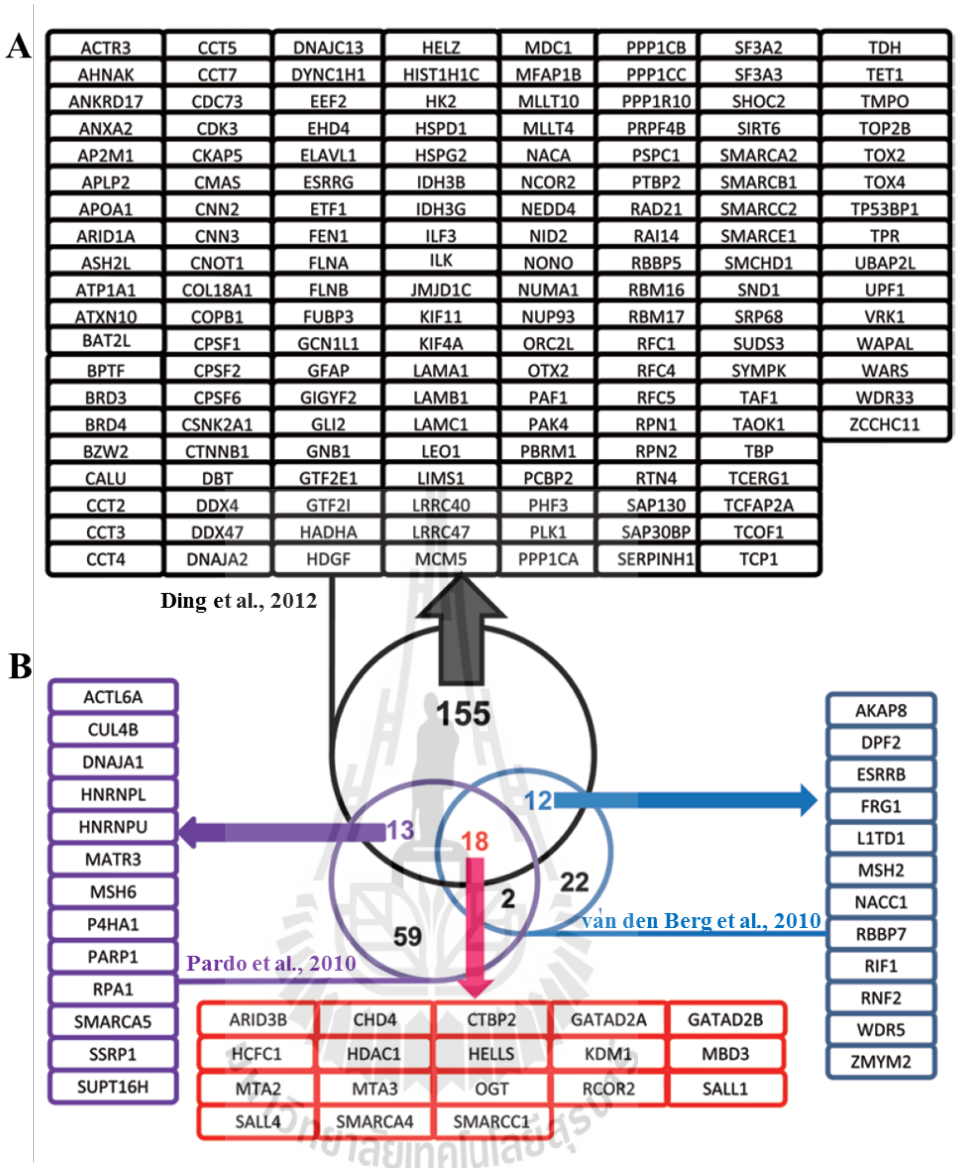

Figure 2.3 OCT4-interacting proteins in mouse ESCs. (A) List of the 155 OCT4-associated proteins from AP-MS-based approaches in ZO4B4 ESCs. (B) Comparison of OCT4 interactome with the three published OCT4 network studies. The overlapping proteins between three published studies are listed (Source: Ding et al., 2012).

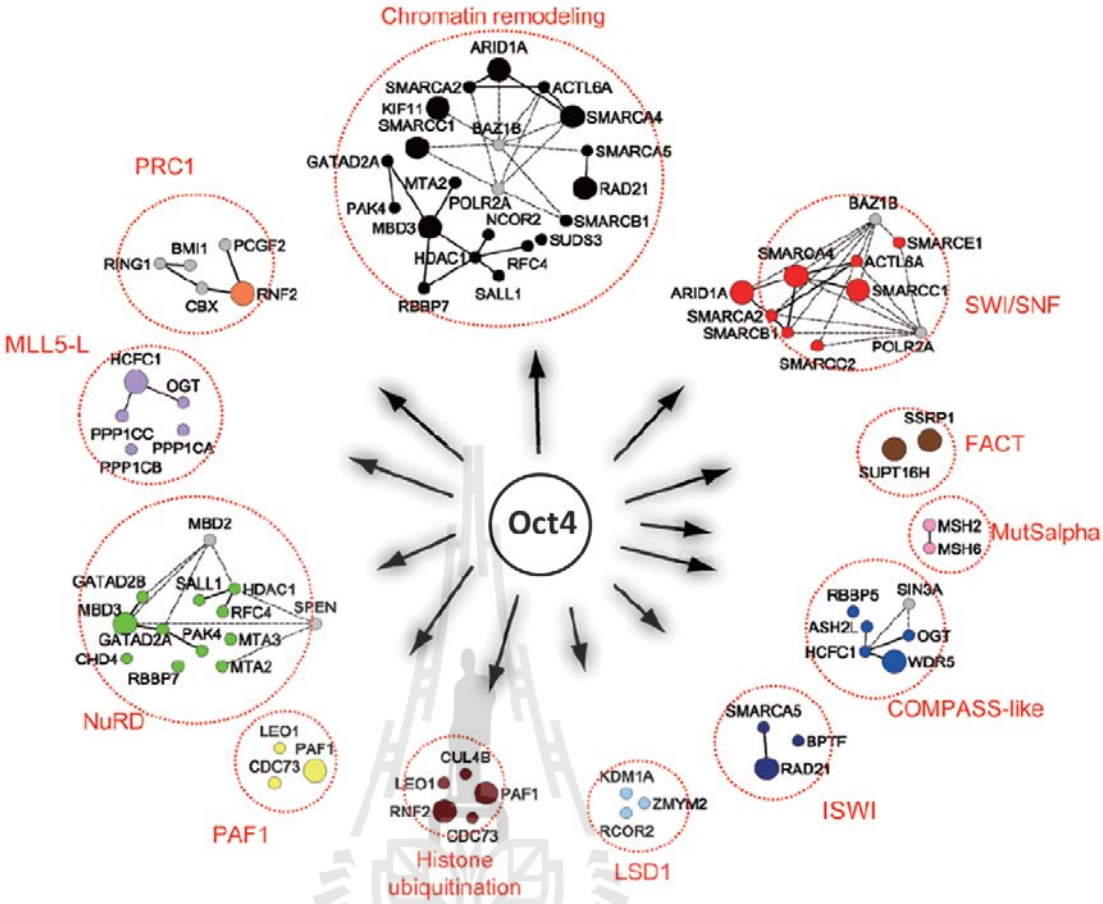

Figure 2.4 Connection of OCT4 with multiple epigenetic regulatory pathways (Source: Ding et al., 2012).

2.4.2 SOX interactome

The SOX2 gene, designated as SRY- (Sex determining Region Y-) box2, encodes a transcription factor belonging to the Sox family of proteins, which bind to DNA through their 79-amino-acid HMG domain (Bullejos et al., 2000). SOX2 is one of the core transcription factors. The mechanism by which SOX2 controls the fate of ESCs is much less well defined. SOX2 expression is dosage sensitive for stem cell maintenance (Kopp et al., 2008).This leads to the investigation of SOX2 protein complexes and its interaction network for understanding stem cell pluripotency. Gao and colleagues have identified the SOX2-interactome in mouse ESC. They first engineered ESC for inducible

expression of FLAG-Strep tagged SOX2 along with OCT4. The result shows that exogenously express tagged-SOX2 and OCT4 together disrupts the self-renewal of ESC and induces their differentiation. In contrast, elevating tagged-SOX2 along with OCT4, Klf4, and c-Myc in ESC does not disrupt their self-renewal or induce differentiation. This suggests that the identification of SOX2 protein complexes in ESCs require careful manipulation of the ectopic expression of the affinity-tagged SOX2 and the maintenance of the self-renewal and pluripotency of ESC depends upon a carefully orchestrated balance of multiple master regulators (Gao et al., 2012). The Gao study demonstrated that among the 71 SOX2-associated proteins are transcription factors, components of chromatin remodeling complexes (e.g. NuRD and SWI/SNF), DNA repair machinery (XRCC1, XRCC5, XRCC7, RBPJ, TOP2A), and DNA replication machinery (POLB, RPA1, RPA2, RPA3) (Figure 2.5). To understand the function of SOX2-associated proteins, Gao study performed gene ontology analysis and indicated that the SOX2-associated proteins involve into several cellular processing including transcription, chromatin assembly/modification, macromolecule metabolism, development, cell cycle progression/cell division, RNA processing and DNA repair (Figure 2.6). Additionally, many of the SOX2-associated proteins are required for the self-renewal of ESC, including: SALL4, ESRRB, MTA1, and MTA2 and approximately 25% of the SOX2-associated proteins have been shown by others to be part of the OCT4-interactome. Moreover, knockdown SMARCD1, SMARCA5 and PARP1, the SOX2-associated protein that also associate with several pluripotency-associated factors, disrupts the self-renewal of ESCs and induces the differentiation (Gao et al., 2012; Lai et al., 2012). This led to suggest that proteins found to associate with multiple required transcription factors are very likely to be required by undifferentiated ESCs.

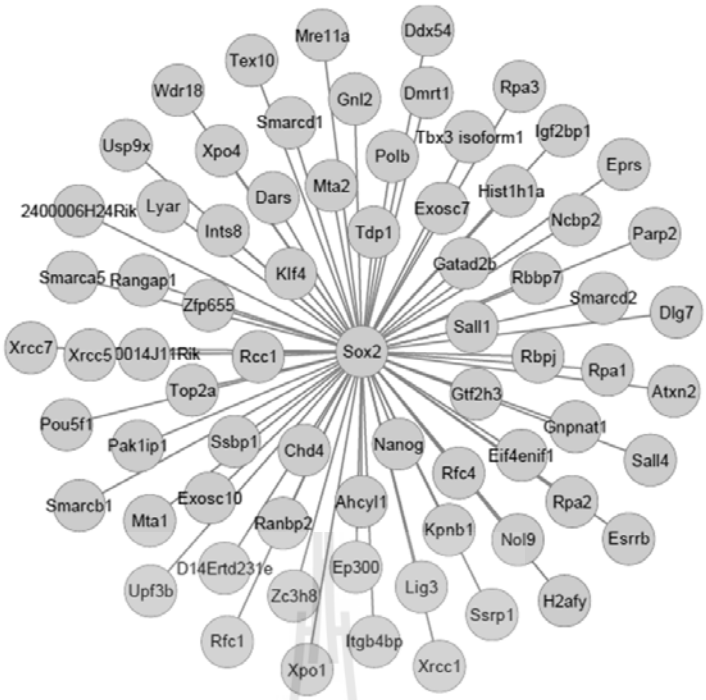

Figure 2.5 SOX2-association proteins in mouse ESCs (Source: modified from Gao et al., 2012).

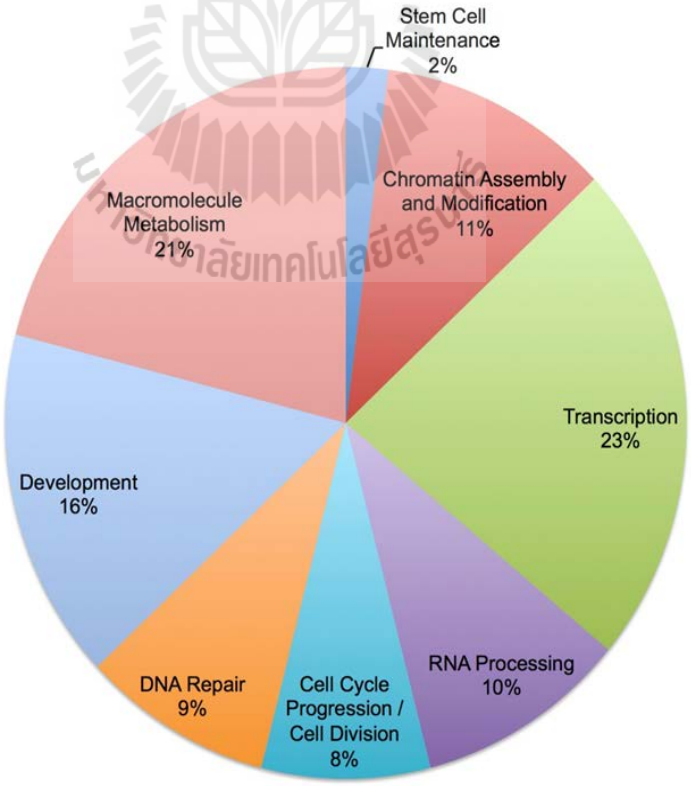

Figure 2.6 Gene ontology classifications of SOX2-associated proteins in mouse ESCs. (Source: Gao et al., 2012).

2.4.3 NANOG interactome

NANOG transcription factor, known as the homeobox protein, plays fundamental roles in early development and stem cell pluripotency (Mitsui et al., 2003). ESCs are sensitive to the dosage of NANOG. The levels of NANOG vary considerably with high levels of NANOG directing efficient self-renewal when enforced expression of this gene relieves ESCs from the LIF requirement for stem cell maintenance in mouse (Chambers et al., 2003). However, the mechanisms by which NANOG delivers this function in ESCs are not fully understood. To understand how pluripotency is programmed and maintained in ESCs, in 2006, the first pioneering protein interaction network in NANOG was studied by Wang and his colleagues. They have utilized a proteomic approach to isolate protein complexes and constructed a protein interaction network surrounding the pluripotency factor NANOG in mouse ESCs (Wang et al., 2006). With a high-affinity biotin/streptavidin (Bio/SA) purification, Wang et al. purified N-terminal FLAG-biotin NANOG and its associated partner protein complexes followed by LC-MS/MS, to identify components of NANOG protein complexes. They have identified a total of 17 proteins of highest confidence that are physically associated with NANOG. A number of factors in the NANOG interactome with both ESC-specific and ubiquitous expression patterns connect to several epigenetic regulatory pathways. These include the histone deacetylase NuRD (P66B and HDAC2), PRC1 (YY1, RNF2/Ring1B and Rybp) and SWI/SNF chromatin remodeling (BAF155) complexes (Wang et al., 2006). The NANOG interactome was further extended in later studies by using biotin-, FLAG tag-, and endogenous NANOG antibody-based affinity approaches (Liang et al., 2008; Costa et al., 2013; Gagliardi et al., 2013). With an improved tagging method and less stringent purification, the Gagliardi (2013) study identified 130 NANOG interactors including transcription factors, chromatin modifying complexes, phosphorylation and

ubiquitination enzymes, basal transcriptional machinery members and RNA processing factors. (Figure 2.7). This information dramatically expanded the knowledge of the NANOG interactome in ESCs. Interestingly, protein complexes with opposing functions, such as histone deacetylase complexes (NuRD, N-CoR, SIN3A) and acetyltransferase complexes (Tip60–p400, also called NuA4–HAT), the LSD1 demethylase complex and the histone 3 lysine 4 methyltransferase (MLL) complex are observed to interact with NANOG. All these data suggest a NANOG-dependent epigenetic regulation of distinct activated and repressed loci in pluripotent cells (Gagliardi et al., 2013; Huang and Wang, 2014).

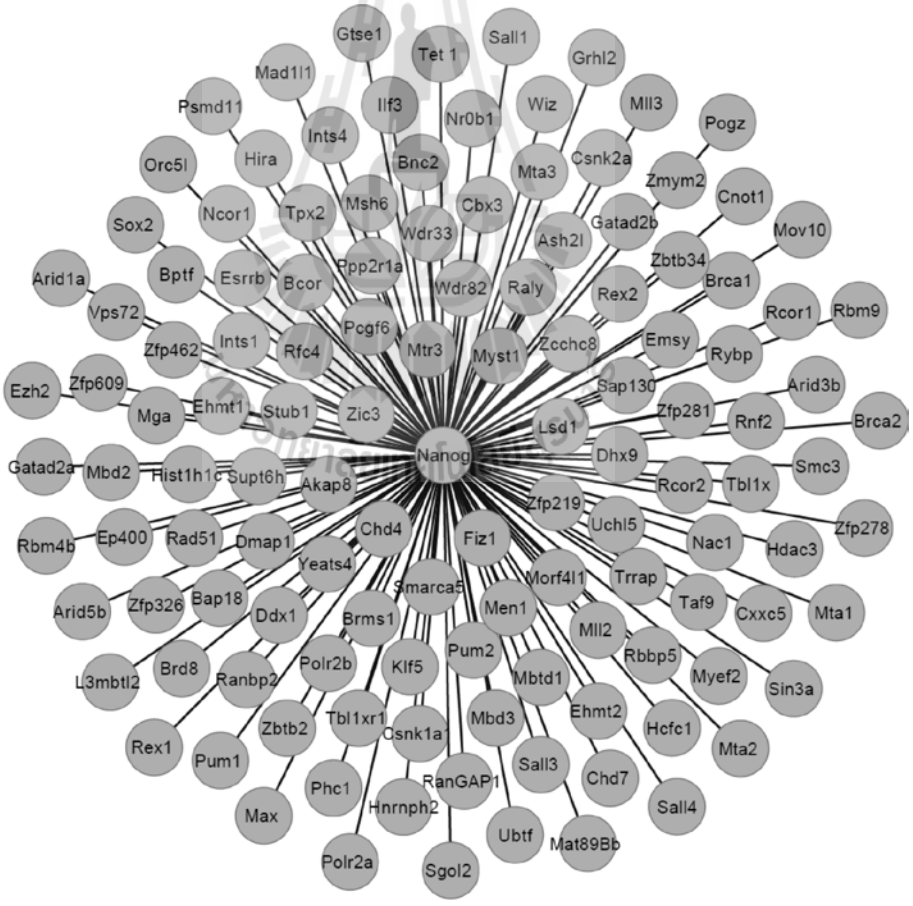

Figure 2.7 NANOG-interacting proteins in mouse ESCs. (Source: modified from Gagliardi et al., 2013).

2.5 Application of recombinant proteins in stem cell researches

In ESC research, the ability to maintain cells in self-renewal stage without losing the pluripotency and no chromosomal aberrations are the main objectives for scientists. Stem cell always balances itself between self-renewal and differentiation. The success of stem cell research depends on the development of optimal cell culture systems. The interest of optimizing stem cell culture is not only because the cells culture are widely used in basic research for studying stem cell biology, but also owing to the potential to the therapeutic applications of cultured stem cells. Therefore, optimization of stem cell cultures required the development of well-defined media supplemented with recombinant growth factors or cytokines. Both expansion and differentiation of stem cells is highly reliant on growth factor and cytokine supplements. There are a number of cytokines and growth factors identified to be crucial for stem cell survival, self-renewal, differentiation, and de-differentiation. Nowadays, more than 150 growth factors and cytokines targeted to stem cell culture are commercially available (http://www. sigmaaldrich.com), including activins, bone morphogenetic proteins (BMPs), epidermal growth factor (EGF), fibroblast-derived growth factors (FGFs), insulin-like growth factor (IGFs), transforming growth factor (TGF), neurotrophic factor, platelet derived growth factors (PDGFs) and vascular endothelial growth factor (VEGF). The high price of growth factors and cytokines is the main cost in stem cell research. Culture of human ESCs on feeder cells in fetal bovine serum (FBS) and basic fibroblast growth factor (bFGF) -containing media is the standard protocol used in many laboratories (Table 2.6). To maintain healthy human ESC cultures, cells must be passaged every 4-7 days. At this time the colonies have reached their maximum size and may merged together. Merged colonies increase the rate of differentiation in the culture. Human ESC research requires careful attention to culture conditions. As shown in Table 2.8, the main cost of human ESC medium preparation is

serum and growth factors. The needs of bFGF and several other growth factors and cytokines, increase the costs of reliable stem cell research. Recently, many laboratories have successfully used recombinant protein technology to produce in-house growth factors and cytokines for stem cell biology (Tomala et al., 2010; Imsoonthornruksa et al., 2011; Song et al., 2013; Rassouli et al., 2013). Several hosts such as, mammalian cells, insect cells, bacteria, yeast and plant seeds are valuable hosts for the production of many recombinant growth factors and cytokines. Eukaryotic expression system is usually chosen for post-translationally modified proteins that are frequently essential for correct protein function since, bacteria are currently unable to incorporate such modifications (Walsh, 2010). In 2013, Magnusdottir reported the production of growth factors and cytokines in barley seeds. This plant does not produce any endotoxins or toxins to human health. (Magnusdottir et al., 2013). There is an inherent limitation to the eukaryotic system with regard to high production cost, sometimes together with low yields. For proteins that are not required to be synthesized in a glycosylated or extensively post-translationally modified form, bacteria are often used as expression system (Terpe, 2006). Bacteria have several advantages. It is an easy system of culture, and very rapid cell growth, the expression can be induced easily using inducer and the purification is quite simple in prokaryotic expression systems. *Escherichia coli* is the most commonly used expression host of a diverse range of proteins. It also has a long pedigree of safe use in laboratories and industry (Huang et al., 2012). *E. coli* is a cost-effective host for recombinant protein production with well characterized physiologically and metabolically (Baneyx, 1999). In an ideal situation, the recombinant protein is expressed from a strong promoter, highly soluble, and recovered in high yield and activity. Unfortunately, it is quite common that the overexpressed recombinant protein is either detrimental to the cell or simply compartmentalized into insoluble inclusion bodies

(Ventura and Villaverde, 2006). A variety of vectors and methods are available to facilitate cloning and expression of target genes to improve the solubility of the target protein in *E. coli*. The effect of medium composition, choice of expression host strains, fusion tags, timing of induction, temperature and duration of induction, inducer concentration and co-expressed in the presence of chaperones should be considered (Papaneophytou and Kontopidis, 2014). Currently, approximately 30% of biopharmaceutical proteins on the market that had been approved by the FDA/EMEA are recombinant protein (Ferrer-Miralles et al., 2009). Many studies have successfully used *E. coli* to produce several human growth factors and cytokines such as epidermal growth factor (Abdull Razis., 2008), basic fibroblast growth factor (bFGF) (Gasparian et al., 2009), interleukin-29 (Le and He, 2006), interleukin-24 (Yang et al., 2007) and leukemia inhibitory factor (Gearing et al., 1989; Samal et al., 1995; Tomala et al., 2010; Imsoonthornruksa et al., 2011). As an example, leukemia inhibitory factor (LIF) is an activator of signal transducer that activates several other activators and appears to be sufficient for maintenance of the mouse ESC pluripotency (Smith et al., 1998). LIF is one of the major costs in stem cell research. Imsoonthornruksa and his colleagues demonstrated that biologically active human LIF fusion proteins could easily be obtained from soluble fraction of *E. coli* protein using single step purification. The study described a simple, quick, cost-effective method to produce recombinant active human LIF that can reduce the cost for medium preparation of approximately 20-40% (Imsoonthornruksa et al., 2011). Another growth factor, bFGF which potently promotes human ESC self-renewal and necessary to supplement into culture to maintain the undifferentiated of human ESCs and iPSCs (Thomson et al., 1998; Amit et al., 2000; Xu et al., 2005). In human ESCs, exogenous bFGF activates the ERK MAPK pathway, which is important for maintaining pluripotency and viability of the human ESCs and iPSCs (Ding et al.,

2011; Kinehara et al., 2013). Inhibition of this signaling transduction pathway results in stem cell differentiation (Li et al., 2007). Several hosts such as *E. coli*, *Pichia pastoris*, insect cell, *S. cerevisiae*, *Bacillus subtilis*, soybean seeds, silkworm (*Bombyx mori L.)* and rice seed have been used for the expressions of recombinant hbFGF with the yield of about 1-100 mg of protein per liter of cell culture (Squires et al., 1988; Hill and Crane-Robinson, 1995; Wu et al., 2001; Ding et al., 2006; Mu et al., 2008; Gasparian et al., 2009; Song et al., 2013; Rassouli et al., 2013; Kwong et al., 2013). The Rassouli's study indicated that there is not a significant difference in the function of in-house produced and commercialized bFGF. This study provided a rapid, cost effective purification of a soluble and biologically active human bFGF protein that can be used in human ESC and iPSC culture (Rassouli et al., 2013). Although, *E. coli* expression system has been remain the first choice, with a number of the advantages. One of the limitations in using *E. coli* relates to the lipopolysaccharide (LPS), also referred to as endotoxin, which is a component of the outer membrane of virtually all Gram-negative bacteria. Such toxic component requires extensive, and expensive, removal during protein purification. However, the studies on hematopoietic stem cells, mesenchymal stem cells and human ESCs have revealed that only traces of endotoxin induce serious adverse effects on stem cell behavior and function (Rinehart and Keville, 1997; Cho et al., 2006; Brandau et al., 2010). Endotoxin removal methods are varied such as ultrafiltration, Triton X phase separation, activated carbon, anion exchange chromatography and immobilized sepharose. The use of these strategies often results in significant yield reduction, increased cost or a loss of the bioactivity of the recombinant proteins. However, *E. coli* produced recombinant proteins are still widely being used, endotoxin contamination stills an important issue for the scientific community. To date, a new *E. coli* strain with a

genetically modified LPS molecule has been created that enables protein expression without the endotoxin (Mamat et al., 2015).

In conclusion, development of the optimal cell culture system for propagation and differentiation of stem cells is highly dependent on growth factor and cytokine supplements. The activity and purity of these growth factors and cytokines is crucial for the study of stem cell biology as well as for future clinical applications. The biological activity testing is also critical point to determine the success of recombinant protein production. Specific cell type have been used to test the biological activity of each growth factor or cytokine such as human leukemic cell line, TF-1 response to human LIF and NIH3T3 cells response to bFGF. If chose the wrong cell type it may led to a false positive or negative. The ability to produce in-house growth factors and cytokines from the published platforms can be reduced the cost in stem cell cultures, accounting for approximately 20-60% depending on the concentration of the growth factors or cytokines produced.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.8 Composition and price of human ESC feeder-dependent culture medium.

Ingredient	Vendor/Cat. No./Price ($)	Final Concentration	Final volume (ml)	Price ($)/500 ml medium
DMEM-F12 (1x; 500ml)	Invitrogen/21331-020/30	1x	387	23.22
Knockout serum replacement (100ml)	Invitrogen/10828-028/68	5%	25	17.00
FBS (500ml)	Hyclone/SH30088.03/325	15%	75	48.75
Non-essential amino acids (100x; 100ml)	Invitrogen/11140-050/26	1x	5	1.30
L-glutamine (200mM; 100ml)	Invitrogen/25030-081/20	1mM	2.5	0.50
β-mercaptoethanol (14.3M; 10ml)	Sigma/M6250/27	0.1mM	0.003	0.01
bFGF (25 ug)	R&D systems/233-FB/185	8 ng/ml	0.16	29.60
Penicillin-streptomycin (100x; 100ml)	Invitrogen/10378-016/25	1x	5	1.25

2.6 References

Aasen, T., Raya, A., Barrero, M. J., Garreta, E., Consiglio, A., Gonzalez, F., Vassena, R., Bilić, J., Pekarik, V., Tiscornia, G., Edel, M., Boué, S., and Izpisúa Belmonte, J.C. (2008). Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. **Nat. Biotechnol.** 26: 1276-1284.

Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. **Nature.** 422: 198-207.

Al-Hakim, A.K., Bashkurov, M., Gingras, A.C., Durocher, D., and Pelletier, L. (2012). Interaction proteomics identify NEURL4 and the HECT E3 ligase HERC2 as novel modulators of centrosome architecture. **Mol. Cell Proteomics.** 11: M111.014233.

Amit, M., Carpenter, M. K., Inokuma, M. S., Chiu, C. P., Harris, C. P., Waknitz, M. A., Itskovitz-Eldor, J., and Thomson, J. A. (2000). Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. **Dev. Biol.** 227: 271-278.

Amit, M., Margulets, V., Segev, H., Shariki, K., Laevsky, I., Coleman, R., and Itskovitz-Eldor, J. (2003). Human feeder layers for human embryonic stem cells. **Biol. Reprod.** 68:2150-2156.

Artlett, C. M., Smith, J. B., and Jimenez, S. A. (1998). Identification of fetal DNA and cells in skin lesions from women with systemic sclerosis. **N. Engl. J. Med.** 338: 1186-1191.

Bader, S., Kuhner, S., and Gavin, A. C. (2008). Interaction networks for systems biology. **FEBS Lett.** 582: 1220-1224.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. **Curr. Opin. Biotechnol**. 10: 411-421.

Bartel, P. L., Roecklein, J. A., SenGupta, D., and Fields, S. A. (1996). Protein linkage map of *Escherichia coli* bacteriophage T7. **Nat. Genet.** 12: 72-77.

Berggard, T., Linse, S., and James, P. (2007). Methods for the detection and analysis of protein-protein interactions. **Proteomics.** 7: 2833-2842.

Blackstock WP, Weir MP (1999) Proteomics: quantitative and physical mapping of cellular proteins. **Trends Biotechnol.** 17:121-127

Blau, H. M., Brazelton, T. R., and Weimann, J. M. The evolving concept of a stem cell: entity or function? **Cell.** 105: 829-841.

Brandau, S., Jakob, M., Hemeda, H., Bruderek, K., Janeschik, S., Bootz, F., and Lang, S. (2010). Tissue-resident mesenchymal stem cells attract peripheral blood neutrophils and enhance their inflammatory activity in response to microbial challenge. **J. Leukoc. Biol.** 88: 1005-10015.

Braun, P., and Gingras, A. C. (2012). History of protein-protein interactions: from egg-white to complex networks. **Proteomics.** 12: 1478-1498.

Brizzard, B. (2008). Epitope tagging. **Biotechniques.** 44: 693-695.

Bullejos, M., Díaz de la Guardia, R., Barragán, M. J., and Sánchez, A. (2000). HMG-box sequences from microbats homologous to the human SOX30 HMG-box. **Genetica.** 110: 157-162.

Cabantous, S., Terwilliger, T. C., and Waldo, G. S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. **Nat. Biotechnol.** 23: 102-107.

Chalkley, R. J., Baker, P. R., Medzihradszky, K. F., Lynn, A. J., and Burlingame, A. L. (2008). In-depth analysis of tandem mass spectrometry data from disparate instrument types. Mol. Cell. **Proteomics.** 7: 2386-2398.

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

Chapman-Smith, A., and Cronan, J. E. Jr. (1999). The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. **Trends Biochem. Sci.** 24: 359-363.

Charbonnier, S., Gallego, O., and Gavin, A. C. (2008). The social network of a cell: recent advances in interactome mapping**. Biotechnol. Annu. Rev.** 14, 1-28

Chimutengwende-Gordon, M., and Khan, W. S. (2012). Advances in the use of stem cells and tissue engineering applications in bone repair. **Curr. Stem Cell Res. Ther.** 7: 122-126.

Cho, H. H., Bae, Y. C., and Jung, J. S. (2006). Role of toll-like receptors on human adipose-derived stromal cells. **Stem Cells.** 24: 2744-2752.

Costa, Y., Ding, J., Theunissen, T. W., Faiola, F., Hore, T. A., Shliaha, P. V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., Das, S., Levasseur, D. N., Li, Z., Xu, M., Reik, W., Silva, J. C., and Wang, J.. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. **Nature.** 495: 370-374.

Coulombel, L. (2004). Identification of hematopoietic stem /progenitor cells: strength and drawbacks of functional assays. **Oncogene.** 23: 7210-7222.

Craig, R., and Beavis, R. C. (2004). TANDEM: matching proteins with tandem mass spectra. **Bioinformatics.** 20: 1466-1467.

Dahéron, L., Opitz, S. L., Zaehres, H., Lensch, M. W., Andrews, P. W., Itskovitz-Eldor, J., and Daley, G. Q. (2004). LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. **Stem Cells.** 22:770-778.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. **Proc Natl Acad Sci USA.** 100: 7480-7485.

Ding, J., Xu, H., Faiola, F., Ma'ayan, A., and Wang, J. (2012). OCT4 links multiple epigenetic pathways to the pluripotency network. **Cell Res.** 22: 155-167.

Ding, S. H., Huang, L. Y., Wang, Y. D., Sun, H. C., and Xiang, Z. H. (2006). High-level expression of basic fibroblast growth factor in transgenic soybean seeds and characterization of its biological activity. **Biotechnol. Lett.** 28: 869-875.

Ding, V. M., Boersema, P. J., Foong, L. Y., Preisinger, C., Koh, G., Natarajan, S., Lee, D. Y., Boekhorst, J., Snel, B., Lemeer, S., Heck, A. J., and Choo, A. (2011). Tyrosine phosphorylation profiling in FGF-2 stimulated human embryonic stem cells. **PLoS One.** 6: e17538.

Drewes, G., and Bouwmeester, T. (2003). Global approaches to protein-protein interactions. **Curr Opin Cell Biol.**15: 199-205.

Dunham, W. H., Mullin, M., and Gingras, A. C. (2012). Affinity-purification coupled to mass spectrometry: basic principles and strategies. **Proteomics.** 12: 1576-1590.

Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C. S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. **Plant J.** 45: 616-629.

Eng, J. K., McCormack, A. L., and Yates, J. R. (1994).An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database**.** **J. Am. Soc. Mass Spectrom.** 5: 976-989.

Evans. M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. **Nature.** 292: 154-156.

Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E., and Villaverde, A. (2009). Microbial factories forrecombinant pharmaceuticals. **Microb. Cell Fact.** 8, 17.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein–protein interactions. **Nature.** 340: 245-246

Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C., Jacq, B., Arpin, M., Bellaiche, Y., Bellusci, S., Benaroch, P., Bornens, M., Chanet, R., Chavrier, P., Delattre, O., Doye, V., Fehon, R., Faye, G., Galli, T., Girault, J. A., Goud, B., de Gunzburg, J., Johannes, L., Junier, M.P., Mirouse, V., Mukherjee, A., Papadopoulo, D., Perez, F., Plessis, A., Rosse, C., Saule, S., Stoppa-Lyonnet, D., Vincent, A., White, M., Legrain, P., Wojcik, J., Camonis, J., and Daviet, L. (2005). A Protein interaction mapping: a Drosophila case study. **Genome Res.** 15: 376-384.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. **Proc. Jpn. Acad. Ser.** B85: 348-362.

Gagliardi, A., Mullin, N. P., Ying Tan, Z., Colby, D., Kousa, A. I., Halbritter, F., Weiss, J. T., Felker, A., Bezstarosti, K., Favaro, R., Demmers, J. Nicolis, S. K., Tomlinson, S. R., Poot, R. A., and Chambers, I. (2013). A direct physical interaction between NANOG and SOX2 regulates embryonic stem cell self-renewal. **EMBO J.** 32: 2231-2247.

Gagliardi, A., Mullin, N. P., Ying Tan, Z., Colby, D., Kousa, A. I., Halbritter, F., Weiss, J. T., Felker, A., Bezstarosti, K., Favaro, R., Demmers, J., Nicolis, S. K., Tomlinson, S. R., Poot, R. A., and Chambers, I. (2013). A direct physical

interaction between NANOG and SOX2 regulates embryonic stem cell self-renewal. **EMBO J.** 32: 2231-2247.

Galarneau, A., Primeau, M., Trudeau, L.E. and Michnick, S.W. (2002) Betalactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein protein interactions. **Nat. Biotechnol.** 20: 619-622.

Gao, Z., Cox, J. L., Gilmore, J. M., Ormsbee, B. D., Mallanna, S. K., Washburn, M. P., and Rizzino, A. (2012). Determination of protein interactome of transcription factor SOX2 in embryonic stem cells engineered for inducible expression of four reprogramming factors**. J. Biol. Chem.** 287: 11384-11397.

Gasparian, M. E., Elistratov, P. A., Drize, N. I., Nifontova, I. N., Dolgikh, D. A., and Kirpichnikov, M. P. (2009). Overexpression in Escherichia coli and purification of human fibroblast growth factor (FGF-2). **Biochemistry (Mosc).** 74: 221-225.

Gavin, A. C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat. C. .M, Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. **Nature.** 415: 141-147.

Ginis, I. Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M. K., Itskovitz-Eldor, J., and Rao, M. S. (2004). Differences between human and mouse embryonic stem cells. **Dev. Biol.** 269: 360-380.

Goel, A., Colcher, D., Koo, J. S., Booth, B. J., Pavlinkova, G., and Batra, S. K. (2000). Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. **Biochim. Biophys. Acta.**1523: 1320.

Graf, T. (2002). Differentiation plasticity of hematopoietic cells. Blood. 99: 3089-3101.

Haase, A., Olmer, R., Schwanke, K., Wunderlich, S., Merkert, S., Hess, C., Zweigerdt, R., Gruh, I., Meyer, J., Wagner, S., Maier, L. S., Han, D. W., Glage, S., Miller, K., Fischer, P., Schöler, H. R., and Martin, U. (2009). Generation of induced pluripotent stem cells from human cord blood. **Cell Stem Cell.** 434-441.

Hills, D., and Crane-Robinson, C. (1995). Baculovirus expression of human basic fibroblast growth factor from a synthetic gene: role of the Kozak consensus and comparison with bacterial expression. **Biochim. Biophys. Acta.** 1260: 14-20.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sørensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., and Tyers, M. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. **Nature.** 415: 180-183.

Hofemeister, H., Ciotta, G., Fu, J., Seibert, P. M., Schulz, A., Maresca, M., Sarov, M., Anastassiadis, K., Stewart, A. F. (2011). Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. **Methods.** 53: 437-452.

Hu, C. D., and Kerppola, T. K. (2003). Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. **Nat. Biotechnol.** 21: 539-545.

Huang, C. J., Lin, H., and Yang, X. (2012). Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. **J. Ind. Microbiol. Biotechnol.** 39: 383-399.

Huang, X., and Wang, J. (2014). The extended pluripotency protein interactome and its links to reprogramming. **Curr. Opin. Genet. Dev.** 28: 16-24.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A. E., and Melton, D. A. (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. **Nat. Biotechnol.** 26, 795-797.

Humphries, A., Graham, T. A., and McDonald, S. A. (2011). Stem cells and inflammation in the intestine. **Recent Results Cancer Res.** 185: 51-63.

Imsoonthornruksa, S., Noisa, P., Parnpai, R., Ketudat-Cairns, M., A simple method for production and purification of soluble and biologically active recombinant human leukemia inhibitory factor (hLIF) fusion protein in *Escherichia coli.* **J. Biotechnol.** 2011, 151, 295-302.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. A. (2001). Comprehensive two-hybrid analysis to explore the yeast protein interactome. **Proc. Natl. Acad. Sci. USA.** 98: 4569-4574.

Jarvik, J. W., and Telmer, C. A. (1998). Epitope tagging. **Annu. Rev. Genet.** 32: 601-618.

Kar, G., Gursoy, A., and Keskin, O. (2009). Human cancer protein-protein interaction network: a structural perspective. **PLoS Comput. Biol.** 5: e1000601.

Kim, D., Kim, C. H., Moon, J. I., Chung, Y. G., Chang, M. Y., Han, B. S., Ko, S., Yang, E., Cha, K. Y., Lanza, R., and Kim, K. S. (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. **Cell Stem Cell.** 4: 472-476.

Kim, J. B., Greber, B., Araúzo-Bravo, M. J., Meyer, J., Park, K. I., Zaehres, H., and Schöler, H. R. (2009). Direct reprogramming of human neural stem cells by OCT4. **Nature.** 461: 649-643.

Kim, J., Cantor, A. B., Orkin, S. H., and Wang, J. (2009). Use of *in vivo* biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. **Nat. Protoc.** 4: 506-517.

Kinehara, M., Kawamura, S., Tateyama, D., Suga, M., Matsumura, H., Mimura, S., Hirayama, N., Hirata, M., Uchio-Yamada, K., Kohara, A., Yanagihara, K., and Furue, M. K. (2013). Protein kinase C regulates human pluripotent stem cell self-renewal. **PLoS One.** 8: e54122.

Koivisto, H., Hyvärinen, M., Strömberg, A. M., Inzunza, J., Matilainen, E., Mikkola, M., Hovatta, O., and Teerijoki, H. (2004)Cultures of human embryonic stem cells: serum replacement medium or serum-containing media and the effect of basic fibroblast growth factor. **Reprod. Biomed. Online.** 9: 330-337.

Kolios, G., and Moodley, Y. (2013). Introduction to stem cells and regenerative medicine. **Respiration.** 85: 3-10.

Kopp, J. L., Ormsbee, B. D., Desler, M., and Rizzino, A. (2008). Small increases in the level of SOX2 trigger the differentiation of mouse embryonic stem cells. **Stem Cells.** 26: 903-911.

Korbling, M., and Estrov, Z. (2003). Adult stem cells fortissue repair – a new therapeutic concept? **N. Engl. J. Med.** 349: 570-582.

Kunisato, A., Wakatsuki, M., Shinba, H., Ota, T., Ishida, I., and Nagao, K. (2011). Direct generation of induced pluripotent stem cells from human nonmobilized blood. **Stem Cells Dev.** 20: 159-168.

Kwong, K. W., Ng, K. L., Lam, C. C., Wang, Y. Y., and Wong, W. K. (2013). Authentic human basic fibroblast growth factor produced by secretion in *Bacillus subtilis.* **Appl. Microbiol. Biotechnol.** 97: 6803-6811.

Lai, Y. S., Chang, C. W., Pawlik, K. M., Zhou, D., Renfrow, M. B., and Townes, T. M. (2012). SRY (sex determining region Y)-box2 (SOX2)/poly ADP-ribose polymerase 1 (Parp1) complexes regulate pluripotency. **Proc. Natl. Acad. Sci. USA.** 109: 3772-3777.

Li, J., Wang, G., Wang, C., Zhao, Y., Zhang, H., Tan, Z., Song, Z., Ding, M., and Deng, H. (2007). MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. **Differentiation.** 75: 299-307.

Li, Y. (2010). Commonly used tag combinations for tandem affinity purification. **Biotechnol. Appl. Biochem.** 55: 73-83.

Li. W., Wei, W., Zhu, S., Zhu, J., Shi, Y., Lin, T., Hao, E., Hayek, A., Deng, H., and Ding, S. (2009). Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. **Cell Stem Cell.** 4: 16-9.

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S. Y., Qin, J., Wong, J., Cooney, A. J., Liu, D., and Songyang, Z. (2008). NANOG and OCT4 associate with unique transcriptional repression complexes in embryonic stem cells. **Nat. Cell Biol.** 10: 731-739.

Liu, H., Ye, Z., Kim, Y., Sharkis, S., and Jang, Y. Y. (2010). Generation of endoderm-derived human induced pluripotent stem cells from primary hepatocytes. **Hepatology.** 51: 1810-1819.

Magnusdottir, A., Vidarsson, H., Björnsson, J. M., and Örvar, B. L. (2013). Barley grains for the production of endotoxin-free growth factors. **Trends Biotechnol.** 31: 572-580.

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., and Hochedlinger, K. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. **Cell Stem Cell.** 3: 340-345.

Mamat, U., Wilke, K., Bramhill, D., Schromm, A. B., Lindner, B., Kohl, T. A., Corchero, J. L., Villaverde, A., Schaffer. L., Head, S. R., Souvignier, C., Meredith, T. C., and Woodard, R. W. (2015). Erratum to: Detoxifying *Escherichia coli* for endotoxin-free production of recombinant proteins. **Microb. Cell. Fact.** 14: 81.

Martí, M., Mulero, L., Pardo, C., Morera, C., Carrió, M., Laricchia-Robbio, L., Esteban, C. R., and Izpisua Belmonte, J.C. (2013). Characterization of pluripotent stem cells. **Nat. Protoc.** 8: 223-253.

Masters, S. C. (2004). Co-immunoprecipitation from transfected cells. **Methods Mol. Biol.** 261: 337-350.

Matsuoka, S., Tsuji, K., Hisakawa, H., Xu, M. j., Ebihara, Y., Ishii, T., Sugiyama, D., Manabe, A., Tanaka, R., Ikeda, Y., Asano, S., and Nakahata, T. (2001). Generation of definitive hematopoietic stem cells from murine early yolk sac and paraaortic planchnopleures by aorta-gonad-mesonephros region-derived stromal cells. **Blood.** 98: 6-12.

McCormick, J. B., and Huso, H. A. (2010). Stem cells and ethics: current issues. **J. Cardiovasc. Transl. Res.** 3: 122-127.

Medzihradszky, K.F. (2005). In-solution digestion of proteins for mass spectrometry. **Methods Enzymol.** 405: 50-65.

Mellacheruvu, D., Wright. Z., Couzens, A. L., Lambert, J. P., St-Denis, N. A., Li, T., Miteva, Y. V., Hauri, S., Sardiu, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N. C., Al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault, M., Lin, Z. Y., Badillo, B. G., Pawson, T., Durocher, D., Coulombe, B., Aebersold, R., Superti-Furga, G., Colinge, J., Heck, A. J., Choi, H., Gstaiger, M., Mohammed, S., Cristea, I. M., Bennett, K. L., Washburn, M. P., Raught, B., Ewing, R. M., Gingras, A. C., Nesvizhskii, A. I. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. **Nat. Methods.** 10: 730-736.

Miernyk, J. A., and Thelen, J. J. (2008). Biochemical approaches for discovering protein-protein interactions**. Plant J.** 53: 597-609.

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

Moorman, N. J., Sharon-Friling, R., Shenk, T. and Cristea, I. M. A. (2010). targeted spatial-temporal proteomics approach implicates multiple cellular trafficking pathways in human cytomegalovirus virion maturation. **Mol. Cell. Proteomics.** 9: 851–860.

Mu, X., Kong, N., Chen, W., Zhang, T., Shen, M., and Yan, W. (2008). High-level expression, purification, and characterization of recombinant human basic fibroblast growth factor in pichia pastoris. **Protein Expres. Purif.** 59: 282-288.

Musunuru, K. (2013). Genome editing of human pluripotent stem cells to generate human cellular disease models. **Dis. Model Mech.** 6: 896-904.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor OCT4. **Cell.** 95: 379-391.

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

Obradovic, S., Rusovic, S., Balint, B., Ristic-Andelkov, A., Romanovic, R., Baskot, B., Vojvodic, D., and Gligic, B. (2004). Autologous bone marrow-derived progenitor cell transplantation for myocardial regeneration after acute infarction. **Vojnosanit. Pregl.** 61: 519-529.

Obrdlik, P., El-Bakkoury, M., Hamacher, T., Cappellaro, C., Vilarino, C., Fleischer, C., Ellerbrok, H., Kamuzinzi, R., Ledent, V., Blaudez, D., Sanders, D., Revuelta, J. L., Boles, E., Andre, B., and Frommer, W.B. (2004). K+ channel interactions detected by a genetic system optimized forsystematic studies of membrane protein interactions. **Proc. Natl. Acad. Sci. USA.** 101: 12242-12247.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germ-line competent induced pluripotent stem cells. **Nature.** 448: 313-317.

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T., and Yamanaka, S. (2008). Generation of mouse induced pluripotent stem cells without viral vectors. **Science.** 322: 949-953.

Olsen, J.V., Ong, S. E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. **Mol. Cell Proteomics.** 3: 608-614.

Papaneophytou, C. P., and Kontopidis, G. (2014). Statistical approaches to maximize recombinant protein expression in *Escherichia coli*: a general review. **Protein Expr. Purif.** 94: 22-32.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M. M., and Choudhary, J. (2010). An expanded OCT4 interaction network: implications for stem cell biology, development, and disease. **Cell Stem Cell.** 6: 382-395.

Park, Y., Kim, J. H., Lee, S. J., Choi, I. Y., Park, S. J., Lee, S. R., Sung, H. J., Yoo, Y. D., Geum, D. H., Choi, C. W., Kim, S. H., and Kimm, B. S. (2011). Human feeder cells can support the undifferentiated growth of human and mouse embryonic stem cells using their own basic fibroblast growth factors. **Stem Cells Dev.** 20: 1901-1910.

Pelletier, J. N., Campbell-Valois, F. X., and Michnick, S. W. (1998). Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. **Proc. Natl. Acad. Sci. USA.** 95: 12141-12146.

Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. **Electrophoresis.** 20: 3551-3567.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification**. Methods.** 24: 218-229.

Ransone, L. J. (1995). Detection of protein-protein interactions by coimmunoprecipitation and dimerization. **Methods Enzymol.** 254: 491-497.

Rassouli, H., Tabe Bordbar, M. S., Rezaei Larijani, M., Pakzad, M., Baharvand, H., and Salekdeh, G. H. (2013), Cloning, expression and functional characterization of in-house prepared human basic fibroblast growth factor. **Cell J.** 14: 282-291.

Remy, I., and Michnick, S. W. (2001). Visualization of biochemical networks in living cells. **Proc. Natl. Acad. Sci. USA.** 98: 7678-7683.

Richards, M., Fong, C. Y., Chan, W. K., Wong, P. C., and Bongso, A. (2002). Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells**. Nat. Biotechnol.** 20:933-936.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. A. (1999). Generic protein purification method for protein complex characterization and proteome exploration. **Nat. Biotechnol.** 17: 1030−1032.

Rinehart, J. J. and Keville, L. (1997). Effects of endotoxin on proliferation of human hematopoietic cell precursors. **Cytotechnology.** 24: 153-159.

Rodriguez, P., Braun, H., Kolodziej, K. E., de Boer, E., Campbell, J., Bonte, E., Grosveld, F., Philipsen, S., Strouboulis, J. (2006). Isolation of transcription factor complexes by *in vivo* biotinylation tagging and direct binding to streptavidin beads. **Methods Mol. Biol.** 338: 305-323.

Roque, A. C., and Lowe, C. R. (2008). Affinity chromatography: history, perspectives, limitations and prospects. **Methods Mol. Biol.** 2008, 421: 1-21.

Rual, J. F.,Venkatesan, K. Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F.P., and Vidal, M. (2005). Towards a proteome-scale map of the human protein-protein interaction network. **Nature.** 437: 1173-1178.

Ruiz, S., Brennand, K., Panopoulos, A. D., Herrerías, A., Gage, F. H., and Izpisua-Belmonte, J. C. (2010). High-efficient generation of induced pluripotent stem cells from human astrocytes. **PLoS One.** 5: e15526.

Rumlova, M., Benedikova, J., Cubinkova, R., Pichova, I. Ruml, T. (2001). Comparison of classical and affinity purification techniques of Mason-Pfizer monkey virus capsid protein: the alteration of the product by an affinity tag. **Protein Expr. Purif.** 23: 75-83.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A. H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. **Nat. Med.** 10: 55-63.

Seet, B. T., Dikic, I., Zhou, M. M., and Pawson, T. (2006) Reading protein modifications with interaction domains. **Nat. Rev. Mol. Cell Biol** 7: 473-483.

Seki, T., Yuasa, S., Oda, M., Egashira, T., Yae, K., Kusumoto, D., Nakata, H., Tohyama, S., Hashimoto, H., Kodaira, M., Okada, Y., Seimiya, H., Fusaki, N., Hasegawa, M., and Fukuda, K. (2010). Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. **Cell Stem Cell.** 7: 11-14.

Slack, J. M. W. (2000). Stem cells in epithelial tissues. Science. 287: 1431-1433.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., and Rogers, D.(1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. **Nature.** 336: 688-690.

Somers, A., Jean, J. C., Sommer, C. A., Omari, A., Ford, C. C., Mills, J. A., Ying, L., Sommer, A. G., Jean, J. M., Smith, B. W., Lafyatis. R., Demierre, M. F., Weiss, D. J., French, D. L., Gadue, P., Murphy, G. J., Mostoslavsky, G., Kotton, D. N. (2010). Generation of transgene-free lung disease-specific human induced

pluripotent stem cells using a single excisable lentiviral stem cell cassette. **Stem Cells.** 28: 1728-1740.

Song, B., Niclis, J. C., Alikhan, M. A., Sakkal, S., Sylvain, A., Kerr, P. G., Laslett, A. L., Bernard, C. A., and Ricardo, S. D. (2011). Generation of induced pluripotent stem cells from human kidney mesangial cells. **J. Am. Soc. Nephrol.** 22: 1213-1220.

Song, J. A., Koo, B. K., Chong, S. H., Kwak, J., Ryu, H. B., Nguyen, M. T., Vu, T. T., Jeong, B., Kim, S. W., and Choe, H. (2013). Expression and purification of biologically active human FGF2 containing the b'a' domains of human PDI in *Escherichia coli. Appl. Biochem. Biotechnol.* 170: 67-80.

Squires, C. H., Childs, J., Eisenberg, S. P., Polverini, P. J., and Sommer, A. (1988). Production and characterization of human basic fibroblast growth factor from *Escherichia coli.* **J. Biol. Chem.** 263: 16297-16302.

Stadtfeld, M., and Hochedlinger, K. (2010). Induced pluripotency: history, mechanisms, and applications. **Genes Dev.** 24: 2239-2263.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. (2005). A human protein-protein interaction network: a resource for annotating the proteome. **Cell.** 122: 957-968.

Stevens, R. C. (2000). Design of high-throughput methods of protein production for structural biology. **Structure.** 8: R177-185.

Sugii, S., Kida, Y., Kawamura, T., Suzuki, J., Vassena, R., Yin, Y. Q., Lutz, M. K., Berggren, W. T., Izpisúa Belmonte, J. C., and Evans, R. M. (2010). Human and

mouse adipose-derived cells support feeder-independent induction of pluripotent stem cells. **Proc. Natl. Acad. Sci. USA.** 107: 3558-3563.

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

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

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. **Appl. Microbiol. Biotechnol.** 72: 211-222

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. **Science.** 282: 1145-1147.

Tomala, M., Lavrentieva, A., Moretti, P., Rinas, U., Kasper, C., Stahl, F., Schambach, A., Warlich, E., Martin, U., Cantz, T., and Scheper, T. (2010) Preparation of bioactive soluble human leukemia inhibitory factor from recombinant *Escherichia coli* using thioredoxin as fusion partner. **Protein Express Purif.** 73: 51-57.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M. Fields, S., and Rothberg, J. M. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. **Nature.** 403: 623-627.

Utikal, J., Maherali, N., Kulalert, W., and Hochedlinger, K. (2009). SOX2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. **J. Cell Sci.** 22: 3502-3510.

van den Berg, D. L., Snoek ,T., Mullin, N. P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R. A. (2010). An OCT4-centered protein interaction network in embryonic stem cells. **Cell Stem Cell.** 6: 369-381.

van den Berg, D. L., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R. A. (2008).Estrogen-related receptor beta interacts with OCT4 to positively regulate NANOG gene expression. **Mol Cell Biol.** 28: 5986-5995.

Ventura, S., and Villaverde, A.(2006). Protein quality in bacterial inclusion bodies. **Trends Biotechnol.** 24: 179-185.

Villalobos, V., Naik, S., and Piwnica-Worms, D. (2007). Current state of imaging protein-protein interactions *in vivo* with genetically encoded reporters. **Annu. Rev. Biomed.** Eng. 9: 321-349.

Walsh, G. (2010). Post-translational modifications of protein biopharmaceuticals. **Drug Discov. Today** 15: 773-780.

Wang, J. (2012). Deciphering protein complexes and protein interaction networks for stem cell pluripotency. **In New Frontiers of Network Analysis in Systems Biology.** Edited by Ma'ayan A, MacArthur BD. Netherlands: Springer; pp97-118

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T.W., and Orkin, S. H. (2006). A protein interaction network for pluripotency of embryonic stem cells. **Nature.** 444, 364-368.

Warren, L., Manos, P. D., Ahfeldt, T., Loh, Y. H., Li, H., Lau, F., Ebina, W., Mandal, P. K., Smith, Z. D., Meissner, A., Daley, G. Q., Brack, A. S., Collins, J. J., Cowan,

C., Schlaeger, T. M., and Rossi, D.J. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. **Cell Stem Cell.** 7: 618-630.

Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., and Humphery-Smith, I. (1995). Progress with geneproduct mapping of the Mollicutes: Mycoplasma genitalium. **Electrophoresis.** 16: 1090-1094.

Westermarck, J., Ivaska, J., and Corthals, G. L. (2013). Identification of protein interactions involved in cellular signaling. **Mol. Cell. Proteomics.** 12: 1752-1763.

Woltjen, K., Michael, I. P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H. K., and Nagy, A. (2009). piggyback transposition reprograms fibroblasts to induced pluripotent stem cells. **Nature.** 458: 766-770.

Wu, X., Kamei, K., Sato, H., Sato, S., Takano, R., Ichida, M., Mori, H.,and Hara, S. (2001). High-level expression of human acidic fibroblast growth factor and basic fibroblast growth factor in silkworm (Bombyx mori L.) using recombinant baculovirus. **Protein Expres. Purif.** 21: 192-200.

Xu, R. H., Peck, R. M., Li, D. S., Feng, X., Ludwig, T., and Thomson, J. A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. **Nat. Methods.** 2: 185-190.

Yao, S., Chen, S., Clark, J., Hao, E., Beattie, G. M., Hayek, A., and Ding, S. (2006). Long-term self-renewal and directed differentiation of human embryonic stem cells in chemically defined conditions. **Proc. Natl. Acad. Sci. USA.** 103: 6907-6912.

Ye, Z., Zhan, H., Mali, P., Dowey, S., Williams, D. M., Jang, Y. Y., Dang, C. V., Spivak, J. L., Moliterno, A. R., and Cheng, L. (2009). Human-induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders. **Blood.** 114: 5473-5480.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. **Science.** 324: 797-801.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. **Science**. 318: 1917-1920.

Zhou, T., Benda, C., Dunzinger, S., Huang, Y., Ho, J. C., Yang, J., Wang, Y., Zhang, Y., Zhuang, Q., Li, Y., Bao, X., Tse, H. F., Grillari, J., Grillari-Voglauer, R., Pei, D., and Esteban, M. A. (2012). Generation of human induced pluripotent stem cells from urine samples. **Nat. Protoc.** 7: 2080-2089.

Zhou, W., and Freed, C. R. (2009). Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells. **Stem Cells.** 27: 2667-2674.

CHAPTER III

AN INTERACTOME OF OCT4 TRANSCRIPTION FACTOR IN HUMAN EMBRYONIC STEM CELLS

3.1 Abstract

The knowledge of the molecular mechanism of human embryonic stem cells (ESCs) pluripotency is essential for manipulation and differentiation of ESCs. OCT4 is one of the key factors in the regulation of pluripotency. To provide the answer of how OCT4 transcription factor is central to human stem cell pluripotency, an affinity-tagged endogenous OCT4 cell line was established via TALEN-mediated gene targeting in human ESCs to express FLAG-Biotin-tagged OCT4. OCT4 protein complexes were purified and identified by affinity purification combined with liquid chromatography-tandem mass spectrometry (LC-MS/MS). A number of OCT4 interacting proteins that involved in many processes such as transcription factors and epigenetic regulators that play important roles in maintaining pluripotency were found. Our results showed that SFPQ and OCT4 act as a transcriptional regulatory module that is critical for pluripotency capacity of human ESCs. This study provided the first survey of OCT4 interactome in human ESCs and it also uncovered new candidate proteins which may function significantly in human stem cell pluripotency and self-renewal.

3.2 Introduction

Self-renewal ability and pluripotency property are two main characters that define embryonic stem cells (ESCs) (Evans and Kaufman, 1981). Both human and mouse ESCs

have therapeutic potential value in regenerative medicine (Prelle et al., 2002). The undifferentiated state of ESCs is maintained by a set of transcription factors including OCT4, SOX2 and NANOG which are the core pluripotency factors for human and mouse ESCs (Orikin, 2005). Pluripotency factors regulated not only the expression of other essential genes but also regulate their own expression by auto- regulatory feedback loops (Boer et al., 2007; Ding et al., 2012). However, the mechanisms that regulate these core factor protein levels have not yet been clarified. The POU transcription factor OCT4 is a central player in ESCs self-renewal and differentiation into specific lineages. It is recognized to be essential for both *in vivo* and *in vitro* early development and maintenance of pluripotency. The loss of OCT4 in mouse embryo causes the failure of inner cell mass development (Nichols et al., 1998). ESCs are sensitive to dosage alterations of OCT4. Overexpression or repression in OCT4 level causes differentiation toward primitive endoderm and mesoderm or trophectoderm lineage, respectively (Niwa et al., 2000), possibly due to the direct repression of the NANOG promoter by excessive OCT4 (Pan et al., 2006). It would be necessary to understand the interaction of transcription factors and epigenetic cofactors that are critical for maintaining as well as directly differentiating pluripotent stem cells. Analyzing the protein interaction network and their regulation can provide useful clues for understanding protein functions. Recently, proteomic approaches have been used to characterize the network of protein-protein interactions controlling pluripotent stem cell functions. Wang and his colleagues were the first to report a protein interaction network surrounding the transcription factor NANOG in mouse ESCs. They suggested that some NANOG interactors may assist NANOG to mediate gene regulation and established connections to several epigenetic regulatory pathways such as histone deacetylase Mi-2/NuRD, PRC1 and SWI/SNF chromatin remodeling complexes (Wang et al., 2006). An OCT4 interactome was also

reported in independent studies of van den Berg et al. (2010), Pardo et al. (2010) and Ding et al. (2012). These results demonstrated that the OCT4 interactors are co-regulated and specifically downregulated upon ESC differentiation. They are also involved in NuRD, SWI/SNF and LSD1 protein complexes which are consistent with the finding in the NANOG interactome. OCT4 binding partners include transcription factors with important roles in ESCs, for examples SALL1, SALL4, HCFC1 and HELLS. These findings suggest that OCT4 is essential for integrating the epigenetic machinery into the pluripotency network and that NANOG and OCT4 interactomes are inherently connected. Recent reports focused almost exclusively on the pluripotency interactome of transcription factors and epigenetic regulators in mouse ESCs (Huang and Wang, 2014). The pluripotency-centered protein interaction network has not been fully explored in human ESCs. To elucidate the human ESC transcriptional network, here we surveyed the OCT4-associated proteins using affinity purification and mass spectrometry approaches to isolate protein complexes and identified a protein interaction network surrounding OCT4. Stable isotope labeling with amino acids in cell culture (SILAC) system was performed to increase the protein identification sensitivity and differentiate it from the nonspecific interacting proteins. Here, *in vivo* biotinylation was applied to human ESCs and biotin tagged OCT4 was recovered from nuclear extracts with one-step streptavidin purification and subjected to LC-MS/MS. The results demonstrated that many proteins in the human OCT4 interactome are components of complexes involved in epigenetic regulation and chromatin remodeling. They include transcription factors essential in self-renewal and differentiation of pluripotency stem cells. Our data here provide a resource that can be used to map the pluripotency landscape of human ESCs and to explore the basic principles underlying stem cell biology.

3.3 Materials and Methods

3.3.1 Cell culture

Human ESCs (Whitehead Institute for Biomedical Research (WIBR) line #29, Lengner et al., 2010) were maintained on mitomycin C inactivated mouse embryonic fibroblast (MEF) feeder layers and cultured in DMEM/F12 (Invitrogen) supplemented with 15% fetal bovine serum (FBS) (Hyclone), 5% KnockOutTM Serum Replacement (KSR) (Invitrogen), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma), 1% penicillin-streptomycin (Invitrogen) and 4 ng/ml human bFGF (R&D systems) at 37°C under 5% CO_2 and 5% O_2 in air (label-free experiment). Cultures were enzymatically passaged every 5 to 7 days with 1.5 mg/ml collagenase type IV (Invitrogen). For the Stable Isotope Labeling by Amino acids in Cell culture (SILAC) labeling experiment, human ESCs were grown for three passages in DMEM/F12 SILAC media with corresponding complete supplements but deficient in both L-lysine and L-arginine (Thermo Fisher Scientific) and supplemented with normal L-arginine and L-lysine (SILAC light) (Thermo Fisher Scientific) or 13C615N4 L-arginine, and 13C615N2 L-lysine (SILAC heavy) (Cambridge Isotope Laboratories). The medium was supplemented with 10% dialyzed FBS for SILAC (Thermo Fisher Scientific), 5% KSR, 1 mM glutamine, 1% nonessential amino acids, 0.1 mM β-mercaptoethanol, 1% penicillin-streptomycin and 4 ng/ml human bFGF.

3.3.2 Targeting of human ESCs using transcription activator-like effector nuclease (TALEN) mediated homologous recombination

AAVS1 and *OCT4* TALEN plasmids were designed and assembled as previously described by Hockemeyer et al., 2011. In this study, *AAVS1* and *OCT4* loci were engineered for gene targeting with biotin ligase BirA and FLAG-Biotin tag to the

intron 1 of *AAVS1* and the N-terminus of *OCT4* loci, respectively (Figure 3.1). Plasmid containing cDNA encoding the biotin ligase of *E. coli* BirA was a generous gift from Dr. Jianlong Wang (Department of Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, NY, USA). The *AAVS1* and *OCT4* targeting vectors were constructed by PCR amplifying arms of homology from the corresponding loci using genomic DNA isolated from human ESCs. The homology arms followed by either biotin ligase of *E. coli* BirA or 3xFLAG-Biotin sequences (Figure 3.1C) were then cloned into pCR®2.1-TOPO vector using standard cloning methods.

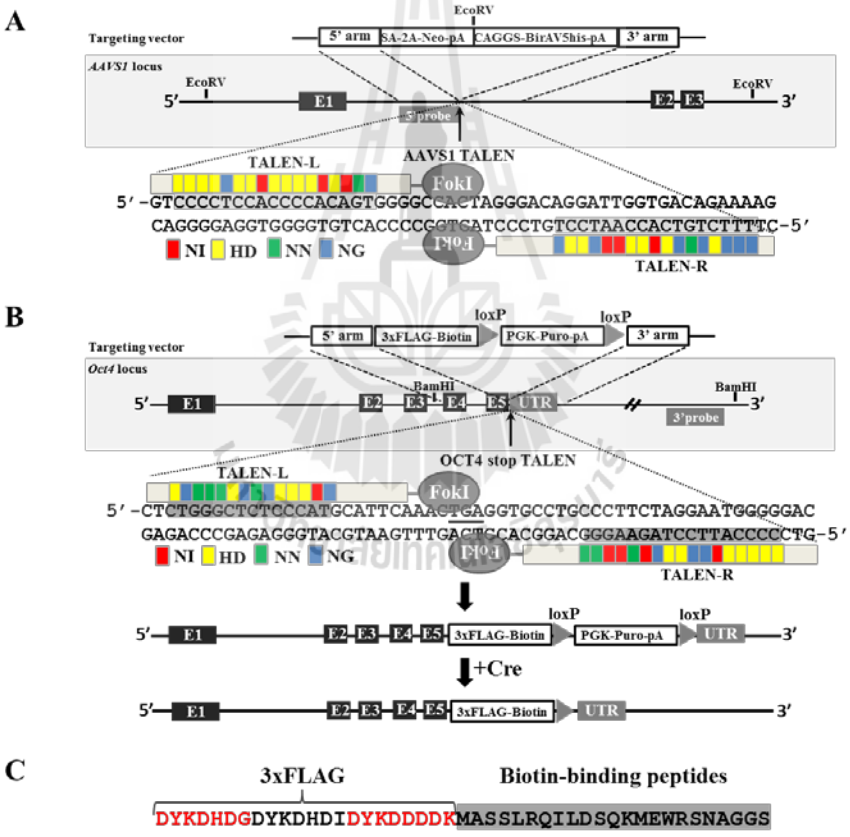

Figure 3.1 Schematic overview depicting the targeting strategy for the *AAVS1* and *OCT4* loci using TALENs directed against the first intron of *AAVS1* (A) and the exon 5 of *OCT4* (B). The binding sites for the TALEN pair used in this study are highlighted. The different donor vectors to targeting the *AAVS1* and *OCT4*

loci were indicated. (C) The sequences of triple FLAG tag and biotin-binding peptide used in this study (modified from Hockemeyer et al., 2011).

3.3.2.1 Gene targeting of biotin ligase BirA to the *AAVS1* locus in human ESCs

Human ESCs were cultured in Rho Kinase (ROCK)-inhibitor (Calbiochem) 24 h prior to electroporation. Cells were harvested using 0.25% trypsin/EDTA solution (Invitrogen) and resuspended in phosphate buffered saline (PBS). Ten million cells were electroporated with 40 μg of targeting vectors and 5 μg of each *AAVS1* TALEN encoding plasmid using Gene Pulser Xcell System, Bio-Rad: 250 V, 500 μF, 0.4 cm cuvettes. Cells were subsequently plated on DR4 MEFs feeder layer with neomycin for selection in human ESC medium supplemented with ROCK-inhibitor for the first 24 h. Individual colonies were picked and expanded after neomycin selection (300 μg/ml) 10 to 14 days after electroporation. Gene targeting analysis was verified by Southern blotting.

3.3.2.2 TALEN-mediated biotinylation of endogenous OCT4 in AAVS1-BirA human ESCs

AAVS1-BirA human ESCs, named WIBR#29BirA were cultured in ROCK-inhibitor 24 h prior to electroporation. Cells were harvested using 0.25% trypsin/EDTA solution and resuspended in PBS. The cells were electroporated with 40 μg of targeting vectors and 5 μg of each OCT4 TALEN encoding plasmid. Cells were subsequently plated on DR4 MEFs feeder layer under puromycin selection in human ESC medium supplemented with ROCK-inhibitor for the first 24 h. Individual colonies were picked and expanded after puromycin selection (0.5 μg/ml) 10 to 14 days after

electroporation. Correct targeted clones were confirmed by Southern blot analysis (EcoRV digested).

3.3.2.3 Removal of PGK-Puro cassette by transient Cre-recombinase expression

WIBR#29BirA with 3xFLAG-Biotin and PGK-Puro cassette in the *OCT4* locus was cultured in ROCK-inhibitor 24 hours prior to electroporation. Cells were harvested using 0.25% trypsin/EDTA solution. Ten million cells were resuspended in PBS and co-transfected with pTurbo-Cre (40 μg; GenBank accession number AF334827) and pEGFP-N1 (10 μg; Clontech) by electroporation. Cells were subsequently plated on MEF feeder layers in human ESC medium supplemented with ROCK inhibitor for the first 24 h. Cre-recombinase-expressing cells were selected with FACS sorting (FACS-Aria; BD Biosciences) of a single cell suspension for EGFP-expressing cells 60 h after electroporation followed by replating at a low density in ROCK inhibitor containing human ESC medium. Individual colonies were picked 10 to 14 days after electroporation. The removal of PGK-Puro cassette and correctly targeted clones, named WIBR#29OCT4FLAG-Bio were confirmed by Southern blot (BamHI digested).

3.3.3 Southern blot

Genomic DNA was extracted from WIBR#29, WIBR#29BirA and WIBR#29OCT4FLAG-Bio by tail lysis buffer (10 mM Tris-Cl (pH8.0), 100 mM NaCl, 10 mM EDTA, 0.5% SDS and 0.4 mg/ml proteinase K) and incubated overnight at 37°C. The next morning an equal volume of isopropyl alcohol was added and centrifuged at 7500 rpm for 5 min. The supernatant was discarded and pellet resuspended in 1 ml of 70% ethanol and centrifuged at 7500 rpm for 5 min. The supernatant was carefully decanted and the pellet was air dried. After that, the pellet was resuspended in 200 μl TE

buffer. 5 μg aliquots were digested overnight with the appropriate enzymes (EcoRV and BamHI for AAVS1 and OCT4 targeting experiments, respectively) at 37°C. EcoRV or BamHI digested genomic DNA was separated on a 0.8% agarose gel followed by capillary transfer to a Hybond N^+ membrane (Amersham Biosciences). For making an AAVS1-specific probe, the targeting vector was double digested with SacI and EcoRI to generate a 643 bp fragment of 5' arm of AAVS1. For OCT4-specific probe, the targeting vector was digested with SbfI and NheI to generate a 697 bp fragment of 5' arm of OCT4. These DNA fragments were labeled by using [$\propto$-^{32}P]dCTP and Prime-It II Random Primer Labeling Kit (Agilent Technologies) according manufacturer's protocol. Hybridization was carried out overnight at 65°C. Post-hybridization washes were done sequentially with two standard saline citrate (SSC) (20xSSC; 3M NaCl in 0.3M sodium citrate (pH 7.0)), 2x SSC and 0.2x SSC along with 0.2% SDS, each for 20 min at 65°C. Blots were exposed at -80°C with X-ray film in film cassette for 24h. Then, the films were developed by an auto-processor in a darkroom.

3.3.4 Immunostaining

Cells in 12-well plate were fixed in PBS supplemented with 4% paraformaldehyde for 20 min at room temperature. The cells were then permeabilized and blocked using 0.2% Triton X-100, 0.1% Tween-20 and 3% donkey serum in PBS for 30 min. Primary antibody against human OCT4 (1:500, mouse monoclonal, Santa Cruz Biotechnology), human SOX2 (1:500, goat polyclonal, R&D Systems) and human NANOG (1:250, goat polyclonal, R&D Systems) was diluted in 0.2% Triton X-100 in PBS and incubated with the samples overnight at 4°C. The cells were treated with appropriate Molecular Probes Alexa Fluor® dye conjugated secondary antibodies (1:500, Invitrogen) and then incubated for 1 h. The nuclei were stained with DAPI for 10 min.

3.3.5 Teratoma formation

WIBR#29BirA and WIBR#29OCT4FLAG-Bio colonies were collected by collagenase treatment and separated from feeder cells by subsequent washes with medium and sedimentation by gravity. The cells were resuspended in 250 µl of PBS and injected subcutaneously in SCID mice. Tumors generally developed within 4-8 weeks and animals were sacrificed before tumor size exceeded 1.5 cm in diameter. Teratomas were isolated after sacrificing the mice and fixed in formalin. After sectioning, teratomas were diagnosed based on hematoxylin and eosin staining.

3.3.6 Western Blot

Cell pellets were lysed on ice in Lysis Buffer (10 mM Tris-HCl (pH 7.4), 2% SDS, 1 mM DTT, 10% glycerol, and 120 mg/ml urea) for 30 min in presence of Complete EDTA-free Protease Inhibitor Cocktail (Roche Diagnostics) and then boiled for 10 min at 95ºC. Samples were fractionated by SDS-PAGE on 4-12% precast minigels (Invitrogen) and electrophoretically transferred to nitrocellulose membranes (GE Healthcare) by wet electrotransfer using a tank blot device (TransBlot, system, Bio-Rad) at 100 volt for 2 h. Blots were blocked with 10% non-fat milk powder in PBS containing 0.1%Tween 20 for 1h at room temperature. Primary antibodies (1:500) used were mouse human OCT4 (mouse monoclonal, Santa Cruz), FLAG (mouse monoclonal-HRP Conjugated, Sigma), Streptavidin-HRP Conjugate (GE Healthcare), human GAPDH (rabbit polyclonal, Proteintech), at concentrations as recommended (1:5,000). Blots were incubated with primary antibodies overnight at 4°C. Subsequently the blots were incubated with anti-mouse or anti-rabbit pig IgG-HRP (H+L) secondary antibody for OCT4 and GAPDH antibodies, respectively (EMD Biosciences) for 1 h at room

temperature. Amersham ECL Western Blotting Detection Reagent (GE Healthcare) was used for visualization.

3.3.7 Affinity purification of OCT4 associated protein complexes

WIBR#29OCT4FLAG-Bio and WIBR#29BirA were cultured in SILAC medium for two passages to incorporate the heavy and light amino acids, respectively. Then, the WIBR#29OCT4FLAG-Bio containing biotin-tagged OCT4 transcription factor and control WIBR#29BirA cells were expanded to fifteen 15 cm diameter dishes in SILAC conditions. Duplicates were performed for SA agarose-based affinity purification of SILAC conditions, one of FLAG antibody pull down (label-free) and one of endogenous OCT4 antibody pull down (label-free). The colonies were collected by collagenase treatment and separated from feeder cells by subsequent washes with medium and sedimentation by gravity. Nuclear extraction and affinity purification of biotin-tagged OCT4-associated protein complexes were performed as previously described (Kim et al., 2009). Briefly, the cell pellets were resuspended in ice-cold nuclear extract buffer A (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.2 mM PMSF and protease inhibitor cocktail (Roche)) and incubated for 10 min on ice. The sample was centrifuged at 4,300 ×g for 5 min at 4°C and the pellet containing nuclei was washed by resuspending with 3 ml of ice-cold nuclear extract buffer A and centrifuging at 25,000 x g for 20 min at 4°C. Then, nuclei were resuspended with 3 ml of ice-cold nuclear extract buffer B (20 mM HEPES, pH 7.9, 25% glycerol (v/v), 0.42 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, and protease inhibitor cocktail) and incubated at 4°C for 30 min. Insoluble materials were pelleted by centrifugation at 25,000 x g for 20 min at 4°C. The supernatant was collected as nuclear extract (NE) and dialyzed against buffer C (20 mM HEPES, pH 7.9, 20% glycerol (v/v),

100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF) at 4°C for 1.5 h. For SILAC experiment, mix equal amounts of total protein from the NE of WIBR#29Bir A and the NE of WIBR#29OCT4FLAG-Bio samples were performed, approximately 100-200 mg of total protein. Then, 1 ml of Protein G agarose (Roche Diagnostic) equilibrated in buffer C containing 0.02% NP40 was added to 3 ml of nuclear extract in 50 ml tubes (BD Falcon) and incubated overnight at 4°C, the extract in the presence of 750 units of Benzonase (Novagen) with continuous mixing to pre-clear. Precleared extract was then transferred to the already equilibrated (with buffer C) streptavidin-agarose beads (Invitrogen), and rotated for 6 h at 4°C. Beads were washed five times for 15 min each with buffer D containing 0.02% NP40. Bound material was eluted by boiling for 5 min in Laemmli buffer, fractionated on a 10% SDS-PAGE, stained with the GelCode™ Blue Safe Protein Stain buffer (Thermo). The gel lanes were horizontally cut into 8-10 pieces and each piece was subjected to digestion with porcine trypsin (Promega) as previously described (Shevchenko et al., 2006). The resulting peptides from each piece were dried down and analyzed by LC-MS/MS

มหาวิทยาลัยเทคโนโลยีสุรนารี

3.3.8 Mass spectroscopy

The samples were reconstituted in 5-10 μl of HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nano-scale reverse-phase HPLC capillary column was created by packing 5μm C18 spherical silica beads into a fused silica capillary (100 μm inner diameter x 12 cm length) with a flame-draw tip. After equilibrating the column each sample was loaded onto the column. A gradient of acetonitrile from 2.5% to 97.5% was used to elute the peptides. As peptides eluted, they were subjected to electrospray ionization and then they entered into a LTQ-Orbitrap-Velos mass spectrometer (ThermoFinnigan) with collision-induced dissociation (CID). Eluting peptide were

detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. MS data were processed by Thermo Proteome Discoverer software with SEQUEST engine and using the Mascot search engine with the IPI human protein sequence database. The biological processing and molecular function of identified proteins were assessed using Gene Ontology database (http://www.geneontology.org).

3.3.9 Verification of OCT4 interactions by co-immunoprecipitation (co-IP)

To validate OCT4-associated proteins in human ESCs, nuclear extracts were prepared from WIBR#29OCT4FLAG-Bio or WIBR#29BirA. Endogenous BEND3, FUS, RBM25, SFPQ and WDR82 were immunoprecipitated with pre-bound BEND3, FUS, RBM25, SFPQ and WDR82 antibodies and rabbit IgG beads were used as a control. Co-IP OCT4 was identified by western blotting with SA-HRP and anti-OCT4 antibodies.

3.3.10 Functional validation of OCT4 interactome by lentivirus expressing shRNAs

VSV-glycoprotein pseudotyped lentiviral vector particles were produced by co-transfection of HEK293T cells with packaging and envelope plasmids and either pLKO.1-shRNA plasmids that targeted to the mRNA of *BEND3*, *FUS*, *RBM25*, *SFPQ* or *WDR82* coding sequences. The shRNA plasmids were a generous gift from Dr. Jianlong Wang. Viral supernatants were harvested after 48 h and filtered through a 0.45 μm membrane. Human ESCs were infected with the lentivirus expressing shRNAs against *BEND3*, *FUS*, *RBM25*, *SFPQ* or *WDR82* in the presence of 6 μg/ml polybrene (Millipore). One day after transduction, cultures were switched to human ESC medium supplemented with ROCK-inhibitor. The next day, the cells were trypsinized and re-seeded onto 6-well plate with mitomycin C inactivated DR4 feeder cells. Infected cells were selected with 1.5 μg/ml of puromycin at day five after transduction and selection

was maintained for 10 days. Colony morphology and proliferation rate were used to determine the effects of gene expression knockdown.

3.4 Results and Discussion

3.4.1 Establishment of transgenic human ESC lines for *in vivo* biotinylation and biotin-tagged OCT4

In this study, *E. coli* BirA biotin ligase gene was inserted into adeno-associated virus site 1 (*AAVS1*) locus of human ESC line WIBR#29 to establish *in vivo* biotinylation cell line, named WIBR#29BirA. Then, the cells expressing BirA biotin ligase were used for targeting of endogenous *OCT4* with biotin-binding peptide sequences at the lysine side chain within the peptides (Figure 3.1C) and named WIBR#29OCT4FLAG-Bio. The *AAVS1* locus is a safe-harbor region for the integration of foreign DNA in the human genome. The transgene can function without perturbation of endogenous gene activity and no phenotypic effects were observed (Sadelain et al., 2011). Transcription activator-like effector nucleases (TALENs) have been used to target genome engineering of human pluripotent stem cells. The TALEN pairs were designed to target the human *AAVS1* and *OCT4* loci (Hockemeyer et al., 2011). Here, *AAVS1* and *OCT4*-TALENs from Hockemeyer study were applied for genome targeting. We targeted the *AAVS1* gene in the first intron of WIBR#29 cells with a gene trapping cassette consisting of a splice acceptor (SA) site immediately upstream of a neomycin resistance (neo) selectable marker gene followed *E. coli* BirA biotin ligase gene that was driven by CAGGS-promoter. The WIBR#29BirA cells were then targeted at the last codon of the *OCT4* gene with targeting vector that generated an in-frame fusion of exon 5 with the triple FLAG tag and biotin-binding peptide (Figure 3.1C), followed by a loxP-flanked puromycin (Puro) gene expressed from the constitutive PGK promoter (PGK-Puro). This

cell was named WIBR#29FLAG-Bio-PGK-Puro.The correctly targeted clones were electroporated with Cre recombinase plasmid. Sixty hours after electroporation, GFP positive cells were isolated using FACS. The looping out of PGK-Puro strategy minimizes any potential impact on the endogenous protein. Due to the concern that the insertion of a drug resistance cassette may interfere with the expression of the targeted gene or neighboring genes, it is necessary to remove the drug-resistance cassette for proper targeting gene expression (Zhu et al., 2015). Southern blot analysis showed correct *AAVS1* and *OCT4* targeted clones and the removal of the PGK-Puro cassette (Figure 3.2A, 3.3A and 3.3C). The efficient targeting of the human *AAVS1* and *OCT4* loci by TALENs was showed in Table 3.1. The efficiency of TALEN-mediated *AAVS1* and *OCT4* targeting in human ESCs was 58% (7/12) and 3.5% (2/57), respectively. All of the clones were targeted only in one allele (heterozygous). The targeting efficiencies were comparable to that of a previous study using similar targeting strategies (Hockemeyer et al., 2011).

Table 3.1 Summary of TALEN gene targeting efficiency.

Cell line targeted	TALEN pair	Donor plasmid	No. of clones analyzed	Targeted + Additional integration	Correct targeted clones		
					Heterozygous	Homozygous	Targeting efficiency (%)
WIBR#29	AAVS1	SA-2A-neo-pA-CAGGS-BirAV5His-pA	12	4	7	0	58
WIBR#29BirA	Oct4Stop	FLAG-Biotin-PGK-Puro-pA	57	0	2	0	3.5

Gene targeting of human ESCs for creating homologous recombination knock-in alleles is still considered inefficient (Goulburn et al., 2011; Irion et al., 2007). The development of TALENs dramatically increased the efficiency of genome targeting. TALENs consist of a sequence specific DNA binding domain and a nonspecific DNA

cleavage module, Fok1. The nucleases are composed of active pairs of scissors in which two monomers bind adjacent target sites separated by a DNA spacer which allows for the formation of an active dimer to introduce targeting DNA double strand breaks (DSBs) (Ul Ain et al., 2015). Nuclease-induced DSBs can be repaired by either nonhomologous end-joining (NHEJ) or homology-directed repair (HDR). NHEJ often results in introduction of insertion and/or deletions (Indels) which can disrupt the translational reading frame of a coding sequence that can lead to knock-out of the target gene in human ESCs (Ding et al., 2013). In contrast, HDR can be employed to efficiently incorporate exogenous sequences that can introduce specific point mutations or to insert desired sequences through recombination of the target locus with exogenously supplied DNA. Recent studies demonstrated that TALENs allow for not only efficient gene inactivation through NHEJ but also enhanced HDR-based gene targeting for functional correction of genetic diseases or insertion of the functional gene into a safe harbor locus in human stem cells (Merkert et al., 2014; Hou et al., 2013; Hockemeyer et al., 2011; Soldner et al., 2011; Yusa et al., 2011). In additional, the TALEN system has provided rapid and simple construction and is now available from the non-profit organization Addgene (www.addgene.org) and commercial companies.

The off-target activity of *AAVS1*- and *OCT4*-TALENs was checked by Southern blot analysis (Figure 3.2A and 3.3A and Table 3.1). The results showed 33% of off-targets were present in *AAVS1*-TALENs (4/12). In OCT4 TALENs, no random integration was found. The off-target effects involved non-specific recognition and digestion of non-targeted region by TALENs. This can be toxic to cells, and is difficult to comprehensively predict and presents a common problem for genome editing. It has been reported that TALENs have minimal off-target effects when compared with other programmed nucleases, zinc finger nucleases or clustered regularly interspaced short

palindromic repeat (CRISPR)/CRISPR-associated (Cas) system (Ding et al., 2013; Hisano et al., 2013).

WIBR#29BirA and WIBR#29OCT4FLAG-Bio lines maintained normal human ESC morphologies (Figure 3.4B and C) and remained pluripotent based on expression of the pluripotency markers OCT4, SOX2, and NANOG and they retained the ability to differentiate into all three germ layers by *in vivo* teratoma formation after injection into immunodeficient mice (Figure 3.2B-C and 3.3B-C). This suggested that TALEN-mediated AAVS1 and OCT4 targeting did not perturb the function of the protein.

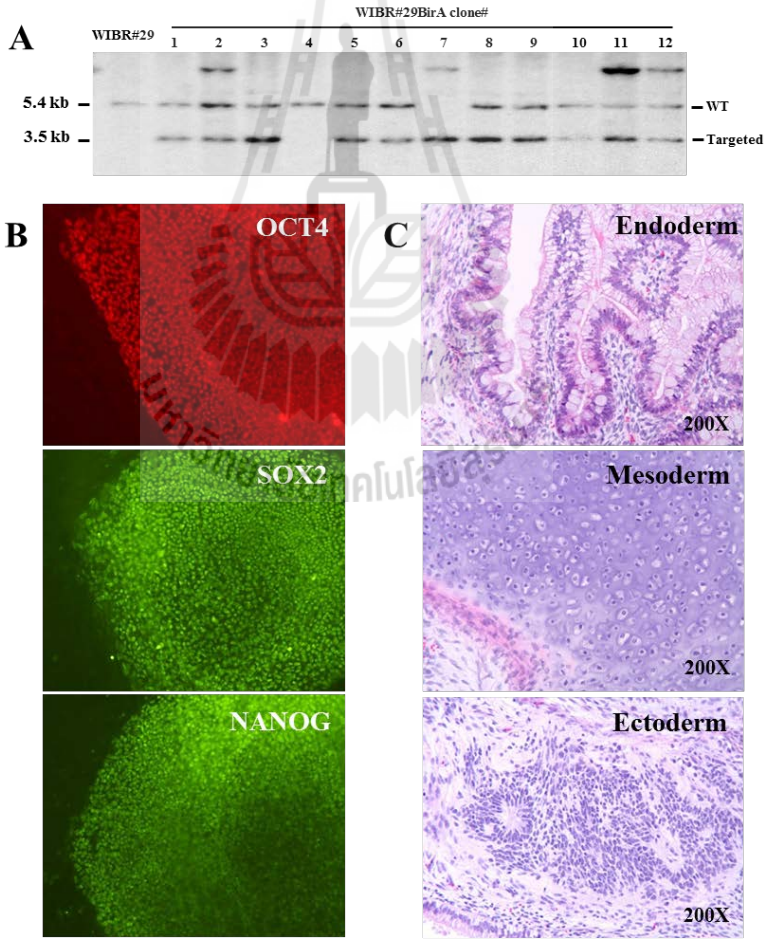

Figure 3.2 Characterization of WIBR#29BirA human ESCs. (A) Southern blot analysis of WIBR#29 human ESC targeted AAVS1 locus with a CAGGS-BirAV5his

pA cassette using TALENs, wild type (WT) = 5.4 kb, targeted = 3.5 kb. (B) Immunofluorescence staining of WIBR#29BirA colony for pluripotency markers OCT4, SOX2, and NANOG. (C) Teratoma formation analysis composed of tissues derived from the three germ layers.

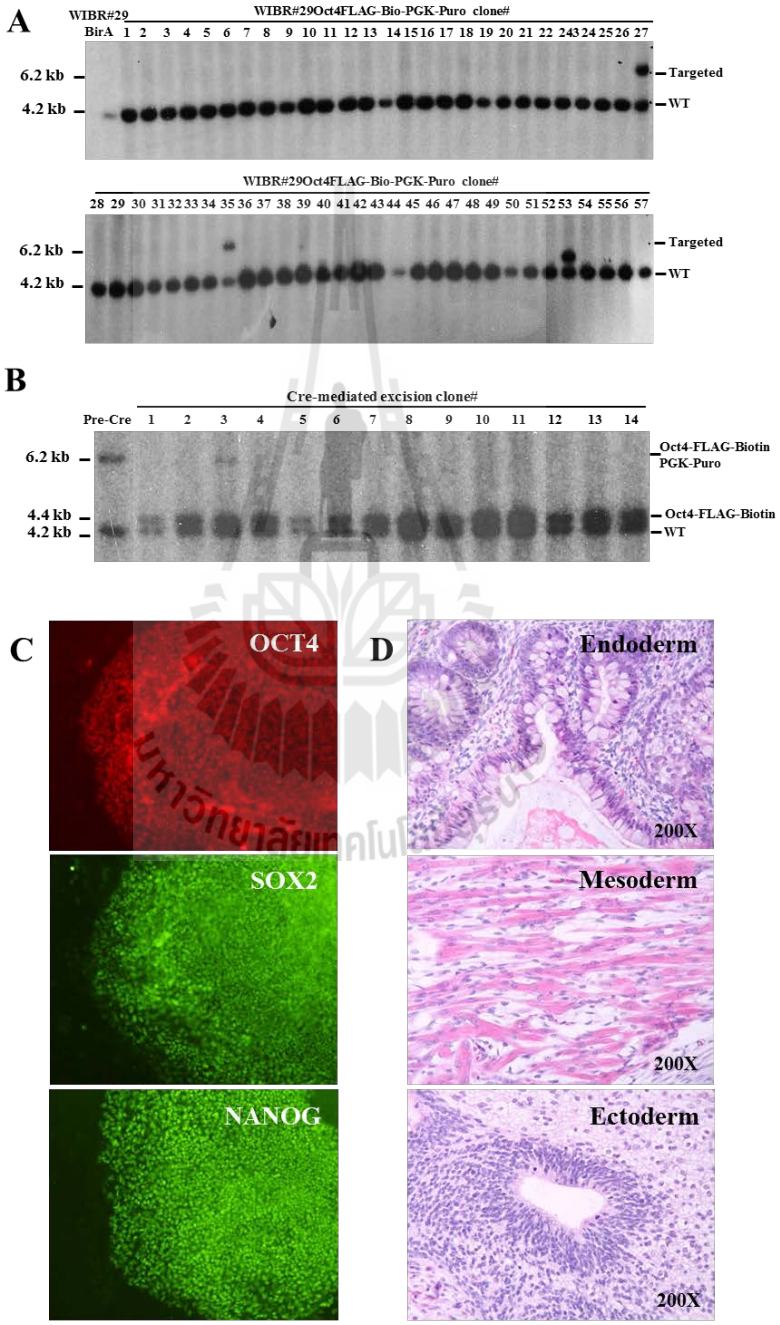

Figure 3.3 Characterization of WIBR#29OCT4FLAG-Bio human ESCs. (A) Southern blot analysis of WIBR#29BirA human ESC targeted in the *OCT4* locus with

a triple-FLAG and biotin-binding peptides and PGK-Puro cassette using TALENs, wild type (WT) = 4.2 kb, targeted = 6.2 kb. (B) Looping out the PGK-Puro cassette in WIBR#29OCT4FLAG-Bio. Pre-Cre control is the WIBR#29OCT4FLAG-Bio-PGK-Puro prior to Cre recombinase plasmid electroporation. Cre-mediated excision results in the removal of the PGK-Puro cassette from the WIBR#29OCT4 FLAG-Bio-PGK-Puro line. (C) Immunofluorescence staining of WIBR#29OCT4 colony for pluripotency markers OCT4, SOX2, and NANOG (40X). (D) Teratoma formation analysis composed of tissues derived from the three germ layers (endoderm, mesoderm and ectoderm).

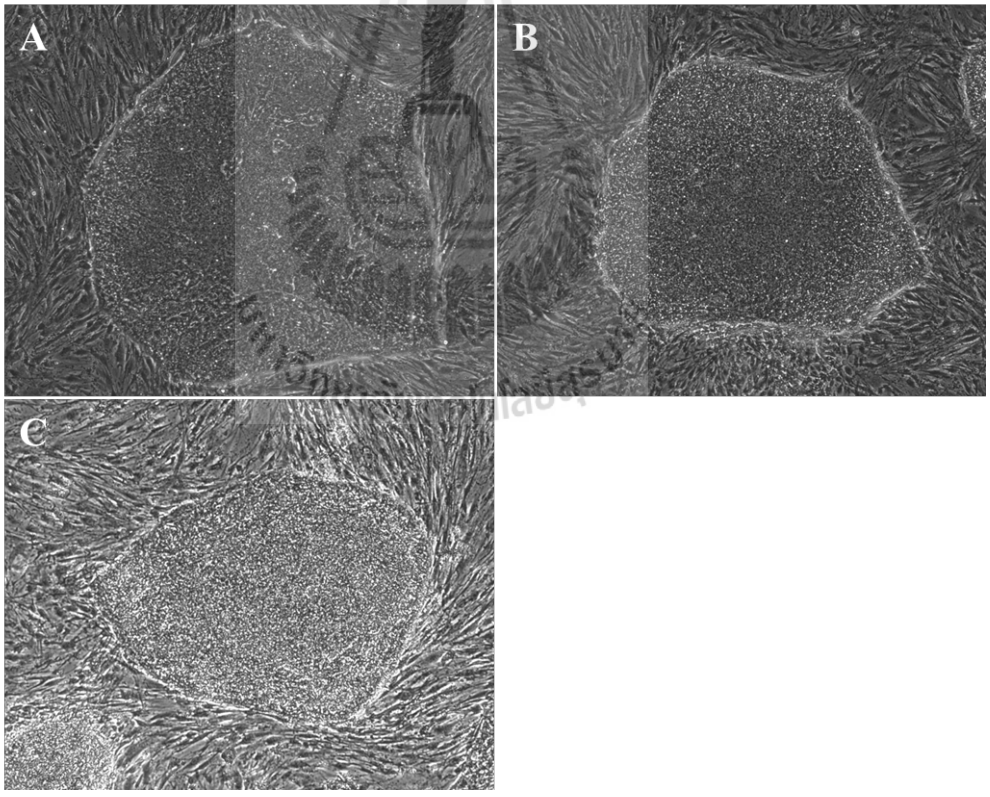

Figure 3.4 Phase contrast photographs of human ESCs (40X). Colonies of undifferentiated WIBR#29 (A), WIBR#29BirA (B) and WIBR#29OCT4 FLAG-Bio (C) grew on mytomycin-inactivated MEFs.

Further western blot analysis was performed to check the expression of biotinylated OCT4 of WIBR#29BirA and WIBR#29OCT4FLAG-Bio using anti-OCT4, anti-FLAG and anti-Biotin antibodies (Figure 3.5), thus indicating efficient biotinylation of the tagged protein. This study obtained only heterozygous biotinylated *OCT4* allele encoding 3xFLAG-Biotin tagged OCT4 (50kDa) and the wild type allele wtOCT4 of 45kDa. The result revealed that the WIBR#29OCT4FLAG-Bio contained biotin tag as a 50 kDa band (lane 2 and 3) which was absent in the WIBR#29BirA having no tag (lane 1). The targeted allele was expressed at lower level than the wild type allele. Previous studies used biotinylation-proteomics approach to identify OCT4 and NANOG interactomes in mouse ESCs. They also obtained expression of biotin tagged transcription factors at about 20% of the endogenous level. However, it was still within sufficient range to recover the biotinylated OCT4 and NANOG from mouse ESC nuclear extract with streptavidin agarose (Ding et al., 2012; Wang et al., 2006). The result demonstrated that targeting and modification of the *OCT4* locus with TALENs results in stable transgene expression and no discernible off-target phenotypes. This evidence indicates specificity and safety of using TALENs for genome editing in human ESCs.

3.4.2 Identification of OCT4 interacting proteins in human ESCs

To survey and identify proteins associated with biotinylated OCT4, complexes were isolated from nuclear extracts of human ESCs expressing either BirA plus biotinylated OCT4 or BirA alone: WIBR#29OCT4FLAG-Bio or WIBR#29BirA, respectively. Here, we used SILAC and label-free conditions to perform a relative quantitation of interactor proteins. Four independent experiments including two experiments of streptavidin pull down (SILAC), one experiment of FLAG antibody pull down (label-free) and one experiment of endogenous OCT4 antibody pull down (label-

free) were performed in duplicates. MS data were analyzed and the spectra were compared to determine peptide abundance in WIBR#29OCT4FLAG-Bio relative to WIBR#29BirA. Proteome Discoverer generated the protein abundance ratios between WIBR#29OCT4FLAG-Bio versus WIBR#29BirA proteins of SILAC (heavy-labeled/ light-labeled) and label-free conditions.

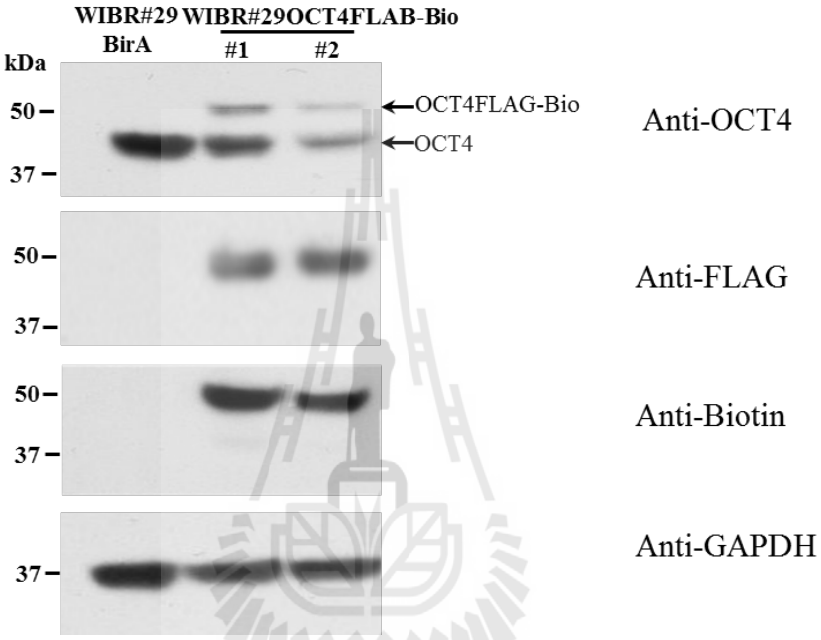

Figure 3.5 Expression of 3xFLAG-Biotin tagged transcription factor OCT4 in WIBR#29BirA, WIBR#29OCT4FLAG-Bio clone # 1 and 2 analyzed by Western blotting against OCT4, FLAG or Biotin antibodies. GAPDH used as loading control.

The ratios with less than 2.0 were removed from the lists. Background proteins (carboxylases and their coenzymes; ribosomal and splicing factors; cytoskeletal proteins such as actins, tubulins and keratins; etc.) were also excluded from the lists. We analyzed four independent experiments and representations of the identified proteins by threshold ratio value of ≥ 2.0 or $\propto (2/0)$ were considered to be in the OCT4 interacting proteins lists (Figure 3.6). The final list of OCT4 interacting proteins, Figure 3.7 and Appendix A,

included only those proteins detected in at least two independent experiments (Figure 3.6) with protein abundance ratio over WIBR#29BirA control at least 2.0 or ∝ (2/0).

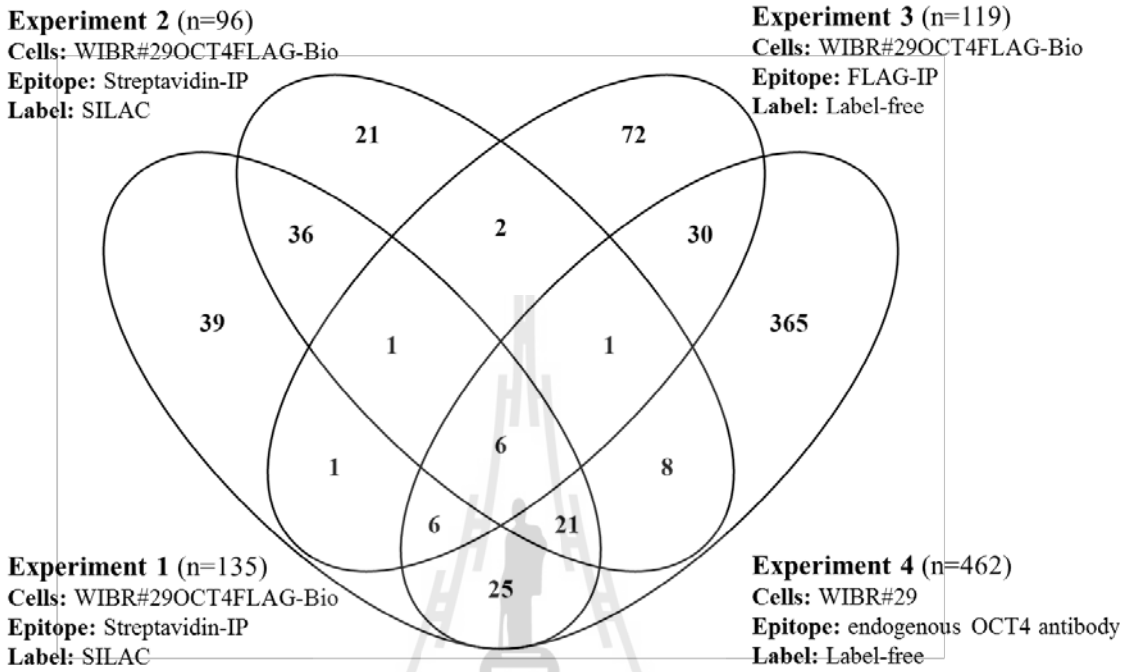

Figure 3.6 The number of OCT4 interacting proteins of the four experiments. The threshold ratio value of 2.0 or ∝ (2/0) of identified proteins was selected for the lists of OCT4 interacting proteins. The overlapping proteins in at least two independent experiments were considered as high-confidencec OCT4 interacting proteins.

From the Figure 3.7, we identified 142 interaction partners of OCT4. We then performed a functional annotation analysis using the PANTHER database (Thomas et al., 2003). The data set was classified into known protein complexes and functional categories. We found an enrichment of gene ontology (GO) terms such as proteins associated with binding activity and catalytic activity involved in transcription regulation, translation regulation, RNA splicing, helicase, kinase, ligase, isomerase, hydrolase, oxidoreductase and transferase activities (Figure 3.8 and Appendix B). This GO analysis

indicates that OCT4 is engaged with multiple cellular mechanisms (Appendix B). OCT4 interacted with many transcription factors (EWSR1, FUS, GTF3C1, SALL1, SALL4, SMARCA1, SOX2, ZFR, ZMYM2, ZMYM3), translation factor activity (ABCF1, DDX17, DDX19A, DDX39B, DDX41, DDX42, DDX47, DDX6, DHX15, EFTUD2, EIF4A1, EIF4A3, GTPBP1, HELLS, SNRNP200, SNRNP70), DNA repair (MSH6, RAD21, XRCC5, XRCC6), RNA processing (ADAR, CPSF1, CSTF3, FUBP3, GIGYF2, IGF2BP1, IGF2BP2, NCBP1, PCBP2, PRPF19, PRPF40A, PRPF6, PTBP2, PUF60, RBM14, RBM39, SF3A1, SF3A3, SF3B1, SFPQ, U2AF2), replication (MCM5, RFC2, RFC3, RFC4, SMC1A, SMC2, SMC3, SMC4) and glycosyl transferase (PARP1). In addition, a number of factors in the OCT4 interactome are connected to several regulatory pathways include the nucleosome remodeling deacetylase (NuRD; CHD4, MTA1, HDAC1, HDAC2, RBBP4), SWI/SNF chromatin remodeling (ACTL6A, SMARCA4, SMARCB1, SMARCC1, SMARCE1) and FACT (SUPT16H) complexes. The activities of the interactors provide a hint of the possible mechanisms that OCT4 employs. Recent studies in mouse ESCs identified the interactome of a number of pluripotency-associated proteins, such as OCT4, SOX2, NANOG, SALL4, ESRRB, DAX1, NAC1, REX1, ZFP281, TCFCP2L1 and ZFP281 (Huang and Wang, 2014). A protein interactome study conducted in human ESCs has not been reported yet. Like the previous studies in mouse ESCs, we observed that OCT4 associates with a number of different proteins to form various functional complexes in the nucleus including transcriptional regulators and chromatin binding proteins involved in DNA replication, recombination and repair. This observation emphasizes the importance of coordination among pluripotency factors and the critical role of OCT4 in the maintenance of a pluripotent state to form auto-regulatory and cross-regulatory loops (Liang et al., 2008; Pardo et al., 2010; van den Berg et al., 2010). OCT4 is a transcription factor expected to

interact with other transcription factors in a modular fashion to effect transcription regulation.

Here, the majority of the OCT4 associated proteins showed binding (GO:00005488; 39%) and catalytic (GO: 0003024; 36%) activities involved transcriptional regulation, including helicases, ligases and RNA processing. This might reflect a tendency of OCT4 to recruit several enzymes to modify themselves or their associated proteins as machinery for gene regulation. Modification of OCT4 activity levels and association with various enzymes gives OCT4 the ability to modify DNA or other proteins (Xu et al., 2004; Wei et al., 2007; Issad and Kuo, 2008). Moreover, OCT4-associated helicases may be recruited to OCT4 mediated transcriptional sites to maintain genome stability. Consistent with this hypothesis was a report that proliferation of ESCs and embryos required a rapid rate of DNA replication and transcription, suggesting that the helicases are needed to keep the genome stable when replication and transcriptional complexes collide (Koledova et al., 2010; Pomerantz and O'Donnell, 2010). NANOG and SOX2 as well are known to be the OCT4 partners that are part of the same protein complexes (Wang et al., 2006; Liang et al., 2008). Here, we did not detect the transcription factor NANOG that might be because the NANOG-OCT4 interactions are too weak, requiring crosslinking for efficient detection, or because NANOG protein is resistant to trypsin digests. Additional, reduced expression of NANOG compared to OCT4 may make it harder to detect NANOG in OCT4 purifications (Zhang et al., 2007; van den Berg et al., 2010; Pardo et al., 2010). SALL4 is a member of Spalt-like family of transcriptional cofactors that also associate with the Mi-2/NuRD complex in mammals (Lauberth and Rauchman, 2006). Yang's study demonstrated that SALL4 is essential for regulation of pluripotency and self-renewal of ESCs through both transcriptional and epigenetic controls including direct interaction with OCT4 and NANOG (Yang et al.,

2010). Our results also revealed that OCT4 associated with the three FET (FUS, EWSR1, and TAF15) families of RNA binding proteins, originally discovered as N-terminal partners of fusion oncoproteins and as components of transcription or splicing complexes. FET-proteins associate with a number of factors involved in transcriptional regulation, pre-mRNA splicing, RNA transport, translation, miRNA processing, and DNA repair (Kovar, 2011). PARP1 and SUPT16H were also identified as associated with OCT4 and recent studies have shown that PARP1 is a multifunctional protein involved in transcriptional regulation, epigenetics, and apoptosis (Wang et al., 2009; Caiafa et al., 2009; Krishnakumar and Kraus, 2010). SUPT16H is a member of FACT complex that is conserved in the chromatin-remodeling complex implicated in DNA replication and regulated transcription. Recently, the FACT complex has been shown to be essential for opening of chromatin and pluripotency of ESCs and for cell reprogramming (Gaspar-Maia et al., 2009). Additionally, OCT4 can both activate and repress transcriptional targets in mouse and human ESCs. Many reports including this study found that OCT4 interactomes contain chromatin-modifying complexes, such as NuRD and SWI/SNF, which may be recruited to prevent differentiation of pluripotent cells by repression of lineage-specific transcription factors. (Wang et al., 2006; Liang et al., 2008; van den Berg et al., 2010; Pardo et al., 2010; Ding et al., 2012). These findings suggest that links between epigenetic modifying complexes and pluripotency transcriptional factors have an essential role in stem cell pluripotency during normal development and cell reprogramming (Ho et al., 2009; Singhal et al., 2010).

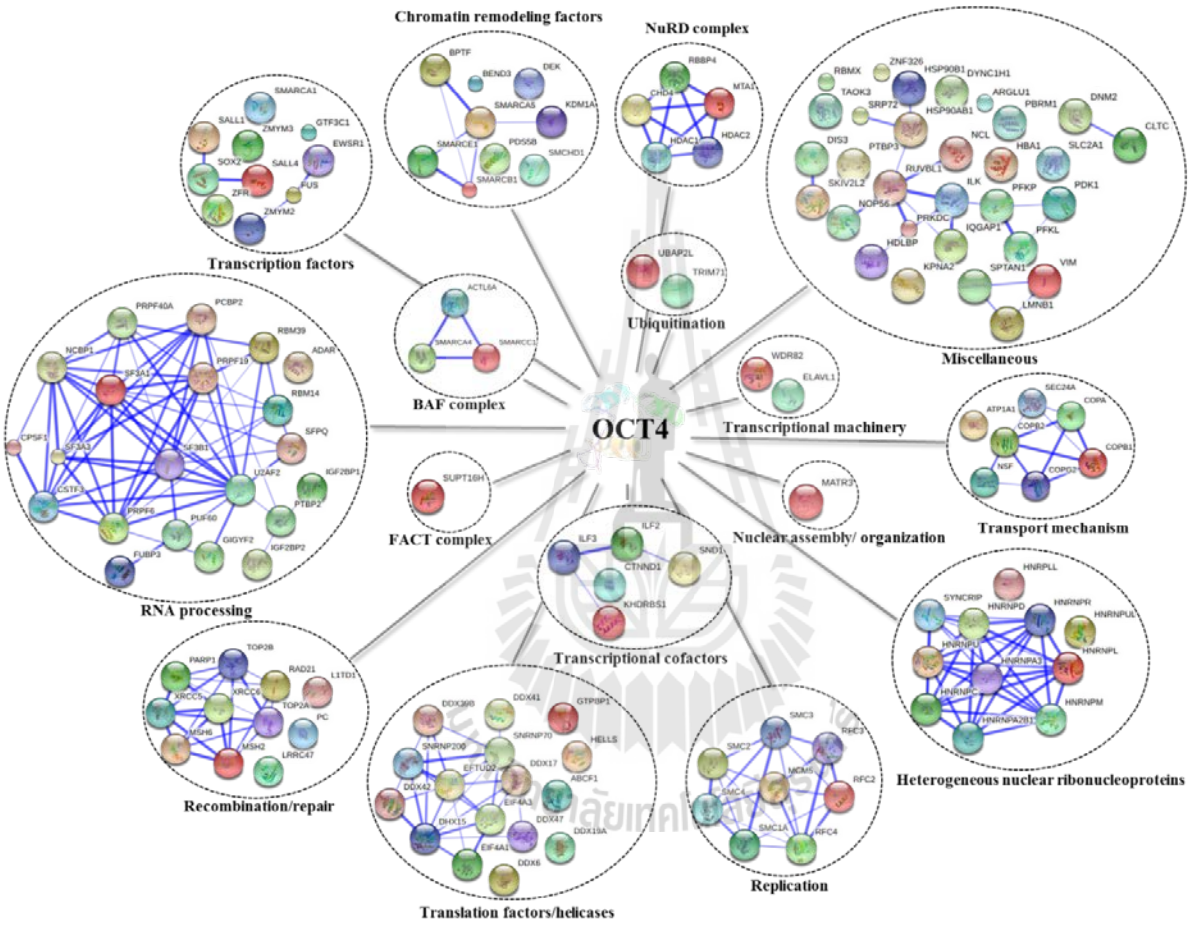

Figure 3.7 OCT4 interactor proteins as identified by mass spectrometry analysis were grouped in classes. Protein-protein interactions of each grouped was predicted by STRING database.

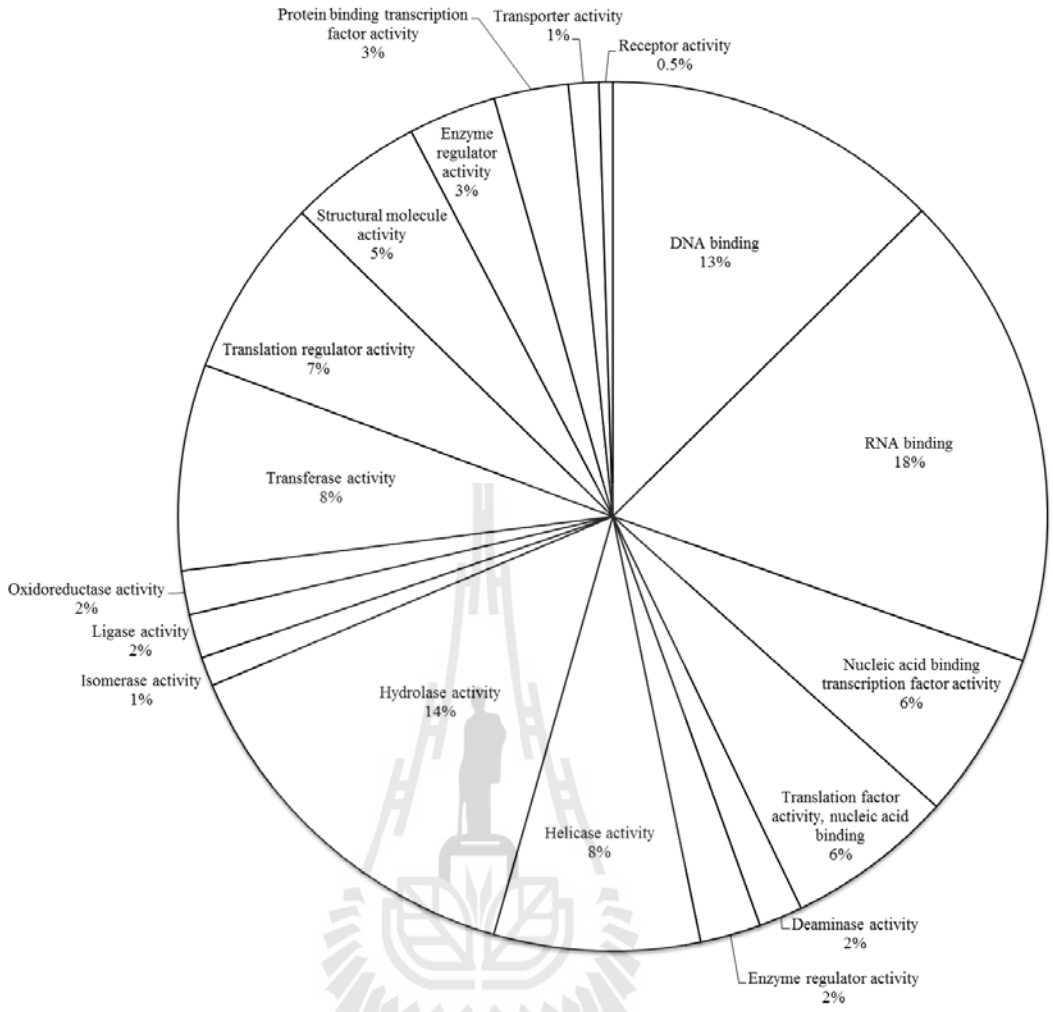

Figure 3.8 Analysis of the OCT4 interactome for gene ontology terms molecular function.

In comparing the dataset from this study with the recently published mouse OCT4 interactome study (Ding et al., 2012), we found 43 proteins that overlapped between human and mouse OCT4 interactome (Figure 3.9). The overlapping proteins are involved in many cellular pathways include NuRD (CHD4, HDAC1), SWI/SNF (ACTL6A, SMARCA4), FACT (SUPT16H) complexes, chromatin remodeling factors (BPTF, SMARCA5, SMARCB1, SMARCE1, SMCHD1), transcription factors (ILF3, SALL1, SALL4, SND1, ZMYM2, SMARCC1) and transcription and translation regulators (ATP1A1, COPB1, CPSF1, DDX47, FUBP3, GIGYF2, HELLS, L1TD1,

LRRC47, MCM5, MSH2, MSH6, PARP1, PCBP2, PTBP2, RAD21, RFC4, SF3A3, TOP2B, UBAP2L) that play important roles in ESCs and are associated with multiple transcription factors required in ESCs. In addition, this study uncovered 99 novel OCT4 interacting proteins. These proteins may be specific for human ESCs. We next examined the presence of these interactors in more detail by preforming OCT4 immunoprecipitation followed by immunoprecipitation (co-IP) experiments. The co-IP results were done to confirm the physical association of OCT4 with novel partners including BEND3, FUS, RMB25, SFPQ and WDR82. However, in this study only a small number of interactome proteins were confirmed by co-IP because most commercial available antibodies show low specificity.

มหาวิทยาลัยเทคโนโลยีสุรนารี

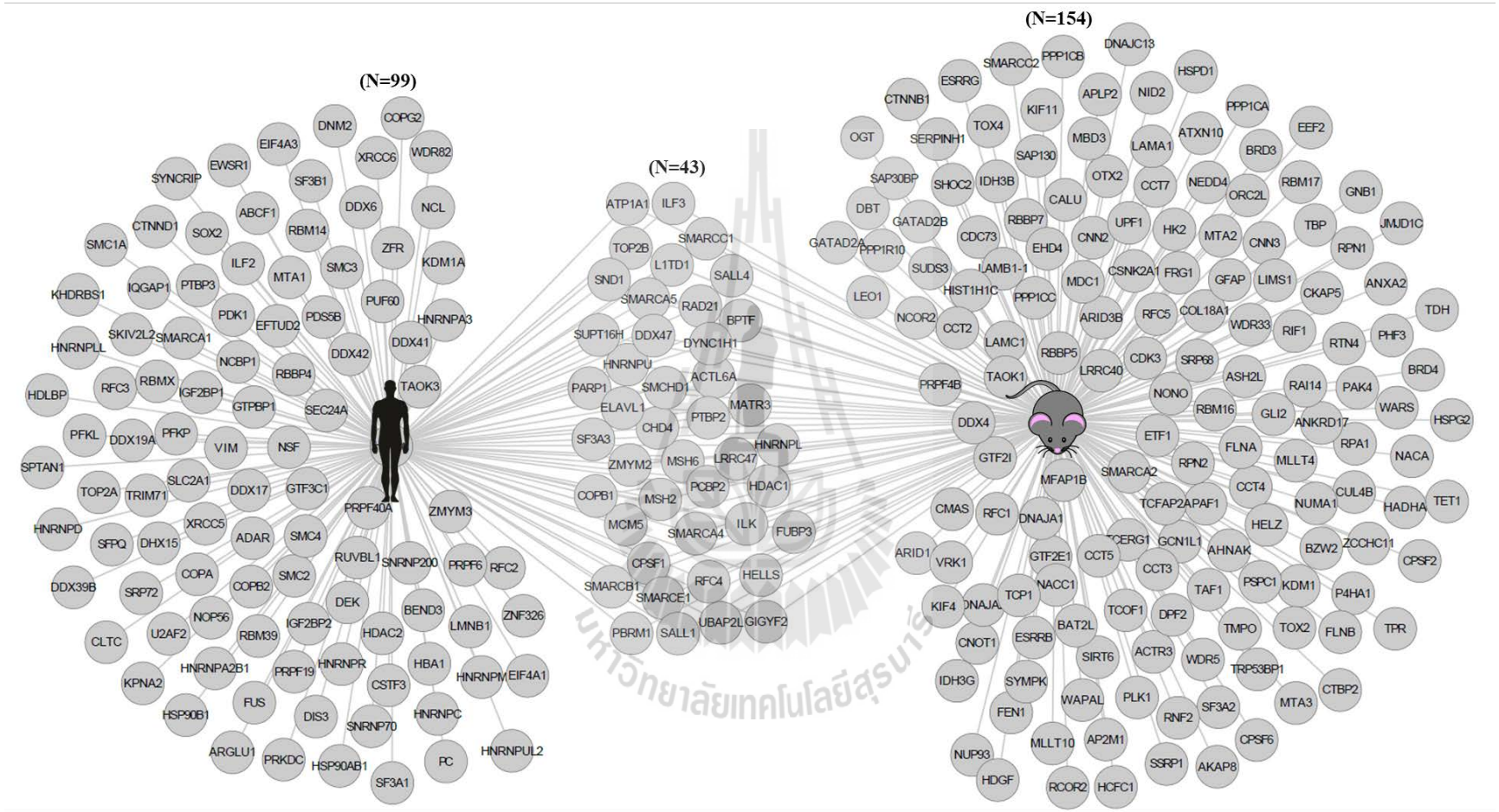

Figure 3.9 Comparison of human OCT4 interacting proteins with the published mouse OCT4 interactome.

3.4.3 Functional validation of OCT4 interacting proteins for human ESC pluripotency

As OCT4 is a key transcription factor and essential for self-renewal and pluripotency of both mouse and human ESCs, it is expected to interact with multiple functional modules involved in different areas of cellular regulation to effect transcription regulation. Here, we were interested in the candidates with a role in transcriptional regulation. Functional investigation of BEND3, FUS, RBM25, SFPQ and WDR82 was not previously examined in human ESCs. BEND3 is a transcriptional repressor that associates with the NuRD complex, possibly via recruiting the histone deacetylases (HDACs) and the transcription factor sal-like 4 (SALL4) (Saksouk et al., 2014). BEND3 heterochromatin-associated protein has an important role in regulating chromatin structure and transcription. The overexpression of BEND3 causes excessive heterochromatinization and transcriptional shut-down (Sathyan et al., 2011). BEND3 is also involved in rDNA silencing, the loss of BEND3 increases histone H3K4 trimethylation and decreases rDNA promoter DNA methylation (Khan et al., 2015). FUS is a member of the FET family (FUS, EWS and TAF15), involved in multiple cellular functions, including RNA processing and transport, transcriptional regulation. FUS is required for self-renewal capacity of hematopoietic stem cells (Sugawara et al., 2010). RBM25 belongs to a family of RNA-binding proteins and functions in pre-mRNA processing. RBM25 overexpression correlates with increased apoptotic cell death, promoting the selection of the proapoptotic Bcl-xS 5′ ss, whereas knockdown of RBM25 shifts the balance toward the antiapoptotic Bcl-xL form (Zhou et al., 2008). WDR82 is involved in initiation of transcription and required for the targeting of SETD1A-mediated histone H3-Lys4 trimethylation complex to transcription start sites of transcribed genes (Lee and Skalnik, 2008). Downregulated WDR82 caused mouse ESCs to begin to lose

pluripotency and differentiate (Van Hoof et al., 2006). SFPQ is a DNA- and RNA binding protein involved in pre-mRNA modification due to its binding to the pre-mRNA splicing factor polypyrimidine tract binding protein (Patton et al., 1993). SFPQ is implicated in central cellular processes including DNA repair, transcriptional regulation, mRNA processing as part of the spliceosome complex and transport (Shav-Tal and Zipori, 2002). In this study, we explored the consequences of knockdown of the candidates, *BEND3, FUS, RBM25, WDR82* and *SFPQ* genes. We employed inhibition of the endogenous RNAs by short hairpin RNA (shRNA). If the candidates influence OCT4 activity, its functional loss would affect ESC pluripotency. As shown in Figure 3.10, we found that BEND3, FUS, RMB25, WDR82 knockdown had no effect on human ESCs. In contrast, SFPQ knockdown promoted the differentiation and death of the cells.

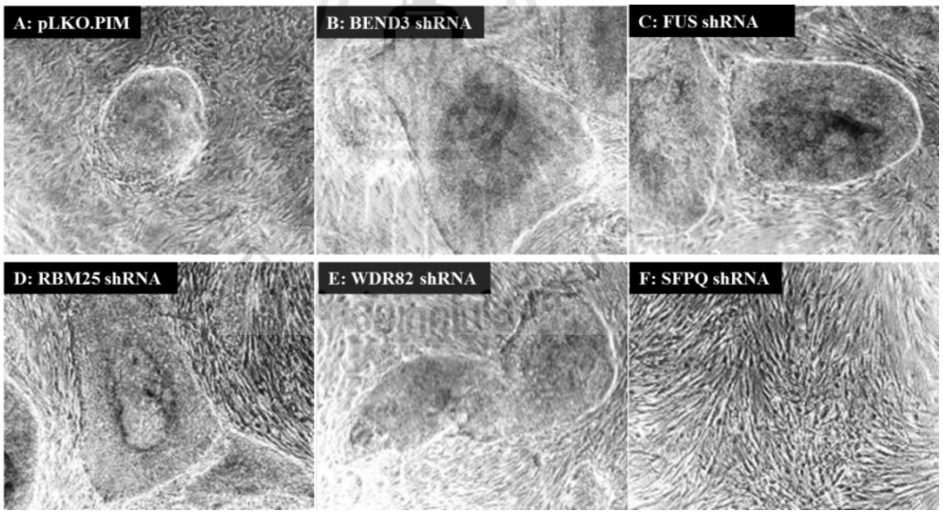

Figure 3.10 Functional validations by shRNA knockdown of the OCT4 interacting proteins. The human ESCs were transfected with lentiviral vectors expressing shRNA constructs against BEND3 (B), FUS (C), RBM25 (D), WDR82 (E), SFPQ (F) genes and non-targeting shRNA (pLKO.PIM) served as a negative control (A).

Moreover, we repeated the knockdown of SFPQ in two different human ESC lines with feeder-free condition on matrigel in mTesr1 medium by three different shRNAs (Figure 3.11). Human ESCs can be cultured either on MEFs in KSR+FBS+FGF or on matrigel in commercial mTesr1 medium. We wanted to see if the phenotype for SFPQ knockdown was specific to one culture system or not so we repeated the knockdown in cells grown on matrigel in mTesr1. The cells showed rapid cell death using two independent shRNAs (#1 and #2), indicating that the phenotype is very robust (Figure 3.11B and C). The third shRNA did not cause cell death because this shRNA did not reduce SFPQ expression according to the Western blot (Figure 3.11D). This may be due to the target sequences of shRNA #1 and #2 being located at the coding sequence of SFPQ mRNA. However, the targeting sequence of shRNA#3 was located at the 5' untranstranslated region (UTR) of the gene.

The result of this study showed that knockdown BEND3, FUS, RBM25, WDR82 proteins has no effects on the phenotype of human ESCs. This may due to compensation by another protein in the same family or related genes (Zhang, 2012). Taken together, the contributions of these factors may be only critical at different developmental stages or regulated and expressed in a cell type-specific manner. In contrast, knockdown of SFPQ protein associated with OCT4 can lead to human ESCs dysfunction and death. A possible explanation for this phenotype is that SFPQ/PSF serves as an interacting partner with the RAD51 paralogs (RAD51D, RAD51C and XRCC2) and non-POU domain-containing octamer-binding protein (NONO/p54nrb). These complexes are multifunctional nuclear proteins that are implicated in a wide variety of regulatory roles in key processes within the cells. Previous studies demonstrated that deficiency of SFPQ leads to sister chromatid cohesion defects and chromosome instability, and that the SFPQ/NONO complex affected messenger

ribonucleoprotein (mRNP, mRNA with bound proteins) dynamics within the cell (Rajesh et al., 2011; Snijders et al., 2015). In addition, we also found SFPQ associated with SOX2 and NANOG in human ESCs (unpublished data). The phenotype of SFPQ is quite striking and the OCT4/SFPQ/NONO complex may be essential for human ESC pluripotency.

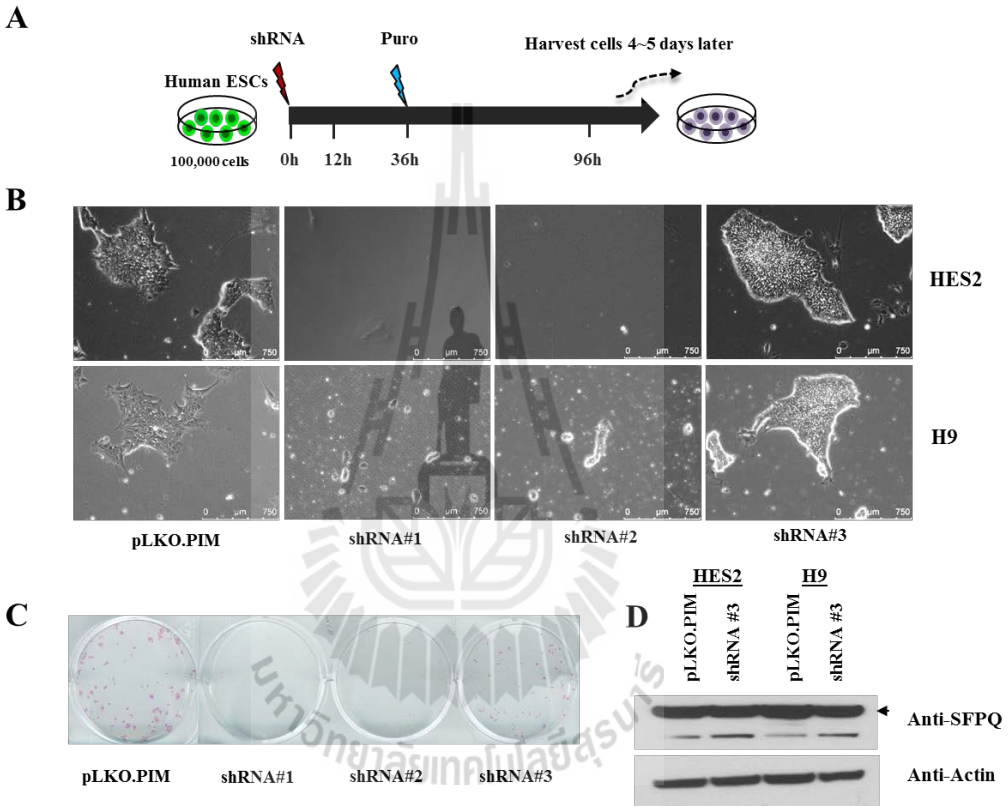

Figure 3.11 Knockdown of the OCT4 interacting protein SFPQ in human ESCs. A) Schematic of shRNA lentivirus transfection. B) Five days after the two different cell lines were transfected with virus expressing different shRNA constructs (#1, #2 and #3) against SFPQ. Non-targeting shRNA (pLKO.PIM) served as a negative control. C) The cells were fixed and stained for alkaline phosphatase. D) Western blot of SFPQ protein after knockdown by shRNA#3.

In conclusion, the OCT4 interacting protein dataset of human ESCs was generated via affinity purification-mass spectrometry methods. We identified 142 high-confidence OCT4 interactors that are implicated in transcriptional regulation, DNA repair, splicing and RNA transport. The network is linked to epigenetic regulatory pathways including NuRD, SWI/SNF, FACT complexes that play important roles in maintaining pluripotency. Functional validation of the candidates by shRNA revealed that SFPQ associated with OCT4 protein may provide a novel PPI complex approach to simultaneously activate multiple transcriptional regulatory, DNA repair and chromosome maintenance pathways that plays a crucial function for maintaining human ESC pluripotency. The OCT4 interactors provide a resource to investigate mechanisms of OCT4 function. Further analysis of the listed candidates at the functional level and through systems biology approaches can greatly assist in the understanding of the human ESC pluripotency network.

3.5 References

Boer, B., Kopp, J., Mallanna, S., Desler, M., Chakravarthy, H., Wilder, P. J., Bernadt, C., and Rizzino, A. (2007). Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of Sox2:Oct-3/4 target genes. **Nucleic Acids. Res.** 35: 1773-1786.

Caiafa, P., Guastafierro, T., and Zampieri, M. (2009). Epigenetics: Poly(ADP-ribosyl)ation of PARP-1 regulates genomic methylation patterns. **FASEB J.** 23: 672-678.

Ding, J., Xu, H., Faiola, F., Ma'ayan, A., and Wang, J. (2012). Oct4 links multiple epigenetic pathways to the pluripotency network. **Cell Res.** 22: 155-167.

Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T., Trevisan, M., Gupta, R. M., Moisan, A., Banks, E., Friesen, M., Schinzel, R. T., Xia, F., Tang, A., Xia, Y., Figueroa, E., Wann, A., Ahfeldt, T., Daheron, L., Zhang, F., Rubin, L. L., Peng, L. F., Chung, R. T., Musunuru, K., and Cowan, C. A. (2013). A TALEN genome-editing system for generating human stem cell-based disease models. **Cell Stem Cell** 12: 238-251.

Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. **Nature** 292: 154-156.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M. J., Heidersbach, A., Ramalho-Santos, J., McManus, M. T., Plath, K., Meshorer, E., and Ramalho-Santos, M. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. **Nature** 460: 863-868.

Goulburn, A.L., Alden, D., Davis, R.P., Micallef, S.J., Ng, E.S., Yu, Q.C., Lim, S.M., Soh, C.L., Elliott, D.A., Hatzistavrou, T., Bourke, J., Watmuff, B., Lang, R. J., Haynes, J. M., Pouton, C. W., Giudice, A., Trounson, A. O., Anderson, S. A., Stanley E. G., and Elefanty, A. G. (2011). A targeted NKX2.1 human embryonic stem cell reporter line enables identification of human basal forebrain derivatives. **Stem Cells** 29: 462-473.

Hisano Y, Ota S, Arakawa K, Muraki M, Kono N, Oshita K, Sakuma T, Tomita M, Yamamoto T, Okada Y, Kawahara A. (2013). Quantitative assay for TALEN activity at endogenous genomic loci. **Biology Open.** 2: 363-367.

Ho, L., Jothi, R., Ronan, J. L., Cui, K., Zhao, K., and Crabtree, G. R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential

component of the core pluripotency transcriptional network. **Proc Natl Acad Sci USA.** 106: 5187-5191.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., Cost, G. J., Zhang, L., Santiago, Y., Miller, J. C., Zeitler, B., Cherone, J. M., Meng, X., Hinkley, S. J., Rebar, E. J., Gregory, P. D., Urnov, F. D., and Jaenisch, R. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. **Nat. Biotechnol.** 29: 731-734.

Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.F., Sontheimer, E.J., and Thomson, J.A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. **Proc. Natl. Acad. Sci. USA.** 110, 15644-15649.

Huang, X., and Wang, J. (2014). The extended pluripotency protein interactome and its links to reprogramming. **Curr. Opin. Genet. Dev.** 28: 16-24.

Irion, S., Luche, H., Gadue, P., Fehling, H.J., Kennedy, M., and Keller, G. (2007). Identification and targeting of the ROSA26 locus in human embryonic stem cells. **Nat. Biotechnol.** 25: 1477-1482.

Issad, T., and Kuo, M. (2008). O-GlcNAc modification of transcription factors, glucose sensing and glucotoxicity. **Trends Endocrinol. Metab.** 19: 380-389.

Khan, A.,Giri, S., Wang, Y., Chakraborty, A., Ghosh, A. K., Anantharaman, A., Aggarwal, V., Sathyan, K. M., Ha, T., Prasanth, K. V., and Prasanth, S. G. (2015). BEND3 represses rDNA transcription by stabilizing a NoRC component via USP21 deubiquitinase. **Proc Natl Acad Sci USA.** 112: 8338-8343.

Kim, J., Cantor, A. B., Orkin. S. H., and Wang, J. (2009).Use of *in vivo* biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. **Nat. Protoc.** 4: 506-517.

Koledova, Z., Kafkova, L. R., Calabkova, L., Krystof, V., Dolezel, P., and Divoky, V. (2010). Cdk2 inhibition prolongs G1 phase progression in mouse embryonic stem cells. **Stem Cells Dev.** 19:181-194.

Kovar, H. (2011). Dr. Jekyll and Mr. Hyde: The Two Faces of the FUS/EWS/TAF15 Protein Family. **Sarcoma** 2011: 837474.

Krishnakumar, R., and Kraus, W. L. (2010). The PARP side of the nucleus: Molecular actions,physiological outcomes, and clinical targets. **Mol. Cell** 39: 8-24.

Lauberth, S. M., and Rauchman, M. (2006). A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. **J. Biol. Chem.** 281: 23922-23931.

Lee, J. H., and Skalnik, D. G. (2008). Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. **Mol. Cell Biol.** 28: 609-618.

Lengner, C. J., Gimelbrant, A. A., Erwin, J. A., Cheng, A. W., Guenther, M. G., Welstead, G.G, Alagappan, R., Frampton, G. M., Xu, P., Muffat, J., Santagata, S., Powers, D., Barrett, C. B., Young, R. A., Lee, J. T., Jaenisch, R., and Mitalipova, M. (2010). Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. **Cell** 141: 872-883.

Liang J., Wan M., Zhang Y., Gu P., Xin H., Jung S. Y., Qin J., Wong J., Cooney A. J., Liu D., Songyang Z. (2008) Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. **Nat. Cell Biol.** 10, 731–739

Merkert, S., Wunderlich, S., Bednarski, C., Beier, J., Haase, A., Dreyer, A. K., Schwanke, K., Meyer, J., Göhring, G., Cathomen, T., Martin, U. (2014). Efficient designer nuclease-based homologous recombination enables direct PCR screening

for footprintless targeted human pluripotent stem cells. **Stem Cell Reports** 2:107-118.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D.,Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. **Cell** 95: 379-391.

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

Orkin, S.H. (2005). Chipping away at the embryonic stem cell network. **Cell** 122: 828-830.

Pan, G., Li, J., Zhou, Y., Zheng, H., and Pei, D. (2006). A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. **FASEB J.** 20: 1730-1732.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M. M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. **Cell Stem Cell** 6: 382-395.

Patton, J. G., Porro, E. B , Galceran, J., Tempst, P., and Nadal-Ginard, B. (1993). Cloning and characterization of PSF, a novel pre-mRNA splicing factor. **Genes Dev**. 7: 393-406.

Pomerantz, R. T., and O'Donnell, M. (2010). What happens when replication and transcription complexes collide? **Cell Cycle** 9: 2535-2541.

Prelle, K., Zink, N., and Wolf, E. (2002). Pluripotent stem cells-model of embryonic development, tool for gene targeting, and basis of cell therapy. **Anat. Histol. Embryol.** 31:169-186.

Rajesh, C., Baker, D. K., Pierce, A. J., and Pittman, D. L. (2011). The splicing-factor related protein SFPQ/PSF interacts with RAD51D and is necessary for homology-directed repair and sister chromatid cohesion. **Nucleic Acids Res.** 39:132-145.

Sadelain, M., Papapetrou, E. P., and Bushman, F. D. (2011). Safe harbours for the integration of new DNA in the human genome. **Nat. Rev. Cancer** 12: 51-58.

Saksouk, N., Barth, T. K., Ziegler-Birling, C., Olova, N., Nowak, A., Rey, E. Mateos-Langerak, J., Urbach, S., Reik, W., Torres-Padilla, M. E., Imhof, A., Déjardin, J., and Simboeck, E. (2014). Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. **Mol. Cell** 56: 580-594.

Shav-Tal, Y., and Zipori, D. (2002). PSF and p54(nrb)/NonO–multi-functional nuclear proteins. **FEBS Lett.** 531: 109-114.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. **Nat. Protoc.** 1: 2856-2860.

Singhal, N., Graumann, J., Wu, G., Araúzo-Bravo, M. J., Han, D. W., Greber, B., Gentile, L., Mann, M., and Schöler, H. R. (2010). Chromatin-remodeling components of the BAF complex facilitate reprogramming. **Cell** 141: 943-955.

Snijders, A. P., Hautbergue, G. M., Bloom, A., Williamson, J. C., Minshull, T. C., Phillips, H. L., Mihaylov, S. R., Gjerde, D. T., Hornby, D. P., Wilson, S. A., Hurd, P. J., and Dickman, M. J. (2015). Arginine methylation and citrullination of splicing factor proline- and glutamine-rich (SFPQ/PSF) regulates its association with mRNA. **RNA** 21: 347-359.

Soldner, F., Laganie`re, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., Zhang, L., Guschin, D.,

Fong, L. K., Vu, B. J., Meng, X., Urnov, F. D., Rebar, E. J., Gregory, P. D., Zhang, H. S., and Jaenisch, R.(2011). Generation of isogenic pluripotent stem cells differing exclusively at two early on set Parkinson point mutations. **Cell** 146: 318-331.

Sugawara T, Oguro H, Negishi M, Morita Y, Ichikawa H, Iseki T, Yokosuka O, Nakauchi H, Iwama A. (2010). FET family proto-oncogene Fus contributes to self-renewal of hematopoietic stem cells. **Exp Hematol.** 38: 696-706.

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

Ul Ain, Q., Chung, J. Y., and Kim, Y. H. (2015). Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. **J. Control Release** 205:120-127.

van den Berg, D. L., Snoek ,T., Mullin, N. P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R. A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. **Cell Stem Cell** 6: 369-381.

Van Hoof, D., Passier, R., Ward-Van Oostwaard, D., Pinkse, M. W., Heck, A. J., Mummery. C. L., and Krijgsveld, J. (2006). A quest for human and mouse embryonic stem cell-specific proteins. **Mol. Cell Proteomics** 5: 1261-1273.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T.W., and Orkin, S. H. (2006). A protein interaction network for pluripotency of embryonic stem cells. **Nature** 444: 364-368.

Wang, Y., Dawson,V. L., and Dawson, T. M. (2009). Poly(ADP-ribose) signals tomitochondrial AIF: A key event in parthanatos. **Exp. Neurol.** 218: 193-202.

Wei, F., Schöler, H. R., and Atchison, M. L. (2007). Sumoylation of Oct4 enhances its stability, DNA binding, and transactivation. **J. Biol. Chem.** 282: 21551-21560.

Xu, H. M., Liao, B., Zhang, Q. J., Wang, B. B., Li, H., Zhong, X. M., Sheng, H. Z., Zhao, Y. X., Zhao, Y. M., Jin, Y. (2004). Wwp2, an E3 ubiquitin ligase that targets transcription factor Oct-4 for ubiquitination. **J. Biol. Chem.** 279: 23495-23503.

Yang, J., Gao, C., Chai, L., and Ma,Y. (2010). A novel SALL4/OCT4 transcriptional feedback network for pluripotency of embryonic stem cells. **PLoS One.** 5: e10766.

Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordo´n˜ez, A., Hannan, N.R., Rouhani, F.J.,Darche, S., Alexander, G., Marciniak, S. J., Fusaki, N., Hasegawa, M., Holmes, M. C., Di Santo, J. P., Lomas, D. A., Bradley, A., and Vallier, L. (2011). Targeted gene correction of a1-antitrypsin deficiency in induced pluripotent stem cells. **Nature** 478: 391-394.

Zhang, J. (2012). Genetic redundancies and their evolutionary maintenance. **Adv. Exp. Med. Biol.** 751: 279-300.

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

Zhou, A., Ou, A. C., Cho, A, Benz, E.J. Jr., and Huang, S. C. (2008). Novel splicing factor RBM25 modulates Bcl-x pre-mRNA 5' splice site selection. **Mol. Cell Biol.** 28: 5924-5936.

Zhu, Z., Verma, N., González, F., Shi, Z, D., and Huangfu, D. (2015). A CRISPR/Cas-mediated selection-free Kknockin strategy in human embryonic stem cells. **Stem Cell Reports** 4:1103-1111.

CHAPTER IV

EXPRESSION AND PURIFICATION OF RECOMBINANT HUMAN BASIC FIBROBLAST GROWTH FACTOR FUSION PROTEINS AND THEIR USES IN HUMAN STEM CELL CULTURE

4.1 Abstract

To reduce the cost of cytokines and growth factors in stem cell research, a simple method for the production of soluble and biological active human basic fibroblast growth factor (hbFGF) fusion protein in *Escherichia coli* was established. Under optimal conditions, approximately 60-80 mg of > 95% pure hbFGF fusion proteins (Trx-6xHis-hbFGF and 6xHis-hbFGF) was obtained from 1 liter of culture broth. The purified hbFGF proteins, both with and without the fusion tags, were biologically active, which were confirmed by their ability to stimulate proliferation of NIH3T3 cells. The fusion proteins also have the ability to support several culture passages of undifferentiated human embryonic stem cells and induce pluripotent stem cells. This work provides a low-cost and uncomplicated method for the production and purification of biologically active hbFGF fusion proteins.

4.2 Introduction

Pluripotent stem cells can potentially be used in variety of medical fields, including disease biology, drug discovery, gene therapy and cell replacement therapy (Picanco-Castro et al., 2014; Singh et al., 2015). However, the mechanisms involved in

the maintenance and expansion of an undifferentiated state in *in vitro* culture and induce differentiation *in vitro* and *in vivo* still need to be clarified before use in therapeutic applications. The optimization of the cell culture system is one of the key factors for the success of stem cell research. Both expansion and differentiation of stem cells require supplementation with growth factors and cytokines (van der Sanden et al., 2010). Various growth factors and cytokines used in stem cell research, such as activins, bone morphogenetic proteins (BMPs), epidermal growth factor (EGF), fibroblast-derived growth factors (FGFs) and insulin-like growth factor (IGFs), are expensive. One of the growth factors added to culture medium for regulating human stem cell proliferation and self-renewal is basic FGF (bFGF) (Ding et al., 2010). In human ESCs, exogenous bFGF activates the ERK MAPK pathway, and inhibition of this signaling transduction pathway results in stem cell differentiation (Li et al., 2007). Growth factors and cytokines are generally the main cost of stem cell culture. To reduce the cost, several studies have used in-house production of growth factors and cytokines (Imsoonthornruksa et al., 2011; Song et al., 2013). In the present study, the use of both hexa-histidine (6xHis) fusion and thioredoxin (Trx)-6xHis fusion strategies for expression and purification of recombinant human bFGF (hbFGF) in *E. coli* were explored. The target proteins were purified with heparin-sepharose affinity chromatography and immobilized metal affinity chromatography (IMAC). Biological activity assays showed that the hbFGF fusion proteins were able to promote murine fibroblast cell line NIH3T3 proliferation and maintain the undifferentiated state of human ESCs and iPSCs in a manner comparable to commercial hbFGF. Therefore, we have demonstrated a simple, convenient and cost-effective system for expression and purification of bioactive soluble hbFGF.

4.3 Material and Methods

4.3.1 Expression vector construction

In this study, two expression constructs were produced using the pET28a(+) and pET32a(+) (Novagen, Germany) expression vectors. The hbFGF gene was amplified from the pWPI_SPbFGF vector (Addgene, USA) using the bFGF pET28 forward primer: CCGGATCCATGGCAGCCGGGAGC; the bFGF pET32 forward primer: AAGAGCTC ATGGCAGCCGGGAGC; and the reverse primer: GGCTCGAGTCAGCTCTTAGCAG AAGACATTGG (restriction sites underlined). The amplified fragments were cleaved by *BamH*I/*Xho*I and *Sac*I/*Xho*I (NEB, USA) and then ligated into the *BamH*I/*Xho*I and *Sac*I/*Xho*I digested pET28 and pET32 vectors, respectively, with T4 DNA ligase (New England Biolabs, USA). After the ligation, the vectors were transformed into *E. coli* strain DH5α. Positive clones were selected by colony PCR (Zon et al., 1987). The constructs, called pET28/hbFGF and pET32/hbFGF, were subsequently verified by sequencing.

4.3.2 Recombinant protein expression and purification

The vectors with correct inserts were transformed into *E. coli* expression strains BL21(DE3) (Novagen) and ArcticExpress™ RIL (Stratagene, USA). For recombinant fusion protein expression, selected clones were grown overnight and then diluted 1:100 in fresh lysogeny broth (LB) (Bertani, 1952) containing the appropriate antibiotics and grown at 37°C until the optical density of 600 nm reached 0.6. Recombinant fusion protein expression was then induced by the addition of isopropyl-β-D-thiogalactoside (IPTG).The cells were shaken at 15 or 20°C for 6 h or longer and then harvested by centrifugation at 4000×g for 20 min. The cell pellets were kept frozen at -70°C until use. Prior to purification, the cell pellets were thawed and resuspended in

lysis/binding buffer containing 50 mM sodium phosphate buffer (pH 8.0) (Sambrook et al., 1989), 1 mM phenylmethylsulfonylfluoride (PMSF), 150 mM NaCl, 0.1% Triton X-100, 10% glycerol, 2 mM imidazole and 200 μg ml^{-1} lysozyme. The cells were disrupted on ice for 5 min by sonication using a GE 100 ultrasonic processor (60-80% amplitude, 60% cycle duty (sonicates 10 s, stop 5 s then repeat the cycle) for 15 min) and separated into soluble and insoluble fractions by centrifugation at 11,700×g for 10 min. The soluble protein fractions were recovered and filterd through 0.45 μm polyethersulfone membrane and applied to a heparin-sepharose affinity chromatography (HiTrap Heparin HP,GE Healthcare, Sweden) or an IMAC (HisTrap HP,GE Healthcare) column according to the manufacturer's instructions. Fractions containing Trx-6xHis-hbFGF were pooled and dialyzed against phosphate buffered saline (PBS) buffer. Then, the Trx-His6-hbFGF fusion proteins were cleaved with in-house enterokinase protease (Kupradit et al., 2008) in an 8 h reaction at 28°C, and the Trx-6xHis and hbFGF fragments were separated by a second HisTrap HP IMAC step. The purified 6xHis-hbFGF, Trx-6xHis-hbFGF and hbFGF were filter sterilized (0.2 μm) and stored at 4°C until further use. With our standard laboratory technique, the endotoxin level is lower than 1 ng/μg protein as determined by the LAL method. The total cell lysate, soluble and insoluble fractions and elution fractions were collected and analyzed with 15% SDS-PAGE followed by Coomassie Brilliant Blue staining. The protein concentrations were determined according to the Bradford method. To quantify the concentration of the fusion proteins in the mixture, densitometric scanning of the SDS-PAGE gel stained with Coomassie Brilliant Blue was evaluated with Quantity One software (Bio Rad, USA). The samples were subjected to immunoblotting after electro-transfer to a nitrocellulose membrane (Bio Rad) using a His-tag rabbit polyclonal antibody (GenScript, USA) followed by a goat

anti-rabbit secondary antibody conjugated to alkaline phosphatase (Sigma, USA). Then, the membranes were developed with NBT/BCIP substrate (Sigma).

4.3.3 MTT cell proliferation assay

The functional activities of the 6xHis-hbFGF, Trx-6xHis-hbFGF and hbFGF recombinant proteins were measured by their ability to promote the proliferation of mouse embryonic fibroblasts in serum-free conditions compared to the activity of commercial hbFGF (PHG0026, Invitrogen, USA). Various dilutions of the recombinant and commercial hbFGF (1.7, 3.4, 6.9, 14 and 27 nM) were added to $5x10^3$ cells that had been seeded in 96-well microtiter plates in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen) without fetal calf serum and incubated at 37°C, 5% CO_2 in air. After 72 h of incubation with the recombinant or commercial hbFGF, 10 µl of 12 mM 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (Sigma) in PBS was added to each well and the incubation was continued for another 4 h. The culture medium, including the MTT solution, was removed from the wells, and 100 µl of 10% SDS in 10 mM HCl was added to each well and mixed to dissolve the crystals. The absorbance of each well was read at 570 nm. From the absorbance values, cell proliferation was determined and presented as a percentage of the proliferation obtained with cells that were cultured in 10% FCS. Triplicate of each dilution of the hbFGF supplemented was assayed.

4.3.4 Maintenance of the undifferentiated state of human pluripotent stem cells in culture

hESCs (Pruksananonda et al., 2012) and hiPSCs (Nishishita et al., 2012) were grown on mitotically inactivated human foreskin-derived fibroblasts plated in a 0.1% gelatin-coated dish in Knock out DMEM (Invitrogen) supplemented with 20% Knock out

serum replacement, 1% glutamax®, 0.55 mM β-mercaptoethanol, 1% nonessential amino acids (Invitrogen, 11140-050), 50 U penicillin ml^{-1} (Invitrogen), 50 μg streptomycin ml^{-1} (Invitrogen), and 8 ng hbFGF ml^{-1} (Invitrogen) or 6xHis-hbFGF, Trx-6xHis-hbFGF fusion proteins. Subsequently, hESC and hiPSC colonies with ES cell morphology were selected for subculture and maintained by standard methods. After 8-10 passages, in addition to morphology, the hESCs and hiPSCs were characterized by the expression of embryonic stem cell markers by qPCR analysis and immunocytochemical staining.

4.3.5 Alkaline phosphatase and immunocytochemical staining

The hESC and hiPSC colonies were fixed in 4% paraformaldehyde. The alkaline phosphatase (AP) activity was detected using an AP detection kit (Sigma) following the manufacturer's instructions. For the OCT4 antibody reaction, the colonies were treated with 0.2% Triton X-100 (Sigma) DPBS (Invitrogen) for 15 min. All colonies were incubated with blocking buffer consists of 5% fetal bovine serum (HyClone) in DPBS for 1 h. The samples were then thoroughly washed with DPBS and incubated with primary antibodies. All primary antibodies purchased from Abcam, USA include rat antibodies against stage-specific embryonic antigen (SSEA)-3 (ab162286), mouse anti-SSEA-4 (ab16287) and goat anti-OCT4 (ab27985) were diluted 1:200 with PBS without calcium and magnesium. The colonies were incubated overnight at 4°C with the diluted primary antibodies. For the detection of the antibodies, the cells were incubated with 1:200 dilutions of rabbit anti-rat-FITC (ab8520), goat anti-mouse-Cy3 (Milipore, AP181C) and rabbit anti-goat-FITC (ab97009) antibodies at room temperature for 1 h. Then, the colonies were stained with Hoechst 33342 (Sigma). The colonies were examined under a fluorescence microscope.

4.3.6 Quantitative polymerase chain reaction

Total RNA from hESCs and hiPSCs colonies were extracted with the GeneJET RNA purification kit and reverse-transcribed to cDNA with the RevertAid H Minus First Strand cDNA Synthesis Kit according to the manufacturer's instructions (Thermo Fisher Scientific, USA). The mRNA expression of *OCT4*, *SOX2*, *NANOG*, *CDX2, BRACHYURY* and *PAX6* was quantified with the SYBR-green-based real-time quantitative-PCR (qPCR) method using the Bio-Rad Chromo4 real-time PCR detection system (Bio-Rad) with the primer sets shown in the Table 4.1. The amplification reactions contained 10 ng cDNA, 1x KAPA SYBR® FAST qPCR Kit solution (KAPA Biosystems, USA), and 0.25 μM of each specific primer to obtain a final volume of 20 μl. All of the reactions were performed in triplicate, and the thermal cycling conditions were 3 min at 95°C, followed by 45 cycles of 95°C for 20 s and 62°C for 1 min.

Table 4.1 Primer sequences for real-time qPCR.

Primer	Sequence (5' to 3')
OCT4-F	TGGGCTCGAGAAGGATGTG
OCT4-R	GCATAGTCGCTGCTTGATCG
SOX2-F	CCCCCGGCGGCAATAGCA
SOX2-R	TCGGCGCCGGGGAGATACAT
NANOG-F	CAGAAGGCCTCAGCACCTAC
NANOG-R	GTCACTGGCAGGAGAATTTGG
CDX2-F	CCTCCGCTGGGCTTCATTCC
CDX2-R	TGGGGGTTCTGCAGTCTTTGGTC
BRACHYURY-F	TGCTTCCCTGAGACCCAGTT
BRACHYURY-R	GATCACTTCTTTCCTTTGCATCAAG
PAX6-F	CAGCTCGGTGGTGTCTTTG
PAX6-R	AGTCGCTACTCTCGGTTTA
GAPDH-F	GTCAACGGATTTGGTCGTATTG
GAPDH-R	CATGGGTGGAATCATATTGGA

4.3.7 Statistical analysis

The percentage of growth of the NIH3T3 cells is presented as the mean ± SD. Relative quantification of the target gene expression levels is presented as fold differences. Statistical analysis of the data was evaluated in the Completely Randomized Design (CRD) with Statistical Analysis System (SAS Inst. Inc., Cary, NC, USA). Analysis of variance (ANOVA) was used to determine the significant differences between samples. Means comparison with Duncan's Multiple Range Test (DMRT) was used to determine the difference between means at $P < 0.05$.

4.4 Results and Discussion

4.4.1 Expression and purification of hbFGF fusion proteins

The expression vectors were constructed base on the pET28a(+) and pET32a(+) backbone (Figure 4.1A). The resulting pET28/hbFGF and pET32/hbFGF were verified by sequencing. The sequence of the recombinant plasmids were as expected (Figure 4.1B and C) without any mutation when compare to the nucleotide sequence of previously reported plasmid pWPI_SPbFGF and the GenBank Accession no. M27968. The recombinant pET28/hbFGF vector encoded hbFGF fused with a 34 amino acid extra N-terminal sequence, MGSSHHHHHHSSGVPRGSHMASMTGGQQMGRGS, which contained a hexa-histidine (6xHis)-tag used for heterologous protein purification by immobilized metal affinity chromatography (IMAC). Additionally, pET32/hbFGF construct encoded a fusion protein in-frame with an N-terminal thioredoxin (Trx) domain, linker sequence of the pET32a(+) with the 6xHis-tag followed by an enterokinase recognition site followed by the hbFGF sequence. The ExPASY Protein Parameters Tools Analysis predicted the theoretical molecular mass and the pI of the 6xHis-hbFGF (189 amino acid), the Trx-6xHis-hbFGF (324 amino acid) fusion proteins

and the hbFGF (166 amino acid) as 20,740 Da/pI=9.8, 35,417 Da/pI=7.2 and 18,360 Da/pI= 9.4, respectively (Figure 4.1).

The effects of the induction temperature, IPTG concentration and induction times on the production of 6xHis-hbFGF and Trx-6xHis-hbFGF fusion proteins from the pET28/hbFGF and pET32/hbFGF expression vectors in *E. coli* BL21(DE3) and ArcticExpress strains as well as the total soluble protein were evaluated. In all cases, the fusion protein expression at 20°C was greater than the expression at 15°C. Increasing the IPTG concentration from 0.1 to 1 mM did not affect the expression level of the fusion proteins. In fact, the Trx-6xHis-hbFGF fusion protein was expressed without IPTG induction in both the BL21(DE3) and ArcticExpress cells (Figure 4.2). Improvement of the fusion protein expression level was obtained by increasing the induction times (Figure 4.3). As shown in Figure 4.4A and B, more than 95% of the produced 6xHis-hbFGF fusion protein was in the soluble form and less than 3% of the protein was in the insoluble form.

After the HisTrap HP affinity purification, approximately 60-80% of the 6xHis-hbFGF fusion protein was lost in the binding step (flow through fractions), and only 10-24% could be harvested when the fusion protein was eluted with 100, 250 and 500 mM imidazole. In contrast, when HiTrap Heparin HP affinity purification was used, almost 100% of the 6xHis-hbFGF fusion protein bound to the column. The elution of the fusion protein was performed using a high salt concentration (2 M NaCl). For Trx-6xHis-hbFGF, the results were similar when expressed in both *E. coli* BL21(DE3) and *E. coli* ArcticExpress strains (Figure 4.4C and D). Approximately 200 ml of each *E. coli* culture provided 1 g of cells, from which purification with HiTrap Heparin HP yielded approximately 12 mg of the 6xHis-hbFGF or 14 mg of the Trx-6xHis-hbFGF fusion protein (Table 4.2). In contrast, only low amounts of both fusion proteins were purified

through HisTrap HP chromatography (Table 4.3). The HiTrap Heparin HP purification step resulted in a high purity of > 95% for both fusion proteins. The purified Trx-6xHis-hbFGF fusion protein was dialyzed against PBS buffer to remove the NaCl and imidazole, and then an enterokinase protease reaction was performed to separate the hbFGF from the N-terminal Trx-6xHis-tag. To purify the hbFGF from the enterokinase and the Trx-6xHis-tag, HisTrap HP chromatography was performed. The recombinant in-house enterokinase and the Trx-6xHis-tag contained 6xHis tags. Therefore, the proteins bound to the column, and the cleaved mature hbFGF was washed out in the flow through fraction. The fusion tags were successfully removed from the hbFGF, and the product could be observed on an SDS-PAGE gel (Figure 4.5A). The purified 6xHis-hbFGF, Trx-6xHis-hbFGF fusion proteins and the hbFGF displayed molecular masses of 21, 35 and 18 kDa on an SDS-PAGE gel, respectively. The results are compatible with the theoretical molecular masses of the proteins, and all of the proteins were immunoreactive when analyzed by western blotting using rabbit anti-His polyclonal antibody, except for the purified hbFGF due to the removal of the 6xHis tag (Figure 4.5A).

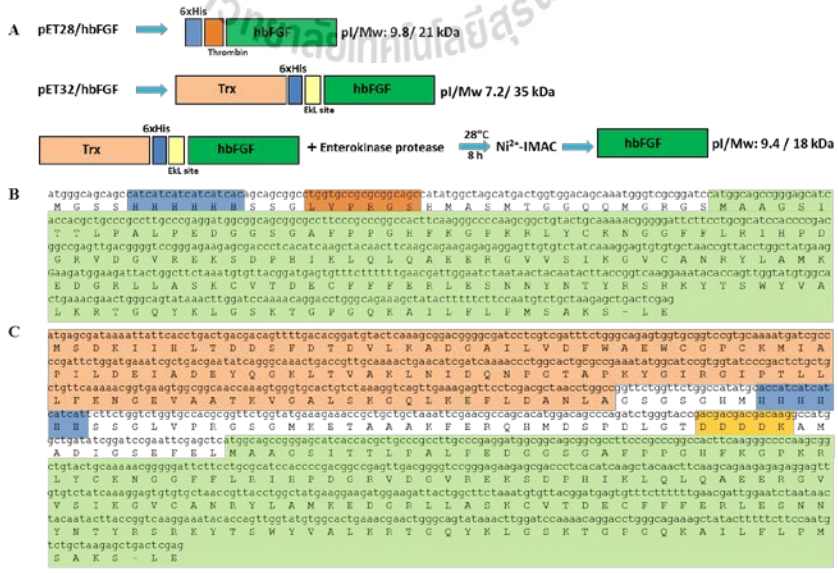

Figure 4.1 Schematic diagram of the recombinant proteins. pET28a(+) and pET32a(+) vector backbones were used to construct the expression vectors (A).

pET28/hbFGF and pET32/hbFGF expression vectors were verified by sequencing (B and C). Hexahistidine (6xHis)-tags are at the N-terminus of both recombinant proteins. The thrombin and enterokinase cleavage sites (LVPRGS and DDDDK, respectively) are located between the 6xHis-tag and the thioredoxin (Trx)-6xHis-tag and the hbFGF sequence, to release of mature hbFGF from the 6xHis-hbFGF and the Trx-6xHis-hbFGF fusion proteins, respectively.

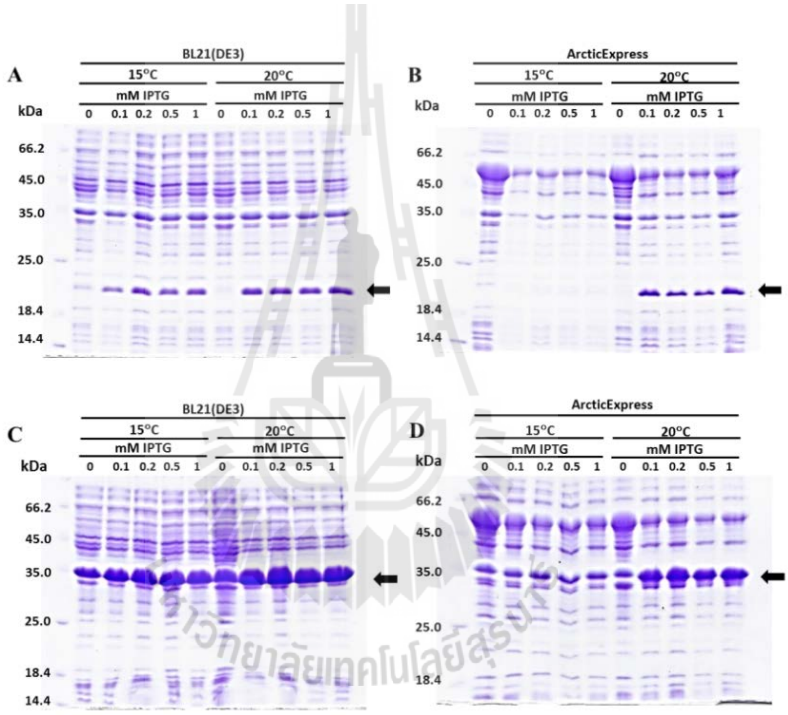

Figure 4.2 Effect of host cell, induction temperature and IPTG concentration on expression of recombinant hbFGF fusion proteins. The pET28/hbFGF (A and B) and pET32/hbFGF (C and D) expression vectors were transformed into *E.coli* BL21(DE3) (A and C) and ArcticExpress strains (B and D). The transformed cells were grown at 37°C to an OD600 of 0.6 and then induced with 0 to 1 mM IPTG for 12 h at 15 or 20°C. The induced cells were harvested, the whole cell lysate was analyzed on 15% SDS-PAGE and the gels stained with Coomassie Brilliant Blue. Proteins from equal amount of

cells were loaded into each lane. The arrows indicate the positions of recombinant fusion proteins.

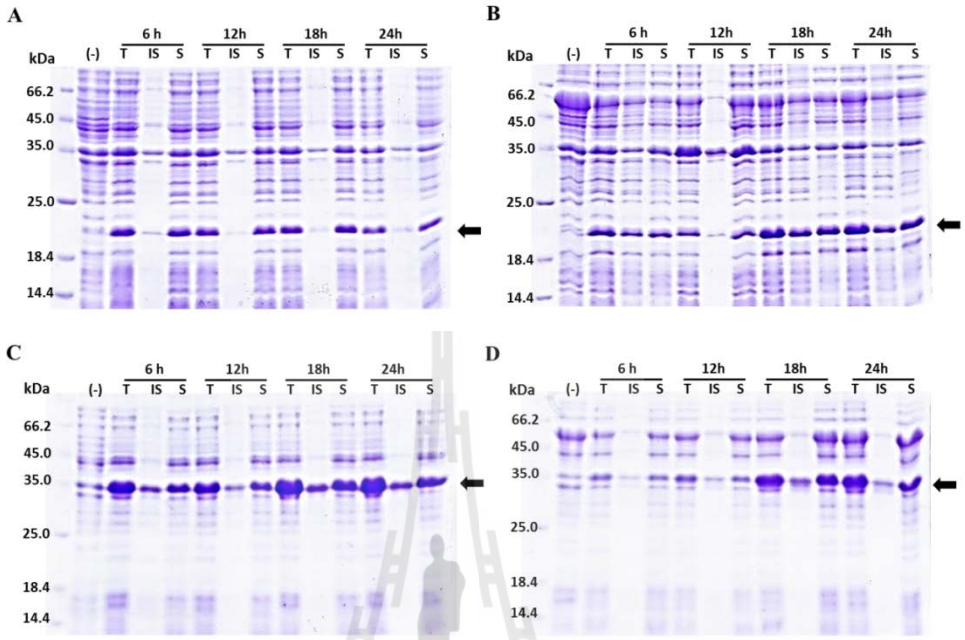

Figure 4.3 The effect of induction times on solubility of recombinant hbFGF fusion proteins. The pET28/hbFGF (A and B) and pET32/hbFGF (C and D) vectors were transformed into *E. coli* BL21(DE3) (A and C) and ArcticExpress(DE3) (B and D) strains. The transformed cells were grown at 37°C to an OD600 of 0.6 and then induced with 0.2 mM IPTG for 6 to 24 h at 20°C. The induced cells were harvested and then the cell lysates were separated into soluble and the insoluble fraction by centrifugation. The solubility of the fusion proteins was analyzed on 15% SDS-PAGE and the gels stained with Coomassie Brilliant Blue. Equal cell concentration was loaded into each lane. The arrows indicate the position of recombinant fusion proteins. (-), total protein before induction; T, total cellular protein after induction; IS, insoluble fraction from induced cells; S, soluble fraction from induced cells.

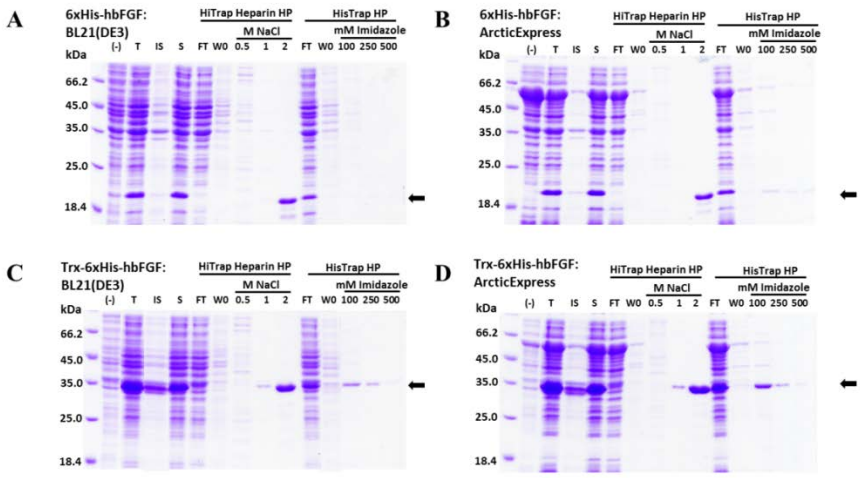

Figure 4.4 SDS-PAGE of recombinant hbFGF fusion proteins at various stages of purification. (A) pET28/hbFGF in *E. coli* BL21(DE3), (B) pET28/hbFGF in *E. coli* ArcticExpress(DE3), (C) pET32/hbFGF in *E. coli* BL21(DE3),(D) pET32/hbFGF in *E. coli* ArcticExpress(DE3). The cells were grown until the OD600 reached 0.6 and were then induced with 0.2 mM IPTG at 20°C for 18 h. The cells (1g wet cell weight) were harvested, and the recombinant proteins were purified on heparin-sepharose affinity chromatography (HiTrap HeparinHP) and Ni-IMAC (HisTrap HP) columns under native conditions. The proteins were analyzed on a 15% SDS-PAGE gel and stained with Coomassie Brilliant Blue R-250. These results represent three independent experiments. The lanes from left to right are protein molecular weight markers (masses marked in kilodaltons); (-), total protein before induction; T, total cellular protein after induction; IS, insoluble fractions from induced cells; S, soluble fractions from induced cells; FT, flow through fractions; W0, initial washing buffer fraction. For the HiTrap Heparin HP column, the fusion proteins were eluted with 0.5, 1 and 2 M NaCl. For HisTrap HP, the fusion proteins were eluted with 100, 250 and 500mM imidazole. The arrows indicate the fusion proteins.

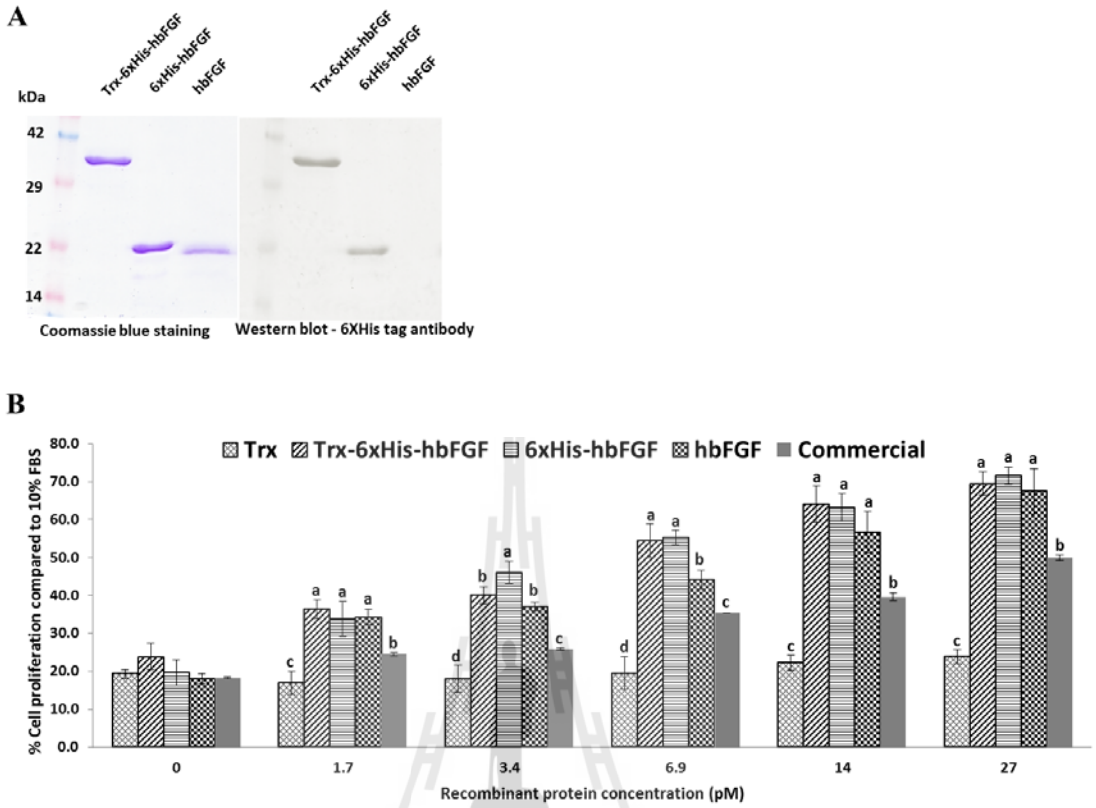

Figure 4.5 Analysis of purified recombinant hbFGF on SDS-PAGE and western blot (A) and biological activity (B). Purified 6xHis-hbFGF and Trx-6xHis-hbFGF fusion proteins and hbFGF were separated on a15% SDS-PAGE gel and were then either stained with Coomassie Brilliant Blue or transferred to a nitrocellulose membrane and identified by immunoblotting with a goat anti-His antibody (A). Comparison of the percentage of growth of NIH3T3 after 72 h incubation with Trx-tag ( ), Trx-6xHis-hbFGF ( ), 6xHis-hbFGF ( ), hbFGF ( ) and commercial hbFGF (Invitrogen) ( ) (B). Different letters (a, b, c, d) within the same concentration indicate significant different ($P < 0.05$).

Table 4.2 Purification of hbFGF fusion proteins from recombinant expression in *E. coli* with different expression vectors and expression hosts with HiTrap Heparin HP column.

Expression vector	Expression Host	Fraction	Total protein (mg)[1]	hbFGF fusion protein (mg)[1,2]	Purity (%)[3]	Yield (%)[4]
pET28/hbFGF	BL21(DE3)	Cell lysate	105.7	13.8	13.1	100
		Soluble protein	84.6	13.6	16.1	98.6
		0.5 M NaCl	3.5	0	0	0
		1 M NaCl	0	0	0	0
		2 M NaCl	13.1	12.4	94.7	90
	ArcticExpress	Cell lysate	117.4	13.4	11.4	100
		Soluble protein	92.8	12.9	13.9	96.3
		0.5 M NaCl	2.8	0	0	0
		1 M NaCl	0	0	0	0
		2 M NaCl	12.9	12.2	94.6	91
pET32/hbFGF	BL21(DE3)	Cell lysate	123.7	18.9	15.3	100
		Soluble protein	96.5	17.6	18.2	93.1
		0.5 M NaCl	3.9	0	0	0
		1 M NaCl	1.8	1.8	100	9.5
		2 M NaCl	15.2	14.8	97.3	78.3
	ArcticExpress	Cell lysate	135.2	21.7	16.1	100
		Soluble protein	107.8	20.3	18.8	93.5
		0.5 M NaCl	3.3	0	0	0
		1 M NaCl	4.7	4.7	100	21.7
		2 M NaCl	14.7	14.2	96.6	65.4

[1] Results are derived from 1 g of wet cell weight. Protein concentrations were estimated by the method of Bradford with bovine serum albumin (BSA) as a standard.

[2] The amount of hbFGF fusion protein was calculated from the ratio of the intensity of the target protein band to the intensity of all protein bands in the fraction. The band intensity was analyzed with Quantity One software.

[3] The % Purity was determined by dividing the amount of the target protein by that of the total proteins in each fraction and multiply by 100.

[4] The yield at each step in the procedure is the amount of the target protein at that step divided by the amount of the target protein in the first step (defined as 100%).

Table 4.3 Purification of hbFGF fusion proteins from recombinant expression in *E. coli* with different expression vectors and expression hosts by IMAC on a HisTrap HP column.

Expression vector	Expression Host	Fraction	Total protein (mg)[1]	hbFGF fusion protein (mg)[1,2]	Purity (%)[3]	Yield (%)[4]
pET28/hbFGF	BL21(DE3)	Cell lysate	108.3	10.2	9.4	100
		Soluble protein	84.6	9.6	11.3	94.1
		100 mM Imidazole	0.5	0.4	80	3.9
		250 mM Imidazole	0.3	0.3	100	2.9
		500 mM Imidazole	0.3	0.3	100	2.9
	ArticExpress	Cell lysate	115.7	12.6	10.9	100
		Soluble protein	95.4	11.9	12.5	94.4
		100 mM Imidazole	1.2	0.7	58.3	5.5
		250 mM Imidazole	0.5	0.5	100	4
		500 mM Imidazole	0.4	0.4	100	3.1
pET32/hbFGF	BL21(DE3)	Cell lysate	119.1	14.9	12.5	100
		Soluble protein	96.5	14.4	14.9	96.6
		100 mM Imidazole	3.5	3.1	88.6	20.8
		250 mM Imidazole	1.2	1.2	100	8.1
		500 mM Imidazole	0.5	0.5	100	3.4
	ArticExpress	Cell lysate	125.1	15.7	12.5	100
		Soluble protein	102.2	15.4	15.1	98.1
		100 mM Imidazole	4.2	3.5	83.3	22.3
		250 mM Imidazole	0.9	0.9	100	5.7
		500 mM Imidazole	0.4	0.4	100	2.5

[1] Results are derived from 1 g of wet cell weight. Protein concentrations were estimated by the method of Bradford with bovine serum albumin (BSA) as a standard.

[2] The amount of hbFGF fusion protein was calculated from the ratio of the intensity of the target protein band to the intensity of all protein bands in the fraction. The band intensity was analyzed with Quantity One software.

[3] The % Purity was determined by dividing the amount of the target protein by that of the total proteins in each fraction and multiply by 100.

[4] The yield at each step in the procedure is the amount of the target protein at that step divided by the amount of the target protein in the first step (defined as 100%).

Several hosts, such as *E. coli*, *Pichia pastoris*, insect cells, *Saccharomyces cerevisiae*, *Bacillus subtilis*, soybean seeds, silkworms (*Bombyx mori L.*) and rice seeds, have been used for expressing recombinant hbFGF with yields of approximately 1-100 mg of protein per liter of cell culture (Squires et al., 1988; Hill and Crane-Robinson, 1995; Wu et al., 2001; Ding et al., 2006; Mu et al., 2008; Gasparian et al., 2009; Song et al., 2013; Rassouli et al., 2013; Kwong et al., 2013). *E. coli* is the most favorable expression host for the expression of hbFGF because of the availability of versatile vector systems, host strains that facilitate high-level expression of recombinant proteins, and rapid growth on inexpensive media. Yet high level expression of recombinant proteins in *E. coli* can result in misfolding and the formation of insoluble aggregates known as inclusion bodies, which leads to biologically inactive proteins. Efficient strategies have been described for optimizing heterologous protein expression in *E. coli*, including the use of expression vectors, host strains, and fusion partners to increase solubility and stability. Expression at reduced temperatures, co-expression of molecular chaperones, varied inducer concentrations, consideration of codon usage and targeting the protein to export to the periplasm have also been tried (Sorensen and Mortensen, 2005; Esposito and Chatterjee, 2006; Rosano and Ceccarelli, 2009). However, the results of this study demonstrated that the *E. coli* hosts BL21(DE3) and ArcticExpress produced soluble target proteins at similar levels, which suggests that the presence of the chaperones and extra tRNA for the rare codons R, I, and L in the ArcticExpress cells was not necessary for the expression of the soluble hbFGF. No improvement in expression was obtained by decreasing the induction temperature from 20°C to 15°C. However, the Trx-6xHis-bFGF showed an approximately 2-fold higher expression than the expression of the 6xHis-hbFGF (Figure 4.2).

Expressed recombinant hbFGF in *E. coli* has been purified by various systems, including 6xHis, maltose binding protein (MBP), glutathione S-transferase (GST), Trx and b′a′ domains of human protein disulfide isomerase (PDI) (Lemaître et al., 1995; Sheng et al., 2003; Song et al., 2013; Rassouli et al., 2013). These fusion tags not only promote the solubility of the recombinant proteins but also increase the expression level and enhance the stability of the fused proteins. In this study, we compared two different *E. coli* fusion expression strategies (Trx- and 6xHis-tagged expression) under various conditions intended to increase their efficiency of producing soluble recombinant hbFGF fusion protein. The Trx-tagged expression in combination with the low induction temperatures (20°C) resulted in higher expression of Trx-6xHis-hbFGF compared to 6xHis-hbFGF. The *Trx* gene is believed to help with the efficient transcription and translation initiation to improve the protein yield and the Trx protein also acts as an intramolecular chaperone mediating the formation of its fusion partner. This mediation may be why the expression level of Trx-6xHis-hbFGF was higher than the expression of 6xHis-hbFGF. The 6xHis tags were fused in-frame at the N-termini of both fusion proteins to use in the purification step for the fusion proteins. The fusion proteins were purified by IMAC. However, we found that only approximately 20-60% of the expressed fusion proteins bind to the Ni^{2+} IMAC. This result was inconsistent with previous studies by Rassouli et al., 2013 and Gasparian et al., 2009. Several purification processes have been established for the production of recombinant hbFGF from *E. coli*, including single-step purification with heparin affinity chromatography or immobilized metal ion affinity chromatography, combination of cation exchange column chromatography and heparin affinity chromatography and combination of continuous bed or expanded bed chromatography and heparin affinity chromatography. It is known that hbFGF contains a unique heparin-binding domain that is required for the formation of a

stable tri-molecular signaling complex with the hbFGF ligand and receptor tyrosine kinases (Dowd et al., 1999). In this study, the use of an initial Ni^{2+} IMAC purification step was not effective because part of the hbFGF was lost, which resulted in low protein recovery. However, the use of heparin-sepharose affinity chromatography significantly improved the hbFGF fusion protein recovery. Additionally, when the purified Trx-6xHis-hbFGF was treated with enterokinase at 28°C for 8 h, aggregation of the protein occurred and led to the loss of the target protein, which might be due to protein instability after the tag removal.

4.4.2 Biological activity of recombinant hbFGF on the NIH3T3 cell proliferation assay

An MTT cell proliferation assay using the NIH3T3 cell line was used to measure the biological activity of the purified 6xHis-hbFGF, the Trx-6xHis-hbFGF fusion proteins and the hbFGF generated by removal of the tag. In the assay, NIH3T3 cells were switched from medium containing fetal bovine serum (FBS) to serum-free medium containing either the purified recombinant proteins or commercial hbFGF (Invitrogen, 13256-029). The results of five independent experiments were used to determine the activity of the recombinant hbFGF in the assay. The results were calculated as the percentage of treated NIH3T3 cell growth compared to the cells treated with media containing 10% FBS alone (Figure 4.5B).The results showed that the NIH3T3 cells treated with the recombinant Trx-tag (negative control) can proliferate at only approximately 20% of the rate of the cells grown in 10% FBS. This result indicated that the Trx-tag alone had no effect on the proliferation of the cells. However, cells treated with the recombinant proteins produced in this study can proliferate in a dose-dependent manner with effective doses (ED_{50}) of approximately 6.2, 6.3 and 12.4 pM for the 6xHis-

hbFGF and Trx-6xHis-hbFGF fusion proteins and the hbFGF, respectively. The result demonstrated that the recombinant hbFGF generated in this study has a significantly higher potency (ED_{50} 6 to 12 pM; approximately 0.1-0.2 ng hbFGF ml^{-1}) than the commercially available hbFGF (Invitrogen) (ED_{50} 27 pM). This finding was consistent with previous reports of the mitogenic activity of recombinant hbFGF, which had an ED_{50} of approximately 0.15-3 ng ml^{-1} (Gasparian et al., 2009; Song et al., 2013). Moreover, the purified hbFGF produced in this study showed higher activity when compared to the activity of the commercial hbFGF. The results demonstrated that the presence of the fusion partner did not interfere with the biological activity of the proteins. This outcome is consistent with previous studies, which indicated that it is not necessary to enzymatically remove the tags after the purification of some growth factor fusion proteins (Imsoonthornruksa et al., 2011; Rassouli et al., 2013).

4.4.3 Recombinant hbFGF fusion proteins support the undifferentiated stage of human pluripotent stem cells

Because our recombinant hbFGF showed similar NIH3T3 cell proliferation activity with the two fusions tagged recombinant proteins, we decided to test the ability of only the purified 6xHis-hbFGF and the Trx-6xHis-hbFGF fusion proteins, to maintain the undifferentiated stage of hESCs and hiPSCs compared to commercial hbFGF. The hESCs and the hiPSCs were grown on mitotically inactive HFF in gelatin-coated dishes and cultured in media supplemented with approximately 0.4 nM 6xHis-hbFGF, Trx-6xHis-hbFGF or commercial hbFGF (8 ng ml^{-1}). The culture media was changed daily. The colonies were passaged every 5-7 days by manual cutting with 23G needles. After 8-10 passages of the culture, the proliferation rate and colony morphology (Baharvand et al., 2006) of the hESCs and hiPSCs showed no differences between the cells cultured

with the purified 6xHis-hbFGF or Trx-6xHis-hbFGF fusion proteins or the commercial hbFGF (Figure 4.6A and B). The colonies of hESCs and hiPSCs cultured in media containing the fusion proteins were positive for the enzyme alkaline phosphatase, the POU family transcription factor OCT4, and the stage-specific embryonic antigens SSEA-3 and SSEA-4, which confirms that the cells maintained pluripotent characteristics (Figure 4.6A and B). Negative control was also performed to check for the auto-fluorescence. This was done by eliminating the primary antibody but add only the secondary antibody, no fluorescent signals was observed (data not shown). Relative gene expression analysis of the pluripotency-associated genes OCT4, SOX2 and NANOG found no significant differences ($P>0.05$) in the expression levels between the cells that were cultured in the presence of the purified 6xHis-hbFGF or Trx-6xHis-hbFGF fusion proteins or the commercial hbFGF (Figure 4.7A and B). Additionally, the expression of the differentiation markers PAX6, CDX2 and BRACHYURY was not detected (Figure 4.7A and B). This study demonstrated that our purified 6xHis-hbFGF and Trx-6xHis-hbFGF fusion proteins could be used to maintain and expand several passages of undifferentiated hESCs and iPSCs in culture just as well as the commercial hbFGF from Invitrogen.

In conclusion, this study, we have demonstrated a simple and efficient procedure for the production of soluble hbFGF fusion proteins, namely 6xHis-bFGF and Trx-6xHis-hbFGF, in *E. coli*, which yields approximately 60-80 mg l^{-1} of culture that can be simply purified to a protein with a purity of > 95%. The fusion proteins exhibit biological activity and can be used directly in human stem cell culture without removing the fusion tags. The purified hbFGF fusion proteins were applicable to human stem cell cultivation. This study described an easy and cost-effective method for large-scale

production of biologically active hbFGF fusion protein, which is needed for large-scale cultivation of human stem cells for biomedical applications.

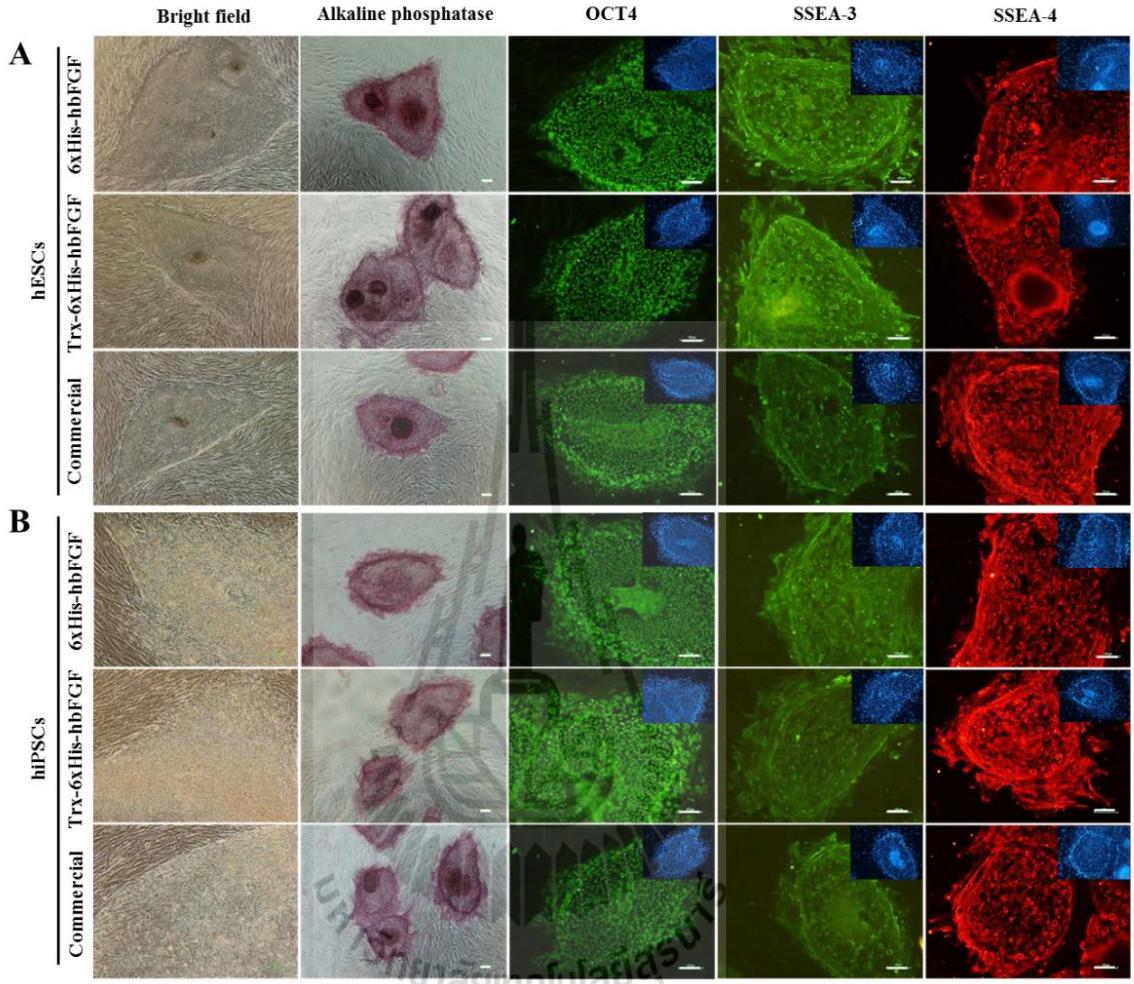

Figure 4.6 Recombinant hbFGF fusion proteins support several passages of undifferentiated human ESCs and iPSCs. The culture media supplemented with the 6xHis-bhFGF, Trx-6xHis-hbFGF or commercial hbFGF maintained the morphology of undifferentiated hESCs (A) and iPSCs (B) colonies. The cells were positive for alkaline phosphatase activity, and immunostaining of expanded colonies revealed expression of nuclear pluripotency (OCT4) and cell-surface markers (SSEA-3 and SSEA-4). Insert boxes show Hoechst 33342 nuclear staining. Scale bars, 100 μm.

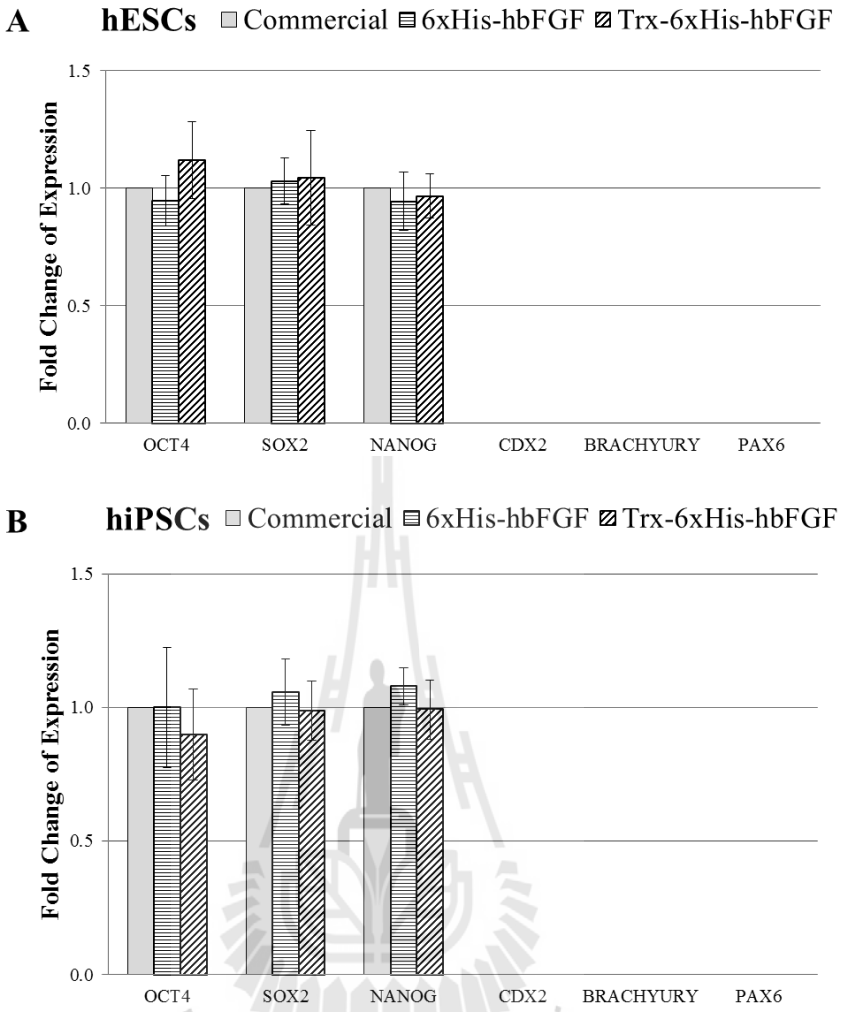

Figure 4.7 qPCR quantitative analysis of pluripotency-associated and the differentiation markers expression in hESCs (A) and iPSCs (B) that were cultured in the media supplemented with commercial hbFGF (□), 6xHis-hbFGF (▤) and Trx 6xHis-hbFGF (▨) for 8-10 passages. These results are representative of three independent experiments. The values were standardized to GAPDH expression. The data are presented as the mean ± SD values indicate non significant different (P >0.05).

4.5 References

Baharvand, H., Ashtiani, S. K., Taee, A., Massumi, M., Valojerdi, M. R., Yazdi, P. E., Moradi, S. Z., and Farrokhi, A. (2006). Generation of new human embryonic stem cell lines with diploid and triploid karyotypes. **Dev. Growth Differ.** 48: 117-128.

Bertani, G. (1952). Studies on Lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. **J. Bacteriology** 62: 293-300.

Ding, S. H., Huang, L. Y., Wang, Y. D., Sun, H. C., and Xiang, Z. H. (2006). High-level expression of basic fibroblast growth factor in transgenic soybean seeds and characterization of its biological activity. **Biotechnol Lett.** 28: 869-875.

Ding, V. M., Ling, L., Natarajan, S., Yap, M. G., Cool, S. M., and Choo, A. B. (2010). FGF-2 modulates Wnt signaling in undifferentiated hESC and iPS cells through activated PI3-K/GSK3beta signaling. **J. Cell Physiol.** 225: 417-428.

Dowd, C. J., Cooney, C. L., and Nugent, M. A. (1999). Heparin sulfate mediates bFGF transport through basement membrane by diffusion with rapid reversible binding. **J. Biol. Chem.** 274: 5236-5244.

Esposito, D., and Chatterjee, D. K. (2006). Enhancement of soluble protein expression through the use of fusion tags. **Curr. Opin. Biotechnol.** 17: 353-358.

Gasparian, M. E., Elistratov, P. A., Drize, N. I., Nifontova, I. N., Dolgikh, D. A., and Kirpichnikov, M. P. (2009). Overexpression in *Escherichia coli* and purification of human fibroblast growth factor (FGF-2). **Biochemistry (Mosc)** 74: 221-225.

Hills, D., and Crane-Robinson, C. (1995). Baculovirus expression of human basic fibroblast growth factor from a synthetic gene: role of the Kozak consensus and comparison with bacterial expression. **Biochim. Biophys. Acta** 1260: 14-20.

Imsoonthornruksa, S., Noisa, P., Parnpai, R., Ketudat-Cairns, M. (2011). A simple method for production and purification of soluble and biologically active recombinant human leukemia inhibitory factor (hLIF) fusion protein in Escherichia coli. **J. Biotechnol.** 151: 295-302.

Kupradid, C., Charoenrat, T., and Ketudat-Cairns, M. Bovine enterokinase light chain production by *Pichia pastoris*: Effect of induction temperature. **Thai J. Biotechnol.** 8: 99-105.

Kwong, K. W., Ng, K. L., Lam, C. C., Wang, Y. Y., and Wong, W. K. (2013). Authentic human basic fibroblast growth factor produced by secretion in *Bacillus subtilis*. **Appl. Microbiol. Biotechnol.** 97: 6803-6811.

Lemaître, G., Laaroubi, K., Soulet, L., Barritault, D., and Miskulin, M. (1995). Production and purification of active FGF2 via recombinant fusion protein. **Biochimie.**77: 162-166.

Li, J., Wang, G., Wang, C., Zhao, Y., Zhang, H., Tan, Z., Song, Z., Ding, M., and Deng, H. (2007). MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. **Differentiation** 75: 299-307.

Mu. X., Kong, N., Chen, W., Zhang, T., Shen, M., and Kirpichnikov, M. P. (2008). High-level expression, purification, and characterization of recombinant human basic fibroblast growth factor in *pichia pastoris*. **Protein Expr. Purif.** 59: 282-288.

Nishishita, N., Shikamura, M., Takenaka, C., Takada, N., Fusaki, N., and Kawamata, S. (2012). Generation of virus-free induced pluripotent stem cell clones on a synthetic matrix via a single cell subcloning in the naïve state. **PLoS One** 7: e38389.

Picanco-Castro, V., Moreira, L. F., Kashima, S., and Covas, D. T. (2014). Can pluripotent stem cells be used in cell-based therapy? **Cell Reprogram.** 16: 98-107.

Pruksananonda, K., Rungsiwiwut, R., Numchaisrika, P., Ahnonkitpanit, V., Isarasena, N., and Virutamasen, P. (2012). Eighteen-year cryopreservation does not negatively affect the pluripotency of human embryos: evidence from embryonic stem cell derivation. **Biores. Open Access** 1: 166-173.

Rassouli, H., TabeBordbar, M. S., RezaeiLarijani, M., Pakzad, M., Baharvand, H., and Salekdeh, G. H. (2013). Cloning, expression and functional characterization of in-house prepared human basic fibroblast growth factor. **Cell J.** 14: 282-291.

Rosano, G. L., and Ceccarelli, E. A. (2009). Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. **Microb. Cell Fact.** 8: 41.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). **Molecular Cloning: A Laboratory Manual**. 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. volume 3, apendix B.12.

Sheng, Z., Chang, S. B., and Chirico, W. J. (2003). Expression and purification of a biologically active basic fibroblast growth factor fusion protein. **Protein Expr. Purif.** 27: 267-271.

Singh, V. K., Kalsan, M., Kumar, N., Saini, A., and Chandra, R. (2015). Induced pluripotent stem cells: applications in regenerative medicine, disease modeling, and drug discovery. **Front Cell Dev. Biol.** 3: 2.

Song, J. A., Koo, B. K., Chong, S. H., Kwak, J., Ryu, H. B., Nguyen, M. T., Vu, T. T., Jeong, B., Kim, S. W., and Choe, H. (2013). Expression and purification of

biologically active human FGF2 containing the b'a' domains of human PDI in Escherichia coli. **Appl. Biochem. Biotechnol.** 170: 67-80.

Sorensen, H. P., and Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. **Microb. Cell Fact.** 4: 1.

Squires, C. H., Childs, J., Eisenberg, S. P., Polverini, P. J., and Sommer, A. (1988). Production and characterization of human basic fibroblast growth factor from Escherichia coli. **J. Biol. Chem.** 263: 16297-16302.

Stewart, E. J., Aslund, F., and Beckwith, J. (1998). Disulfide bond formation in the Escherichia coli cytoplasm: an *in vivo* role reversal for the thioredoxins. **EMBO J**. 17: 5543-5550.

van der Sanden, B., Dhobb, M., Berger, F., and Wion, D. (2010). Optimizing stem cell culture. **J. Cell Biochem.** 111: 801-807.

Wu, X., Kamei. K., Sato, H., Sato, S. I., Takano, R., Ichida, M., Mori, H., and Hara, S. (2001). High-level expression of human acidic fibroblast growth factor and basic fibroblast growth factor in silkworm (*Bombyx mori L.*) using recombinant baculovirus. **Protein Expres. Purif.** 21: 192-200.

Zon, L. I., Dorman, D. M., and Orkin, S. H. (1989). The polymerase chain reaction colony miniprep. **Biotechniques** 7: 696-698.

CHAPTER V

OVERALL CONCLUSION

Stem cell biology is one of the biggest research fields of recent interest. It studies the properties of stem cells and develops tools and therapeutics techniques through modification and engineering of the cells for their potential use in medicine. Like completing a jigsaw puzzle, prior to the use of stem cells in the near future in clinical medicine as well as for drug testing, modeling, and treating disease, scientists need to answers many questions and also need to fully understand, many pieces of the puzzle such as how to establish pluripotent stem cell and maintain pluripotency and renewal properties of the stem cells, how pluripotency gene functions in regulating self-renewal and maintaining and promoting pluripotency and how to control the differentiation of stem cells into specialized cells, just for examples. These puzzles needed to be solved prior to the application of stem cells into clinical approach. The results of this study have help put some of the pieces of the jigsaw together to get a clearer picture of human stem cell biology.

The first experiment was done on OCT4, the transcription factor that plays important roles in mouse and human stem cell self-renewal, pluripotency, and somatic cell reprogramming. There has not been any reported on OCT4 associated protein complexes in human ESCs. To understand how OCT4 function in controlling maintaining pluripotency and self-renewal, we established an *in vivo* biotinylation system for streptavidin affinity purification of protein complexes interacting with OCT4 in human ESCs. With the TALENS technology, we successfully established the genetic targeting

of biotin ligase *BirA* to the *AAVS1* locus in human ESCs (WIBR#29BirA) and TALENS-mediated biotinylation of endogenous OCT4 in WIBR#29BirA, respectively. The targeted cells remained pluripotent based on expression of OCT4, SOX2 and NANOG markers and teratomas formation. To identify OCT4 interacting proteins, the targeted cells were culture in SILAC or label-free conditions followed by nuclear protein extraction and then streptavidin beads were used to capture biotinylated OCT4 and its associated proteins. The eluted proteins were separated on SDS-PAGE. Each lane was excised into multiple gel bands followed by in-gel tryptic digestion and peptide identification by LC-MS/MS. Mass spectrometry analysis resulted in the identification of 142 OCT4 interacting proteins. Gene ontology analyses revealed that the OCT4 interacting proteins can be classify into severed groups, including transcription factors, translation factors, chromatin remodeling, DNA repair, RNA processing, glycosyl transferase as well as epigenetic regulators including NuRD, SWI/SNF and FACT complexes that have been known to influence the self-renewal and pluripotency of ESCs. Some of the interactors (BEND3, FUS, RBM25, SFPQ and WDR82) were selected as candidates to test the function of OCT4 interacting proteins by shRNA knockdown experiments. The result demonstrated that SFPQ knockdown disrupted the pluripotency of human ESCs and induced cell death. Our result provided the first survey of the OCT4 interactome in human ESCs that integrated the protein-protein interaction landscape in human ESCs. We also provide the *in vivo* biotinylation system in human ESCs line that is useful and is an alternative tool to establish affinity tagged of other transcription factors (SOX2 and NANOG) for future investigation. Further analysis of the listed candidates at the functional level and system biology can greatly assist in the understanding of the human ESCs pluripotency network. The OCT4 interactors provided a resource to investigate molecular mechanisms of OCT4 function. The mechanism of investigation

will be also important to fulfill the understanding of human ESCs self-renewal and the maintenance of pluripotency. Moreover, the newly OCT4 associated proteins will be potential reprogramming factors for somatic cell reprogramming.

The study of stem cell biology, condition for maintenance of pluripotency of human ESCs in culture required hbFGF. hbFGF is now commercially available but still very expensive. The cost of cytokines and growth factors is one of the biggest issue for many laboratories. In this experiment, we demonstrated an uncomplicated, fast and low-cost preparation of homemade soluble biologically active hbFGF fusion proteins (Trx-6xHis-hbFGF and 6xHis-hbFGF) expressed in *E. coli*. The target proteins were purified with heparin-sepharose affinity chromatography and immobilized metal affinity chromatography. Under optimal conditions, approximately 60-80 mg of hbFGF fusion proteins was obtained from 1 liter culture broth. The fusion proteins exhibit biological activity via promotion of murine fibroblast NIH3T3 cell line proliferation. They can also be used directly in human ESCs and iPSCs culture without removing the fusion tags in a manner comparable to commercial hbFGF. We also demonstrated a simple purification step that able to produce recombinant hbFGF with $USD 0.88/mg of recombinant hbFGF fusion protein. It is cost effective when compare with commercial available that is approximately $2,800 - 21,000/mg of recombinant hbFGF. The use of the produced fusion proteins was found to significantly reduce the cost of media preparation in our laboratory approximately 30-60%. Our research described a simple, convenient and cost-effective system for expression and purification of biologically active soluble hbFGF, suggesting that it will have widespread usage as a tool in the production of several other growth factors and cytokines for stem cell research.

APPENDICES

APPENDIX A

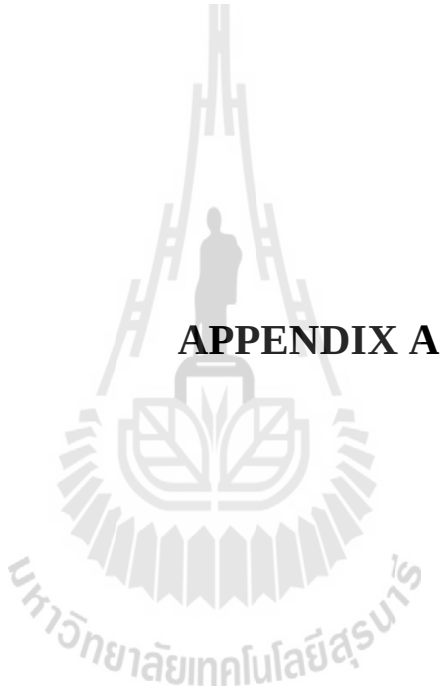

Table 1. OCT4 interacting proteins as identified by mass spectrometry*.

Gene names	Protein IDs	Description	Protein abundance ratio			
			Exp.1[a]	Exp.2[b]	Exp.3[c]	Exp.4[d]
Bait						
OCT4	Q01860	POU domain, class 5, transcription factor 1	11.5	8.3	16.0	27.9
NuRD complex						
CHD4	F5GWX5	Chromodomain-helicase-DNA-binding protein 4	2.4	-	-	5.8
HDAC1	Q5TEE2	Histone deacetylase 1 (Fragment)	-	3.0	-	∝
HDAC2	B3KRS5	Histone deacetylase	2.2	-	2.2	4.0
MTA1	H0Y4T7	Metastasis-associated protein MTA1	3.4	-	-	9.0
RBBP4	H0YDK2	Histone-binding protein RBBP4 (Fragment)	2.2	3.6	2.0	3.0
Chromatin remodeling factors						
BEND3	Q5T5X7	BEN domain-containing protein 3	2.0	2.6	2.0	11.5
BPTF	F5GXF5	Nucleosome-remodeling factor subunit BPTF (Fragment)	-	-	3.0	13.0
DEK	P35659	Protein DEK	2.4	2.2	-	-
KDM1A	O60341	Lysine-specific histone demethylase 1A	2.0	-	5.0	19.0
PDS5B	Q9NTI5	Sister chromatid cohesion protein PDS5 homolog B	-	-	2.3	9.7
SMARCA5	O60264	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	2.2	-	-	3.0
SMARCB1	C9JTA6	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1	-	2.1	2.0	2.0
SMARCE1	J3QKS7	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1 (Fragment)	2.1	5.8	-	-
SMCHD1	A6NHR9	chromosomes flexible hinge domain-containing protein 1	-	-	2.4	4.8

Table 1. Continued.

Gene names	Protein IDs	Description	Protein abundance ratio			
			Exp.1[a]	Exp.2[b]	Exp.3[c]	Exp.4[d]
BAF complex						
ACTL6A	C9JQT2	Actin-like protein 6A (Fragment)	2.1	-	-	6.0
SMARCA4	P51532-5	Isoform 5 of Transcription activator BRG1	2.4	3.3	-	9.4
SMARCC1	Q92922	SWI/SNF complex subunit SMARCC1	2.2	-	-	8.4
FACT complex						
SUPT16H	Q9Y5B9	FACT complex subunit SPT16	2.2	2.2	-	-
Nuclear assembly/organization						
MATR3	P43243	Matrin-3	2.1	-	-	2.3
Ubiquitination						
TRIM71	Q2Q1W2	E3 ubiquitin-protein ligase TRIM71	3.0	5.0	∝	10.0
UBAP2L	Q14157-4	Isoform 4 of Ubiquitin-associated protein 2-like	3.9	4.1	-	-
Transcriptional machinery						
ELAVL1	Q15717	ELAV-like protein 1	2.2	2.1	-	-
WDR82	Q6UXN9	WD repeat-containing protein 82	2.3	2.3	-	-
Transcription factors						
EWSR1	H7BY36	RNA-binding protein EWS (Fragment)	3.1	2.1	-	-
FUS	P35637-2	Isoform Short of RNA-binding protein FUS	4.3	3.3	5.4	-
GTF3C1	Q12789	General transcription factor 3C polypeptide 1	-	-	3.4	25.0
SALL1	H3BSM9	Sal-like protein 1 (Fragment)	3.8	-	5.0	5.0
SALL4	Q9UJQ4	Sal-like protein 4	2.1	2.5	-	3.3
SMARCA1	P28370-2	Probable global transcription activator SNF2L1	-	2.0	-	2.3

Table 1. Continued.

Gene names	Protein IDs	Description	Protein abundance ratio			
			Exp.1[a]	Exp.2[b]	Exp.3[c]	Exp.4[d]
Transcription factors (continued)						
SOX2	P48431	Transcription factor SOX-2	9.5	-	-	2.0
ZFR	Q96KR1	Zinc finger RNA-binding protein	4.1	∝	-	-
ZMYM2	Q9UBW7	Zinc finger MYM-type protein 2	3.3	-	-	6.8
ZMYM3	Q14202	Zinc finger MYM-type protein 3	3.5	-	-	9.0
Transcription cofactors						
CTNND1	O60716-24	Isoform 3 of Catenin delta-1	3.7	-	-	5.5
ILF2	Q12905	Interleukin enhancer-binding factor 2	2.6	2.6	-	
ILF3	Q12906	Interleukin enhancer-binding factor 3	2.3	2.3	-	4.2
KHDRBS1	Q07666-2	Isoform 2 of KH domain-containing, RNA-binding, signal transduction-associated protein 1	2.4	2.2	-	-
SND1	Q7KZF4	Staphylococcal nuclease domain-containing protein 1	2.2	2.2	-	2.2
RNA processing						
ADAR	P55265-5	Isoform 5 of Double-stranded RNA-specific adenosine deaminase	2.2	3.8	-	2.0
CPSF1	Q10570	Cleavage and polyadenylation specificity factor subunit 1	2.0	-	-	4.2
CSTF3	Q12996	Cleavage stimulation factor subunit 3	3.0	-	-	11.0
FUBP3	Q96I24	Far upstream element-binding protein 3	2.8	-	2.0	-
GIGYF2	Q6Y7W6-4	Isoform 3 of PERQ amino acid-rich with GYF domain-containing protein 2	2.1	2.3	-	-
IGF2BP1	Q9NZI8	Insulin-like growth factor 2 mRNA-binding protein 1	2.1	2.3	3.0	6.0

Table 1. Continued.

Gene names	Protein IDs	Description	Protein abundance ratio			
			Exp.1[a]	Exp.2[b]	Exp.3[c]	Exp.4[d]
RNA processing (continued)						
IGF2BP2	Q9Y6M1-5	Isoform 5 of Insulin-like growth factor 2 mRNA-binding protein 2	2.8	-	-	3.0
NCBP1	Q09161	Nuclear cap-binding protein subunit 1	2.2	-	-	2.0
PCBP2	Q15366-6	Isoform 6 of Poly(rC)-binding protein 2	2.6	2.7	-	-
PRPF19	H0YGF3	Pre-mRNA-processing factor 19 (Fragment)	4.2	2.5	-	-
PRPF40A	O75400-2	Isoform 2 of Pre-mRNA-processing factor 40 homolog A	2.4	2.3	-	-
PRPF6	O94906-2	Isoform 2 of Pre-mRNA-processing factor 6	2.6	2.1	-	-
PTBP2	Q9UKA9	Polypyrimidine tract-binding protein 2	2.4	-	4.0	2.1
PUF60	Q9UHX1-4	Poly(U)-binding-splicing factor PUF60	2.1	-	∝	-
RBM14	Q96PK6	RNA-binding protein 14	2.0	2.0	-	-
RBM39	H0Y4X3	RNA-binding protein 39 (Fragment)	2.5	2.5	-	4.0
SF3A1	Q15459	Splicing factor 3A subunit 1	2.9	2.5	-	-
SF3A3	E7EUT8	Splicing factor 3A subunit 3	2.1	2.1	-	-
SF3B1	O75533	Splicing factor 3B subunit 1	2.3	2.1	-	-
SFPQ	P23246	Splicing factor, proline- and glutamine-rich	2.0	2.0	-	2.0
U2AF2	P26368-2	Isoform 2 of Splicing factor U2AF 65 kDa subunit	2.1	2.5	-	-
Translation factors/helicases						
ABCF1	Q8NE71	ATP-binding cassette sub-family F member 1	2.2	3.0	-	-
DDX17	Q92841	Probable ATP-dependent RNA helicase DDX17	2.1	-	-	2.1
DDX19A	I3L0H8	ATP-dependent RNA helicase DDX19A	3.0	-	-	2.0

Table 1. Continued.

Gene names	Protein IDs	Description	Protein abundance ratio			
			Exp.1[a]	Exp.2[b]	Exp.3[c]	Exp.4[d]
Translation factors/helicases (continued)						
DDX39B	Q13838	Spliceosome RNA helicase DDX39B	2.3	-	-	3.1
DDX41	Q9UJV9	Probable ATP-dependent RNA helicase DDX41	2.2	2.5	-	-
DDX42	Q86XP3-2	Isoform 2 of ATP-dependent RNA helicase DDX42	2.2	2.2	-	2.3
DDX47	Q9H0S4-2	Isoform 2 of Probable ATP-dependent RNA helicase DDX47	2.0	2.8	-	4.0
DDX6	P26196	Probable ATP-dependent RNA helicase DDX6	2.5	2.4	-	-
DHX15	O43143	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	2.4	2.6	-	6.0
EFTUD2	Q15029-2	Isoform 2 of 116 kDa U5 small nuclear ribonucleoprotein component	2.5	2.3	-	-
EIF4A1	P60842	Eukaryotic initiation factor 4A-I	2.4	2.1	-	2.6
EIF4A3	P38919	Eukaryotic initiation factor 4A-III	2.1	-	-	2.3
GTPBP1	O00178	GTP-binding protein 1	-	-	2.0	3.0
HELLS	Q9NRZ9-4	Lymphoid-specific helicase	-	2.3	-	15.0
SNRNP200	O75643	U5 small nuclear ribonucleoprotein 200 kDa helicase	2.4	4.1	-	-
SNRNP70	P08621-2	Isoform 2 of U1 small nuclear ribonucleoprotein 70 kDa	2.5	2.2	-	-
Replication						
RFC2	P35250-2	Isoform 2 of Replication factor C subunit 2	2.4	4.2	-	-
RFC3	P40938-2	Isoform 2 of Replication factor C subunit 3	2.1	2.0	-	2.0
RFC4	P35249	Replication factor C subunit 4	-	-	3.0	3.3
MCM5	B1AHB1	DNA replication licensing factor MCM5	∝	-	-	3.0

Table 1. Continued.

Gene names	Protein IDs	Description	Protein abundance ratio			
			Exp.1[a]	Exp.2[b]	Exp.3[c]	Exp.4[d]
Replication (continued)						
SMC1A	Q14683	Structural maintenance of chromosomes protein 1A	2.4	2.3	5.5	4.5
SMC2	O95347	chromosomes protein 2	-	-	2.0	6.3
SMC3	Q9UQE7	Structural maintenance of chromosomes protein 3	2.2	-	-	4.6
SMC4	Q9NTJ3	chromosomes protein 4	-	-	2.5	28.0
Recombination/repair						
MSH2	E9PHA6	DNA mismatch repair protein Msh2	2.3	2.7	-	7.6
MSH6	P52701	DNA mismatch repair protein Msh6	2.7	2.9	-	3.7
PARP1	P09874	Poly [ADP-ribose] polymerase 1	2.4	2.2	-	-
TOP2A	P11388	DNA topoisomerase 2-alpha	2.0	2.2	-	2.7
TOP2B	E9PCY5	DNA topoisomerase 2 (Fragment)	2.1	2.0	-	4.2
XRCC5	C9JZ81	X-ray repair cross-complementing protein 5 (Fragment)	3.0	-	∝	12.0
XRCC6	B1AHC9	X-ray repair cross-complementing protein 6	2.5	-	-	4.0
LRRC47	Q8N1G4	Leucine-rich repeat-containing protein 47	2.4	2.9	-	3.3
PC	P11498	Pyruvate carboxylase, mitochondrial	2.7	2.1	-	
RAD21	O60216	Double-strand-break repair protein rad21 homolog	4.0	-	-	6.0
L1TD1	Q5T7N2	LINE-1 type transposase domain-containing protein 1	-	-	2.0	3.0
Heterogeneous nuclear ribonucleoproteins						
HNRNPA2B1	P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	2.2	3.3	-	-
HNRNPA3	P51991-2	Heterogeneous nuclear ribonucleoprotein A3	2.2	2.1	-	-
HNRNPC	G3V4W0	Heterogeneous nuclear ribonucleoproteins C1/C2 (Fragment)	2.2	2.0	-	-

Table 1. Continued.

Gene names	Protein IDs	Description	Protein abundance ratio			
			Exp.1[a]	Exp.2[b]	Exp.3[c]	Exp.4[d]
Heterogeneous nuclear ribonucleoproteins (continued)						
HNRNPD	D6RAF8	Heterogeneous nuclear ribonucleoprotein D0	-	2.2	-	2.0
HNRNPL	P14866	Heterogeneous nuclear ribonucleoprotein L	2.3	2.1	-	5.5
HNRNPLL	C9JJZ7	Heterogeneous nuclear ribonucleoprotein L-like (Fragment)	2.4	-	-	7.0
HNRNPM	P52272	Heterogeneous nuclear ribonucleoprotein M	-	-	2.4	4.3
HNRNPR	O43390	Heterogeneous nuclear ribonucleoprotein R	3.1	2.3	-	3.0
HNRNPU	Q00839-2	Isoform Short of Heterogeneous nuclear ribonucleoprotein U	2.1	2.1	-	-
HNRNPUL2	Q1KMD3	Heterogeneous nuclear ribonucleoprotein U-like protein 2	2.3	2.8	-	-
SYNCRIP	O60506-4	Isoform 4 of Heterogeneous nuclear ribonucleoprotein Q	2.1	-	-	6.0
Transport mechanism						
ATP1A1	P05023	Sodium/potassium-transporting ATPase subunit alpha-1	-	-	∝	3.0
COPA	P53621	Coatomer subunit alpha	1.9	-	-	2.2
COPB1	P53618	Coatomer subunit beta	2.6	2.8	-	6.5
COPB2	B4DZI8	Coatomer protein complex, subunit beta 2 (Beta prime), isoform CRA_b	2.6	-	-	2.1
COPG2	Q9UBF2	Coatomer subunit gamma-2	-	-	2.0	12.0
NSF	P46459	Vesicle-fusing ATPase	-	-	4.0	11.0
SEC24A	O95486	Protein transport protein Sec24A	-	-	2.0	14.0

Table 1. Continued.

Gene names	Protein IDs	Description	Protein abundance ratio			
			Exp.1[a]	Exp.2[b]	Exp.3[c]	Exp.4[d]
Miscellaneous						
ARGLU1	Q9NWB6	Arginine and glutamate-rich protein 1	2.2	2.3	-	2.0
CLTC	Q00610-2	Isoform 2 of Clathrin heavy chain 1	2.5	2.1	-	-
DIS3	F2Z2C0	Exosome complex exonuclease RRP44	-	-	2.0	9.5
DNM2	P50570-3	Dynamin-2	2.6	-	-	8.2
DYNC1H1	Q14204	Cytoplasmic dynein 1 heavy chain 1	-	-	4.4	3.5
HBA1	P69905	Hemoglobin subunit alpha	-	-	∝	2.5
HDLBP	H0Y394	Vigilin (Fragment)	-	-	2.0	3.0
HSP90AB1	P08238	Heat shock protein HSP 90-beta	2.5	3.0	-	
HSP90B1	P14625	Endoplasmin	2.0	2.0	-	3.0
ILK	Q13418	Integrin-linked protein kinase	-	-	2.0	11.0
IQGAP1	P46940	Ras GTPase-activating-like protein IQGAP1	2.0	2.5	3.3	-
KPNA2	P52292	Importin subunit alpha-2	-	2.4	-	2.5
LMNB1	E9PBF6	Lamin-B1	-	-	3.5	12.0
NCL	P19338	Nucleolin	3.0	7.7	-	-
NOP56	H0Y653	Nucleolar protein 56 (Fragment)	2.0	-	2.4	-
PBRM1	Q86U86-4	Isoform 4 of Protein polybromo-1	2.0	-		5.0
PDK1	F8WC75	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 1, mitochondrial		-	99.0	5.0
PFKL	P17858	6-phosphofructokinase, liver type	2.7	-	2.0	8.0
PFKP	Q01813	ATP-dependent 6-phosphofructokinase, platelet type	-	-	8.0	6.0
PRKDC	P78527	DNA-dependent protein kinase catalytic subunit	-	-	2.7	2.7

Table 1. Continued.

Gene names	Protein IDs	Description	Protein abundance ratio			
			Exp.1[a]	Exp.2[b]	Exp.3[c]	Exp.4[d]
Miscellaneous (continued)						
PTBP3	O95758	Polypyrimidine tract-binding protein 3	-	-	2.0	2.0
RBMX	H0Y6E7	RNA-binding motif protein, X chromosome, N-terminally processed (Fragment)	2.4	2.3	-	-
RUVBL1	Q9Y265	RuvB-like 1	-	-	∝	4.0
SKIV2L2	P42285	Superkiller viralicidic activity 2-like 2	-	2.4	2.2	
SLC2A1	P11166	Solute carrier family 2, facilitated glucose transporter member 1	2.6	-	-	4.0
SPTAN1	Q13813-3	Isoform 3 of Spectrin alpha chain, non-erythrocytic 1	2.3	2.6	-	-
SRP72	O76094-2	Isoform 2 of Signal recognition particle subunit SRP72	2.1	-	-	2.0
TAOK3	H0YF68	Serine/threonine-protein kinase TAO3 (Fragment)	2.5	-	-	6.0
VIM	P08670	Vimentin	4.0	7.3	2.5	
ZNF326	Q5BKZ1	DBIRD complex subunit ZNF326	2.3	2.6	-	-

[a]SILAC condition on SA-IP experiment, protein abundance ratio of WIBR#29OCT4FLAG-Bio vs WIBR#29BirA.

[b]SILAC condition on SA-IP experiment, protein abundance ratio of WIBR#29OCT4FLAG-Bio vs WIBR#29BirA.

[c]Lable-free condition on FLAG-IP experiment, protein abundance ratio of WIBR#29OCT4FLAG-Bio vs WIBR#29BirA.

[d]Lable-free condition, protein abundance ratio of WIBR#29 against endogenous OCT4 antibody vs WIBR#2 against IgG antibody.

*The threshold ratio value ≥ 2.0 or ∝ (2/0) of identified proteins was selected for the lists of OCT4 interacting proteins. The detected proteins in at least two independent IP-MS experiments were consider as confidential OCT4 interacting proteins.

APPENDIX B

Table 2. Functional categories of OCT4 interacting proteins.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
ABCF1	HUMAN|HGNC=70|UniProtKB=Q8NE71	ATP-binding cassette sub-family F member 1	ATP-binding cassette (ABC) transporter translation elongation factor hydrolase
ACTL6A	HUMAN|HGNC=24124|UniProtKB=O96019	Actin-like protein 6A	actin and actin related protein
ADAR	HUMAN|HGNC=225|UniProtKB=P55265	Double-stranded RNA-specific adenosine deaminase	DNA binding protein RNA binding protein deaminase kinase activator defense/immunity protein
ARGLU1	HUMAN|HGNC=25482|UniProtKB=Q9NWB6	Arginine and glutamate-rich protein 1	-
ATP1A1	HUMAN|HGNC=799|UniProtKB=P05023	Sodium/potassium-transporting ATPase subunit alpha-1	cation transporter ion channel hydrolase
BEND3	HUMAN|HGNC=23040|UniProtKB=Q5T5X7	BEN domain-containing protein 3	-
BPTF	HUMAN|HGNC=3581|UniProtKB=Q12830	Nucleosome-remodeling factor subunit BPTF	acetyltransferase chromatin/chromatin-binding protein
CHD4	HUMAN|HGNC=1919|UniProtKB=Q14839	Chromodomain-helicase-DNA-binding protein 4	DNA helicase helicase
CLTC	HUMAN|HGNC=2092|UniProtKB=Q00610	Clathrin heavy chain 1	vesicle coat protein
COPA	HUMAN|HGNC=2230|UniProtKB=P53621	Coatomer subunit alpha	vesicle coat protein
COPB1	HUMAN|HGNC=2231|UniProtKB=P53618	Coatomer subunit beta	vesicle coat protein
COPB2	HUMAN|HGNC=2232|UniProtKB=P35606	Coatomer subunit beta'	vesicle coat protein

Table 2. Continued

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
COPG2	HUMAN|HGNC=2237|UniProtKB=Q9UBF2	Coatomer subunit gamma-2	vesicle coat protein
CPSF1	HUMAN|HGNC=2324|UniProtKB=Q10570	Cleavage and polyadenylation specificity factor subunit 1	damaged DNA-binding protein mRNA polyadenylation factor
CSTF3	HUMAN|HGNC=2485|UniProtKB=Q12996	Cleavage stimulation factor subunit 3	mRNA polyadenylation factor
CTNND1	HUMAN|HGNC=2515|UniProtKB=O60716	Catenin delta-1	intermediate filament binding protein
DDX17	HUMAN|HGNC=2740|UniProtKB=Q92841	Probable ATP-dependent RNA helicase DDX17	RNA helicase helicase
DDX19A	HUMAN|HGNC=25628|UniProtKB=Q9NUU7	ATP-dependent RNA helicase DDX19A	RNA helicase helicase
DDX39B	HUMAN|HGNC=13917|UniProtKB=Q13838	Spliceosome RNA helicase DDX39B	RNA helicase helicase
DDX41	HUMAN|HGNC=18674|UniProtKB=Q9UJV9	Probable ATP-dependent RNA helicase DDX41	RNA helicase helicase
DDX42	HUMAN|HGNC=18676|UniProtKB=Q86XP3	ATP-dependent RNA helicase DDX42	RNA helicase helicase
DDX47	HUMAN|HGNC=18682|UniProtKB=Q9H0S4	Probable ATP-dependent RNA helicase DDX47	RNA helicase helicase
DDX6	HUMAN|HGNC=2747|UniProtKB=P26196	Probable ATP-dependent RNA helicase DDX6	RNA helicase helicase
DEK	HUMAN|HGNC=2768|UniProtKB=P35659	Protein DEK	-

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
DHX15	HUMAN|HGNC=2738|UniProtKB=O43143	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	RNA helicase helicase
DIS3	HUMAN|HGNC=20604|UniProtKB=Q9Y2L1	Exosome complex exonuclease RRP44	endoribonuclease exoribonuclease nuclease hydrolase
DNM2	HUMAN|HGNC=2974|UniProtKB=P50570	Dynamin-2	hydrolase small GTPase microtubule family cytoskeletal protein
DYNC1H1	HUMAN|HGNC=2961|UniProtKB=Q14204	Cytoplasmic dynein 1 heavy chain 1	hydrolase microtubule binding motor protein
EFTUD2	HUMAN|HGNC=30858|UniProtKB=Q15029	116 kDa U5 small nuclear ribonucleoprotein component	ribonucleoprotein translation elongation factor translation initiation factor hydrolase G-protein
EIF4A1	HUMAN|HGNC=3282|UniProtKB=P60842	Eukaryotic initiation factor 4A-I	RNA helicase helicase
EIF4A3	HUMAN|HGNC=18683|UniProtKB=P38919	Eukaryotic initiation factor 4A-III	RNA helicase helicase
ELAVL1	HUMAN|HGNC=3312|UniProtKB=Q15717	ELAV-like protein 1	-
EWSR1	HUMAN|HGNC=3508|UniProtKB=Q01844	RNA-binding protein EWS	transcription factor DNA binding protein mRNA splicing factor

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
FUBP3	HUMAN|HGNC=4005|UniProtKB=Q96I24	Far upstream element-binding protein 3	mRNA splicing factor ribonucleoprotein enzyme modulator
FUS	HUMAN|HGNC=4010|UniProtKB=P35637	RNA-binding protein FUS	transcription factor DNA binding protein mRNA splicing factor
GIGYF2	HUMAN|HGNC=11960|UniProtKB=Q6Y7W6	PERQ amino acid-rich with GYF domain-containing protein 2	-
GTF3C1	HUMAN|HGNC=4664|UniProtKB=Q12789	General transcription factor 3C polypeptide 1	-
GTPBP1	HUMAN|HGNC=4669|UniProtKB=O00178	GTP-binding protein 1	translation elongation factor translation initiation factor hydrolase G-protein
HBA1	HUMAN|HGNC=4824|UniProtKB=P69905	Hemoglobin subunit alpha	transfer/carrier protein
HBA1	HUMAN|HGNC=4823|UniProtKB=P69905	Hemoglobin subunit alpha	-
HDAC1	HUMAN|HGNC=4852|UniProtKB=Q13547	Histone deacetylase 1	reductase nucleic acid binding deacetylase
HDLBP	HUMAN|HGNC=4857|UniProtKB=Q00341	Vigilin	-
HELLS	HUMAN|HGNC=4861|UniProtKB=Q9NRZ9	Lymphoid-specific helicase	-
HNRNPA2B1	HUMAN|HGNC=5033|UniProtKB=P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	-

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
HNRNPA3	HUMAN|HGNC=24941|UniProtKB=P51991	Heterogeneous nuclear ribonucleoprotein A3	-
HNRNPC	HUMAN|HGNC=5035|UniProtKB=P07910	Heterogeneous nuclear ribonucleoproteins C1/C2	mRNA processing factor ribonucleoprotein
HNRNPD	HUMAN|HGNC=5036|UniProtKB=Q14103	Heterogeneous nuclear ribonucleoprotein D0	-
HNRNPL	HUMAN|HGNC=5045|UniProtKB=P14866	Heterogeneous nuclear ribonucleoprotein L	-
HNRNPLL	HUMAN|HGNC=25127|UniProtKB=Q8WVV9	Heterogeneous nuclear ribonucleoprotein L-like	-
HNRNPM	HUMAN|HGNC=5046|UniProtKB=P52272	Heterogeneous nuclear ribonucleoprotein M	ribonucleoprotein
HNRNPR	HUMAN|HGNC=5047|UniProtKB=O43390	Heterogeneous nuclear ribonucleoprotein R	ribosomal protein
HNRNPU	HUMAN|HGNC=5048|UniProtKB=Q00839	Heterogeneous nuclear ribonucleoprotein U	-
HNRNPUL2	HUMAN|HGNC=25451|UniProtKB=Q1KMD3	Heterogeneous nuclear ribonucleoprotein U-like protein 2	-
HSP90AB1	HUMAN|HGNC=5258|UniProtKB=P08238	Heat shock protein HSP 90-beta	Hsp90 family chaperone
HSP90B1	HUMAN|HGNC=12028|UniProtKB=P14625	Endoplasmin	Hsp90 family chaperone
IGF2BP1	HUMAN|HGNC=28866|UniProtKB=Q9NZI8	Insulin-like growth factor 2 mRNA-binding protein 1	mRNA splicing factor ribonucleoprotein enzyme modulator

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
IGF2BP2	HUMAN|HGNC=28867|UniProtKB=Q9Y6M1	Insulin-like growth factor 2 mRNA-binding protein 2	mRNA splicing factor ribonucleoprotein enzyme modulator
ILF2	HUMAN|HGNC=6037|UniProtKB=Q12905	Interleukin enhancer-binding factor 2	transcription cofactor
ILF3	HUMAN|HGNC=6038|UniProtKB=Q12906	Interleukin enhancer-binding factor 3	DNA binding protein RNA binding protein deaminase kinase activator defense/immunity protein
ILK	HUMAN|HGNC=6040|UniProtKB=Q13418	Integrin-linked protein kinase	non-receptor serine/threonine protein kinase non-receptor tyrosine protein kinase non-receptor serine/threonine protein kinase non-receptor tyrosine protein kinase
IQGAP1	HUMAN|HGNC=6110|UniProtKB=P46940	Ras GTPase-activating-like protein IQGAP1	G-protein modulator
KDM1A	HUMAN|HGNC=29079|UniProtKB=O60341	Lysine-specific histone demethylase 1A	DNA methyltransferase oxidase DNA binding protein
KHDRBS1	HUMAN|HGNC=18116|UniProtKB=Q07666	KH domain-containing, RNA-binding, signal transduction-associated protein 1	transcription cofactor mRNA splicing factor
KPNA2	HUMAN|HGNC=6395|UniProtKB=P52292	Importin subunit alpha-1	transfer/carrier protein

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
L1TD1	HUMAN|HGNC=25595|UniProtKB=Q5T7N2	LINE-1 type transposase domain-containing protein 1	-
LMNB1	HUMAN|HGNC=6637|UniProtKB=P20700	Lamin-B1	structural protein intermediate filament
LRRC47	HUMAN|HGNC=29207|UniProtKB=Q8N1G4	Leucine-rich repeat-containing protein 47	nucleic acid binding aminoacyl-tRNA synthetase
MATR3	HUMAN|HGNC=6912|UniProtKB=P43243	Matrin-3	DNA binding protein RNA binding protein
MCM5	HUMAN|HGNC=6948|UniProtKB=P33992	DNA replication licensing factor MCM5	DNA helicase helicase hydrolase
MSH2	HUMAN|HGNC=7325|UniProtKB=P43246	DNA mismatch repair protein Msh2	DNA binding protein
MSH6	HUMAN|HGNC=7329|UniProtKB=P52701	DNA mismatch repair protein Msh6	DNA binding protein
MTA1	HUMAN|HGNC=7410|UniProtKB=Q13330	Metastasis-associated protein MTA1	chromatin/chromatin-binding protein histone
NCBP1	HUMAN|HGNC=7658|UniProtKB=Q09161	Nuclear cap-binding protein subunit 1	mRNA splicing factor
NCL	HUMAN|HGNC=7667|UniProtKB=P19338	Nucleolin	-
NOP56	HUMAN|HGNC=15911|UniProtKB=O00567	Nucleolar protein 56	ribonucleoprotein
NSF	HUMAN|HGNC=8016|UniProtKB=P46459	Vesicle-fusing ATPase	-
PARP1	HUMAN|HGNC=270|UniProtKB=P09874	Poly [ADP-ribose] polymerase 1	glycosyltransferase DNA ligase DNA ligase

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
PBRM1	HUMAN|HGNC=30064|UniProtKB=Q86U86	Protein polybromo-1	-
PC	HUMAN|HGNC=8636|UniProtKB=P11498	Pyruvate carboxylase, mitochondrial	ligase
PCBP2	HUMAN|HGNC=8648|UniProtKB=Q15366	Poly(rC)-binding protein 2	mRNA splicing factor ribonucleoprotein enzyme modulator
PDK1	HUMAN|HGNC=8809|UniProtKB=Q15118	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 1, mitochondrial	protein kinase
PDS5B	HUMAN|HGNC=20418|UniProtKB=Q9NTI5	Sister chromatid cohesion protein PDS5 homolog B	chromatin/chromatin-binding protein
PFKL	HUMAN|HGNC=8876|UniProtKB=P17858	6-phosphofructokinase, liver type	carbohydrate kinase
PFKP	HUMAN|HGNC=8878|UniProtKB=Q01813	6-phosphofructokinase type C	carbohydrate kinase
PRKDC	HUMAN|HGNC=9413|UniProtKB=P78527	DNA-dependent protein kinase catalytic subunit	nucleotide kinase non-receptor serine/threonine protein kinase nucleic acid binding nucleotide kinase non-receptor serine/threonine protein kinase
PRPF19	HUMAN|HGNC=17896|UniProtKB=Q9UMS4	Pre-mRNA-processing factor 19	mRNA splicing factor
PRPF40A	HUMAN|HGNC=16463|UniProtKB=O75400	Pre-mRNA-processing factor 40 homolog A	-

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
PRPF6	HUMAN|HGNC=15860|UniProtKB=O94906	Pre-mRNA-processing factor 6	mRNA splicing factor
PTBP2	HUMAN|HGNC=17662|UniProtKB=Q9UKA9	Polypyrimidine tract-binding protein 2	-
PTBP3	HUMAN|HGNC=10253|UniProtKB=O95758	Polypyrimidine tract-binding protein 3	-
PUF60	HUMAN|HGNC=17042|UniProtKB=Q9UHX1	Poly(U)-binding-splicing factor PUF60	-
RAD21	HUMAN|HGNC=9811|UniProtKB=O60216	Double-strand-break repair protein rad21 homolog	-
RBBP4	HUMAN|HGNC=9887|UniProtKB=Q09028	Histone-binding protein RBBP4	receptor
RBM14	HUMAN|HGNC=14219|UniProtKB=Q96PK6	RNA-binding protein 14	-
RBM39	HUMAN|HGNC=15923|UniProtKB=Q14498	RNA-binding protein 39	-
RBMX	HUMAN|HGNC=9910|UniProtKB=P38159	RNA-binding motif protein, X chromosome	-
RFC2	HUMAN|HGNC=9970|UniProtKB=P35250	Replication factor C subunit 2	nucleotidyltransferase DNA-directed DNA polymerase
RFC3	HUMAN|HGNC=9971|UniProtKB=P40938	Replication factor C subunit 3	nucleotidyltransferase DNA-directed DNA polymerase
RFC4	HUMAN|HGNC=9972|UniProtKB=P35249	Replication factor C subunit 4	nucleotidyltransferase DNA-directed DNA polymerase
RUVBL1	HUMAN|HGNC=10474|UniProtKB=Q9Y265	RuvB-like 1	-

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
SALL1	HUMAN|HGNC=10524|UniProtKB=Q9NSC2	Sal-like protein 1	zinc finger transcription factor nucleic acid binding
SALL4	HUMAN|HGNC=15924|UniProtKB=Q9UJQ4	Sal-like protein 4	zinc finger transcription factor nucleic acid binding
SEC24A	HUMAN|HGNC=10703|UniProtKB=O95486	Protein transport protein Sec24A	vesicle coat protein
SF3A1	HUMAN|HGNC=10765|UniProtKB=Q15459	Splicing factor 3A subunit 1	mRNA splicing factor
SF3A3	HUMAN|HGNC=10767|UniProtKB=Q12874	Splicing factor 3A subunit 3	mRNA splicing factor
SF3B1	HUMAN|HGNC=10768|UniProtKB=O75533	Splicing factor 3B subunit 1	mRNA splicing factor
SFPQ	HUMAN|HGNC=10774|UniProtKB=P23246	Splicing factor, proline- and glutamine-rich	mRNA splicing factor
SKIV2L2	HUMAN|HGNC=18734|UniProtKB=P42285	Superkiller viralicidic activity 2-like 2	-
SLC2A1	HUMAN|HGNC=11005|UniProtKB=P11166	Solute carrier family 2, facilitated glucose transporter member 1	-
SMARCA1	HUMAN|HGNC=11097|UniProtKB=P28370	Probable global transcription activator SNF2L1	-
SMARCA4	HUMAN|HGNC=11100|UniProtKB=P51532	Transcription activator BRG1	DNA helicase helicase

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
SMARCA5	HUMAN|HGNC=11101|UniProtKB=O60264	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	-
SMARCB1	HUMAN|HGNC=11103|UniProtKB=Q12824	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1	DNA binding protein
SMARCC1	HUMAN|HGNC=11104|UniProtKB=Q92922	SWI/SNF complex subunit SMARCC1	transcription cofactor chromatin/chromatin-binding protein
SMARCE1	HUMAN|HGNC=11109|UniProtKB=Q969G3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1	HMG box transcription factor signaling molecule chromatin/chromatin-binding protein
SMC1A	HUMAN|HGNC=11111|UniProtKB=Q14683	Structural maintenance of chromosomes protein 1A	chromatin/chromatin-binding protein hydrolase
SMC2	HUMAN|HGNC=14011|UniProtKB=O95347	Structural maintenance of chromosomes protein 2	chromatin/chromatin-binding protein hydrolase
SMC3	HUMAN|HGNC=2468|UniProtKB=Q9UQE7	Structural maintenance of chromosomes protein 3	chromatin/chromatin-binding protein hydrolase

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
SMC4	HUMAN|HGNC=14013|UniProtKB=Q9NTJ3	Structural maintenance of chromosomes protein 4	chromatin/chromatin-binding protein hydrolase
SMCHD1	HUMAN|HGNC=29090|UniProtKB=A6NHR9	Structural maintenance of chromosomes flexible hinge domain-containing protein 1	-
SND1	HUMAN|HGNC=30646|UniProtKB=Q7KZF4	Staphylococcal nuclease domain-containing protein 1	transcription cofactor nucleic acid binding
SNRNP200	HUMAN|HGNC=30859|UniProtKB=O75643	U5 small nuclear ribonucleoprotein 200 kDa helicase	-
SNRNP70	HUMAN|HGNC=11150|UniProtKB=P08621	U1 small nuclear ribonucleoprotein 70 kDa	mRNA splicing factor
SOX2	HUMAN|HGNC=11195|UniProtKB=P48431	Transcription factor SOX-2	HMG box transcription factor nucleic acid binding
SPTAN1	HUMAN|HGNC=11273|UniProtKB=Q13813	Spectrin alpha chain, non-erythrocytic 1	non-motor actin binding protein
SRP72	HUMAN|HGNC=11303|UniProtKB=O76094	Signal recognition particle subunit SRP72	-
SUPT16H	HUMAN|HGNC=11465|UniProtKB=Q9Y5B9	FACT complex subunit SPT16	transcription cofactor nucleic acid binding
SYNCRIP	HUMAN|HGNC=16918|UniProtKB=O60506	Heterogeneous nuclear ribonucleoprotein Q	ribosomal protein
TAOK3	HUMAN|HGNC=18133|UniProtKB=Q9H2K8	Serine/threonine-protein kinase TAO3	-

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
TOP2A	HUMAN|HGNC=11989|UniProtKB=P11388	DNA topoisomerase 2-alpha	DNA topoisomerase isomerase enzyme modulator
TOP2B	HUMAN|HGNC=11990|UniProtKB=Q02880	DNA topoisomerase 2-beta	DNA topoisomerase isomerase enzyme modulator
TRIM71	HUMAN|HGNC=32669|UniProtKB=Q2Q1W2	E3 ubiquitin-protein ligase TRIM71	-
U2AF2	HUMAN|HGNC=23156|UniProtKB=P26368	Splicing factor U2AF 65 kDa subunit	mRNA splicing factor
UBAP2L	HUMAN|HGNC=29877|UniProtKB=Q14157	Ubiquitin-associated protein 2-like	-
VIM	HUMAN|HGNC=12692|UniProtKB=P08670	Vimentin	structural protein intermediate filament
WDR82	HUMAN|HGNC=28826|UniProtKB=Q6UXN9	WD repeat-containing protein 82	methyltransferase
XRCC5	HUMAN|HGNC=12833|UniProtKB=P13010	X-ray repair cross-complementing protein 5	-
XRCC6	HUMAN|HGNC=4055|UniProtKB=P12956	X-ray repair cross-complementing protein 6	DNA helicase helicase
ZFR	HUMAN|HGNC=17277|UniProtKB=Q96KR1	Zinc finger RNA-binding protein	DNA binding protein mRNA processing factor deaminase kinase activator defense/immunity protein
ZMYM2	HUMAN|HGNC=12989|UniProtKB=Q9UBW7	Zinc finger MYM-type protein 2	transcription factor kinase inhibitor

Table 2. Continued.

Gene Symbol	Gene ID	Gene Name	PANTHER Protein Class
ZMYM3	HUMAN|HGNC=13054|UniProtKB=Q14202	Zinc finger MYM-type protein 3	transcription factor kinase inhibitor
ZNF326	HUMAN|HGNC=14104|UniProtKB=Q5BKZ1	DBIRD complex subunit ZNF326	-

มหาวิทยาลัยเทคโนโลยีสุรนารี

BIOGRAPHY

Mr. Sumeth Imsoonthornruksa was born in 1981. He received a B. Sc. degree in Biotechnology from the Silapakorn University in 2003 and spent a year working as a research assistant in the group of Assoc. Prof. Dr. Rangsun Parnpai. A year later, he started to study master degree in Biotechnology at Suranaree University of Technology (SUT) with a scholarship from Thailand Research Fund-Master Research Grants (Grant No. MRG-WII495S021) and SUT. He graduated in September 2008. During his years as a graduate student, he worked in the field of animal biotechnology including animal cloning, *in vitro* fertilization, embryos freezing and molecular biology techniques. In 2012, he started his Ph. D. training in Molecular Biology and Protein Engineering Laboratory, School of Biotechnology, SUT under the supervision of Assoc. Prof. Dr. Mariena Ketudat-Cairns with funding from SUT-Ph.D. Scholarship Program (SUT-PhD/07/2554). In 2014, he visited Whitehead Institute, Massachusetts Institute of Technology, MA, USA as a Ph. D. visiting research scholar under supervision of Prof. Dr. Rudolf Jaenisch and Dr. Thorold W. Theunissen. Currently, his research interests focus on protein interaction networks of pluripotency factors in human stem cells, cellular reprogramming and recombinant protein productions for stem cells research. Currently, he has 7 first author and 8 co-authors in international journals and more than 20 oral and poster presentations in national and international conferences for his 12 years as student and research assistant at Suranaree University of Technology.