筑　波　大　学

博　士（医　学）　学　位　論　文

Locus-specific isolation of the *Nanog* chromatin identifies regulators relevant to pluripotency of mouse embryonic stem cells and reprogramming of somatic cells

（*Nanog* 領域に対する部位特異的クロマチン分離によるマウス ES 細胞の多能性制御や iPS 細胞誘導に関わる因子の同定）

2022

筑波大学大学院博士課程人間総合科学研究科

BURRAMSETTY ARUN KUMAR

INDEX

ABSTRACT

Even after decades of studying pluripotency, the lack of defining conditions for maintaining pluripotency long-term and efficient production of induced pluripotent stem cells (iPSCs) cells indicates our inadequacy in understanding pluripotency. Employing subtle methods of screening to identify pluripotency-related factors could be one reason. Identification of clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 lets researchers approach genome loci with ease and comfort, and utilization of dCas9 (deactivated Cas9) enables locus-specific pull-down of DNA, RNA as well as proteins, called CRISPR affinity purification *in situ* of regulatory elements (CAPTURE) method. It helps to understand regulatory mechanisms comprehensively at any given locus. To identify proteins/mechanisms that associate with pluripotency; I utilized CAPTURE method and baited *Nanog* promoter locus in mouse embryonic stem cells (mESCs). dCas9-mediated pull down of *Nanog* promoter revealed more than 300 proteins including some known *Nanog* promoter binding proteins like TRIM28, THRAP3, and BCLAF1 suggesting the ability of dCas9 to purify interacting proteome at the promoter regions. Based on gene ontology most of the purified proteins were RNA binding proteins (RBPs) with functions related to RNA splicing, mRNA processing and gene expression. Six proteins that were enriched with functions in aspects of gene regulation were chosen (BCLAF1, FUBP1, MSH6, PARK7, PSIP1, and THRAP3) and tested for their role in pluripotency. Chromatin immunoprecipitation (ChIP) of the candidate proteins showed enrichment of the *Nanog* promoter in a quantitative PCR (qPCR) analysis. Upon knockdown, these proteins exhibited various levels of regulation on self-renewal; differentiation; embryoid body (EB) development and somatic cell reprogramming, with revealing FUBP1 as an important regulator of pluripotency gene expression in mESCs. PARK7 and THRAP3 were needed to decrease pluripotency gene expression in the absence of LIF. PARK7 and THRAP3 also have inhibitory effects on somatic cell reprogramming. Knockdown of the candidate proteins diminished the growth potential of EBs, and regulated the expression of primitive endoderm marker (*Gata6*) and other developmental markers tested

Hence, unlike previous approaches, my *Nanog* locus-based proteome approach identified many new proteins regulating pluripotency. These results re-instigate that CAPTURE of *Nanog* promoter successfully identified new regulatory proteins involved in *Nanog* gene transcription and pluripotency, suggesting the robustness of this approach in unbiased identification of regulatory proteome at a given locus.

ABBREVIATIONS

2i	2 Inhibitory medium
AKT	Protein kinase B
BCLAF1	BCL2 Associated Transcription Factor 1
BHK	Baby Hamster Kidney
BirA	Biotin Ligase
BRD4	Bromodomain Containing 4
c-MYC	cellular Myelocytomatosis
CAPTURE	CRISPR Affinity Purification insitu of Regulatory Elements
CDH1	Cadherin 1
cDNA	Complementary DNA
CDX2	Caudal-type Homeobox transcription factor 2
ChIP	Chromatin Immunoprecipitation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAPI	4′,6-diamidino-2-phenylindole
dCas9	Deactivated Cas9
DD	Destabilization Domain
DMD	Duchene Muscular Dystropy
DMEM	Dulbecco Minimal Essential Medium
DNMT1	DNA methyltransferase 1
DTT	Dithiotreitol
EB	Embryoid Body
ECC	Embryonic Carcinoma Cell
EDTA	Ethylene Diamine Tetra Acetic acid
EGFP	Enhanced Green Fluorescent Protein
EMSA	Electrophoretic Mobility Shift Assay
enChIP	engineered DNA binding molecule-mediated chromatin immunoprecipitation
EPCAM	Epithelial Cell Adhesion Molecule
ERK	Extracellular signal-regulated kinase
ESC	Embryonic Stem Cell
ESRRB	Estrogen Related Receptor Beta
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
FDA	U.S. Food and Drug Administration
FGF	Fibroblast Growth factor
FOXA1	Forkhead Box A1
FOXA2	Forkhead Box A2
FUBP1	Far Upstream Binding Protein 1
Gbx2	Gastrulation Brain Homeobox 2
GO	Gene Ontology

gp130	Glycoprotein 130
gRNA	guide RNA
Gsk3	Glycogen synthase kinase
H3K27me3	Histone 3 lysine 27 trimethylation
HEK293T	Human Embryonic Kidney 293T
hESC	Humans ESC
hKO	Humanised Kusabira Orange
HMG	High Mobility Group
HRP	Horse Radish Peroxidase
HSC	Hematopoietic Stem Cells
iChIP	insertional chromatin immunoprecipitation
ICM	Inner Cell Mass
iPSC	induced Pluripotent Stem Cell
JAK	Janus Kinase
KLF2	Kruppel like zinc finger transcription factor 2
KLF4	Kruppel like zinc finger transcription factor 4
KR	Keima Red
LC-MS-MS	Liquid chromatography with tandem mass spectrometry
LIF	Leukemia Inhibitory Factor
LIFR	LIF receptor
MAPK	Mitogen Activated Protein Kianse
MEF	Mouse embryonic Fibroblasts
MEK	MAPK/ERK kinase
Meox1	Mesenchyme Homeobox 1
mESC	Mouse ESCs
MET	Mesenchymal to Epithelial Transition
mRNA	messenger RNA
MSC	Mesenchymal Stem Cells
MSH6	MutS Homolog 6
NANOG	Nanog Homeobox
ORF	Open reading Frame
PARK7	Parkinsonism Associated deglycase
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PGK	Phosphoglycerate Kinase
pH	Potential of Hydrogen
PI	Propidium Iodide
PI3K	Phosphatidylinositol-4,5-bisphosphate 3 kinase
PICh	proteomics of isolated chromatin segments
Plat-E	Platinum E
POU5F1 (OCT4)	POU Class 5 Homeobox 1
PRC2	Polycomb Repressive Complex 2

PRDM14	PR/SET Domain 14
PSC	Pluripotent Stem Cell
PSIP1	PC4 and SFRS1 associated proteein 1
PVDF	Polyvinylidene fluoride
qPCR	quantitative polymerase chain reaction
ReKO	Rex1-hKO
RNA-Seq	RNA sequencing
SALL4	Spalt Like Transcription Factor 4
SDS	Sodium Dodecyl Sulphate
SEM	Standard Error Mean
SeV	Sendai virus
SeVdp	Sendai Virus defective persistant
SH2	Src Homology 2
SNAIL1	Snail Family Transcriptional Repressor 1
SOX2	SRY-Box Transcription Factor 2
SP1	Specificity Protein Transcription factor 1
SP3	Specificity Protein Transcription factor 3
SSEA1	Stage Specific Embryonic Antigen 1
STAT3	signal transducer and activator of transcription
TAL	Transcription activator like
TALE	Transcription activator-like effector
TBS	Tris Buffer Saline
TBS-T	Tris Buffer Saline with Tween 20
TBX3	T-Box Transcription Factor 3
TE	Trophectoderm
TFCP2L1	Transcription Factor CP 2 like 1
THRAP3	Thyroid Hormone Receptor Associated Protein 3
TINC	TALE-mediated isolation of nuclear chromatin
TP53	Tumor Protein P53
TRIM28	Tripartite Motif Containing 28
ZEB2	Zinc Finger E-Box Binding Homeobox 2
ZFP143	Zinc finger protein 143
ZFP42 (REX1)	Zinc Finger Protein 42 Homolog

1. INTRODUCTION

1.1 Background

From the time Hippocrates laid foundations of medicine, medicine has advanced in various technological aspects and found safer ways of treatment for a variety of diseases. Recent medical innovation is progressing toward utilizing the potential of cellular potencies in not only the repair but also the regeneration of damaged, aged, or abused cells/tissues in the body. These regenerative approaches are of prime importance in biomedical research and thus grab the attention of developmental biologists and biomedical engineers across the globe. Regenerative medicine may be applicable to various diseases related to lifestyle, and many U.S. Food and Drug Administration (FDA) approved therapies are currently available in the market (Resource Table 1) (Mao and Mooney, 2015).

1.1.1 Cellular potencies

Each organism is a complex system coordinated by different types of cells, and all of these cells are derived from a single-celled zygote. Thus, during embryonic development, zygotes pass through various potencies to generate cells of diverse functions. Briefly, these potencies include totipotency, pluripotency, multipotency, and unipotency.

a) Totipotency:

Totipotency is the ability of the cell to develop into all the cells of the embryo properly, and in the case of mammals, the extraembryonic placenta. Single and 2-celled embryos can be best exemplified as totipotent stem cells. These cells generate a population of extraembryonic tissue (trophectoderm (TE)), and inner cell mass (ICM) from which pluripotent stem cells (PSCs) are derived. TE surrounds ICM and supports its development.

b) Pluripotency:

As the name, this potency indicates the ability of the cell to develop into various cells of a given organism. During development, when the zygote reaches the blastocyte stage a pluripotent population of cells called ICM were formed. These ICM cells can be the best example of PSCs. These cells can be able to generate all cell types of the body except for extra-embryonic lineages (like TE). ICM cells when cultured *in vitro* were termed embryonic stem cells (ESCs) either from mice are mouse embryonic stem cells (mESCs) or from humans are human embryonic stem cells (hESCs). Other types of PSCs

include embryonic carcinoma cells (ECCs), induced pluripotent stem cells (iPSCs) etc. Thus PSCs attracted the attention of biomedical researchers recently.

c) Multipotency:

These cells often reside as progenitor cells in tissues or organs and generate a limited number of cell types in a particular cellular niche/lineage. Mesenchymal stem cells (MSCs) and hematopoietic stem cells (HSCs) are one of the best examples of multipotent stem cells. Early therapies were developed using these cells to prove the potential of cellular therapies in medical applications. However, these therapies often require cells in large quantities and suffer from immune rejection.

d) Unipotency:

These cells also termed as precursor cells, possess very limited differentiation capabilities, often generating only one cell type. Because of their ability to self-renew these cells were regarded as stem cells. Epidermal stem cells and germ-line stem cells can be best represented as unipotent stem cells.

1.1.2 Challenges with stem cell therapies

While stem cell therapies appear promising for regenerative medicine, there are still challenges to their safe applications in clinical practice. For example, adult stem cells are confined to a specific niche in their resident tissues, and obtaining them in large quantities from patients is technically difficult. Maintenance and expansion of adult stem cells under *in vitro* conditions without losing their potency is technically challenging. Invasive methods to obtain adult stem cells may also hinder their use for therapeutic purposes. In addition, isolation of hESCs poses ethical issues because their derivation involves the inevitable destruction of human embryos. Indeed, use of hESCs raised a wide outrage from humanitarian and religious organizations. In this respect, the discovery of iPSCs obviated many ethical issues that accompany the use of human PSCs. Since then, iPSCs became the major source of PSCs for application in regenerative medicine.

Another challenge faced by stem cell and biomedical scientists is to obtain a sufficient number of iPSCs and differentiated cells, which should be used to repair the damaged tissues. Management of large-scale cultures of these cells and training of technical personnel are no small issues for successful cell-based therapies. In recent years,

many biotech companies have started automating the large-scale culture of cells to meet the growing demands of these cells. Currently, cell-based therapies are only practical for those diseases in which a small number of cells suffices for effective treatment, such as retinal disorder, Parkinson's disease, and repair of damaged heart muscles, and potentially Type I diabetes.

When stem cells are used for clinical practice, there is a serious concern regarding the safety of the cells derived from them, especially with regard to tumorigenesis. Initial methodologies of iPSC generation utilized vectors that become integrated into the genome. These vectors are largely replaced by non-integrating vectors such as episomal plasmids (Okita et al., 2008). The use of Sendai virus (SeV) is also promising because of its absence of genome integration and higher efficiency of reprogramming, together with the automatic elimination of the vector in fully pluripotent iPSCs (Fusaki et al., 2009; Nakanishi and Otsu, 2012; Nishimura et al., 2011). In any case, extensive and laborious characterizations of patient derived iPSCs are an essential part of establishing iPSCs that can be applied in clinical practice.

Typically, somatic cells generated from PSCs contain partially or incompletely differentiated PSCs that still possess proliferative capacity. This raises the possibility of tumor formation within a small population of cells transplanted to the patient. In particular, iPSCs are prone to this type of incomplete differentiation if they have not achieved full pluripotency when generated from somatic cells. Because of their reduced pluripotency, these partially reprogrammed iPSCs could compromise the outcome of iPSC-based therapy by tumor formation. Thus, understanding its mechanisms and increasing the efficiency of reprogramming is of utmost importance. In practice, purification of cells differentiated from PSCs is an important step for the application of cell-based therapy.

1.2 Induced pluripotent stem cells

The introduction of four reprogramming factors OCT4, KLF4, SOX2, c-MYC (OKSM) in somatic cells lead to the generation of mice (Takahashi and Yamanaka, 2006) and human (Takahashi et al., 2007) iPSCs. These iPSCs resemble ESCs and share similar properties at the molecular transcriptome and epigenetic levels. Soon after their discovery, iPSCs replaced ESCs in various applications of regenerative medicine. iPSC generation or acquisition of pluripotency involves complex processes of chromatin opening (or decondensation), epigenetic reprogramming, and pluripotency network establishment. As

compared with reprogramming using cell fusion or nuclear transfer, transduction of transcription factors using retroviral or other vector-based methods is far more efficient and robust for applications in regenerative medicine.

1.2.1 Mechanism of reprogramming

iPSC generation starts with the initial binding of reprogramming factors on various distal sequence motifs that may be located within condensed chromatin, followed by decondensation of the bound chromatin regions. The decondensation is mediated by OCT4, KLF4, and SOX2, either alone or in combination, but not by c-MYC. Because of their ability to alter the chromatin configuration to allow other factors to bind the genomic DNA, these OCT4, KLF4, and SOX2 are termed pioneer factors. Once decondensation begins, these proteins allow chromatin reorganization by recruitment of transcriptional regulators, followed by multiple protein-protein interactions. Subsequently, the enhancers and promoters of target genes form a loop structure, which ultimately leads to the activation of transcription (Basu and Tiwari, 2021).

On the other hand, c-MYC occupies open chromatin regions on genes and usually modulates RNA-Polymerase II pause release and regulates the complete transcription of these sites (Young, 2011). Thus, the initial phases comprise chromatin remodeling and involve the loss of the repressive Histone 3 lysine 27 trimethylation (H3K27me3) mark. In the subsequent phases, reprogramming factors along with the polycomb repressive complex 2 (PRC2) repress the expression of lineage-specific genes. Especially many mesenchymal-related transcription factors like ZEB2 and SNAIL1 are suppressed and epithelial markers like CDH1 and EPCAM are activated resulting in mesenchymal to epithelial transition (MET) (Basu and Tiwari, 2021; Krause et al., 2016). In the late phase, repressive marks on pluripotency-related genes are erased, causing activation of the endogenous pluripotency network. Those iPSCs that successfully initiated the pluripotency network are of high quality and can give rise to chimeras when introduced into the blastocyst of developing embryos (Krause et al., 2016).

1.2.2 iPSC therapies currently available for regenerative medicine

Successful application of patient-specific iPSCs in regenerative medicine in Japan include:

- A Japanese woman received a reprogrammed cornea developed from the iPSCs generated using the skin cells of the patient. This cornea significantly improved her eyesight (Cyranoski, 2019).

- In another case, neuronal precursor cells were generated from iPSCs that are reprogrammed from skin cells of Parkinson's disease patient. When surgically implanted, these neuronal precursor cells are expected to mature and generate dopamine-producing neurons. The patient has shown no adverse reactions so far after the surgery (Cyranoski, 2018a)

- Currently, research is going on to expand the application of iPSC therapies to treat cardiac diseases (Cyranoski, 2018b), spinal injuries (Nagoshi et al., 2019), and muscle diseases like Duchene muscular dystrophy (DMD) (Kupatt et al., 2021).

1.3 Understanding Pluripotency Network

Better quality and efficient iPSCs generation could be possible to revoke the understanding of pluripotency maintenance in detail. Since ESCs are the conventional source of pluripotency, revisiting the knowledge on ESC pluripotency and finding the knowledge gap is essential for improving our understanding of development, differentiation, and somatic cell reprogramming.

1.3.1 Characteristics of pluripotent stem cells

PSCs invariably exhibit two distinct properties; namely, self-renewal and pluripotency. Self-renewal is the ability of stem cells to generate daughter cells with the same potencies through cell division. Thus, self-renewal helps the propagation of stem cells until they receive definitive cues for differentiation. This property is exploited during *in vitro* culture and maintenance of PSCs. Pluripotency is the ability of stem cells to differentiate into all types of cells that comprise an organism. When transplanted or grafted, PSCs generate all three germ layers; for example, mESCs & iPSCs give rise to teratomas when transplanted ectopically or contribute to chimeras upon injection into blastocysts. PSCs are characterized by their expression of surface markers; for example, stage-specific embryonic antigen-1 (SSEA1) and alkaline phosphatase in the case of mESCs, and these marker proteins enable easy identification of PSCs by simple staining methods. On the other hand, the differentiation of PSCs was well studied by making embryoid bodies (EBs). EBs recapitulate the early events of embryo development *in vitro.*

1.3.2 Signaling in pluripotent stem cells:

The early knowledge of maintaining pluripotency *in vitro* stems from ECCs or ESCs cultured with feeder cells, in which leukemia inhibitory factor (LIF) was found to be crucial for ESC self-renewal. LIF is a member of the interleukin-6 cytokine family and activates the Janus kinase (JAK)-signal transducer and activator of transcription 3 (STAT3) pathway to maintain self-renewal of ESCs (Niwa et al., 1998). LIF is a glycosylated polypeptide of about Mr 20,000, and its primary structure is conserved across mammals, including several N-linked glycosylation sites and cysteine residues, whose disulfide bonds are the major determinants of the biological activity of LIF across species. LIF is secreted by various tissues, and it exerts its effect by binding to its cell surface receptor (Metcalf, 1991). LIF receptor (LIFR) is a heteromeric complex with LIFR-β and glycoprotein 130 (gp130) receptor chain(s). LIFR-β has high specificity but low affinity to LIF. Receptor activation occurs only when the gp130 receptor associates with LIFR-β upon LIF binding. Once heterodimerized, activation of LIF signaling leads to the phosphorylation of JAK family of tyrosine kinases. These JAKs in turn phosphorylate tyrosine residues on the cytoplasmic domain of gp130. Then, these phosphotyrosine residues recruit Src homology 2 (SH2) domain-containing proteins to JAKs. This interaction activates three possible signaling cascades: STAT transcription factors, Ras mitogen-activated protein kinase (MAPK), and phosphatidylinositol-4,5-bisphosphate 3 kinase (PI3K)- protein Kinase B (AKT) pathways (Chen et al., 2014) (Illustration 1).

Upon activation, STAT proteins dimerize and translocate to the nucleus to exert their action as a transcription factor. In ESCs, STAT3 is the main protein downstream of LIF signaling. Dimerized STAT3 translocates to the nucleus and induces the expression of several genes related to pluripotency, including *Klf4, Gbx2,* and *c-Myc.* Some recent studies suggest that the primary target of STAT3 could be *Tfcp2l1* because overexpression of *Tfcp2l1* conferred self-renewal of ESCs in the absence of LIF (Hackett and Surani, 2014). ESCs cultured in LIF/serum medium form domed colonies are competent in forming colonies (clonogenicity), and express *Oct4*, *Sox2*, and *Nanog* (heterogeneously) (Hackett and Surani, 2014). Parallelly, LIF mediated activation of MAPK/extracellular signal-regulated kinase (ERK) pathway by fibroblast growth factor 4 (FGF4), which has differentiation-inducing abilities (Kunath et al., 2007; Stavridis et al., 2007) can be blocked by using a 2 inhibitor (2i) medium (PD0325901, an inhibitor of the MAPK/ERK kinase (MEK)/ERK pathway, and CHIR99021 an inhibitor of glycogen synthase kinase 3 (Gsk3),

conferring a ground state of pluripotency to mouse ESCs (Ying et al., 2008). It is of note that ESCs in the LIF/serum condition fluctuate between alternate states and are prepared to differentiate as soon as they receive appropriate cues.

1.3.3 Transcription machinery in Pluripotent stem cells:

Downstream to the signaling molecules, transcriptional networks control pluripotency. A plethora of transcription factors was identified as regulators of pluripotency. For example, the computational model predicted that a minimal set of 12 factors (OCT4, SOX2, NANOG, ESRRB, SALL4, KLF2, KLF4, GBX2, TFCP2L1, and STAT3) govern the program of pluripotency and self-renewal in ESCs (Dunn et al., 2014). These factors form an intricate and autoregulatory network that links signaling molecules with transcriptional regulation in ESCs (Dunn et al., 2014). Some of these factors are irreplaceable (OCT4 & SOX2) (Hackett and Surani, 2014) while others (NANOG; ESRRB; SALL4; KLF4; Zfp42 (REX1)) are not essential (Dunn et al., 2014; Ng and Surani, 2011). OCT4, SOX2, and NANOG form an intricate core pluripotency network. This core pluripotency network autoregulates its own promoters and occupies promoters of various pluripotency-related genes.

OCT4 is a Pit, Oct, and Unc (POU) family of transcription factors that contain a low-affinity POU-specific domain (POU$_S$) and a high-affinity POU homeodomain (POU$_{HD}$) (Klemm and Pabo, 1996). It is expressed in primary germ cells, unfertilized oocytes, and ESCs, offering clues to its transcriptional activity in early embryogenesis (Scholer et al., 1989; Schöler et al., 1989), and loss of OCT4 was shown to impair the development of ICM during embryo development (Nichols et al., 1998). Like other OCT family proteins, OCT4 can activate expression of its target genes by binding to the octameric motif containing the AGTCAAAT consensus sequence (Palmieri et al., 1994; Schöler et al., 1990). *Oct4* expression is maintained along the early cleavage events and then becomes restricted in ICM when it disappears in the TE lineage. This may be in part due to the association of OCT4 and a caudal-type homeobox transcription factor 2 (CDX2), a TE lineage-inducing transcription factor (Niwa et al., 2000), although a recent single-cell RNA-sequencing analysis of the blastocyte indicates that a different mechanism may operate in the developing embryo (Stirparo et al., 2021). OCT4 protein modulates various other genes of developmental importance by binding to their regulatory regions; for

example, it induces the expression of FGF4 (Yuan et al., 1995), represses fork head box A1 (FOXA1) and A2 (FOXA2) proteins (Pan et al., 2002), and regulates primitive endoderm formation (Palmieri et al., 1994).

SOX2 is another member of the core pluripotency network. It belongs to the high mobility group (HMG) DNA-binding family of proteins. Since their HMG domain is the conserved domain also observed in the testis-determining factor Sry, they are termed Sox (Sry-related HMG box factor) (Bowles et al., 2000). SOX proteins interact with DNA through its HMG domain. *Sox2* is expressed uniformly along with *Oct4* in ESCs, and SOX2 activates the expression of OCT4 (Nichols et al., 1998), one of the early known pieces of evidence of the auto-regulatory loop of the core pluripotency network. OCT4-SOX2 heterodimer binds to the regulatory sequences to promote transcriptional activity (Dailey et al., 1994; Yuan et al., 1995). Structural studies on related proteins showed that OCT4-SOX2 heterodimerization is mediated by the direct interaction between POU_S and HMG domains as observed in the complex of OCT1 and SOX2 (Williams et al., 2004).

NANOG is another master regulator of the core pluripotency network. This protein was first described independently in 2003 by two groups (Chambers et al., 2003; Mitsui et al., 2003). It is dispensable to pluripotency/self-renewal maintenance but an important factor in the development stages for the ICM to acquire a pluripotent state. The name, Nanog, is derived from the Irish and Scottish mythology describing the land of eternal health and young Tir na nOg (Land of Youth in Irish) because of its ability to contribute to the self-renewal ability of ESCs (Chambers et al., 2003). Over-expression of *Nanog* itself is sufficient to bypass LIF for the self-renewal of ESCs (Chambers et al., 2003).

In addition to these well-known factors for pluripotency, a subset of ancillary factors are also expressed in ESCs to support and coordinate the core pluripotency network. These include SALL4, ESRRB, KLF2, KFL4, TBX3, and TFCP2L1. Most of these factors act directly by associating with the core pluripotency network (van den Berg et al., 2010; Wang et al., 2006). PRDM14 represses FGF4 signaling and DNA methylation in ESCs (Hackett and Surani, 2014). PI3K-Akt pathway activates *Tbx3* and *Nanog* (Niwa et al., 2009). Thus, the pluripotency is maintained and regulated by a network comprised of OCT4, SOX2, and NANOG (core pluripotency network) supported by external signals and additional factors acting as co-factors.

1.4 HYPOTHESIS

One of the major hurdles for generating iPSCs is the activation of the pluripotency program, without which the quality of generated iPSCs remain poor. Hence, finding ways to activate this pluripotency program should improve the efficiency of iPSC generation as well as the quality of iPSCs. Reprogramming involves the introduction of OCT4, KLF4, SOX2, and c-MYC in somatic cells. When the self-sustained pluripotency network is established, endogenous *Nanog* is usually expressed in the derived iPSCs. However, studies also showed that a mere overexpression of *Nanog* as one of the exogenous factors is not sufficient to improve the efficiency of reprogramming. Hence, there is a possible link between the events leading to activation of the endogenous *Nanog* gene and the successful establishment of high-quality iPSCs. Thus, I hypothesize that the understanding the regulation of *Nanog* gene expression would provide clues to its induction during reprogramming.

1.5 STRATEGY

Recent advances in molecular biology introduced various approaches to understand a regulatory proteome of specific gene locus. These approaches could offer comprehensive and unbiased identification of regulatory mechanisms of *Nanog* gene expression.

1.5.1 Locus-specific approach

Proteins associated with a specific DNA sequence have been studied recently using yeast one-hybrid and nucleic acid affinity isolation with a bait DNA sequence. However, these methods did not fully capture the protein composition at specific loci. Hence, proteomics of isolated chromatin segments (PICh) was developed using an antibody against a protein to isolate associated DNA (Déjardin and Kingston, 2009). Later, another method called insertional chromatin immunoprecipitation (iChIP) has been developed (Hoshino and Fujii, 2009). It involves inserting a recognition sequence of a DNA binding protein LexA near the target locus. Then, the targeted locus was isolated using an antibody specific to the binding protein. Proteins occupying the purified locus were identified in a mass spectrometric analysis. However, this method is time-consuming because of the insertion of recognition sequences into the genome (Hoshino and Fujii, 2009). This led to the development of engineered DNA binding molecule-mediated chromatin

immunoprecipitation (enChIP). This method used transcription activator-like (TAL) proteins that recognize telomeres to isolate proteins bound to telomere repeat regions. TALs are tagged with an epitope sequence and immunoprecipitated using antibodies against the specific tag. However, TAL proteins needed to be engineered for each target locus, this step is laborious and time-consuming (Fujita and Fujii, 2013). Another version of enChIP exploits the principle of clustered regularly interspaced short palindromic repeats (CRISPR)-deactivated Cas9(dCas9). A dCas9 tagged with an epitope is bound to the target locus with the help of guide RNA (gRNA) specific to the target site. Then, dCas9-bound locus was immunoprecipitated using the antibody against the tag and the proteins were analysed by mass spectrometry (Illustration 2). Since the manipulation of gRNA was found convenient, enChIP involving dCas9 is suitable for my study (Fujita and Fujii, 2013). Because of its dependency on antibody-mediated precipitation of native chromatin, specificity of enChIP in a genomic scale is not sufficient . Hence, a more efficient method of enChIP involving *in vivo* biotinylation of dCas9 was developed, termed CRISPR affinity purification insitu of regulatory elements (CAPTURE). This method combines the specificity of streptavidin-biotin binding along with the convenience of CRISPR-gRNA design to isolate target chromatin regions (Liu et al., 2017) (Illustration 2).

1.5.2 Known regulators of *Nanog* promoter:

After the discovery of *Nanog*, sequence analysis revealed a composite OCT4-SOX2 binding motif in the proximal promoter region of *Nanog*. Electrophoretic mobility shift assays (EMSA) and chromatin immunoprecipitation (ChIP) assays further confirmed the binding of OCT4 and SOX2 to this site (Kuroda et al., 2005; Rodda et al., 2005). Co-regulators like ZFP143 were also identified to regulate *Nanog* expression by modulating OCT4 binding to the *Nanog* promoter region. siRNA knockdown of ZFP143 decreases *Nanog* promoter activity and self-renewal of mES cells, indicating the dependency of OCT4 on ZFP143 (Chen et al., 2008). On the other hand, a set of specificity protein transcription factors (SP1 and SP3) was found to directly bind to a different site on the *Nanog* promoter and regulate *Nanog* expression (Wu and Yao, 2006). A more recent study showed that protein BRD4 regulate *Nanog* in mESCs (Liu et al., 2014). However, some proteins including TP53 were found to suppress *Nanog* expression in the event of differentiation and DNA damage (Lin et al., 2005). This search for finding regulators of *Nanog* revealed that SALL4 binds to the enhancer region of *Nanog*. Interestingly, SALL4

was found to be able to activate the transcription of *Nanog* in somatic cells (Wu et al., 2006) unlike OCT4 and SOX2 (Kuroda et al., 2005; Rodda et al., 2005). Known *Nanog* promoter regulatory proteins were illustrated (Illustration 3)

One notable phenomenon of *Nanog* transcription is its stochastic promoter induction, monoallelic activation, and expression noise (Ochiai et al., 2014). NANOG is one such regulatory protein, which auto-represses its transcription when its level reaches a defined level (Navarro et al., 2012). Further, *Nanog* showed monoallelic expression in mESC cultures with serum and a biallelic expression in the 2i medium (Ficz et al., 2013; Miyanari and Torres-Padilla, 2012), although heterogeneous expression of *Nanog* is found in 2i cultures as well. This noise is probably caused by both intrinsic and extrinsic factors (Ochiai et al., 2014). It is believed that heterogeneous expression of *Nanog* would allow mESCs to switch from self-renewal to a differentiation program (Chambers et al., 2003, 2007; Navarro et al., 2012; Nichols et al., 2009) Thus, a comprehensive understanding of transcription at the *Nanog* promoter is necessary to shed light on how ESCs maintain self-renewal and pluripotency as well as conversion between the two states.

2. RESEARCH PURPOSE

I aimed to understand the mechanisms and identify regulatory proteins that modulate pluripotency. *Nanog* gene expression is correlated with the acquisition of pluripotency, and NANOG is considered a hallmark/gateway of pluripotency (Chambers et al., 2007; Silva et al., 2009). Hence, I baited the *Nanog* gene promoter region with CAPTURE in mouse ESCs and purified the *Nanog* gene interacting proteome by affinity purification. Further, these proteins were identified by mass spectrometry. Then I evaluated the role of proteins that function related to gene expression in self-renewal and pluripotency of mESC and somatic cell reprogramming. The impact of this research is two-fold.

1. Identifying regulatory proteins of *Nanog* gene transcription might uncover new mechanisms controlling pluripotency, especially outside the core pluripotency network.

2. Understanding regulatory mechanisms of *Nanog* transcription and/or pluripotency network would provide clues to improving reprogramming strategies and bring iPSCs much closer to the application in regenerative medicine.

3. MATERIALS AND METHODS

3.1 Cell culture

EB5 mouse embryonic stem cells (mESC) were obtained from RIKEN BRC (Tsukuba; Japan) and were cultured in Dulbecco's Modified Eagle Medium (DMEM) high glucose (Nacalai Tesque) supplemented with 15% Fetal Bovine Serum (FBS) (Gibco), 100 mM Non-Essential Amino Acids, 50 mM β-Mercaptoethanol, 1000 U/mL LIF, and 1% Penicillin/Streptomycin solution. For making EBs, mouse ES cell medium without LIF was used. NIH3T3 cells, HEK293T cells, Plat-E cells, and mouse embryonic fibroblast (MEF) cells were cultured and maintained in DMEM medium (DMEM High Glucose supplemented with 10% FBS and 100 U/mL Penicillin/Streptomycin solution). For differentiation, EB5/ReKO (EB5 Rex1-hKO knock in) were cultured in a DMEM medium with 5 µM of retinoic acid (Sigma). All cultures were maintained at 37ºC, with 5% CO_2.

3.2 Cloning and plasmid construction

Vectors expressing dCas9 (pEF1a-FB-dCas9-puro) and *E.coli* Biotin ligase (BirA) (pEF1a-BirA-V5-neo) were purchased from Addgene. Neomycin-resistant gene and Keima Red (KR) were cloned from pcDNA3 (Invitrogen) and complementary DNA (cDNA) of Sendai virus defective persistent (SeVdp) vector SeVdp (KR/Bsr/EGFP/KO) encoding KR gene, Blastidicn resistance (Bsr) gene, enhanced green fluorescent protein gene (EGFP), and a Kusabira-Orange (KO) gene (Nishimura et al., 2011), respectively. Further, the open reading frames (ORFs) of these genes were amplified by polymerase chain reaction (PCR), and the amplified products were gel purified by a QIAEX II Gel Extraction Kit (Qiagen). Then the purified fragments were cloned into SeVdp vector cDNA using In-Fusion HD Cloning Kit (TAKARA) to obtain SeVdp(dCNBR). On the other hand, gRNA (including scaffold) or short hairpin RNA (shRNA) were cloned into pMCsΔYY1-U6-Puro (Aizawa et al., 2022) containing U6 promoter with puromycin resistant gene expressed under phosphoglycerate kinase (PGK) promoter. Briefly, annealed gRNA and shRNA fragments were inserted into the BbsI and EcoRI-BamHI sites, respectively. The oligonucleotide sequence of gRNA (Table 1) and shRNA (Table 2) used in this study were tabulated.

For a lentiviral vector expressing far upstream binding protein 1 (*Fubp1)*, the mouse *Fubp1* gene fused with a FLAG-tag was cloned by PCR into pCW57.1 (Addgene). Primers used for the PCR (Table 3) were tabulated.

For checking Cas9 recruitment efficiency of gRNAs, I inserted gRNA into pX330-U6-Chimeric-BB-CBh-hspCas9 (Addgene) plasmid at BbsI site. An EGFP donor vector was generated by amplifying 410 bp of *Nanog* promoter by PCR using the primers (Table 4) and infused into pCAG-EGxxFP (Addgene) at SalI restriction enzyme site.

3.3 Production of Sendai viral vector

SeVdp vectors were produced as described previously (Nishimura et al., 2011). Briefly, 2 µg of SeVdp cDNA and other plasmids (NP, P, M, F, HN, and L) encoding for Sendai virus (SeV) proteins were transfected into BHK/T7/151M(SE) cells in a 6-well plate using Lipofectamine LTX Plus reagent (Invitrogen). After one day of incubation, cells were trypsinized and seeded onto a 100 mm plate and cultured at 32ºC for 5 days. On the 6th day, cells were dissociated using cell dissociation buffer (Gibco) and plated onto 6-well plates. These were further incubated at 37ºC, 5% CO_2 for 24 hours. Then these cells were transfected again with SeV structural proteins (M, F, and HN) to enhance the packaging and release of viral particles into the culture medium. After 24 hours, culture plates were transferred to 32ºC for 5 days. Finally, the culture supernatant containing viral particles was collected and filtered through a 0.45µm cellulose acetate filter, snap-frozen in liquid N_2, and stored at -80ºC until use.

3.4 Production of Retroviral vector

Constructed retroviral vector plasmids were transfected into Plat-E cells by Lipofectamine 2000 (Invitrogen). Briefly, 2 µg of the plasmid was mixed with 10 µL of Lipofectamine 2000 and 500 µL of Opti-MEM medium (Invitrogen) and incubated for 5 minutes at room temperature. After incubation, the transfection mixture was layered on 2 x 10^5 Plat-E cells in each well of the 6-well plate. Further, these cells were cultured overnight with DMEM High Glucose supplemented with 10% FBS. The medium was changed to DMEM High Glucose supplemented with 10% FBS and 1% Penicillin/Streptomycin solution on the other day and the cells were left undisturbed for the next two days. On day 3, the medium containing the viral vectors was collected by passing supernatant through a 0.45µm cellulose acetate filter. Thus, the collected retroviral particles were stored at -80ºC until further use.

3.5   Transduction of viral vectors

Target cells were transduced with the respective retroviral vector based on the purpose of each experiment. In short, a 1:1 mixture of retroviral vector and the respective medium was mixed and supplemented with a final concentration of 8 µg/mL Polybrene. This mixture was added to respective cells and was incubated at 37ºC, 5% CO_2 for 14-16 hours, and the medium was replaced with a fresh culture medium. After 24 hours of adding fresh culture medium, cells were trypsinised and transferred to a medium containing respective drugs. Infected cells were obtained after 2-3 days or 7 days of selection with Puromycin (2 µg/mL) or Neomycin (600-1200 µg/mL), respectively. In the case of mES cells, the transduction protocol was modified to increase the transduction efficiency. Cells were centrifuged at 1500 rpm for 40 minutes at room temperature after the addition of the transduction mixture, then transferred to the incubator (Bui et al., 2019).

A 1:1 mixture of SeV viral particle with respective culture medium was left to sit on the cells for 16 hours at 32ºC for virus infection to the target cells. Infected cells were selected by Neomycin (800 µg/mL for mESCs and 1000 µg/mL for NIH3T3 cells) drug.

3.6   CAPTURE of *Nanog* promoter

mESCs or NIH3T3 cells expressing dCas9, BirA, and gRNA were expanded under described culture conditions. About 1 x 10^8 cells were collected with formaldehyde fixation. Briefly, to 7 mL of culture medium, 1% Formaldehyde was added and incubated at room temperature for 10 minutes. Then the formaldehyde was quenched with 0.125 M Glycine for 5 minutes while shaking at room temperature. Now, the cells were washed twice with ice-cold PBS and collected using a cell scraper in a 15 mL centrifuge tube by centrifugation at 2500 rpm for 5 minutes. These cell pellets can be stored at -80ºC until further use. To isolate chromatin, cells were resuspended in 10 mL of cell lysis buffer (25 mM tris hydrochloride (Tris-HCl), 85 mM KCl, 0.1% Triton X-100, pH 7.4, freshly added 1 mM dithiotreitol (DTT), and 1:200 proteinase inhibitor cocktail (Sigma)) and rotated at 4ºC for 15 minutes on a rotary mixer. To isolate nuclei, cells were centrifuged at 2,300 x g for 5 minutes at 4ºC, and the supernatant was carefully removed without disturbing the pellet. The pellet was suspended in 5 mL nuclear lysis buffer (50 mM Tris-HCl, 10 mM ethylenediaminetetraacetic acid (EDTA), 4% sodium dodecyl sulphate (SDS), pH 7.4, freshly added 1 mM DTT, and 1:200 proteinase inhibitor cocktail) and incubated for 10 minutes at room temperature. Nuclei suspension was then mixed with 15 mL of 8 M urea

buffer and centrifuged at 16,100 x g for 25 minutes at room temperature. Nuclei pellets were then resuspended in 5 mL nuclear lysis buffer and mixed with 15 mL of 8M urea buffer and centrifuged at 16,100 x g for 25 minutes at room temperature. The samples were washed twice with 5 mL nuclear lysis buffer and mixed with 15 mL of 8 M urea buffer, followed by centrifugation at 16,100 x g for 25 minutes at room temperature. Pelleted chromatin was then washed twice with 5 mL cell lysis buffer. Chromatin pellet was resuspended in 5 mL of IP binding buffer without NaCl (20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.1% Nonidet P (NP)-40, 10% glycerol, freshly added proteinase inhibitor) and aliquoted into Eppendorf tubes. The chromatin suspension was then subjected to sonication to an average size of 300 - 500 bp on the Branson Sonifier 450 ultrasonic processor (10% amplitude, 0.5 seconds on 1 second off for 1 minute). The fragmented chromatin was centrifuged at 16,100 x g for 25 minutes at 4°C. The supernatant was combined and 150 mM NaCl was added to the sheared chromatin. To prepare the streptavidin beads for affinity purification, 25 μL of streptavidin magnetic beads (Life Technologies) were washed 3 times with 1 mL of IP binding buffer and added to the soluble chromatin. After overnight incubation at 4°C, the streptavidin beads were collected using a magnetic stand. The beads were then washed 5 times with 1 mL of IP binding buffer with 150~300 mM NaCl. The collected chromatin was analyzed in CAPTURE-quantitative PCR (qPCR) or proteomic approach.

3.7   CAPTURE-qPCR

The *Nanog* promoter region was captured as described above. After, overnight incubation and washing, streptavidin beads were eluted using 80 μL of SDS elution buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0) and incubated at 85°C for 10 minutes. The eluted fraction was separated from beads using a magnetic stand. The eluted samples were incubated at 65°C overnight to reverse cross-linking. DNA fragments were purified Phenol/Chloroform/Isoamyl alcohol treatment followed by ethanol precipitation. Primers targeting *Nanog* promoter (Table 5) were used for qPCR analysis, and enrichment was quantified relative to non-specific gRNA (gRNA4)-expressing samples.

3.8   CAPTURE-proteomics

For proteomics analysis, the beads were washed twice in IP binding buffer without NP-40 and treated with trypsin (Promega) overnight at 37°C, followed by purification using SPE C-TIP-T300 (Wako). Peptide analysis by liquid chromatography tandem mass

spectrometry (LC/MS/MS) was performed as described previously (Bui et al., 2019; Chronis et al., 2017).

For western blotting, beads were eluted for proteins by directly adding 2X sample buffer (50 mM Tris-HCl (pH 6.8), 2% SDS, 10% Glycerol, 0.1 mg/mL bromophenol blue, 5% 2-mercaptoethanol) and incubating at 100ºC for 5 minutes. Thus, eluted proteins were loaded through an 8% SDS-PAGE gel in a 2X sample buffer at 70-120 volts until separated. Later, these gels were layered with PVDF, and proteins were transferred onto the membrane using a semi-dry blotting system. Further, the poly vinylidene difluoride (PVDF) membrane was blocked in 5% skim milk prepared in tris buffered saline with tween 20 (TBS-T) (25 mM Trizma base, 140 mM NaCl, 0.0005% KCl, 0.1% Tween 20) for a minimum of one hour at room temperature. Then the membrane was washed followed by incubation at 4ºC overnight with primary antibody (diluted in TBS-T to a final concentration of 1 µg/mL). These blots were washed twice followed by incubation with horse radish peroxidase (HRP) conjugated secondary antibody (diluted in TBS-T to a final concentration of 1 µg/mL) at room temperature for 2 hours. Finally, blots were developed using chemiluminescence and visualized under Fusion Fx7.Edge system (Vilber Lourmat).

3.9 Gene Ontology (GO) analysis

mRNA expression profiles of proteins identified by the proteomics were obtained from published data (Chronis et al., 2017), and mESC/MEF ratio of the profiles was plotted by a custom R script that is available on GitHub (https://github.com/akikuno/iPS-proteomics). Proteins expressing highly in mESC (mESC/MEF > 2) were selected for further analyses. GO Term Finder (Boyle et al., 2004) was used for GO analysis of the highly expressed proteins.

3.10 Fluorescence-Activated Cell sorting (FACS)

For FACS, cells were collected by trypsinization, and by gentle pipetting, a single-cell suspension was made. Cells were passed through a 1 µm nylon mesh filter to remove any clumps. The filtered suspension was supplemented with 2 µg/mL Propidium Iodide (PI) dye to distinguish live cell from dead cells. PI(-) cells were sorted as live cells by MoFlo XDP (Beckman).

3.11 RNA extraction & reverse transcription

RNA was extracted from RNeasy Mini Kit (QIAGEN) unless mentioned by following the manufacturer's guidelines. Once eluted, RNA was mixed with 1 µL of 10 mM deoxyribonucleotide triphosphate (dNTP) mix and 1 µL of 50 ng/µL random primers, followed by heating at 65°C for 5 minutes and on ice for 5 minutes. 5X First-Strand Buffer, 0.1 M DTT, RNaseOUT™ Recombinant RNase Inhibitor, and SuperScript™ III RT were added to the mixture, which was incubated at 25°C for 5 minutes and 50°C for 60 minutes. The reaction was inactivated by heating at 70°C for 15 minutes, and the synthesized cDNA was stored at -20°C until use.

3.12 qPCR analysis

PCR was performed in a 10 µL reaction containing 1 µL of 10-fold dilution of cDNA, 0.04 µL each of 50 µM forward and reverse primers, and 5 µL of GoTaq qPCR master mix (Promega), using 7500 Fast Real-Time PCR system (Applied Biosystem). A housekeeping gene, TATA-binding protein (TBP), was used as an internal control. Primers for qPCR are listed in Table 6

3.13 Microscopy & Image analysis

Whole well good images of EB5/ReKO cells with shRNA's were obtained by multi-point capture setting in BZ-X710 (Keyence) microscope and the raw images were stitched to get a final whole well image. Images were captured in brightfield and fluorescence filters separately. Fluorescent images were analyzed for intensities using the Image Analyzer program in BZ-X710 (Keyence). Briefly, for each image measurement area was set to cover the whole well. To remove background, threshold fluorescence was set as hue:255; brightness 40-125; tolerance:20. Then edge smoothening was applied, and the areas larger than 1500 µm^2 were extracted as colonies. Then the Image Analyser program automatically perform quantification of colony numbers and fluorescence intensities. This data was further analyzed in Microsoft-Excel worksheet.

Graphing of the analyzed data was prepared using Graphpad Prism v8.0. Immunofluorescent images were captured using AxioObserver microscope (Zeiss) and the data was processed using Axiovision software (Zeiss).

3.14 Immunofluorescence

To visualize the target protein in the cells, cells growing in monolayer were gently washed with phosphate-buffered saline (PBS) and incubated for 10 minutes at room temperature in the presence of 3.7% formaldehyde. The fixed cells were washed again with PBS and incubated for 10 minutes at room temperature in PBS containing 0.1% NH_4Cl. The cells were then permeabilized by incubation for 5 minutes at room temperature in the presence of 0.1% Triton X-100. The cells were then washed with PBS, overlaid with primary antibody diluted in 0.1% Saponin (in PBS) to the required dilution, and incubated for 60 minutes at room temperature. The cells were washed with PBS, and, after the addition of a secondary antibody, the cells were incubated for 30 minutes. Nuclei were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) using Fluoro-KEEPER Antifade Reagent (Nacalai).

3.15 Statistical analysis

Replicate samples were analyzed for statistical significance using a student *t-test*. Fluorescence intensity measurements were tested for significance groups using one-way analysis of variance (ANOVA) in GraphPad Prism *v*8.0. A value of $P < 0.05$ was regarded as statistically significant.

4. RESULTS

4.1 Generation of cells expressing CAPTURE components

To identify proteins that regulate *Nanog* gene transcription in mESCs, I relied on the previously described CAPTURE method (Liu et al., 2017). To that end, dCas9 and BirA genes were cloned into a SeV vector backbone along with genes for neomycin resistance and KR fluorescence. A FLAG tag and a biotin acceptor site were added to the N-terminus of the dCas9 protein (Figure 1A). The obtained SeVdp(dCNBR) Sendai virus was transduced into EB5 (mESC) and NIH3T3 cells using the protocol described before (Nishimura et al., 2014) (Figure 1A). After 3 days of transduction, infected cells were selected by neomycin treatment to obtain EB5/dCNBR and NIH3T3/dCNBR cells. Thus, obtained cells showed KR fluorescence as expected (Figure 1B). Then, the expression of dCas9 and BirA as well as *in vivo* biotinylation in EB5/dCNBR and NIH3T3/dCNBR cells were verified with affinity conjugation of with Avidin-488. This data confirmed the expression of dCas9, BirA, and proper biotinylation of dCas9 *in vivo* (Figure 1C). Then, biotin-labeled dCas9 was affinity purified using Streptavidin Dynabeads (Thermo Fisher Scientific) and the dCas9 enrichment was confirmed by western blotting (Figure 1D). Thus, I obtained cells that can be utilized for capturing the locus-specific proteome.

4.2 gRNAs efficiently recruit Cas9 to the target site

Because of its important roles in pluripotency maintenance and acquisition, the *Nanog* promoter was used as a bait for locus-specific proteome capture. To recruit dCas9 and purify the *Nanog* promoter region, three specific gRNAs (gRNA1, 2, and 3) were designed using the CRISPRdirect (https://crispr.dbcls.jp/) tool. Care was taken to avoid gRNAs interfering with the OCT4-SOX2 consensus binding region on the *Nanog* promoter (Figure 2A). The ability of these gRNAs to recognize the target and recruit Cas9 protein was screened as described previously (Mashiko et al., 2013). All gRNAs generated EGFP fluorescence by homologous recombination, (Figure 2B) confirming their ability to recruit Cas9 protein to their target site. Thus, I used all three gRNAs to purify the *Nanog* promoter along with a gRNA against the yeast *Gal4* gene (Liu et al., 2017) as negative control (gRNA4).

4.3 Enrichment efficiency of dCas9 on the *Nanog* promoter in mESCs and fibroblasts

For dCas9-based purification of the *Nanog* promoter, retrovirus expressing individual gRNAs under the U6 promoter (Figure 3A) were transduced to EB5/dCNBR and NIH3T3/dCNBR cells as described previously (Bui et al., 2019). After infection and selection of infected cells by puromycin, the morphology (Figure 3B) of EB5/dCNBR cells indicate that the dCas9-gRNA complex did not interfere with the mESC properties. Further, the target enrichment efficiency of the gRNAs was verified by affinity purifying dCas9 bound to chromatin using Streptavidin Dynabeads and quantifying *Nanog* promoter by CAPTURE-qPCR. The enrichment efficiency of gRNA2 was higher than the other two (Figure 3C). Interestingly, fibroblast cells (NIH3T3) showed enrichment of *Nanog* promoter with gRNA1 (Figure 3D) indicating the possible difference in dCas9 accessibility based on their respective chromatin status (Yarrington et al., 2018). Because gRNA2 did not exhibit transcriptional interference of *Nanog* gene (Figure 3E) while enriching *Nanog* promoter efficiency, I used gRNA2 for the CAPTURE-proteomics experiment. Thus, using the CAPTURE system, I successfully collected *Nanog* promoter DNA.

4.4   CAPTURE purification of *Nanog* promoter interacting proteome

Now to identify proteins that may regulate *Nanog* expression, I isolated soluble chromatin from approximately $1x10^8$ EB5/dCNBR cells expressing gRNA2 or gRNA4 and subjected them to affinity purification using Streptavidin Dynabeads (Figure 4A). Briefly, chromatin was treated with Streptavidin Dynabeads (Thermo Fisher Scientific) and processed as described in materials and methods (Section 3.6). Then, the beads were subjected to on-bead trypsin digestion to identify proteins using LC/MS/MS. Enriched proteins were filtered out among gRNA2 and gRNA4 to obtain a total of 325 proteins (Figure 4B) (Table 7) specific to the *Nanog* promoter. Of these 325 proteins, some were previously known to occupy the *Nanog* promoter including Tripartite motif-containing 28 (TRIM28) protein (Hu et al., 2009; Seki et al., 2010), Thyroid Hormone Receptor Associated Protein 3 (THRAP3), and BCL2 Associated Transcription Factor 1 (BCLAF1) (Knaupp et al., 2020) (Figure 4B). However, well-known *Nanog* promoter-occupying proteins including OCT4 and SOX2 were not enriched in my CAPTURE proteome. Thus, I isolated previously known and new proteins from the *Nanog* promoter region. To verify the relevance of the candidate proteins in pluripotency, I compared the expression levels of these proteins between MEFs and mESCs from the published RNA-Seq data (Chronis et al., 2017). This revealed that nearly 34% of the isolated proteins have a higher expression

(2-fold) in mESCs over MEFs (Figure 4C), suggesting that the isolated proteins may have a function in pluripotency. GO analysis of the highly expressed proteins, using GO Term Finder (Boyle et al., 2004) showed that enriched proteins have functions related to RNA metabolism, splicing, gene expression, and cellular homeostasis (Figure 5A). These results indicate that the isolated proteins have distinct roles in pluripotency and mESC homeostasis. To explore novel regulatory mechanisms in *Nanog* gene transcription and pluripotency, I choose proteins that have functions related to aspects of gene regulation (Figure 5A). Briefly, proteins from metabolic processes were excluded first to obtain 73 proteins that includes proteins from RNA splicing, RNA processing, ribo-nucleoproteins, and gene expression. Of 73 proteins, 22 proteins showed DNA binding or DNA binding transcription factor binding activity. Six proteins among the 22 have previously described roles in transcriptional regulation including BCLAF1 (Vohhodina et al., 2017), Far upstream binding protein 1 (FUBP1) (Hoang et al., 2019), Parkinsonism Associated Deglycase (PARK7) (Vasseur et al., 2009), PC4 And SRSF1 Interacting Protein 1 (PSIP1) (Singh et al., 2017; Sono et al., 2018), and THRAP3 (Sono et al., 2018). In addition, a DNA repair protein MutS Homolog 6 (MSH6) has been shown to cooperate with OCT4 in mice (Ding et al., 2011; Pardo et al., 2010) and human ES cells (Huang et al., 2021). Because of their functions in gene expression (Table 7), I wanted to understand its role in *Nanog* gene regulation. To re-confirm the occupancy of the candidate proteins on the *Nanog* promoter, a ChIP-qPCR was performed with antibodies against each chosen protein in mESCs. Observed *Nanog* promoter enrichment suggests the occupancy of these proteins except for FUBP1 (Figure 5B). The reason for no enrichment in FUBP1 could be the compatibility of the antibody used for ChIP. Hence, I expressed FUBP1 fused with 3xFLAG under the control of doxycycline inducible system and verified its expression in a western blot (Figure 5C). When ChIP was performed using an anti-FLAG antibody, FUBP1 showed enrichment of *Nanog* promoter was significantly higher than the background (Figure 5D). Thus, I have identified known and new proteins that bind to the *Nanog* promoter using the CAPTURE method. Now, I ought to understand the function of these proteins on the *Nanog* gene transcription and pluripotency.

4.5 Knockdown of candidate proteins affects the self-renewal ability of mESC

To monitor pluripotency in mESCs during analyses of *Nanog* expression and pluripotency regulation by candidate proteins, I used a pluripotency reporter cell line which had been produced in our lab. In this mESC cell line, *Zfp42* (*Rex1*) ORF is replaced with

humanized KO (hKO) fluorescent gene because *Zfp42* (*Rex1)* expression is tightly correlated to pluripotency. Thus, these cells exhibit hKO fluorescence in pluripotency maintenance conditions. These mES were named as EB5/ReKO

Next, to determine the regulatory role of candidate proteins on *Nanog* gene expression and the self-renewal ability of mESCs, I designed a pair of retroviruses that express shRNA against each candidate protein and transduced them into EB5/ReKO cells (Figure 5A). shRNA's against *Oct4* (shOct4) and Luciferase (shLuc) was used as positive and negative control, respectively. Two shRNA for each candidate protein and one *Oct4* that showed a significant knockdown of the candidate gene were used in the further analysis (Figure 6A). Then the effect of knockdown on the pluripotency properties of the EB5/ReKO cell line was tested. First, I checked if the downstream effect of knockdown was noticeable on day 5, and no difference in hKO fluorescence (Figure 6B) or expression of pluripotency genes, *Nanog* and *Oct4,* was observed on day 5 after shRNA transduction (Figure 6C & 6D). The self-renewal ability was quantified based on colony number and hKO fluorescence intensity as described in Materials and Methods (Figure 6E). Thus, to assess the function of candidate proteins on the self-renewal ability of mESCs, 1×10^3 shRNA expressing EB5/ReKO cells were sorted using an MD-Cell sorter onto a 0.1% Gelatin coated 96-well tissue culture plate on day 5 after knockdown (Figure 7A). Except for shRNA against *Bclaf1* and *Fubp1,* cells expressing shRNA against all candidate proteins formed a significantly smaller number of colonies as compared to the background (Figure 7B & 7C), indicating loss of the colony-forming ability of mESCs. However, the intensity of hKO fluorescence shows a significant increase in fluorescence levels in knockdown of *Thrap3*, *Psip1*, *Park7*, and *Oct4* (Figure 7D). This increase in fluorescence could be due to the compact shape of the survived mESC colonies. qPCR analyses of available colonies on day 7 showed a significant loss of pluripotency gene expression (*Nanog, Oct4, and Rex1*) in *Fubp1* knockdown cells compared to control (Figure 8A-C) along with the loss of hKO fluorescence (Figure 7D). Knockdown of *Psip1* showed loss of *Nanog* (Figure 8A) and *Oct4* (Figure 8B) gene expression, although not significant. Knockdown of *Thrap3* exclusively showed a reduction of *Nanog* (Figure 8A) but no other pluripotency gene expression. These results recapitulate the significance of the candidate proteins in regulating *Nanog* gene expression as well as pluripotency of mESC. Although the candidate proteins occupy *Nanog* promoter, knockdown of these proteins did not exert a significant effect on *Nanog* transcription except for *Fubp1*. On the other hand, most

knockdowns reduced colony numbers except for *Fubp1*. This indicates that FUBP1 has a functional role in pluripotency maintenance in mESCs.

4.6 Effect of the candidate proteins on mESC exit from pluripotency

Next, I asked if the candidate proteins have any role in mESC exit from pluripotency when LIF is removed. To this extent, I sorted 1×10^3 shRNA-expressing EB5/ReKO cells onto each well of gelatin-coated 96 well plates using an MD-Cell sorter and cultured with mES medium with LIF. After 2 days, the medium was replaced with mES medium without LIF and cultured for seven days (Figure 9A). EB5/ReKO cells reduced hKO fluorescence (Figure 9B), and pluripotency gene expression in control shRNA cells by 52%, 68%, and 55.5% in *Nanog* (Figure 9C)*, Oct4* (Figure 9D)*,* and *Rex1* (Figure 9E) genes indicating that mESCs exited from pluripotency. Interestingly, upon the knockdown of *Park7* or *Thrap3* mESCs did not exit from pluripotency as efficiently as the control cells (Figure 9C & 9D) suggesting that PARK7 and THRAP3 are necessary for switching off self-renewal program and initiating differentiation when LIF is removed. Thus, from the *Nanog* promoter capture, I have identified novel regulators of mESC self-renewal maintenance and differentiation initiation.

4.7 Effect of the candidate proteins on differentiation via EB formation

Knockdown of candidate proteins showed various effects on self-renewal and early phase of differentiation of mESCs. To further explore the role of the candidate proteins on embryonic development, I used shRNA expressing EB5/ReKO cells to test their effect on the formation of EBs, which recapitulates some aspects of early embryogenesis. Five days after transduction of each shRNA expressing vector and drug selection, the cells were trypsinized and hKO$^+$ cells were sorted into 96-well v-bottom plates to allow formation of EBs (Figure 10A). These EB's were cultured for the next 7 days with intermittent medium change. Growth of EB, as estimated by their lateral diameter, showed significantly slower growth in EBs in which *Thrap3*, *Psip1*, and *Msh6* were knocked down, like knockdown of *Oct4* (Figure 10B-D), which indicates their supportive roles in EB development. I then analyzed the expression of differentiation markers including *Fgf5* a marker for epiblast development; *Gata6,* a marker for primitive endoderm; and mesenchyme homeobox 1 *(Meox1),* a marker for mesodermal differentiation. No consistent or significant difference in the *Fgf5* gene was observed either on day 4 or day 7 EBs (Figure 11A). A significant

decrease in *Gata6* expression was observed for all gene knockdown on day 7 EB (Figure 11B), indicating that they may be important for primitive endodermal formation. Knockdown of *Thrap3*, *Psip1* and *Oct4* significantly decreased *Meox1* expression (Figure 11C), indicating their roles in the differentiation of the mesodermal lineage. The candidate proteins thus exhibited various levels of regulation on mESC self-renewal, differentiation initiation, and embryonic development implying that CAPTURE enables the purification of regulatory proteins in mESCs.

4.8 Effect of the candidate proteins on somatic cell reprogramming

Reprogramming of fibroblasts to iPSCs enables an understanding of various regulatory mechanisms of pluripotency. To study the significance of candidate proteins in somatic cell reprogramming, I used the Stage Specific Sendai virus based (3S) reprogramming system (Figure 12A) (Nishimura et al., 2014). In this system, a Sendai viral vector SeVdp(fKg-OSM) expresses four reprogramming factors and KLF4 was fused with the destabilization domain (DD) and FLAG tag, which can be stabilized by adding Shield1 (Nishimura et al., 2014). Immunostaining of FLAG enabled to monitor the infectivity of the SeVdp(fKg-OSM) viral vector (Figure 12B). A low amount of KLF4 in the absence of Shield1 generates a low pluripotent iPSCs, and 100 nM of Shield1 maintains the protein level of KLF4 same as wild type, enabling generating of iPSC cells with full pluripotency (Figure 12C).

To test the roles of the candidate proteins in reprogramming, I infected MEFs with retrovirus expressing shRNA specific to each candidate mRNA. shRNA-expressing MEFs were selected for 5 days (Figure 13A) and reprogrammed for 28 days with SeVdp(fK-OSM) vector by adding either sh0 or sh100 nM of Shield1 into the medium. At 28 days of reprogramming, iPS colonies were analyzed for the expression of pluripotency genes (Figure 13A). Knockdown of *Park7* induced expression of pluripotency genes *Nanog, Oct4* and *Rex1* in both Low-K and High-K iPSCs (Figure 13B). *Knockdown of Psip1* and *Thrap3* induced expression of *Nanog* (Figure 13B) both in Low-K and High-K cells. Whereas knockdown of *Msh6* significantly reduced the expression of *Nanog* even in High-K cells (Figure 13B). Interestingly, knockdown of *Fubp1*, significantly induced *Rex1* expression (Figure 13B). Overall, knockdown of candidate proteins in somatic cell reprogramming suggests that PARK7, PSIP1, and THRAP3 could be reprogramming roadblocks.

Hence, from my analysis on proteins identified from the *Nanog* promoter proved to exhibit different levels of regulation from self-renewal maintenance to embryonic development and somatic cell reprogramming.

5. DISCUSSION

In this study we purified proteins that have a role in *Nanog* transcription and pluripotency maintenance, development, and reprogramming using a dCas9-based approach. This is the first of its kind locus-based proteomic approach to an important pluripotency promoter using the CRISPR-based approach (Knaupp et al., 2020; Liu et al., 2017). Affinity purification resulted in 325 proteins, of which the majority of highly expressed (>log2) proteins have functions related to RNA splicing and mRNA metabolism. These results revealed a different proteome from the recent *Nanog* proteome purified using transcription activator like effector (TALE) based approach (Knaupp et al., 2020), indicating the differences in the approaches and the insufficiency of either approach in purifying a complete protein complex from a promoter site. We applied a single gRNA-based proteome purification, which might have revealed a rare interactome of the *Nanog* promoter (Liu et al., 2017). Nevertheless, we identified proteins BCLAF1, FUBP1, MSH6, PARK7, PSIP1, and THRAP3; which are highly expressed in mESCs with function in gene expression and showed binding to the *Nanog* promoter.

My data (Figure 7 & 8) indicate that the FUBP1 could be an essential factor in the expression of pluripotency genes in mESCs. However, knockdown of *Fubp1* has no effect on the number of mESC colonies formed; and EB size compared with the control agreeing that the absence of FUBP1 has no differences in cell-cycle progression in mESCs (Wesely et al., 2017). Recent studies implied that the DNA repair couples transcription regulation is prominent in mESCs (Fong et al., 2011). One study showed that a DNA repair complex directly coordinates and acts as a co-activator for OCT4-SOX2 heterodimer in mESCs (Fong et al., 2011). MSH6 purified in this study has been implicated in co-purified with OCT4 in mice (Ding et al., 2011; Pardo et al., 2010) and human ESCs (Huang et al., 2021). However, in this study, we reported first time the function of this protein in *Nanog* gene transcription and pluripotency. MSH6 decreases transcription of methylation rich sites by recruiting DNA Methyltransferase 1 (DNMT1) and the knockdown of either MSH6 or DNMT1 abolishes the transcription inhibition at sites of oxidative damage (Ding et al., 2016) a phenomenon partly true to the increased expression of some genes after *Msh6* knockdown in my study. PARK7, PSIP1, and THRAP3 showed functions related to switching off pluripotency gene expression when LIF was removed. These proteins also behaved as barriers to the induction of pluripotency gene expression during somatic cell reprogramming. Candidate proteins from *Nanog* promoter exhibited a role in embryo development via the EB formation experiment. These proteins seemed to be important for

proper growth as evident from the smaller EBs. The retarred growth of EBs can be attributed to the loss of nutrient assimilation because the EBs with knockdown proteins cannot express *Gata6* a primitive endoderm marker. Primitive endoderm was shown to be a vital extraembryonic layer contributing to the yolk sac formation and as a nutrient absorption source for embryo development (Ross & Boroviak, 2020). Thus my data re-implying the function of these proteins in the transcriptional outcome of pluripotent cells.

In other cell types, functions of these candidate proteins were reported to be involved in RNA-related cellular processes and have functional similarities with RNA binding proteins (RBPs). FUBP1 is a single-strand DNA and RNA binding protein, that has functions in regulating transcription (Hoang et al., 2019; Rabenhorst et al., 2015; Wesely et al., 2017); splicing (Elman et al., 2019) and exon inclusion *in vivo* (Li et al., 2013). BCLAF1 and THRAP3 functions related to mRNA splicing and export (Vohhodina et al., 2017). PSIP1 mediates alternative splicing by interacting with methylated histone H3K36 and splicing factors (Pradeepa et al., 2012). Actively transcribing chromatin regions are the hotspot for RBPs interaction. RBPs were recently found to co-ordinate with transcription factors and influence transcriptional outcomes (Xiao et al., 2019). Functional genomic studies revealed that many RBPs were enriched from chromatin (Ren et al., 2021; Xiao et al., 2019) and thus they establish cross-talk between transcription and post-transcriptional control via chromatin binding (Du and Xiao, 2020). Proteins analyzed in this study must have such functions as observed from their *Nanog* promoter occupancy and regulation in the transcriptional output of pluripotency-related genes

Identification of a different set of proteins from a recent locus-specific proteome of *Nanog* promoter using TALEs indicates that neither of the methods is sufficient to comprehensively purify the proteome of a particular locus. However, the absence of OCT4, SOX2, KLF4, and SALL4 in my method and the presence of these known *Nanog* regulators in the TALE-mediated isolation of nuclear chromatin (TINC) method (Knaupp et al., 2020) reflect the sensitivity differences in these two approaches. It is surprising to note that the previous studies that used CAPTURE for proteome analysis have never discussed the quality of their MS data (Zhang et al., 2020, 2021). But, according to Wierer & Mann, 2016, this decreased sensitivity in MS identification is an inherent phenomenon in locus-specific proteome approaches especially mediated by dCas9. Previous research shows that TALEs can purify both the strands of target DNA while dCas9 is strand-specific (Fujita

and Fujii, 2013). Secondly, I used only a single guide RNA, unlike TINC which used two TALEs targeting the *Nanog* promoter (Knaupp et al., 2020) adding to the decreased sensitivity of this approach. It was shown that in human cells dCas9/Cas9 is displaced by the FACT complex in either of the strands as an intrinsic mechanism to safeguard nucleosomes and that the residence of dCas9 was increased by the knockdown of FACT (Wang et al., 2020). As the outcomes of the CAPTURE method were dependent on the residence of dCas9 on the target locus, I suggest future studies consider such caveats in applying this method. Albeit high-abundance proteins like RBPs could have masked the identification or blocked the purification of transcription factors in this dCas9-based approach. Nevertheless, binding data of the candidate proteins to *Nanog* promoter DNA and regulation of transcription and pluripotency clarifies the specificity of CAPTURE approach used in this study.

Thus, I have isolated *Nanog* promoter bound proteins using CAPTURE; and identified novel regulators of pluripotency and reprogramming as claimed. In a way, I succeeded in picking candidates that might have regulation with pluripotency, but I could not enrich the complete network at the *Nanog* promoter. This suggests that although CAPTURE is efficient and easy to apply, its success could depend majorly on the target site. Increasing the sensitivity of mass spectrometry with the adoption of some labeling methods would have helped us identify the comprehensive identification of transcription regulation complex at the *Nanog* promoter.

6. CONCLUSION

In conclusion, dCas9 mediated purification of *Nanog* promoter locus identified novel regulatory proteins that are relevant to pluripotency maintenance and somatic cell reprogramming.
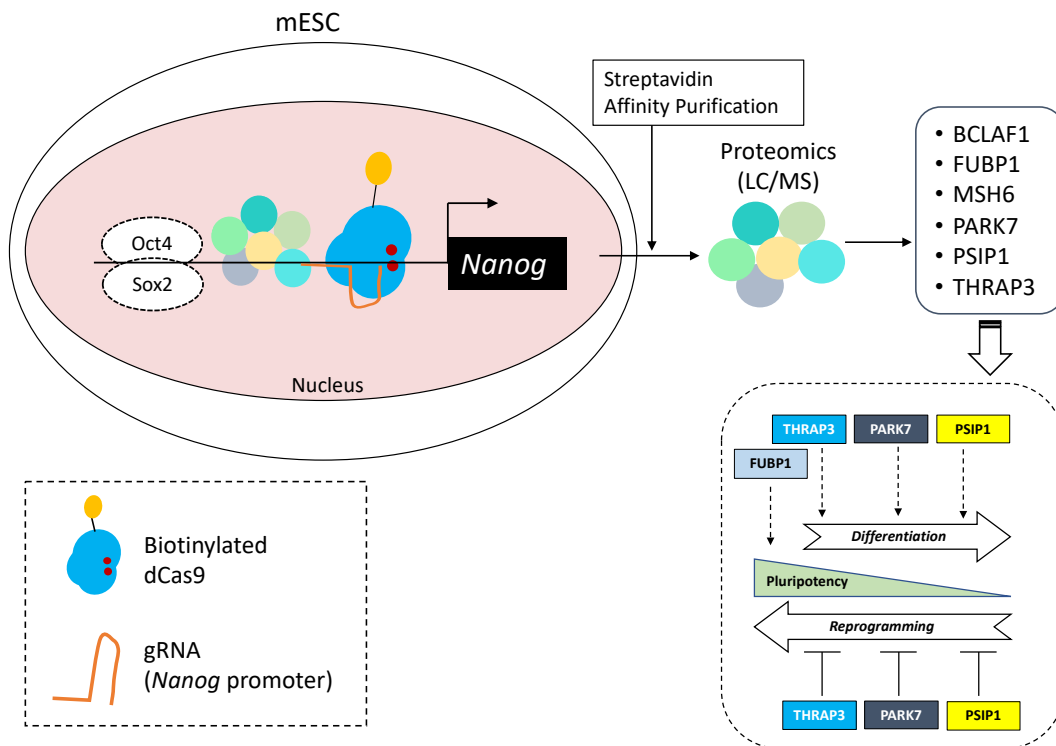
7. ILLUSTRATIONS
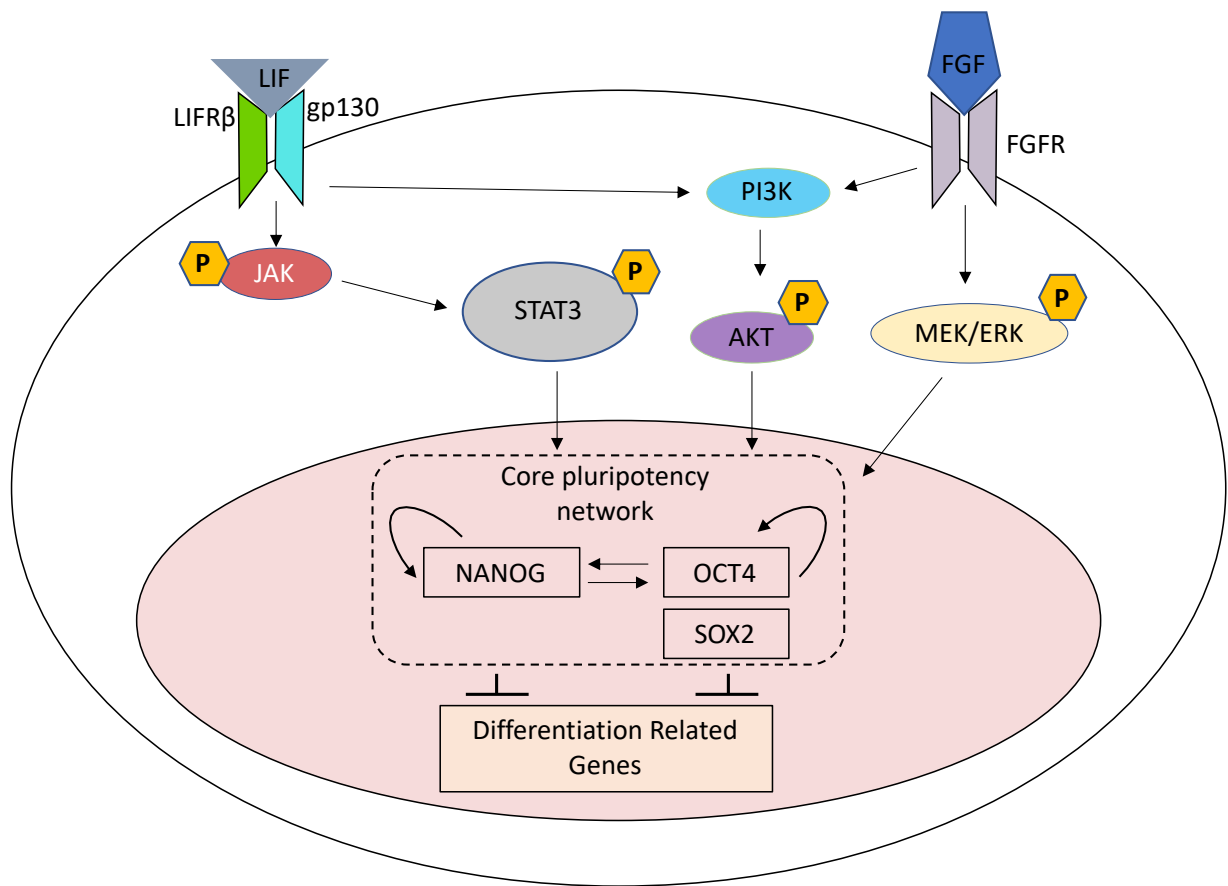
- Pluripotency is regulated by external signaling molecules and an auto regulatory internal transcriptional network.
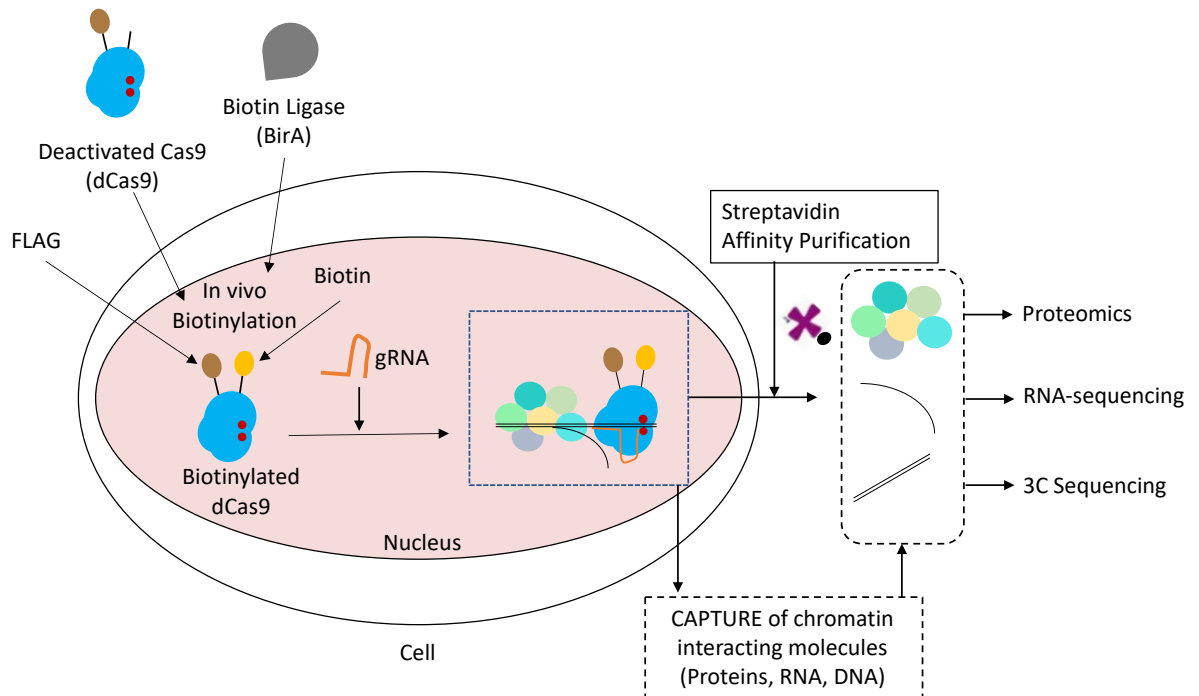
- dCas9 is biotinylated in vivo through biotin ligase enzyme.
- Target specific gRNA guides biotinylated dCas9 to target chromatin region.
- dCas9-gRNA complex on the and the target chromatin region is fixed in situ.
- Target region is purified via streptavidin affinity purification and the chromatin interacting molecules can be identified.

Illustration 3: Regulation of *Nanog* gene expression through promoter binding proteins
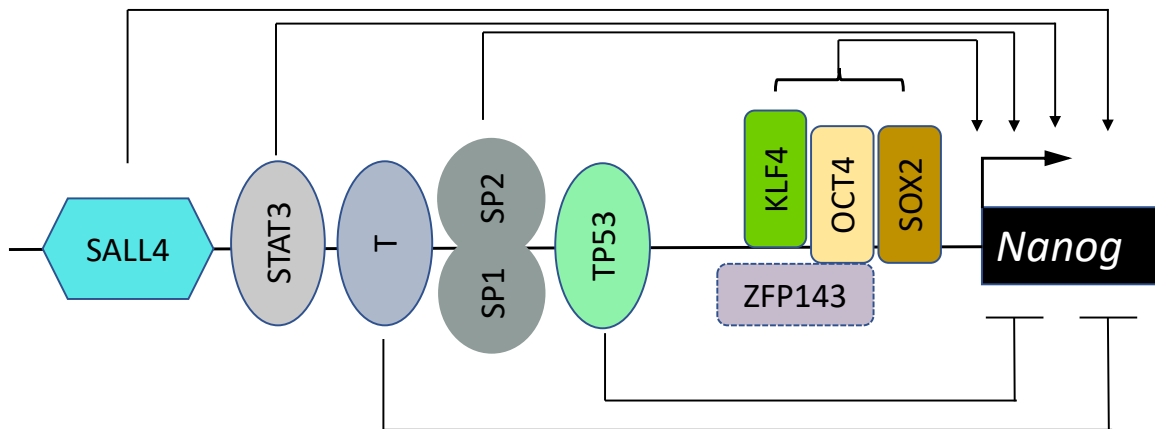
- *Nanog* is regulated in a context dependent manner by a variety of promoter binding proteins

8. FIGURES

Figure 1: Generating mES cells for CAPTURE
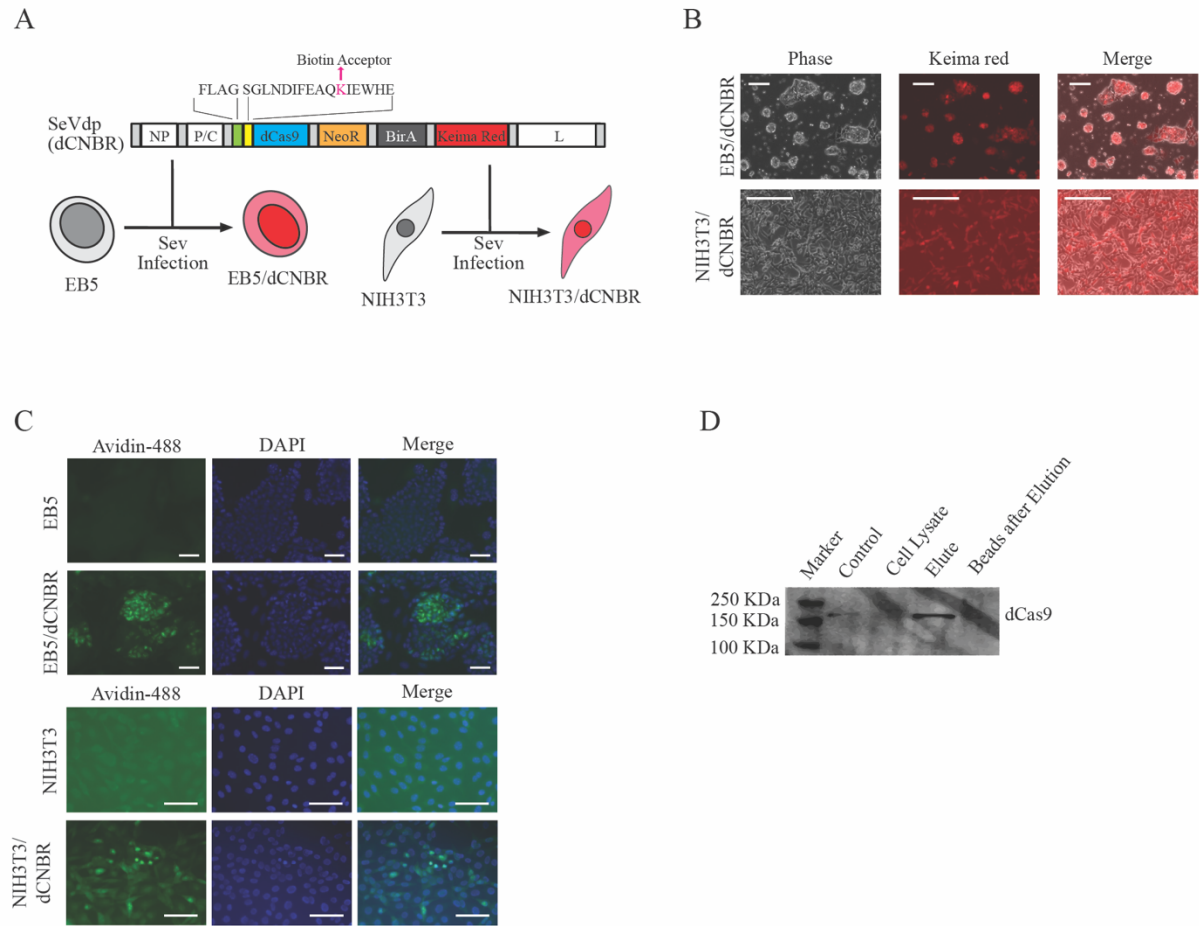
A. SeV vector (SeVdp(dCNBR)) was used to express proteins for CAPTURE including dCas9; Neomycin resistance (NeoR); Biotin Ligase (BirA) and Keima Red fluorescent gene. dCas9 is fused with FLAG and Biotin Acceptor sequences on the N-terminal. N, P, and L genes encode Sendai viral proteins N, P, and L, respectively. EB5 and NIH3T3 cells after the vector infection were named as EB5/dCNBR and NIH3T3/dCNBR respectively.

B. EB5/dCNBR and NIH3T3/dCNBR cells expressing Keima Red from the vector after Neomycin selection. Scale bars represent 200 μm.

C. Green fluorescence of Avidin-488, binding to the biotinylated dCas9 in the nuclei of infected cells. Top and bottom panels represent EB5/dCNBR and NIH3T3/dCNBR cells respectively. Scale bars represent 200 μm.

D. Western blot of affinity purified dCas9 protein by streptavidin beads. Cell lysate from NIH3T3 transfected with FLAG-tagged Cas9 was used as control.

Figure 2: Designed gRNAs can recruit Cas9 to the target site.
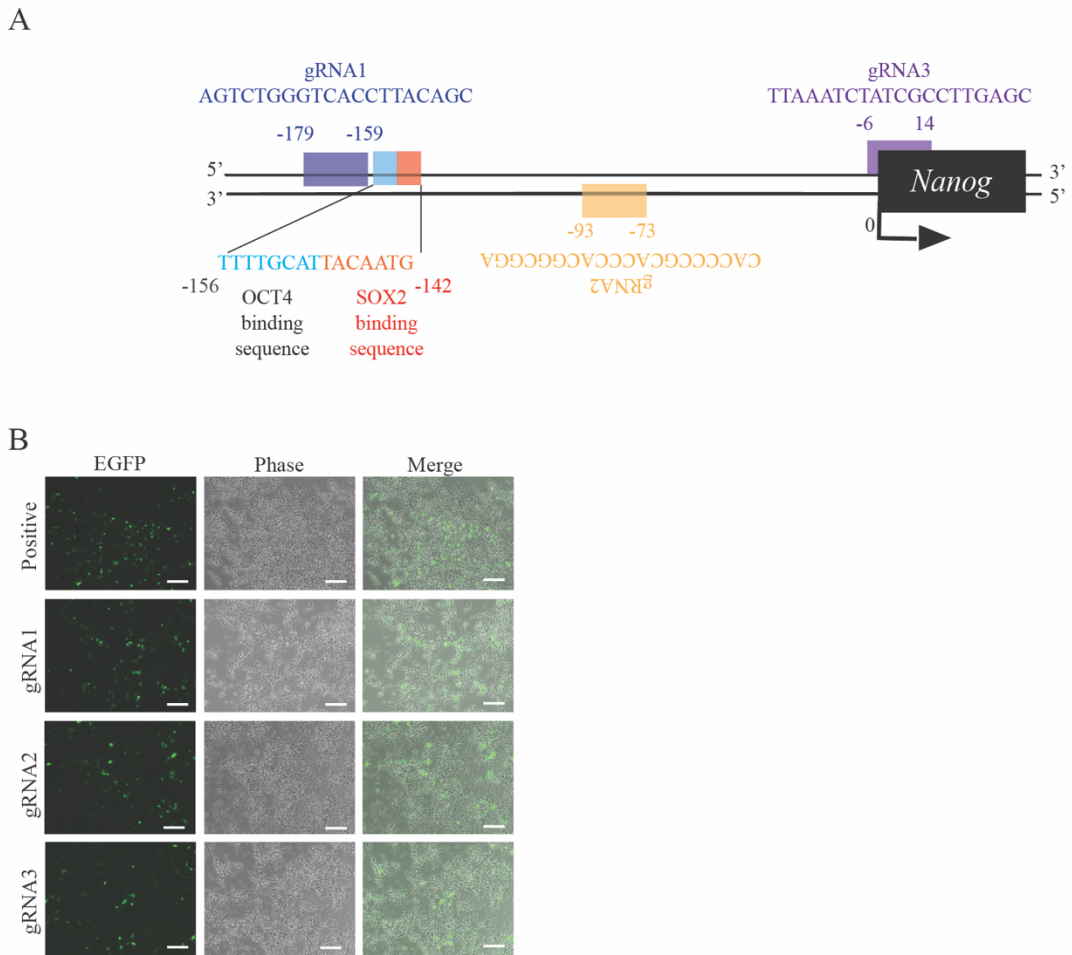
A. Position of gRNAs on *Nanog* promoter region along with the indicated binding sites of OCT4-SOX2 proteins.

B. Three days after transfection of donor and gRNA expressing plasmids, HEK293T cells express EGFP fluorescence representing the functioning of gRNAs. The positive control is the plasmid with control sequences described previously (Mashiko et al., 2013). Scale bars, 50 μm

Figure 3: dCas9 has cell type-specific target enrichment efficiency
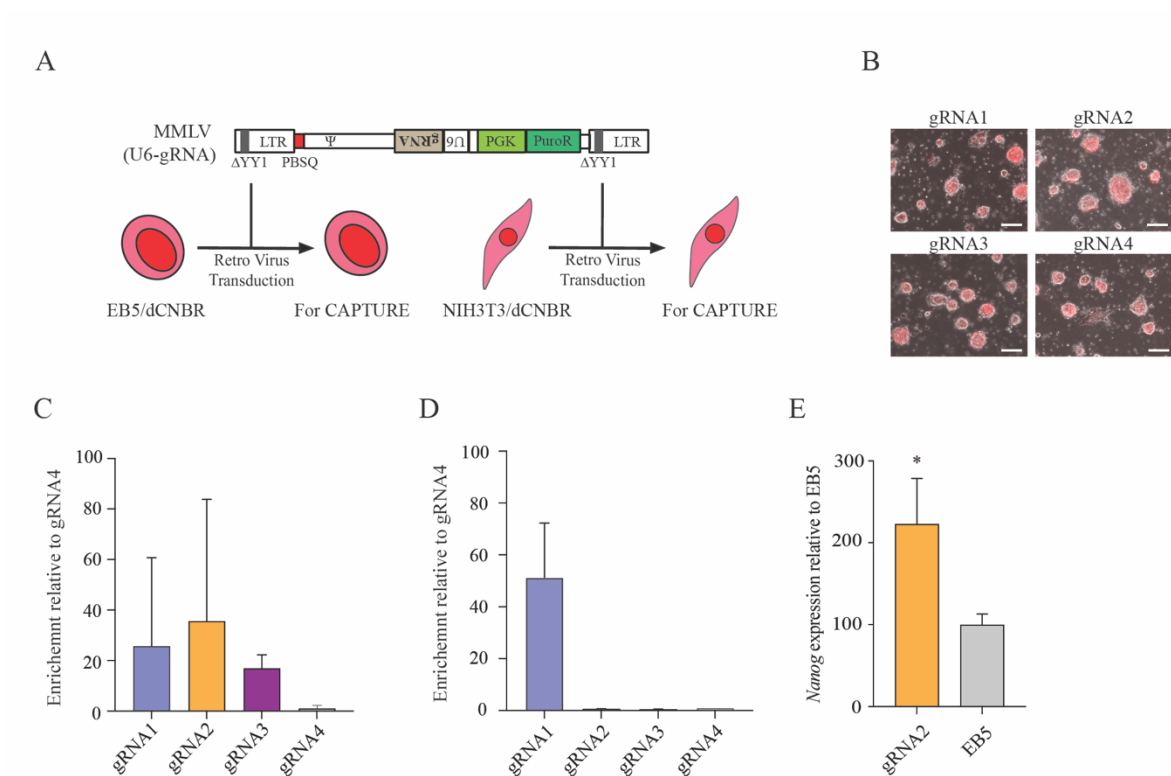
A. Scheme of generating cells expressing gRNA with SeVdp(dCNBR).

B. Morphology of EB5/dCNBR cells after the transduction of gRNA expressing retrovirus. Scale bars, 200 μm

C. Enrichment of *Nanog* promoter in EB5/dCNBR cells after affinity purification of dCas9. Enrichment was quantified using ChIP-qPCR of *Nanog* promoter region and shown as relative to gRNA4.

D. Enrichment of *Nanog* promoter in NIH3T3/dCNBR cells after affinity purification of dCas9. Enrichment was quantified using ChIP-qPCR of *Nanog* promoter region and shown as relative to gRNA4.

E. mRNA expression of the *Nanog* gene in gRNA2 expressing EB%/dCNBR cells. Data is normalized to TBP and shown as relative to *Nanog* expression in mESC. Data represent the mean ± SEM of three independent experiments.

Figure 4: Purification of proteins binding to the *Nanog* promoter in mESCs
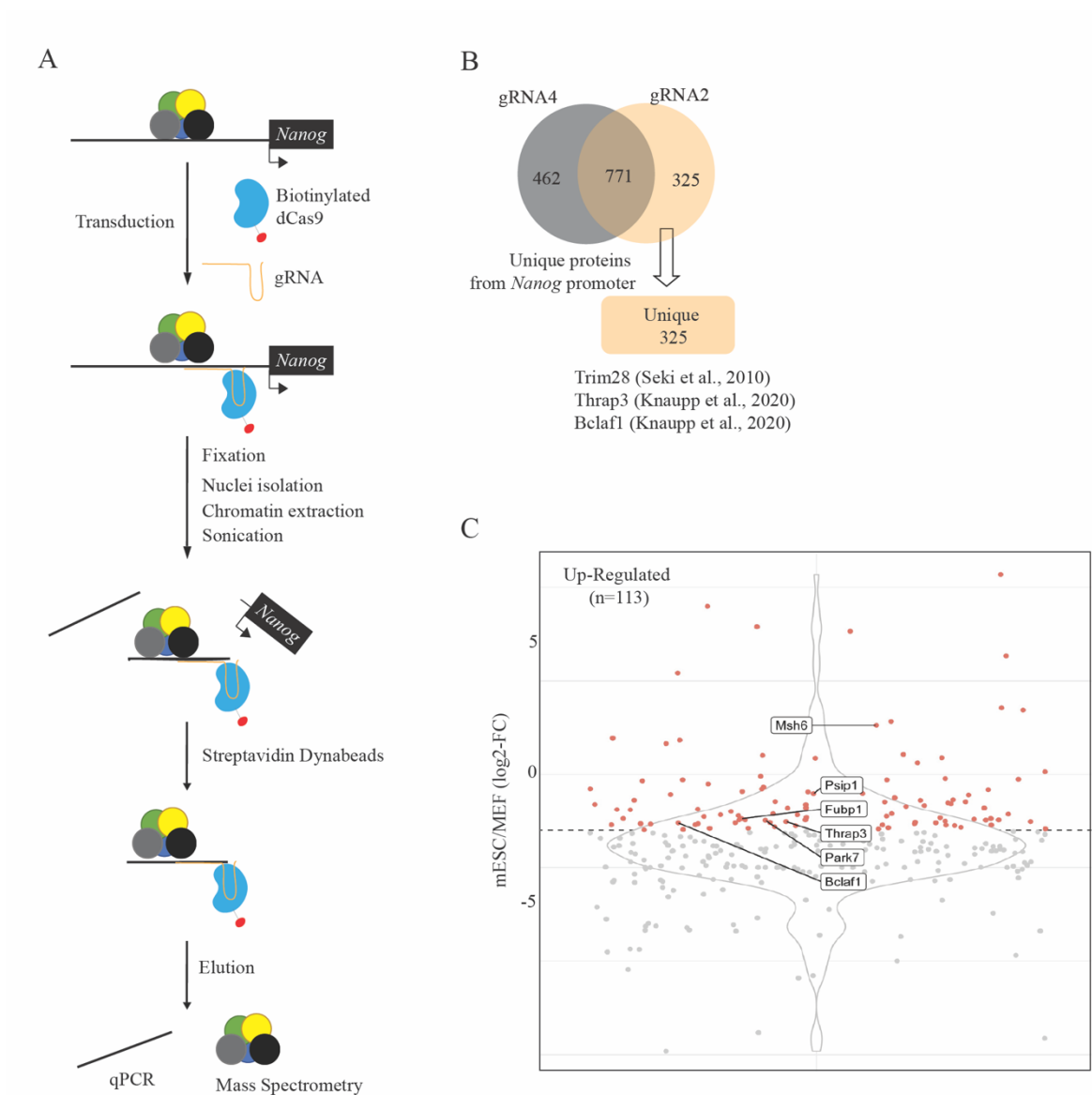
A. Scheme of CAPTURE and isolation of proteins from *Nanog* promoter in mESC.

B. Venn diagram of identified proteins by LC/MS/MS using gRNA2 and gRNA4. Proteins were identified from three independent experiments. Totally, 325 proteins were obtained as unique proteins for gRNA2.

C. The published RNA-seq data (Chronis et al., 2017) was used to obtain differential expression data of each unique protein in MEFs and mESCs. Log2-fold change was calculated and shown in a violin plot. Highly expressed (> 2-fold) genes in mESC were selected for further analyses. The 6 candidate proteins are highlighted.

Figure 5: Purified proteins have gene regulatory functions
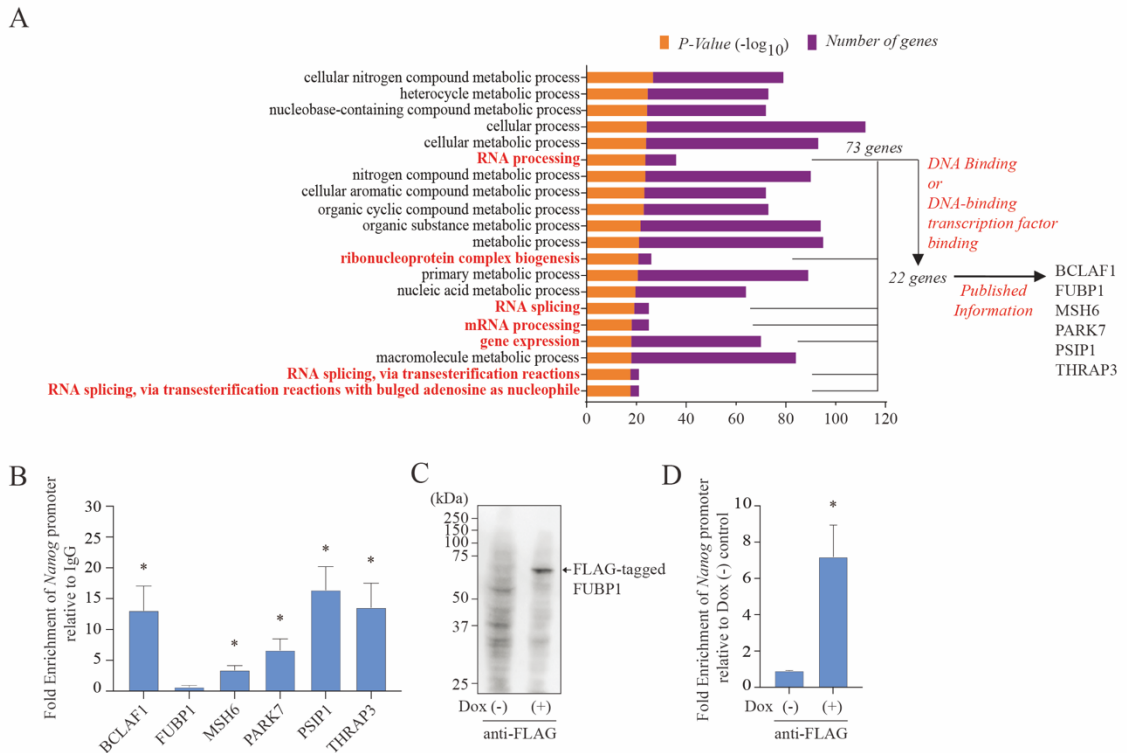
A. Top 20 enriched GO terms from GO Term Finder (Boyle et al., 2004) along with the number of genes corresponding to each GO term were plotted. Scheme of selection of candidate proteins involved in gene regulatory function. Total of 73 genes enriched in GO terms including RNA processing, ribonucleotide complex biogenesis, RNA splicing, and gene expression were narrowed down to six proteins as indicated.

B. Binding of the 6 candidate proteins to the *Nanog* promoter. ChIP assays of chromatin from EB5 cells were performed using an antibody against each protein. Data represent the mean ± SEM of three independent experiments.

C. Expression of FLAG-tagged FUBP1. EB5 cells were transduced with a lentiviral vector (LV(TO-3F-Fubp1)) expressing FLAG-tagged FUBP1 using a doxycycline-inducible expression system. Whole cell lysate from the cells cultured with or without 2 µg/mL doxycycline for 5 days were prepared and subjected to western blotting using anti-FLAG antibody.

D. Binding of the FUBP1 to the *Nanog* promoter. ChIP assay of chromatin from cells in **C** was performed using anti-FLAG antibody. Data are represented as the means ± SEM of three independent experiments.

Figure 6: Knockdown of candidate proteins have no immediate effect on mESC properties
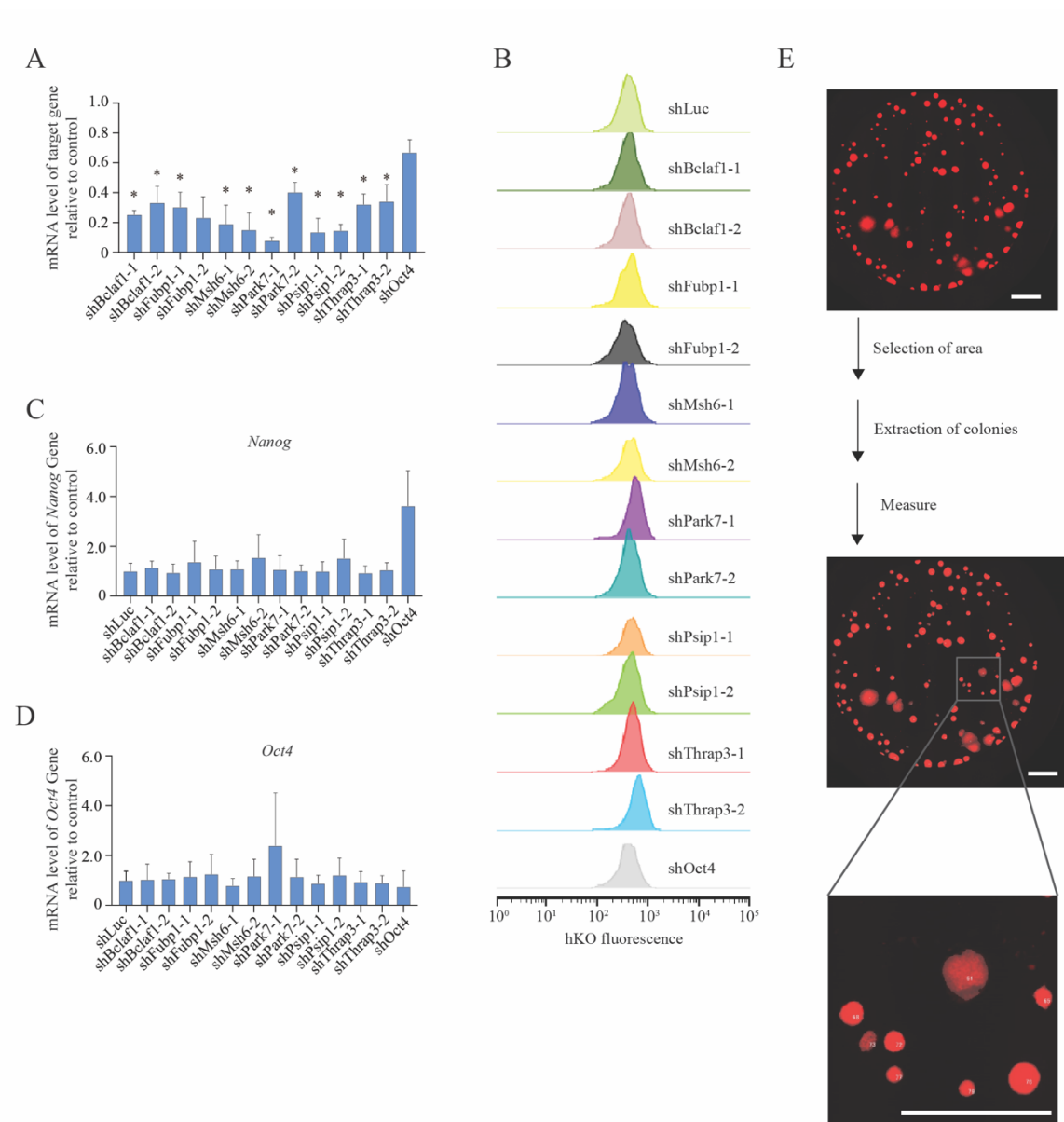
A. Knockdown of candidate gene expression by shRNA. EB5/ReKO cells were transduced with a retroviral vector expressing the indicated shRNA followed by puromycin selection. mRNA levels of each candidate gene were determined 5 days after transduction. Data represent mean ± SEM from three independent experiments. *$P < 0.05$ versus EB5 cells treated with control shRNA (shLuc).

B. hKO fluorescent level in *Rex1* reporter cells with shRNA treatment. EB5/ReKO cells were transduced as **A**. Five days after transduction, cells were analyzed for hKO expression using MoFlo XDP. Representative FACS histograms are shown.

C. Pluripotency of cells after shRNA treatment for 5 days. EB5/ReKO cells were transduced as **A**. *Nanog* mRNA levels in the cells were determined 5 days after transduction. Data are represented as the means ± SEM of three independent experiments.

D. Pluripotency of cells after shRNA treatment for 5 days. EB5/ReKO cells were transduced as **A**. *Oct4* mRNA levels in the cells were determined 5 days after transduction. Data are represented as the means ± SEM of three independent experiments.

E. Pipeline for image analyses of mESC colonies. mESC colonies expressing hKO were analyzed as described in Materials and Methods. Scale bars, 1,000 μm.

Figure 7: Changes in self-renewal properties of mESCs after knockdown of candidate proteins
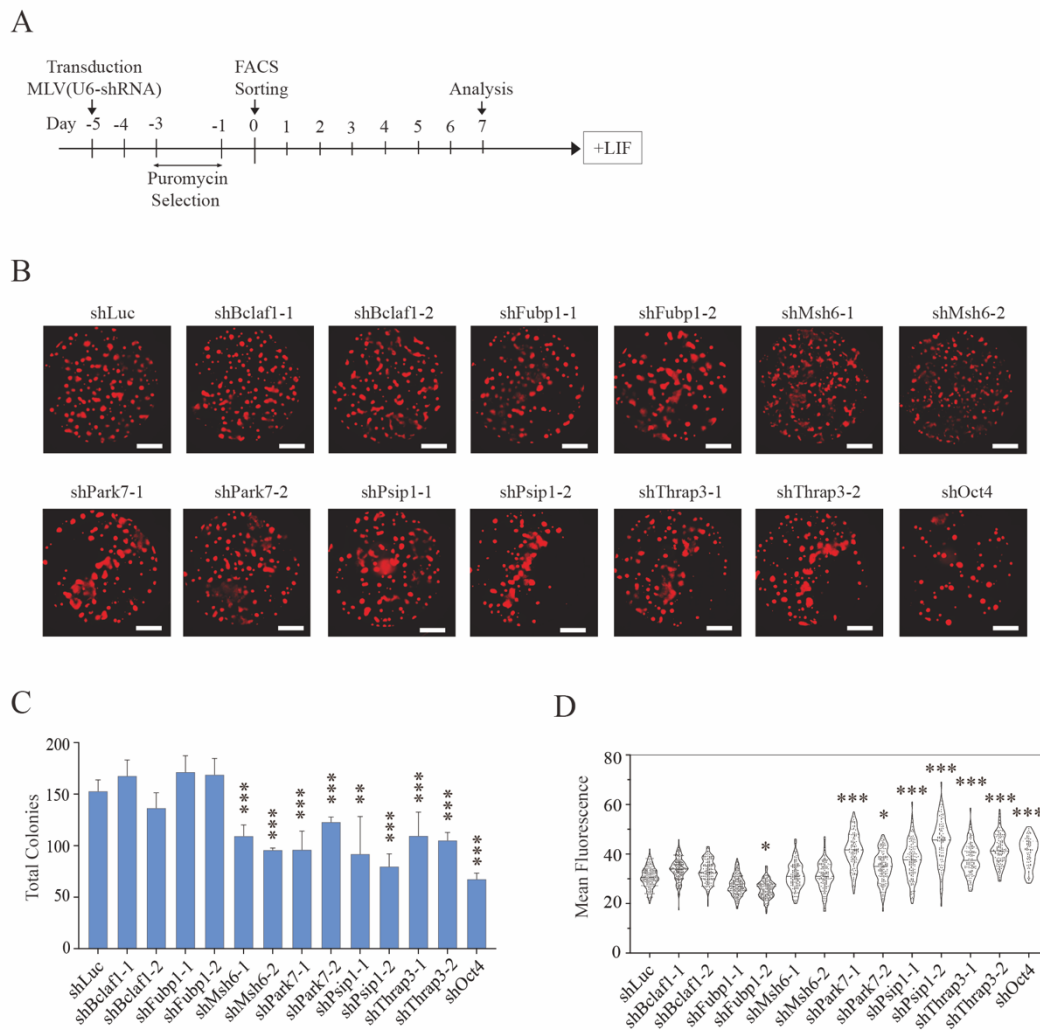
A. Experimental scheme of pluripotency analysis. EB5/ReKO cells were transduced with a retroviral vector expressing each shRNA followed by puromycin selection for 2 days. Five days after transduction, 1,000 cells/well were sorted to a flat-bottom 96-well plate for culture in mESC medium with LIF.

B. Whole well images of hKO (+) EB5/ReKO colonies cultured with LIF. EB5/ReKO cells were treated with indicated shRNA and cultured in mESC medium with LIF. Whole well images were collected 7 days after sorting. Representative images are shown. Scale bars, 1,500 μm.

C. Colony numbers of EB5/ReKO cells treated with shRNA. Colony numbers were counted from the images collected in B. Data represent the mean ± SEM from five independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).

D. Mean fluorescent intensity of EB5/ReKO cells treated with shRNA. hKO fluorescent intensity in each colony was measured from the images collected in B. Data represent the mean ± SEM from total colonies in each shRNA. *P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).
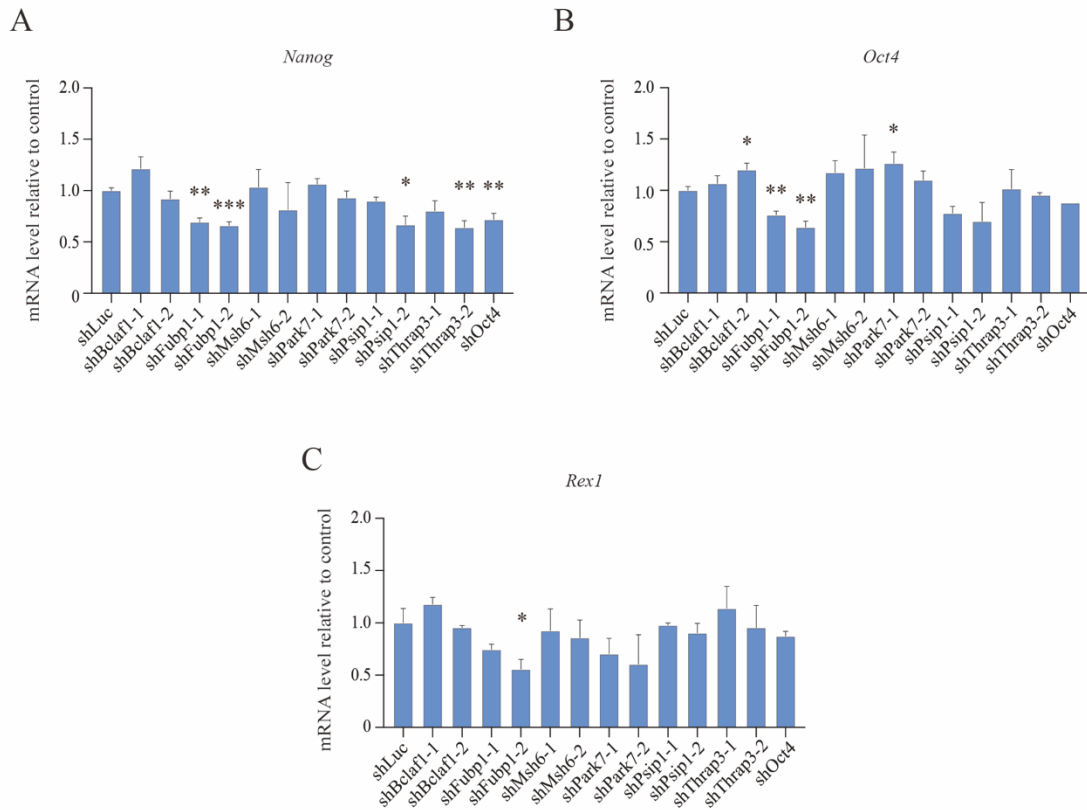
A. mRNA level of pluripotency-related genes. *Nanog* mRNA levels in the EB5/ReKO cells prepared as **Figure 7A** were determined 7 days after cell sorting. Data represent the mean ± SEM of three independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).

B. mRNA level of pluripotency-related genes. *Oct4* mRNA levels in the EB5/ReKO cells prepared as **Figure 7A** were determined 7 days after cell sorting. Data represent the mean ± SEM of three independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).

C. mRNA level of pluripotency-related genes. *Rex1* mRNA levels in the EB5/ReKO cells prepared as **Figure 7A** were determined 7 days after cell sorting. Data represent the mean ± SEM of three independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).

Figure 9: Comparative analysis on the pluripotency in the presence and absence of LIF after knockdown
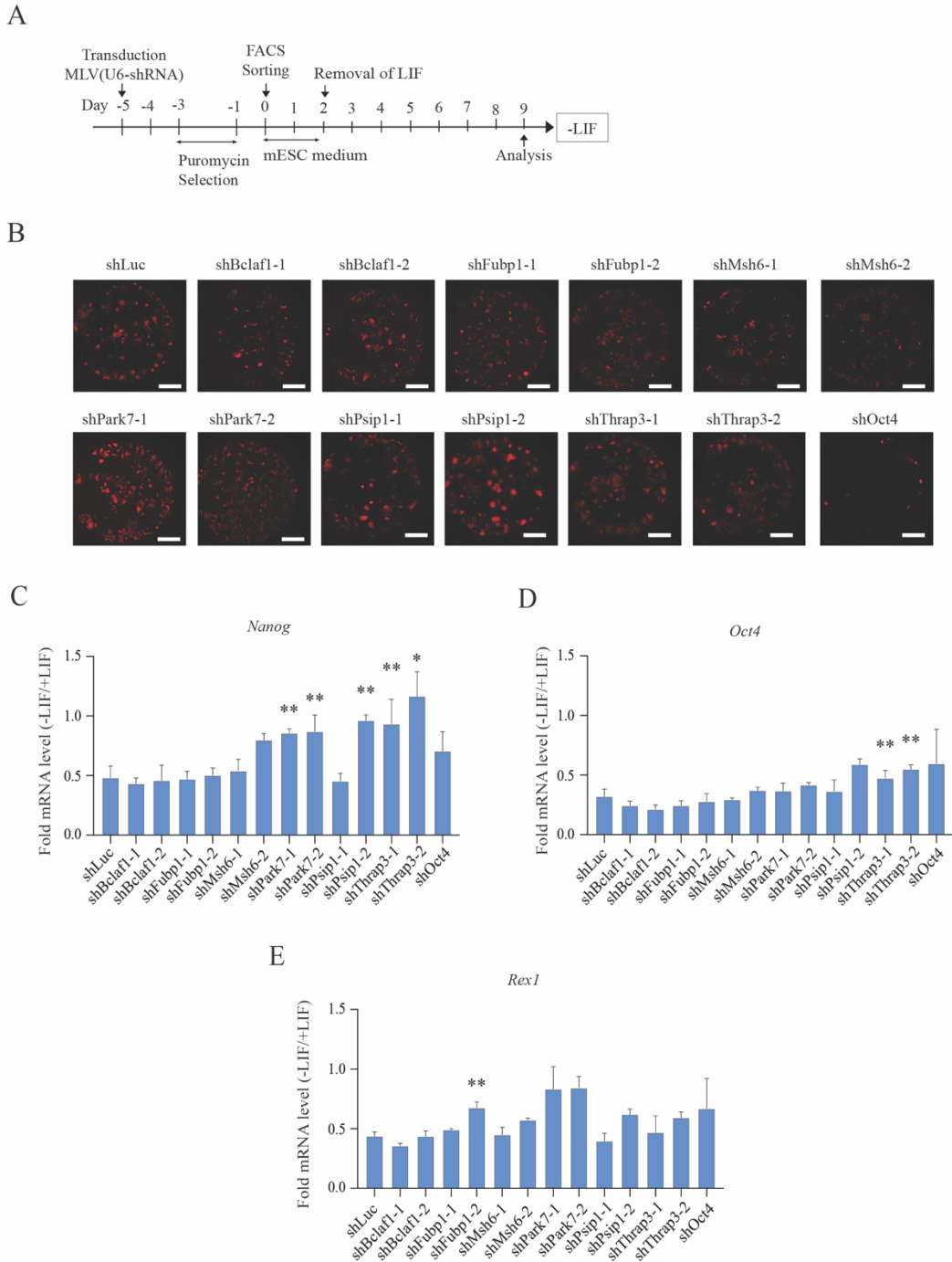
A. Experimental scheme of differentiation analyses. EB5/ReKO cells were transduced with retroviral vector expressing each shRNA followed by puromycin selection for 2 days. Five days after transduction, 1,000 cells/well were sorted to a flat-bottom 96-well plate for culture in mESC medium without with LIF for first 2 days, then without LIF for 7 days.

B. Whole well images of hKO (+) EB5/ReKO colonies cultured without LIF. EB5/ReKO cells treated with indicated shRNA and cultured in mESC medium without LIF. Whole well images were collected 9 days after cell sorting. Representative images are shown. Scale bars, 1,500 μm.

C. mRNA level of pluripotency-related genes. *Nanog*, mRNA levels in the EB5/ReKO cells prepared as (**A**) were determined 9 days after cell sorting. Data represent the mean ± SEM of three independent experiments. P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).

D. mRNA level of pluripotency-related genes. *Oct4* mRNA levels in the EB5/ReKO cells prepared as (**A**) were determined 9 days after cell sorting. Data represent the mean ± SEM of three independent experiments. P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).

**E.** mRNA level of pluripotency-related genes. *Rex1* mRNA levels in the EB5/ReKO cells prepared as (**A**) were determined 9 days after cell sorting. Data represent the mean ± SEM of three independent experiments. P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).
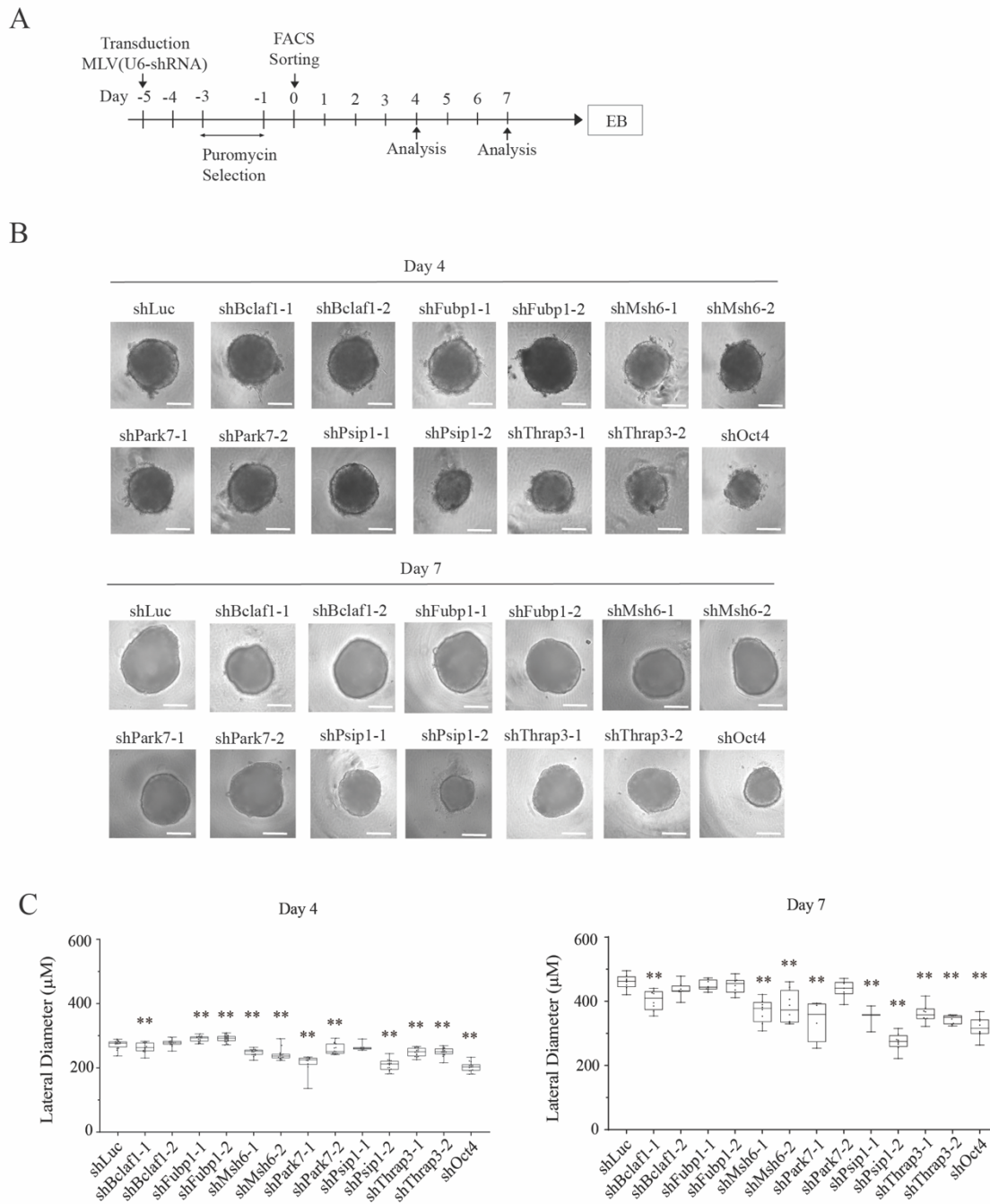
A. Experimental scheme of differentiation analysis. EB5/ReKO cells were transduced with retroviral vector expressing each shRNA followed by puromycin selection for 2 days. Five days after transduction, 1,000 cells/well were sorted to a V-bottom 96-well plate for EB formation (EB).

B. Images of EBs. Representative images of EB from EB5/ReKO cells treated with indicated shRNA and cultured in EB differentiation medium were taken at indicated date. Scale bars, 150 μm.

C. Size of EBs. Lateral diameters of EB from EB5/ReKO cells prepared as A were measured 4 days or 7 days after sorting. Data represent the mean ± SEM of 16 (Day 4) or 8 (Day 7) EBs. P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).

Figure 11: Effect of knockdown on EB development
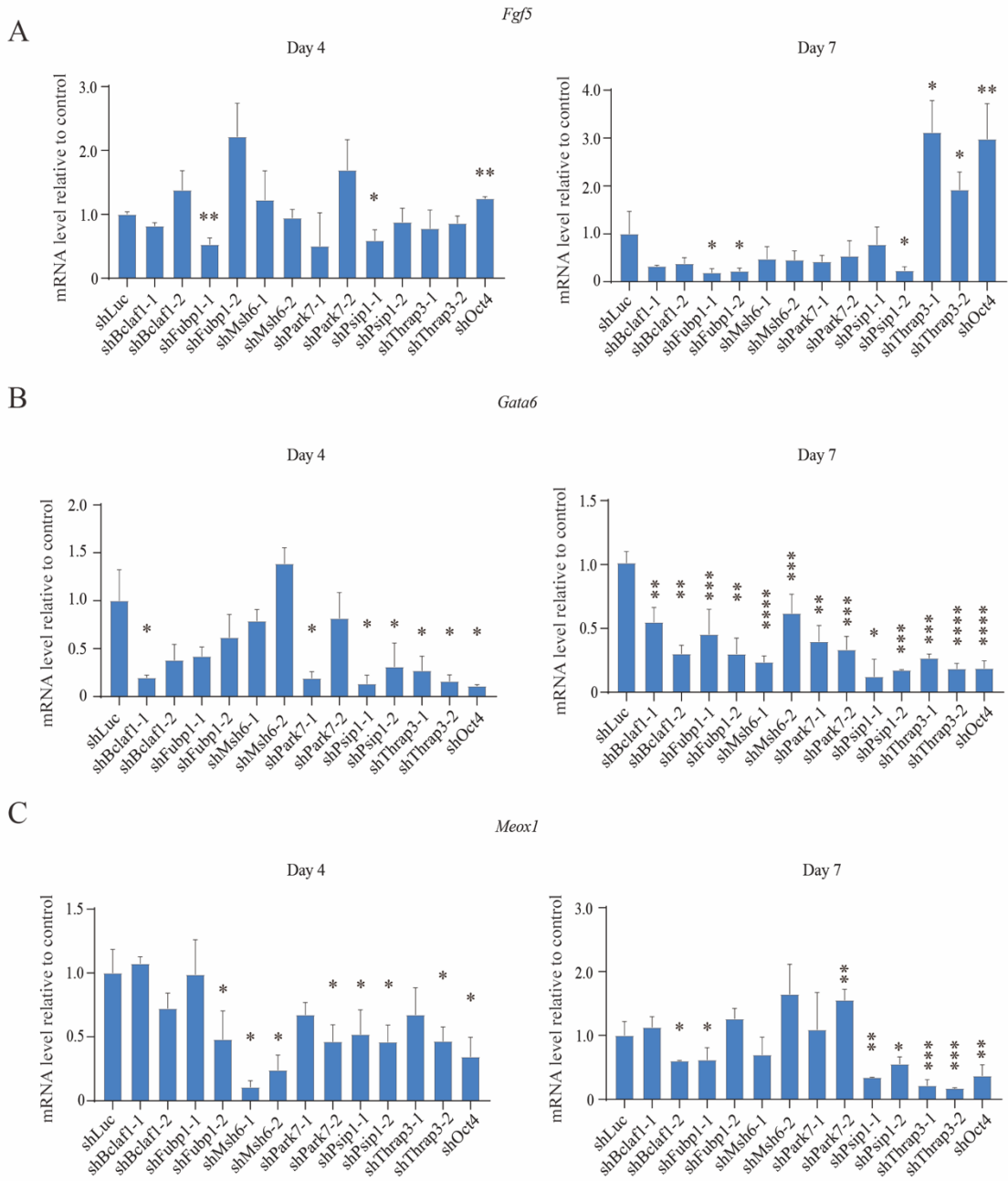
A-C. mRNA level of lineage marker genes. *Fgf5* (**A**), *Gata6* (**B**), or *Meox1* (**C**) mRNA levels in the EB from EB5/ReKO cells prepared as Figure 10A were determined 4 days or 7 days after cell sorting. Data represent the mean ± SEM of three independent experiments. P < 0.05, **P < 0.01, ***P < 0.001 versus EB5/ReKO cells treated with control shRNA (shLuc).

Figure 12: SeV mediated somatic cell reprogramming
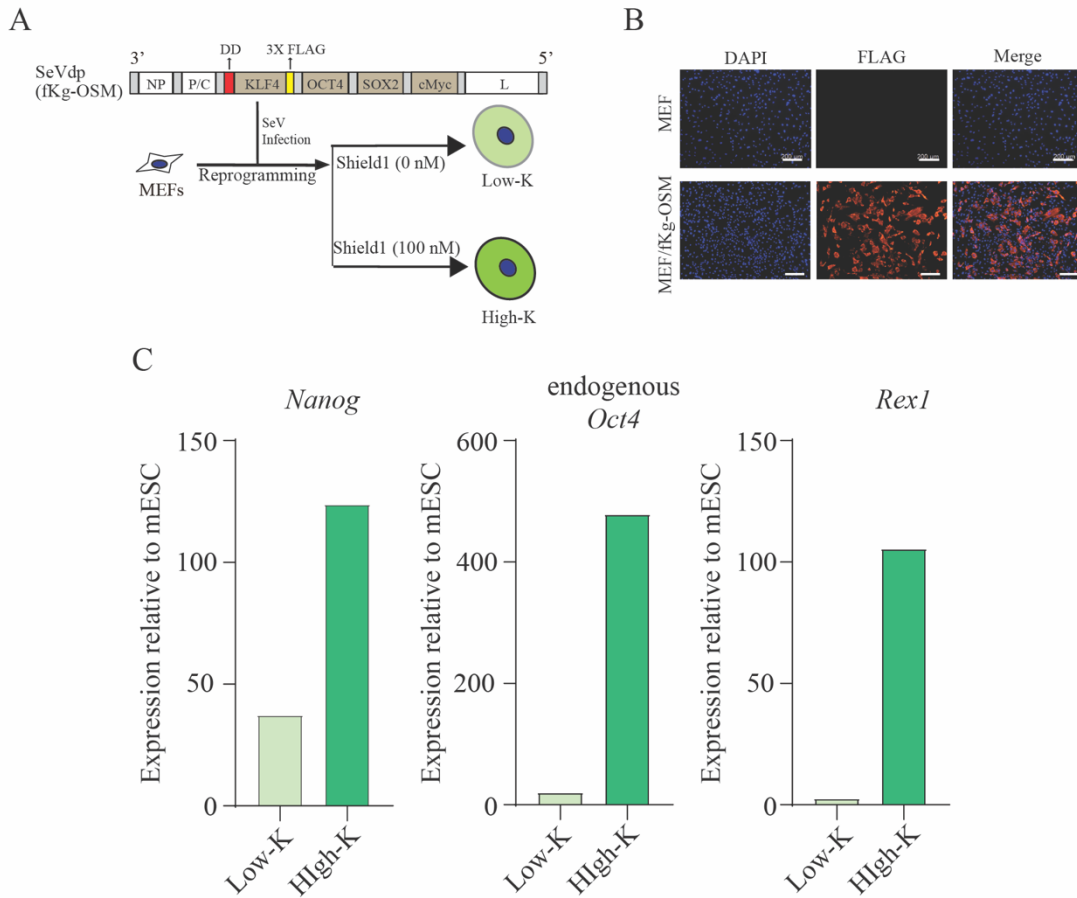
A. Schematic representation of iPSC generation using SeVdp(fKg-OSM) and cultured with (High-K) or without (Low-K) 100 nM Shield1.

B. Infectivity of SeVdp(fKg-OSM) verified from the immunostaining with anti-FLAG antibody conjugated with AlexaFlour 555-conjugated secondary antibodies. Top panel is MEF cell without infection and bottom panel with SeV infection. Scale bars, 200 μm

C. mRNA expression level of pluripotency-related genes *Nanog*, endogenous *Oct4* and *Rex1* in Low-K and High-K cells were determined, and the expression was plotted relative to mESC.

Figure 13: Effect of knockdown of candidate proteins on somatic cell reprogramming
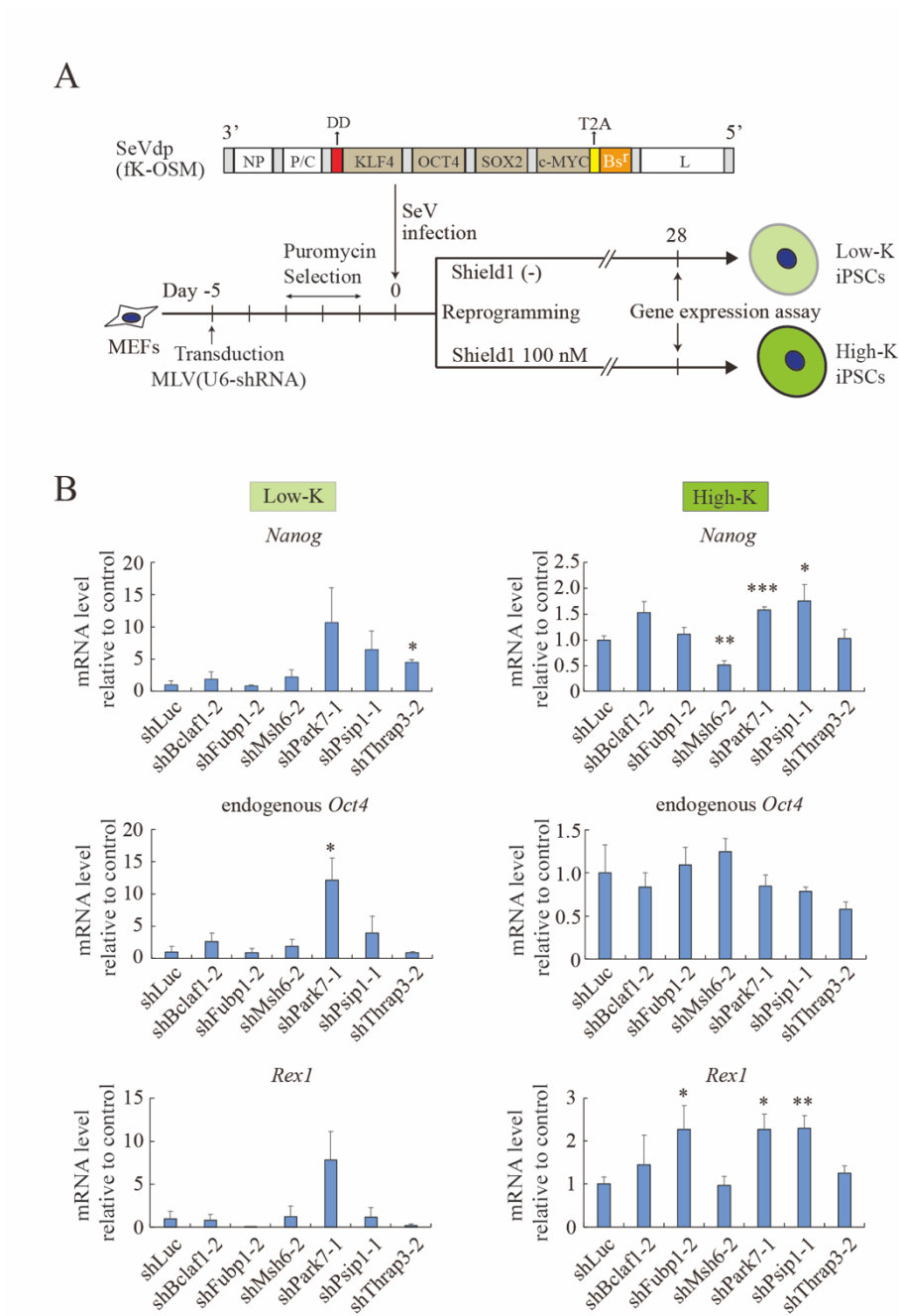
A. Schematic representation of iPSC production with knockdown of candidate proteins. MEFs were transduced with retroviral vector expressing each shRNA. After puromycin selection for 3 days, the selected MEFs were infected with SeVdp(fK-OSM) and cultured with (High-K) or without (Low-K) 100 nM Shield1.

B. mRNA level of pluripotency-related genes. MEFs were treated with indicated shRNA and reprogrammed as described in (**A**). *Nanog*, endogenous *Oct4* and *Rex1* mRNA level were determined at day 28 of reprogramming. Data represent the mean ± SEM of three independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 versus reprogrammed MEFs treated with control shRNA (shLuc).

9. TABLES

Resource Table 1: List of FDA-approved cell based therapies (Mao and Mooney, 2015)

Category	Name	Biological agent	Approved use
Biologics	laViv	Autologous fibroblasts	Improving nasolabial fold appearance
	Carticel	Autologous chondrocytes	Cartilage defects from acute or repetitive trauma
	Apligraf, GINTUIT	Allogeneic cultured keratinocytes and fibroblasts in bovine collagen	Topical mucogingival conditions, leg and diabetic foot ulcers
	Cord blood	Hematopoietic stem and progenitor cells	Hematopoietic and immunological reconstitution after myeloablative treatment
Cell-based medical devices	Dermagraft	Allogenic fibroblasts	Diabetic foot ulcer
	Celution	Cell extraction	Transfer of autologous adipose stem cells
Biopharmaceuticals	GEM 125	PDGF-BB, tricalcium phosphate	Periodontal defects
	Regranex	PDGF-BB	Lower extremity diabetic ulcers
	Infuse, Infuse bone graft, Inductos	BMP-2	Tibia fracture and nonunion, and lower spine fusion
	Osteogenic protein-1	BMP-7	Tibia nonunion

Table 1: Sequence of gRNAs used in this study

gRNA1 (sgNanog-1)	5'-CACC GCTGTAAGGTGACCCAGACT-3'
	5'-AAAC AGTCTGGGTCACCTTACAGC-3'
gRNA2 (sgNanog-2)	5'-CACC GTGGGGCGTGGGTGCCGCCT-3'
	5'-AAAC AGGCGGCACCCACGCCCCAC-3'
gRNA3 (sgNanog-3)	5'-CACC GCTCAAGGCGATAGATTTAA-3'
	5'-AAAC TTAAATCTATCGCCTTGAGC-3'
gRNA4 (sgGal4)	5'-CACC GAACGACTAGTTAGGCGTGTA-3'
	5'-AAAC TACACGCCTAACTAGTCGTTC-3'

Table 2: Sequence for shRNAs used in this study

shBclaf1-1	5'-GATCCGGATCTGGCTCTGTTGGAAATTTCAAGAGAATTTCCAACAGAGCCAGATCCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGGATCTGGCTCTGTTGGAAATTCTCTTGAAATTTCCAACAGAGCCAGATCCG-3'
shBclaf1-2	5'-GATCCGGTTCATCATGTGAAAGAATTCAAGAGATTCTTTCACATGATGAACCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGGTTCATCATGTGAAAGAATCTCTTGAATTCTTTCACATGATGAACCG-3'
shFubp1-1	5'-GATCCGCTGGTACATCATTGAATTCATTCAAGAGATGAATTCAATGATGTACCAGCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGCTGGTACATCATTGAATTCATCTCTTGAATGAATTCAATGATGTACCAGCG-3'
shFubp1-2	5'-GATCCGGACAGGTTGATTATACAATTCAAGAGATTGTATAATCAACCTGTCCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGGACAGGTTGATTATACAATCTCTTGAATTGTATAATCAACCTGTCCG-3'
shMsh6-1	5'-GATCCGAGGTGCATGAGGCTTATTTATTCAAGAGATAAATAAGCCTCATGCACCTCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGAGGTGCATGAGGCTTATTTATCTCTTGAATAAATAAGCCTCATGCACCTCG-3'
shMsh6-2	5'-GATCCGCTACTGAGTAAGATTCATTTCAAGAGAATGAATCTTACTCAGTAGCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGCTACTGAGTAAGATTCATTCTCTTGAAATGAATCTTACTCAGTAGCG-3'
shPark7-1	5'-GATCCGCAGTGTAGCCGTGATGTAATTTCAAGAGAATTACATCACGGCTACACTGCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGCAGTGTAGCCGTGATGTAATTCTCTTGAAATTACATCACGGCTACACTGCG-3'
shPark7-2	5'-GATCCGCTGCAGTCTTTAAGAAATTTCAAGAGAATTTCTTAAAGACTGCAGCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGCTGCAGTCTTTAAGAAATTCTCTTGAAATTTCTTAAGACTGCAGCG-3'
shPsip1-1	5'-GATCCGGTTATTGATGAAGATAAATAACCTGACCCATTATTTATCTTCATCAATAACCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGGTTATTGATGAAGATAAATAATGGGTCAGGTTATTTATCTTCATCAATAACCG-3'
shPsip1-2	5'-GATCCGCAATGAGGATGTGACTAAAGTTCAAGAGACTTTAGTCACATCCTCATTGCTTTTTTGGAAAG-3'

	5'-AATTCTTTCCAAAAAAGCAATGAGGATGTGACTAAAGTCTCTTGAACTTTAGTCACATCCTCATTGCG-3'
shThrap3-1	5'-GATCCGGGATATATAGTAATATTATAACCTGACCCATTATAATATTACTATATATCCCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGGGATATATAGTAATATTATAATGGGTCAGGTTATAATATTACTATATATCCCG-3'
shThrap3-2	5'-GATCCGGAGCTGGATGAGCACGATAATTCAAGAGATTATCGTGCTCATCCAGCTCCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGGAGCTGGATGAGCACGATAATCTCTTGAATTATCGTGCTCATCCAGCTCCG-3'
shOct4	5'-GATCCGCCTTAAGAACATGTGTAATTCAAGAGATTACACATGTTCTTAAGGCTTTTTTGGAAAG-3'
	5'-AATTCTTTCCAAAAAAGCCTTAAGAACATGTGTAATCTCTTGAATTACACATGTTCTTAAGGCG-3'

Table 3: Sequence of primers used to amplify ORF of *Fubp1* gene

mFubp1	5'-ATGGCCGACTACTCCACAGTG-3'
	5'-TTATTGGCCCTGAGGTGCTGG-3'

Table 4: Sequence of primer to amplify *Nanog* promoter region

Nanog promoter	5'-TAGCCTGCAGGTCATCCCCCACTTGACCTGAAAC-3'
	5'-TATCGAATTCGTCCAGCCTTCCCACAGAAAGAGC-3'

Table 5: Sequence of primers used in enrichment analysis

CAPTURE-qPCR	5'-GGTCACCTTACAGCTTCTTTTG-3'
	5'-TATTCTCCCAGGCACCCAG-3'
ChIP-qPCR	5'-GCAGGACCTACCCTTTAAATC-3'
	5'-CCCACAGAAAGAGCAAGAC-3'

Table 6: Sequence of primers used for mRNA expression analysis in this study

mThrap3	5'-AGAAAAGCCCAGAGATACACAG-3'
	5'-TATTTTCCCTCAGCCTTGTAGCC-3'
mBclaf1	5'-TGTTGAAGATGACGAAGAGACC-3'
	5'-CAAGTTCTGCTCCCTGTTGC-3'
mPsip1	5'-CGAGGAAGAAAGAGAAAGGCTG-3'
	5'-AGGTCTGCCTCTTGGTTTGG-3'
mPark7	5'-TGCAGTGTAGCCGTGATGTA-3'
	5'-TATGGTCCCTGCGTTTTTGC-3'
mFubp1	5'-GCTACAACCCAAACGAACGG-3'

	5'-AGGAACTGCTTGACCCATTTTC-3'
mMsh6	5'-CGTGCCTCCCAGTTCTTGTG-3'
	5'-GGACAGATTTCCCTTTCTTCCG-3'
mTBP	5'-TATCTGCTGGCGGTTTGGC-3'
	5'-TGAAATAGTGATGCTGGGCAC-3'
mT	5'-GCTTCAAGGAGCTAACTAACGAG-3'
	5'-CCAGCAAGAAAGAGTACATGGC-3'
mNanog	5'-ACCTGAGCTATAAGCAGGTTAAGAC-3'
	5'-GTGCTGAGCCCTTCTGAATCAGAC-3'
mRex1	5'-ATACCACTGACCAAAAAGCAGG-3'
	5'-GCCACTTGTCTTTGCCGTTTTC-3'
mOct4	5'-AGCATTGAGAACCGTGTGAGG-3'
	5'-AACCATACTCGAACCACATCC-3'
mFgf5	5'-AAGTAGCGCGACGTTTTCTTC-3'
	5'-CTGGAAACTGCTATGTTCCGAG-3'
mMeox1	5'-ACAGCCTTCACCAAGGAGC-3'
	5'-CCCCTTCACACGTTTCCA-3'
mGata6	5'-GGCTCTGTCCCTATGACTCC-3'
	5'-TGATGCCCCTACCCCTGAG-3'

Table 7: List of purified proteins from *Nanog* promoter

Accession No.	Gene Symbol	ESC/MEF
Q8BGQ7	Aars	< 2.0
Q5SWU9	Acaca	< 2.0
E9Q4Z2	Acacb	> 2.0
Q07417	Acads	< 2.0
Q9JIX8	Acin1	> 2.0
Q99KI0	Aco2	< 2.0
Q32MW3	Acot10	< 2.0
Q9R0X4	Acot9	< 2.0
P68134	Acta1	< 2.0
P62737	Acta2	< 2.0
P60710	Actb	< 2.0
P68033	Actc1	> 2.0
P63260	Actg1	< 2.0
P45376	Akr1b3	< 2.0
P07724	Alb	< 2.0
O55143	Atp2a2	< 2.0
P56480	Atp5b	> 2.0
Q9DB20	Atp5o	< 2.0
P28658	Atxn10	< 2.0
Q8CI61	Bag4	> 2.0
Q8K019	Bclaf1	> 2.0
P80317	Cct6a	NA
Q61390	Cct6b	> 2.0
P42932	Cct8	> 2.0
Q61081	Cdc37	< 2.0
Q6A068	Cdc5l	> 2.0
P18760	Cfl1	< 2.0
Q68FD5	Cltc	< 2.0
P08122	Col4a2	< 2.0
Q9QZS0	Col4a3	< 2.0
O55029	Copb2	< 2.0
Q9ERK4	Cse1l	> 2.0
P16381	D1Pas1	< 2.0
Q9JII5	Dazap1	> 2.0
Q501J6	Ddx17	< 2.0
Q9JIK5	Ddx21	> 2.0
Q8VDW0	Ddx39	> 2.0
Q9Z1N5	Ddx39b	> 2.0

Q62167	Ddx3x	> 2.0
Q61656	Ddx5	< 2.0
O35286	Dhx15	< 2.0
Q9D2G2	Dlst	< 2.0
Q9QYJ0	Dnaja2	< 2.0
Q9CWR8	Dnmt3l	< 2.0
P10126	Eef1a1	> 2.0
P62631	Eef1a2	< 2.0
O70251	Eef1b2	< 2.0
Q9D8N0	Eef1g	< 2.0
P58252	Eef2	< 2.0
O08810	Eftud2	> 2.0
P63242	Eif5a	> 2.0
Q8BGY2	Eif5a2	< 2.0
P17182	Eno1	> 2.0
P17183	Eno2	< 2.0
P21550	Eno3	< 2.0
Q9DCW4	Etfb	< 2.0
P62862	Fau	< 2.0
P35550	Fbl	> 2.0
Q80WS3	Fbll1	< 2.0
P39749	Fen1	> 2.0
P97807	Fh1	< 2.0
P30416	Fkbp4	> 2.0
P11276	Fn1	< 2.0
Q91WJ8	Fubp1	> 2.0
P97855	G3bp1	> 2.0
P16858	Gapdh	> 2.0
P03975	Gfap	< 2.0
Q61543	Glg1	< 2.0
P50247	Gm4737	NA
P05202	Got2	< 2.0
P24472	Gsta4	> 2.0
P10922	H1f0	< 2.0
Q8R1M2	H2afj	< 2.0

Accession No.	Gene Symbol	ESC/MEF
Q3THW5	H2afv	> 2.0
P27661	H2afx	> 2.0
P0C0S6	H2afz	> 2.0
P84244	H3f3a	< 2.0
P02301	H3f3c	NA
P01942	Hba-a1	< 2.0
Q6NVF4	Helb	> 2.0
P43275	Hist1h1a	> 2.0
P43276	Hist1h1b	< 2.0
P15864	Hist1h1c	< 2.0
P43277	Hist1h1d	< 2.0
P43274	Hist1h1e	< 2.0
Q07133	Hist1h1t	< 2.0
C0HKE4	Hist1h2ac	< 2.0
Q8CGP5	Hist1h2af	> 2.0
Q8CGP6	Hist1h2ah	> 2.0
Q8CGP7	Hist1h2ak	> 2.0
P70696	Hist1h2ba	< 2.0
P68433	Hist1h3g	> 2.0
Q6GSS7	Hist2h2aa1	< 2.0
Q64522	Hist2h2ab	< 2.0
Q64523	Hist2h2ac	< 2.0
Q64524	Hist2h2be	< 2.0
Q8BFU2	Hist3h2a	< 2.0
Q9D2U9	Hist3h2ba	> 2.0
Q8CGP0	Hist3h2bb-ps	< 2.0
P17095	Hmga1	< 2.0
P52927	Hmga2	< 2.0
P63158	Hmgb1	< 2.0
P30681	Hmgb2	> 2.0
O54879	Hmgb3	< 2.0
P09602	Hmgn2	< 2.0
O88569	Hnrnpa2b1	< 2.0
Q99020	Hnrnpab	> 2.0
Q9Z204	Hnrnpc	> 2.0
P61979	Hnrnpk	< 2.0
Q8R081	Hnrnpl	> 2.0
Q9D0E1	Hnrnpm	> 2.0
Q8VEK3	Hnrnpu	> 2.0

P07901	Hsp90aa1	> 2.0
P11499	Hsp90ab1	> 2.0
P08113	Hsp90b1	< 2.0
Q61696	Hspa1a	< 2.0
P17879	Hspa1b	< 2.0
P16627	Hspa1l	< 2.0
P17156	Hspa2	< 2.0
P63017	Hspa8	> 2.0
P14602	Hspb1	> 2.0
P63038	Hspd1	> 2.0
P54071	Idh2	< 2.0
P33175	Kif5a	> 2.0
Q61768	Kif5b	< 2.0
P28738	Kif5c	< 2.0
P04104	Krt1	< 2.0
P08730	Krt13	> 2.0
Q61781	Krt14	> 2.0
Q61414	Krt15	< 2.0
Q9QWL7	Krt17	< 2.0
P19001	Krt19	< 2.0
Q3TTY5	Krt2	< 2.0
Q8K0Y2	Krt33a	< 2.0
Q9D646	Krt34	< 2.0
Q497I4	Krt35	< 2.0
Q6IFX3	Krt40	< 2.0
Q6IFX2	Krt42	< 2.0
Q922U2	Krt5	< 2.0
Q9R0H5	Krt71	< 2.0
Q6NXH9	Krt73	< 2.0
Q6IFZ9	Krt74	< 2.0
Q3UV17	Krt76	< 2.0
Q6IFZ6	Krt77	< 2.0
Q8VED5	Krt79	< 2.0
P11679	Krt8	> 2.0

Accession No.	Gene Symbol	ESC/MEF
Q9ERE2	Krt81	< 2.0
Q99M74	Krt82	< 2.0
P97861	Krt86	< 2.0
P19137	Lama1	> 2.0
P02469	Lamb1	< 2.0
P48678	Lmna	< 2.0
P14733	Lmnb1	> 2.0
Q9D5S7	Lrguk	< 2.0
Q8BGS0	Mak16	> 2.0
P63085	Mapk1	< 2.0
Q3THS6	Mat2a	> 2.0
Q8K310	Matr3	< 2.0
Q99MR8	Mccc1	< 2.0
Q3ULD5	Mccc2	< 2.0
P54276	Msh6	> 2.0
Q7TPV4	Mybbp1a	> 2.0
Q61879	Myh10	< 2.0
Q8VDD5	Myh9	< 2.0
Q8BWZ3	Naa25	> 2.0
Q8BP47	Nars	< 2.0
P09405	Ncl	> 2.0
P10493	Nid1	< 2.0
Q9D6Z1	Nop56	> 2.0
Q6DFW4	Nop58	> 2.0
Q61937	Npm1	> 2.0
P09103	P4hb	< 2.0
P29341	Pabpc1	> 2.0
Q9DCL9	Paics	> 2.0
Q99LX0	Park7	> 2.0
P11103	Parp1	> 2.0
P24610	Pax3	< 2.0
P47239	Pax7	< 2.0
P60335	Pcbp1	> 2.0
Q91ZA3	Pcca	< 2.0
Q99MN9	Pccb	< 2.0
P17918	Pcna	> 2.0
Q05920	Pcx	< 2.0
P35486	Pdha1	< 2.0
Q8QZR7	Pdik1l	< 2.0
P70296	Pebp1	< 2.0

Q9DBJ1	Pgam1	> 2.0
O70250	Pgam2	> 2.0
P09411	Pgk1	< 2.0
P09041	Pgk2	< 2.0
O35129	Phb2	> 2.0
Q8R1K4	Phykpl	NA
P53657	Pklr	< 2.0
P52480	Pkm	NA
P17742	Ppia	> 2.0
Q9CR16	Ppid	> 2.0
Q99KR7	Ppif	> 2.0
Q9D868	Ppih	> 2.0
P35700	Prdx1	> 2.0
Q9JIF0	Prmt1	> 2.0
Q99JF8	Psip1	> 2.0
P62334	Psmc6	< 2.0
Q8VDM4	Psmd2	< 2.0
P17225	Ptbp1	> 2.0
Q8BHD7	Ptbp3	NA
Q9R0Q7	Ptges3	> 2.0
Q3UEB3	Puf60	< 2.0
Q922W5	Pycr1	< 2.0
Q922Q4	Pycr2	> 2.0
P61027	Rab10	< 2.0
Q9DD03	Rab13	< 2.0
Q8K386	Rab15	< 2.0
P62821	Rab1a	NA
Q9D1G1	Rab1b	< 2.0
Q6PHN9	Rab35	< 2.0
P55258	Rab8a	< 2.0
P61028	Rab8b	< 2.0
Q64012	Raly	< 2.0
P62827	Ran	> 2.0

Accession No.	Gene Symbol	ESC/MEF
Q61820	Rasl2-9	NA
Q60972	Rbbp4	> 2.0
Q60973	Rbbp7	< 2.0
O89086	Rbm3	< 2.0
Q8VH51	Rbm39	< 2.0
Q9WV02	Rbmx	< 2.0
Q91VM5	Rbmxl1	NA
Q9WUK4	Rfc2	> 2.0
Q99J62	Rfc4	> 2.0
P56716	Rp1	< 2.0
Q6ZWV3	Rpl10	< 2.0
P53026	Rpl10a	< 2.0
P86048	Rpl10l	< 2.0
Q9CXW4	Rpl11	< 2.0
P35979	Rpl12	< 2.0
P47963	Rpl13	< 2.0
Q9CR57	Rpl14	> 2.0
Q9CZM2	Rpl15	< 2.0
P62717	Rpl18a	> 2.0
P84099	Rpl19	< 2.0
P67984	Rpl22	< 2.0
P62751	Rpl23a	> 2.0
P61358	Rpl27	> 2.0
P14115	Rpl27a	< 2.0
P62911	Rpl32	< 2.0
P17932	Rpl32-ps	NA
P83882	Rpl36a	< 2.0
Q9D8E6	Rpl4	> 2.0
P47962	Rpl5	> 2.0
P47911	Rpl6	> 2.0
P14148	Rpl7	< 2.0
P12970	Rpl7a	> 2.0
P63325	Rps10	< 2.0
P62281	Rps11	< 2.0
P63323	Rps12	< 2.0
P62301	Rps13	< 2.0
P62264	Rps14	< 2.0
P62245	Rps15a	< 2.0
P14131	Rps16	< 2.0
P62270	Rps18	< 2.0
P25444	Rps2	< 2.0
P62267	Rps23	< 2.0
P62852	Rps25	< 2.0

P62855	Rps26	< 2.0
P62983	Rps27a	< 2.0
P62908	Rps3	< 2.0
P97351	Rps3a1	NA
P62702	Rps4x	< 2.0
P62754	Rps6	< 2.0
P62242	Rps8	< 2.0
Q6ZWN5	Rps9	< 2.0
Q9WTM5	Ruvbl2	> 2.0
Q9D1J3	Sarnp	< 2.0
Q8K2B3	Sdha	< 2.0
P19324	Serpinh1	< 2.0
Q99NB9	Sf3b1	< 2.0
Q921M3	Sf3b3	> 2.0
Q62407	Speg	< 2.0
Q62261	Sptbn1	NA
Q6PDM2	Srsf1	> 2.0
P84104	Srsf3	> 2.0
Q8VE97	Srsf4	> 2.0
O35326	Srsf5	< 2.0
Q3TWW8	Srsf6	> 2.0
Q8BL97	Srsf7	> 2.0
Q80ZW0	Stk35	> 2.0
Q69ZM6	Stk36	< 2.0
Q9WUM5	Suclg1	< 2.0
Q920B9	Supt16	NA
Q7TMK9	Syncrip	< 2.0
Q9WVA4	Tagln2	< 2.0
Q8C4J7	Tbl3	> 2.0
P10711	Tcea1	NA
Q9QVN7	Tcea2	< 2.0
O08784	Tcof1	> 2.0
Q8K3F7	Tdh	< 2.0
Q569Z6	Thrap3	> 2.0
P40142	Tkt	> 2.0
Q62318	Trim28	> 2.0
Q923J1	Trpm7	< 2.0

Accession No.	Gene Symbol	ESC/MEF
Q9WTR1	Trpv2	< 2.0
P68369	Tuba1a	< 2.0
P05213	Tuba1b	> 2.0
P68373	Tuba1c	< 2.0
P68368	Tuba4a	> 2.0
A2AQ07	Tubb1	< 2.0
Q7TMM9	Tubb2a	< 2.0
Q9CWF2	Tubb2b	< 2.0
Q9ERD7	Tubb3	< 2.0
Q9D6F9	Tubb4a	NA
P68372	Tubb4b	NA
P99024	Tubb5	< 2.0
Q922F4	Tubb6	< 2.0
Q8BFR5	Tufm	< 2.0
P62984	Uba52	< 2.0
P0CG49	Ubb	> 2.0
P0CG50	Ubc	< 2.0
Q6WKZ8	Ubr2	< 2.0
P52624	Upp1	> 2.0
Q8CGR7	Upp2	< 2.0
Q01853	Vcp	< 2.0
P62960	Ybx1	> 2.0
P62259	Ywhae	> 2.0

10. ACKNOWLEDGEMENT

I am extremely grateful to Professor Koji Hisatake for believing and accepting me to the Laboratory of Gene Regulation as a graduate student. I am also deeply indebted to Professor Ken Nishimura for training me, relentlessly spending time in research discussions, and continuously guiding me throughout my graduation research. Especially research discussions with the professor(s) were productive and improved my research aptitude and perception. It is also not enough to thank Professor Fukuda and Professor Hayashi for their support during research discussions and for helping my research progress.

I am grateful for the explicit technical support received from NIMS Molecule and Material synthesis platform for LC/MS/MS analysis; Dr. Akihiro Kuno for making a visually pleasing violin plot for up-regulated genes; Dr. Yuji Yamazaki in setting up FACS and flow cytometry experiments; Mr. Hamzah for the EB5/ReKO reporter cell lines; Mr. Takumi Kishimoto for supporting reprogramming experiments and Ms. Tomoko Nishimura for her experimental assistance.

I am also deeply indebted to the committee for reviewing and improving my research with their valuable suggestions from time to time.

Special thanks to my lab member Dr. Anh (Hoang Anh) for being by my side as a big brother for your advice and for lifting my spirit in times of need. My life here in Tsukuba, wouldn't have been easier if it is not for Dr. Aizawa (Shiho Aizawa). She always helped me to settle down in this new town. I always remember the talks we had about the future, and graduation while sipping coffee at Starbucks. Then, I should thank Phoung Linh (Dr.Linh) for always checking on me.

I also extend my sincere gratitude to my other lab member; Mr. Yuya Sekiguchi, the very first person I met in Japan. Thank you, Yuya for being my tutor and guiding me through the initial procedures. I thank all my other lab members, Dr. Yen, Ms. Miho

11. REFERENCES

1. Aizawa, S., Nishimura, K., Mondejar, G.S., Kumar, A., Bui, P.L., Tran, Y.T.H., Kuno, A., Muratani, M., Kobayashi, S., Nabekura, T., et al. (2022). Early reactivation of clustered genes on the inactive X chromosome during somatic cell reprogramming. Stem Cell Reports *17*, 53–67.

2. Basu, A., and Tiwari, V.K. (2021). Epigenetic reprogramming of cell identity: lessons from development for regenerative medicine. Clin Epigenetics *13*, 144.

3. Bowles, J., Schepers, G., and Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol *227*, 239–255.

4. Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., and Sherlock, G. (2004). GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics *20*, 3710–3715.

5. Bui, P.L., Nishimura, K., Seminario Mondejar, G., Kumar, A., Aizawa, S., Murano, K., Nagata, K., Hayashi, Y., Fukuda, A., Onuma, Y., et al. (2019). Template Activating Factor-I α regulates retroviral silencing during reprogramming. Cell Rep *29*, 1909-1922.e5.

6. 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.

7. Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

8. Chen, K.G., Mallon, B.S., Johnson, K.R., Hamilton, R.S., McKay, R.D.G., and Robey, P.G. (2014). Developmental insights from early mammalian embryos and core signaling pathways that influence human pluripotent cell growth and differentiation. Stem Cell Res *12*, 610–621.

9. Chen, X., Liou, Y.-C., Ng, H.-H., Chen, X.I., and Fang, F. (2008). Zfp143 regulates Nanog through modulation of Oct4 binding. Stem Cells *26*, 2759–2767.

10. Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative binding of transcription factors orchestrates reprogramming. Cell *168*, 442-459.e20.

11. Cyranoski, D. (2018a). 'Reprogrammed' stem cells implanted into patient with Parkinson's disease. Nature https://doi.org/10.1038/d41586-018-07407-9.

12. Cyranoski, D. (2018b). 'Reprogrammed' stem cells approved to mend human hearts for the first time. Nature *557*, 619–620.

13. Cyranoski, D. (2019). Woman is first to receive cornea made from 'reprogrammed' stem cells. Nature https://doi.org/10.1038/D41586-019-02597-2.

14. Dailey, L., Yuan, H., and Basilico, C. (1994). Interaction between a novel F9-specific factor and octamer-binding proteins is required for cell-type-restricted activity of the fibroblast growth factor 4 enhancer. Mol Cell Biol *14*, 7758–7769.

15. Déjardin, J., and Kingston, R.E. (2009). Purification of proteins associated with specific genomic loci. Cell *136*, 175–186.

16. Ding, J., Xu, H., Faiola, F., Ma'Ayan, A., and Wang, J. (2011). Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res *22*, 155–167.

17. Ding, N., Bonham, E.M., Hannon, B.E., Amick, T.R., Baylin, S.B., and O'Hagan, H.M. (2016). Mismatch repair proteins recruit DNA methyltransferase 1 to sites of oxidative DNA damage. J Mol Cell Biol *8*, 244–254.

18. Du, X., and Xiao, R. (2020). An emerging role of chromatin-interacting RNA-binding proteins in transcription regulation. Essays Biochem *64*, 907–918.

19. Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science *344*, 1156–1160.

20. Elman, J.S., Ni, T.K., Mengwasser, K.E., Jin, D., Wronski, A., Elledge, S.J., and Kuperwasser, C. (2019). Identification of FUBP1 as a long tail cancer driver and widespread regulator of tumor suppressor and oncogene alternative splicing. Cell Rep *28*, 3435-3449.e5.

21. Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351–359.

22. Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. Cell *147*, 120–131. https://doi.org/10.1016/j.cell.2011.08.038.

23. Fujita, T., and Fujii, H. (2013). Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated

chromatin immunoprecipitation (enChIP) using CRISPR. Biochem Biophys Res Commun *439*, 132–136.

24. 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 B Phys Biol Sci *85*, 348–362.

25. Hackett, J.A., and Surani, M.A. (2014). Regulatory principles of pluripotency: from the ground state up. Cell Stem Cell *15*, 416–430.

26. Hoang, V.T., Verma, D., Godavarthy, P.S., Llavona, P., Steiner, M., Gerlach, K., Michels, B.E., Bohnenberger, H., Wachter, A., Oellerich, T., et al. (2019). The transcriptional regulator FUBP1 influences disease outcome in murine and human myeloid leukemia. Leukemia *33*, 1700–1712.

27. Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev *23*, 837–848.

28. Huang, X., Park, K.-M., Gontarz, P., Zhang, B., Pan, J., McKenzie, Z., Fischer, L.A., Dong, C., Dietmann, S., Xing, X., et al. (2021). OCT4 cooperates with distinct ATP-dependent chromatin remodelers in naïve and primed pluripotent states in human. Nat Commun *12*, 5123.

29. Hoshino, A., and Fujii, H. (2009). Insertional chromatin immunoprecipitation: A method for isolating specific genomic regions. J Biosci Bioeng *108*, 446–449.

30. Klemm, J.D., and Pabo, C.O. (1996). Oct-1 POU domain-DNA interactions: Cooperative binding of isolated subdomains and effects of covalent linkage. Genes Dev *10*, 27–36.

31. Knaupp, A.S., Mohenska, M., Larcombe, M.R., Ford, E., Lim, S.M., Wong, K., Chen, J., Firas, J., Huang, C., Liu, X., et al. (2020). TINC— A method to dissect regulatory complexes at single-locus resolution— reveals an extensive protein complex at the Nanog promoter. Stem Cell Reports *15*, 1246–1259.

32. Krause, M.N., Sancho-Martinez, I., and Izpisua Belmonte, J.C. (2016). Understanding the molecular mechanisms of reprogramming. Biochem Biophys Res Commun *473*, 693–697.

33. Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition

of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development *134*, 2895–2902.

34. Kupatt, C., Windisch, A., Moretti, A., Wolf, E., Wurst, W., and Walter, M.C. (2021). Genome editing for Duchenne muscular dystrophy: a glimpse of the future? Gene Ther *28*, 542–548.

35. Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S., Suemori, H., Nakatsuji, N., and Tada, T. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog Gene Expression . Mol Cell Biol *25*, 2475–2485.

36. Li, H., Wang, Z., Zhou, X., Cheng, Y., Xie, Z., Manley, J.L., and Feng, Y. (2013). Far upstream element-binding protein 1 and RNA secondary structure both mediate second-step splicing repression. Proc Natl Acad Sci U S A *110*, E2687–E2695.

37. Lin, T., Chao, C., Saito, S., Mazur, S.J., Murphy, M.E., Appella, E., and Xu, Y. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. Nat Cell Biol *7*, 165–171.

38. Liu, W., Stein, P., Cheng, X., Yang, W., Shao, N.-Y., Morrisey, E.E., Schultz, R.M., and You, J. (2014). BRD4 regulates Nanog expression in mouse embryonic stem cells and preimplantation embryos. Cell Death Differ *21*, 1950–1960.

39. Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Cao, H., Ni, M., Liu, Y., Gu, Z., et al. (2017). In situ capture of chromatin interactions by biotinylated dCas9. Cell *170*, 1028-1043.e19.

40. Mao, A.S., and Mooney, D.J. (2015). Regenerative medicine: Current therapies and future directions. Proc Natl Acad Sci U S A *112*, 14452–14459.

41. Mashiko, D., Fujihara, Y., Satouh, Y., Miyata, H., Isotani, A., and Ikawa, M. (2013). Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. Sci Rep *3*, 3355.

42. Metcalf, D. (1991). The leukemia inhibitory factor (LIF). The International Journal of Cell Cloning *9*, 95–108.

43. 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.

44. Miyanari, Y., and Torres-Padilla, M.-E. (2012). Control of ground-state pluripotency by allelic regulation of Nanog. Nature *483*, 470–473.

45. Nagoshi, N., Tsuji, O., Nakamura, M., and Okano, H. (2019). Cell therapy for spinal cord injury using induced pluripotent stem cells. Regen Ther *11*, 75–80.

46. Nakanishi, M., and Otsu, M. (2012). Development of Sendai Virus vectors and their potential applications in gene therapy and regenerative medicine. Curr Gene Ther *12*, 410–416.

47. Navarro, P., Festuccia, N., Colby, D., Gagliardi, A., Mullin, N.P., Zhang, W., Karwacki-Neisius, V., Osorno, R., Kelly, D., Robertson, M., et al. (2012). OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. EMBO J *31*, 4547–4562.

48. Ng, H.-H., and Surani, M.A. (2011). The transcriptional and signalling networks of pluripotency. Nat Cell Biol *13*, 490–496.

49. Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, 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.

50. Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development *136*, 3215–3222.

51. Nishimura, K., Sano, M., Ohtaka, M., Furuta, B., Umemura, Y., Nakajima, Y., Ikehara, Y., Kobayashi, T., Segawa, H., Takayasu, S., et al. (2011). Development of defective and persistent Sendai virus vector: a unique gene delivery/expression system ideal for cell reprogramming. J Biol Chem *286*, 4760–4771.

52. Nishimura, K., Kato, T., Chen, C., Oinam, L., Shiomitsu, E., Ayakawa, D., Ohtaka, M., Fukuda, A., Nakanishi, M., and Hisatake, K. (2014). Manipulation of KLF4 expression generates iPSCs paused at successive stages of reprogramming. Stem Cell Reports *3*, 915–929.

53. Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev *12*, 2048–2060.

54. Niwa, H., Miyazaki, J.I., 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.

55. Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118–122.

56. Ochiai, H., Sugawara, T., Sakuma, T., and Yamamoto, T. (2014). Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. Sci Rep *4*, 7125.

57. 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.

58. Palmieri, S.L., Peter, W., Hess, H., and Schöler, H.R. (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the firsttwo extraembryonic cell lineages involved in implantation. Dev Biol *166*, 259–267.

59. Pan, G.J., Chang, Z.Y., Schöler, H.R., and Pei, D. (2002). Stem cell pluripotency and transcription factor Oct4. Cell Res *12*, 321–329.

60. 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.

61. Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W.A. (2012). Psip1/Ledgf p52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing. PLoS Genet *8*, e1002717.

62. Rabenhorst, U., Thalheimer, F.B., Gerlach, K., Kijonka, M., Böhm, S., Krause, D.S., Vauti, F., Arnold, H.H., Schroeder, T., Schnütgen, F., et al. (2015). Single-Stranded DNA-Binding Transcriptional Regulator FUBP1 Is Essential for Fetal and Adult Hematopoietic Stem Cell Self-Renewal. Cell Rep *11*, 1847–1855.

63. Ren, Y., Huo, Y., Li, W., He, M., Liu, S., Yang, J., Zhao, H., Xu, L., Guo, Y., Si, Y., et al. (2021). A global screening identifies chromatin-enriched RNA-binding proteins and the transcriptional regulatory activity of QKI5 during monocytic differentiation. Genome Biol *22*, 290.

64. Rodda, D.J., Chew, J.-L., Lim, L.-H., Loh, Y.-H., Wang, B., Ng, H.-H., and Robson, P. (2005). Transcriptional regulation of Nanog by OCT4 and SOX2. J Biol Chem *280*, 24731–24737.

65. Ross, C., and Boroviak, T.E. (2020). Origin and function of the yolk sac in primate embryogenesis. Nat Commun *11*, 3760.

66. Schöler, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N., and Gruss, P. (1989). Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. EMBO J *8*, 2551–2557.

67. Scholer, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N., and Gruss, P. (1989). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. EMBO J *8*, 2543–2550.

68. Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990). New type of POU domain in germ line-specific protein Oct-4. Nature *344*, 435–439.

69. Seki, Y., Kurisaki, A., Watanabe-Susaki, K., Nakajima, Y., Nakanishi, M., Arai, Y., Shiota, K., Sugino, H., and Asashima, M. (2010). TIF1beta regulates the pluripotency of embryonic stem cells in a phosphorylation-dependent manner. Proc Natl Acad Sci U S A *107*, 10926–10931.

70. Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.

71. Singh, D.K., Gholamalamdari, O., Jadaliha, M., Ling Li, X., Lin, Y.-C., Zhang, Y., Guang, S., Hashemikhabir, S., Tiwari, S., Zhu, Y.J., et al. (2017). PSIP1/p75 promotes tumorigenicity in breast cancer cells by promoting the transcription of cell cycle genes. Carcinogenesis *38*, 966–975.

72. Sono, T., Akiyama, H., Miura, S., Deng, J.M., Shukunami, C., Hiraki, Y., Tsushima, Y., Azuma, Y., Behringer, R.R., and Matsuda, S. (2018). THRAP3 interacts with and inhibits the transcriptional activity of SOX9 during chondrogenesis. J Bone Miner Metab *36*, 410–419.

73. Stavridis, M.P., Simon Lunn, J., Collins, B.J., and Storey, K.G. (2007). A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. Development *134*, 2889–2894.

74. Stirparo, G.G., Kurowski, A., Yanagida, A., Bates, L.E., Strawbridge, S.E., Hladkou, S., Stuart, H.T., Boroviak, T.E., Silva, J.C.R., and Nichols, J. (2021). OCT4 induces embryonic pluripotency via STAT3 signaling and metabolic mechanisms. Proc Natl Acad Sci U S A *118*, e2008890118.

75. 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.

76. 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.

77. van den Berg, D.L.C., 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.

78. Vasseur, S., Afzal, S., Tardivel-Lacombe, J., Park, D.S., Iovanna, J.L., and Mak, T.W. (2009). DJ-1/PARK7 is an important mediator of hypoxia-induced cellular responses. Proc Natl Acad Sci U S A *106*, 1111–1116.

79. Vohhodina, J., Barros, E.M., Savage, A.L., Liberante, F.G., Manti, L., Bankhead, P., Cosgrove, N., Madden, A.F., Harkin, D.P., and Savage, K.I. (2017). The RNA processing factors THRAP3 and BCLAF1 promote the DNA damage response through selective mRNA splicing and nuclear export. Nucleic Acids Res *45*, 12816–12833.

80. Wang, A.S., Chen, L.C., Wu, R.A., Hao, Y., McSwiggen, D.T., Heckert, A.B., Richardson, C.D., Gowen, B.G., Kazane, K.R., Vu, J.T., et al. (2020). The histone chaperone FACT induces Cas9 multi-turnover behavior and modifies genome manipulation in human cells. Mol Cell *79*, 221-233.e5.

81. 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.

82. Wesely, J., Steiner, M., Schnütgen, F., Kaulich, M., Rieger, M.A., and Zörnig, M. (2017). Delayed mesoderm and erythroid differentiation of murine embryonic stem cells in the absence of the transcriptional regulator FUBP1. Stem Cells Int *20*

83. Wierer, M., and Mann, M. (2016). Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. Hum Mol Genet *25*, R106–R114.

84. Williams, D.C., Cai, M., and Clore, G.M. (2004). Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kda oct1·sox2· hoxb1-dna ternary transcription factor complex. Journal of Biological Chemistry *279*, 1449–1457.

85. Wu, D.Y., and Yao, Z. (2006). Functional analysis of two Sp1/Sp3 binding sites in murine Nanog gene promoter. Cell Res *16*, 319–322.

86. Wu, Q., Chen, X., Zhang, J., Loh, Y.-H., Low, T.-Y., Zhang, W., Zhang, W., Sze, S.-K., Lim, B., and Ng, H.-H. (2006). Sall4 interacts with Nanog and co-occupies *Nanog* genomic sites in embryonic stem cells. J Biol Chem *281*, 24090–24094.

87. Xiao, R., Chen, J.Y., Liang, Z., Luo, D., Chen, G., Lu, Z.J., Chen, Y., Zhou, B., Li, H., Du, X., et al. (2019). Pervasive chromatin-RNA binding protein interactions enable rna-based regulation of transcription. Cell *178*, 107-121.e18.

88. Yarrington, R.M., Verma, S., Schwartz, S., Trautman, J.K., and Carroll, D. (2018). Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. Proc Natl Acad Sci U S A *115*, 9351–9358.

89. Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

90. Young, R.A. (2011). Control of the Embryonic Stem Cell State. Cell *144*, 940–954.

91. Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev *9*, 2635–2645.

92. Zhang, J., Hua, C., Zhang, Y., Wei, P., Tu, Y., and Wei, T. (2020). KAP1-associated transcriptional inhibitory complex regulates C2C12 myoblasts differentiation and mitochondrial biogenesis via miR-133a repression. Cell Death Dis *11*, 732.

93. Zhang, S., Xu, Y., Xie, C., Ren, L., Wu, G., Yang, M., Wu, X., Tang, M., Hu, Y., Li, Z., et al. (2021). RNF219/α-Catenin/LGALS3 axis promotes hepatocellular carcinoma bone metastasis and associated skeletal complications. Advanced Science *8*, 2001961.