筑　波　大　学

博　士　（医学）　学　位　論　文

ANALYSIS ON TRANSCRIPTIONAL REGULATION IN THE EARLY PHASE OF SOMATIC CELL REPROGRAMMING

(体細胞リプログラミング早期での転写制御の解析)

2022

筑波大学大学院博士課程人間総合科学研究科

LE PHUONG HOANG ANH

TABLE OF CONTENTS

Abstract

The generation of induced pluripotent stem cells (iPSCs) is an innovative approach to generate pluripotent stem cells from somatic cells by introducing reprogramming factors: KLF4, OCT4, SOX4, and c-MYC (hereafter referred to as KOSM). The development of iPSCs led to a massive promise in biomedical applications such as modeling diseases, developing cells for regenerative medicine, and screening for a new drug. Various studies have been conducted to understand multiple aspects of the iPSC generation mechanism and improve its efficiency. However, reprogramming remains an inefficient and time-consuming process for clinical application. Furthermore, detailed mechanisms are not yet fully understood to enable rapid and highly efficient means to generate iPSCs from somatic cells.

Cells significantly increase cell proliferation in the early stage of reprogramming mesenchymal cells such as mouse embryonic fibroblasts. In addition, they undergo marked morphological changes while downregulating mesenchyme-associated genes that presumably define the mesenchymal phenotype. However, the importance of these downregulated genes to reprogramming needs to be better defined. Therefore, this study aims to identify functionally relevant transcription factors that characterize the mesenchymal phenotype and should be immediately downregulated upon reprogramming by KOSM. Moreover, mechanical analyses of the identified gene were performed to understand its functional role during the reprogramming of somatic cells to iPSCs.

Mesenchyme-transcription factors that downregulate during reprogramming were screened to identify 8 candidate genes that may determine the mesenchymal cell identity. The expression of the candidates before reprogramming detrimentally affected the cell survival during reprogramming. Therefore, a retroviral vector harboring human cytomegalovirus upstream open reading frame 2 (uORF2) was generated to reduce their translation and mitigate their negative effect on cell survival. Expression of the selected transcription factors from this vector showed that odd-skipped related 2 (*Osr2*) should be downregulated for efficient reprogramming. Analysis of the epithelial-mesenchymal transition (EMT) model cells identified *Osr2* as a new regulator of EMT and upregulates TGF-β signaling. In the reprogramming model, *Osr2* downregulation resulted in the suppression of TGF-β signaling and allowed activation of Wnt signaling.

In conclusion, the study demonstrates that downregulation of a new EMT regulator, OSR2, reduces TGF-β signaling and activates Wnt signaling, leading to more efficient mesenchymal-epithelial transition (MET) and acquisition of pluripotency.

Abbreviations

Cdh1	Cadherin 1
Cdh2	Cadherin 2
DAPI	4′,6-diamidino-2-phenylindole
DMEM	Dulbecco's modified eagle's medium
dNTP	deoxyribonucleotide triphosphate
DTT	Dithiothreitol
Ebf1	Early B-Cell Factor 1
Ebf3	Early B-Cell Factor 1
EBP1	ErbB3-binding protein 1
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced Green Fluorescent Protein
EMSAs	Electrophoretic mobility shift assays
EMT	Epithelial to Mesenchymal Transition
EpCAM	Epithelial Cell Adhesion Molecule
ESCs	Embryonic stem cells
FBS	Fetal bovine serum
FN1	Fibronectin 1
iPSCs	Induced Pluripotent Stem cells
IRES	Internal ribosome entry sites
KOSM	Klf4, Oct4, Sox2, c-Myc

LIF	Leukemia Inhibitory Factor
MEF	Mouse Embryonic Fibroblast
Meox1	Mesenchyme Homeobox 1
Meox2	Mesenchyme Homeobox 2
MET	Mesenchymal to Epithelial Transition
NMuMG	Normal murine mammary gland
Ocln	Occludin
Oct4	Octamer-binding transcription factor 4 phosphoprotein
Osr2	Odd-Skipped Related 2
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PLAT-E	Platinum-E
Prrx1	Pair related homeobox 1
PSCs	Pluripotent stem cells
qPCR	Quantitative polymerase chain reaction
qRT-PCR	Quantitative reverse transcriptase-polymerase chain reaction
RT-PCR	Reverse transcriptase PCR
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SeVdp	defective and persistent Sendai virus vector
shRNA	short hairpin RNA

Smarcd3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 3
Snai1	Snail Family Transcriptional Repressor 1
Snai2	Snail Family Transcriptional Repressor 2
Sox2	Sex determining region Y-box 2
TBP	TATA-binding protein
TGF-β	transforming Growth Factor Beta
Vim	Vimentin
Zeb1	Zinc Finger E-Box Binding Homeobox 1
Zeb2	Zinc Finger E-Box Binding Homeobox 2
Zic1	Zinc family member 1

CHAPTER 1: INTRODUCTION

1. From Cells to induced Pluripotent Stem Cells

Somatic cells committed to a specific cell lineage can be reprogrammed into an embryonic-like pluripotent state to become induced pluripotent stem cells (iPSCs) [Takahashi and Yamanaka, 2006]. iPSCs may be differentiated into any type of human cells in an unlimited amount for using regenerative medicine. This part of the thesis will describe the concept of the cells, the stem cells, and the original ideas of cell reprogramming, followed by the current knowledge of iPSCs generation, the remaining questions related to reprogramming, and the remaining unknown mechanisms of reprogramming.

1.1. Cell classification

Cells are the fundamental component of all living organisms, composed of many types of cells with specialized functions such as making structures in the body and producing energy from food [Alberts, 2017]. Although cells all have similar elemental compositions, they vary enormously in appearance and function. The genetic makeup in an organism's genome directs the production of various distinct cell types, expressing different sets of genes in a precise and intricate pattern. Except for germ cells, gametocytes, and undifferentiated cells such as precursors, all the cells in an organism are composed of differentiated somatic cells. Somatic cells make up all the internal organs and tissues and function to maintain the organism's homeostasis [Urry et al., 2020]. In general, most human somatic cells are capable of only a limited number of cell divisions before they become senescent [Hayflick and Moorhead, 1961]. Terminally differentiated cells in an organism are replaced by cells generated from a stock of proliferating precursor cells originally derived from stem cells.

1.2. Stem cells and pluripotency

Stem cells possess the ability to produce daughter cells that are identical to their mother cell (self-renewal) and to differentiate into various types of cells (potency), which originate from a single cell (clonality) [Weissman et al., 2001; Smith, 2001; Alberts et al., 2019]. Stem cells occur naturally both in embryos and an adult body. Stem cells may be divided into five categories by their degree of potency: totipotent, pluripotent, multipotent, oligopotent, and unipotent. The most undifferentiated cells are totipotent cells found in early development. Totipotent cells, such as cells derived from the fertilized oocyte and the cells of the first two divisions, can differentiate into both embryonic and extra-embryonic cells and thus are able to generate the embryo proper and extra-embryonic tissues [Weissman, 2000; Singh et al.,

2016; Smith, 2006]. Similar to totipotent stem cells, pluripotent stem cells (PSCs) can generate all types of cells in three primary germ cell layers. However, PSCs lack the potential to contribute to the extra-embryonic cells such as the placenta.

Embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) are typical examples of PSCs [Rosales and Mullen, 2010]. ESCs are derived from the inner cell mass in the blastocyte [Kawase et al., 2000]. In 2012, Sir John Gurdon and Shinya Yamanaka received the Nobel Prize for reprogramming adult cells into PSCs. PSCs generated from somatic cells by introducing transcription factors are called induced pluripotent stem cells [Takahashi and Yamanaka, 2006]. Multipotent stem cells (MSCs) have a more limited spectrum of differentiation than PSCs. They can differentiate into various specialized cell lineages found in a specific tissue or organ. Well-known multipotent cells such as MSCs, which can differentiate into different types of cells, are present in tissues such as adipose tissue, umbilical cord blood, bone marrow, bone, and peripheral blood [Caplan, 1991; Zakrzewski et al., 2019; Smith, 2006]. Oligopotent stem cells can differentiate into limited types of cells. An example of oligopotent stem cells includes hematopoietic stem cells, which can differentiate into lymphoid and myeloid lineages [Moore, 2014; Smith, 2006]. Unipotent stem cells, such as muscle stem cells, produce only cells in a single lineage [Smith, 2006; Biehl and Russell, 2009].

Understanding how cells work and perform their functions improves our knowledge of how living organisms are generated and also helps improve human life and the natural world. For example, researchers working in medical science could develop effective medicines, new vaccines, and new sources of food. In addition, stem cells, which differentiate into various cell types, may be used to cure a disease. For example, a medical therapy using stem cells is established using hematopoietic stem cell transplantation [Müller et al., 2016]. Therefore, stem cells have a vast potential for application in medical fields such as regenerative medicine, disease modeling, and drug development.

1.3. Embryonic Stem Cells versus Adult Stem Cells

In mammals, stem cells can be divided into embryonic and adult stem cells [Jukes et al., 2008; Prochazkova et al., 2015]. Both types of stem cells have the shared capabilities: to undergo unlimited self-renewal and turn into differentiated cell types in the body. Adult stem cells are undifferentiated cells found throughout the body. They can be found in most adult tissues in small amounts. Adult stem cells play a role in the growth, maintenance, and

regeneration of tissues and repairing damaged tissues. They are multipotent and commonly contribute to producing cells in a specific tissue.

In contrast, ESCs have a higher differentiation capacity than adult stem cells. Therefore, they are categorized as PSCs. In other words, ESCs are able to differentiate into all types of cells in an organism. Embryonic stem cells are derived from the undifferentiated inner mass cells within the blastocyst. ESCs are responsible for embryonic and fetal development and growth. They do not exist in the adult body but can be identified as a transient cell population present in early embryos.

Besides ESCs, there are other sources for PSCs that can be used for research and medicine, such as primordial germ cells [Shamblott et al., 1998], cells derived from the fusion between somatic and pluripotent cells [Cibelli et al., 1998; Kawase et al., 2000], cells obtained by single-cell nuclear transfer (SCNT) [Gurdon, 1960], and the latest finding - iPSCs. All pluripotent stem cells possess resemblances in their ability, function, morphology, and similar transcription and epigenetic profiles. Moreover, pluripotent stem cells *in vivo* can differentiate into three germ layers in the adult body (endoderm, mesoderm, and ectoderm). By teratoma formation and a chimeric assay, the developmental potential of stem cells could be determined; they are considered the gold standard criteria for assessing the pluripotency of stem cells [Thomson, 1998; Tarkowski, 1961].

2. The remarkable studies led to the emergence of induced pluripotent stem cells

iPSCs are derived from various types of somatic cells, which are more readily available and free from ethical issues. iPSCs are functionally similar to ESCs, are capable of unlimited proliferation, and have the potential to differentiate into all types of cells in the body. Especially, iPSCs are less likely to be rejected by the immune system due to their genetic relatedness to the recipient. Therefore, iPSCs are at the center of intense research and application in regenerative medicine. The below discoveries that are briefly described show that cell fate can be changed and led to the establishment of iPSCs in 2006 [Takahashi and Yamanaka, 2006].

2.1. Somatic cell nuclear transfer approach

In 1955 King and Briggs developed a technique to transplant the nucleus from one cell to another enucleated cell in the frog species *Rana pipiens* [King and Briggs, 1955]. It was the first experiment that showed nuclear transplantation. Their results showed that nuclei of the somatic cell contain all genetic information, which can be reprogramed to an alternate cell fate. However, their ability for differentiation was restricted depending on the source of nuclei. They concluded that differentiating cells often undergo stable changes, which limit their developmental capacity. The majority of nuclei taken from early embryonic cells have not experienced any change of this kind [King and Briggs, 1955]. Influenced by King and Briggs's discoveries, John Gurdon investigated the transfer of intestine nuclei, considered differentiated cells, into enucleated eggs that then generate a fully adult amphibian. John Gurdon optimized the technique to implant nuclear transfer of eggs, which developed to a tadpole stage under optimal conditions. Then, he further proved that those eggs undergo development, and several eggs can fully develop into adult frogs [Gurdon, 1960, 1962].

King, Briggs, and Gurdon's discoveries are the very first evidence indicating that somatic cell nuclei containing all required genetic makeup could be reprogrammed to a pluripotent state under the optimal conditions where the right signaling factors and environments are present. Their pioneering findings opened a brand-new field of studying nuclear transplantation and cloning. Subsequent studies showed that nuclei from fully differentiated cells in mammals could also be efficiently reprogrammed to produce offspring and develop into adults, as demonstrated by cloned sheep [Wilmut et al., 1997] and mice [Hochedlinger and Jaenisch, 2002]. Moreover, conversion of the potency of somatic cells to that of pluripotent cells was successfully conducted by fusion with ESCs *in vitro* [Tada et al., 2001]. This study demonstrated that nuclei contain all genetic information and can be experimentally reprogrammed into a pluripotent state. However, the limitation of this SNCT experiment is its low efficiency and technical difficulties.

2.2. Forced cell fate change approach

Taylor and Jones first found the forced cell fate change by chemical reagents in 1979 [Taylor and Jones, 1979]. They discovered that 5-azacytidine, a nucleoside analog that inhibits DNA methylation, could transform fibroblasts into myocytes, chondrocytes, and adipocytes. This finding suggested that somatic cells could be trans-differentiation under the epigenetic

approach [Taylor and Jones, 1979]. This result suggested the possibility of altering the cell identity by changing its epigenetic status and opened a new field of biological studies. In 1986, based on the fact that the addition of 5-azacytidine converts fibroblasts into highly similar myoblasts i*n vitro*, studies identified genes that are expressed specifically in myoblasts. When they overexpressed one of the genes, namely *MyoD*, in fibroblasts, they observed that up to 50% of the cell become myoblasts (muscle cells). This finding represented the birth of the idea that critical transcription factors can force cells to change their identity [Lassar et al., 1986; RL et al., 1987]. This discovery prompted searches for other transcription factors that are critical for defining the cell identity; for example, conversion of B cells to macrophages by expression of transcription factor C/EBPα and β [Xie et al., 2004] and generation of chondrocytes from fibroblasts by transcription factor SOX9 [Murakami et al., 2000]. These data indicate that differentiated cells can be forced to change their cell identity by forced expression of a cell type- or lineage-specific transcription factor.

2.3. Isolation and culture Embryonic Stem Cells *in vitro* approach

ESCs can be cultured indefinitely *in vitro* under specific conditions and maintain their pluripotent ability for a long time [Prochazkova et al., 2015]. Due to their capacity for pluripotency, ESCs can contribute to all the different tissues in the body and have more potential for differentiation than adult stem cells. ESCs are ideal sources for cells that cannot be easily obtained from adult tissues such as neural cells. In basic biological research, ESCs can be utilized for studying biological processes during development [Rizzoti et al., 2016]. In biomedical research and clinical application, ESCs could be used for modeling diseases [Hanna et al., 2007; Urbach, 2004], the discovery of new drugs [Mackay-Sim et al., 2011], developing new diagnostic and biomarkers [Rubin and Haston, 2011] as well as regenerative medicine that includes transplantation and cell replacement therapies for numerous diseases such as spinal cord injuries, diabetes, and Parkinson disease [Hescheler, 2019; Müller et al., 2016].

Mouse ESCs (mESCs) are maintained on a feeder layer of mouse embryonic fibroblasts *in vitro* [Sokol, 2011]. Derived from the STO cell line, which has been transformed with neomycin resistance and murine leukemia inhibitory factor (LIF) genes, SNL can be used as the feeder for growing ESCs and iPSCs *in vitro*. On the other hand, in the absence of a feeder cell layer, mESCs can be grown in a medium supplemented with

recombinant LIF and BMP4 [Keller, 2005]. LIF functions through gp130, and the LIF signal is transduced by the Janus kinase/signal transducer and activators of transcription (JAK/STAT), which activates STAT3 and its downstream targets such as c-MYC [Cartwright et al., 2005]. While BMP4 has been linked to the mesenchymal-epithelial transition (MET), its role for ESCs is mediated by inhibition of MAPK pathways to promote self-renewal. Maintaining pluripotent stem cells *in vitro* is essential in stem cell research. It provides the way to deeply study the molecular basis of stemness observed in pluripotent stem cells. However, although stem cells hold great hope for new therapies, it is difficult to isolate and maintain stem cells *in vitro*. In particular, ethical issues and tissue rejection are the main problems that prevent human ESCs 'utilized in research and clinical application.

2.4. Induced pluripotent stem cells approach

The above evidence raised a possibility that cells in late-stage development could be reprogrammed to other cell fate, even toward pluripotent stem cells. Inspired by this possibility, Yamanaka and colleagues made a list of pluripotency-associated genes which may promote cell fate change to induce the pluripotent state [Takahashi and Yamanaka, 2006]. They overexpressed 24 candidate genes in mouse embryonic fibroblasts (MEFs) and successfully generated pluripotent stem cells. Performing sequential removal of 24 candidates, the four essential factors: KLF4, OCT4, SOX2, and c-MYC, were identified as reprogramming factors that generate pluripotent stem cells from MEFs. Moreover, four reprogramming factors could also reprogram tail-tip mouse fibroblasts and human fibroblasts to a pluripotent state [Takahashi and Yamanaka, 2006; Mikkelsen et al., 2008; Takahashi et al., 2007]. Compared with ESCs, iPSCs show similar expression patterns and epigenetic marks and could be maintained in media that supports ESC growth. Also, iPSCs could contribute to all three germ layers and give rise to chimeric mice. Full-term development of viable mice consisting of all iPSC-derived cells finally proved that iPSC has full pluripotency [Maherali et al., 2007; Kang et al., 2009].

Among pluripotent stem cell sources, iPSCs are believed to be best suited for future research [Takahashi and Yamanaka, 2016]. Viable embryos are required and destroyed to obtain ESCs. ESCs raised many ethical issues for stem cell study using human embryonic stem cells[Prochazkova et al., 2015]. SCNT and cell fusion have many limitations, from incomplete reprogramming causing abnormal phenotype in clones, technical difficulties, low efficiency, and limited success in applying to humans and primates [Alberts et al.,

2019]. iPSCs are derived from somatic cells that are more readily available than embryos, and they can be generated from various cell types such as fibroblasts. iPSCs are functionally similar to ESCs, are capable of unlimited proliferation, and have the potential to differentiate into all kinds of cells in the body. Moreover, iPSCs are less likely to be immunologically rejected due to being genetically related to the recipient. Altogether, iPSCs are at the center of intense research and application in regenerative medicine.

3. The mechanism of reprogramming and the hurdles of applying iPSCs to clinical

3.1. Hallmark of reprogramming fibroblast into pluripotent stem cells.

Several studies indicate that the timing of reprogramming varies widely among the types of cells and reprogramming systems. However, all reprogrammed cells undergo multiple changes before becoming iPSCs [David and Polo, 2014]. In the early stage of reprogramming, a range of events occurs, including suppression of somatic cell-related genes; cell morphological changes, cell-matrix adhesion, and cell-cell attachment; the metabolic shift from oxidative phosphorylation to glycolysis; the increased resistance to apoptosis and senescence; acquisition of pluripotency [David and Polo, 2014; Mikkelsen et al., 2008].

The first feature of reprogramming is the downregulation of somatic cell-related genes, such as reducing *Thy1*, *Snai1*, *Cdh2*, and *CD44* expression. The intermediate stage is the upregulation of pluripotency-related genes and ends with the late phase when iPSCs acquire pluripotent independently with exogenous reprogramming factors. This stage is marked by epigenetic profiles of X chromosome reactivation [Maherali et al., 2007]. Different markers could be used to study the mechanism of reprogramming. The early phase is marked by E-cadherin (*Cdh1*), alkaline phosphatase, and stage-specific embryonic antigen 1 (SSEA1) and followed by the intermediate step indicated by the endogenous expression of *Oct4*, *Nanog*, and *Esrrb*. The final stage is characterized by *Sox2*, TRA-1-60, *Rex1*, *Sall4* expression, and X chromosome reactivation [Buganim et al., 2012; Polo et al., 2012]. However, the underlying reprogramming mechanisms are still not fully elucidated, and the iPSC generation is inefficient. Significantly, the understanding of molecular events during the early phase, such as the downregulation of the somatic program, is not as intensively investigated as in the late stage of reprogramming.

3.2. The downregulation of somatic cell-related genes: a promising approach to achieve efficient reprogramming.

As the hallmarks of reprogramming, silencing of the somatic program occurs quickly [Samavarchi-Tehrani et al., 2010; Li et al., 2010; Polo et al., 2012]. A signature of somatic reprogram downregulation is reducing *Thy1* gene expression and CD44 surface marker. THY1 is a surface antigen that is highly expressed in various somatic cells such as fibroblasts. Its loss of expression is considered the marker for the downregulation of somatic programs during reprogramming. When *Thy1* silencing occurs, the cells start to undergo reprogramming. *Thy1*+ cells express lower levels of K, O, S, M proteins and fail to produce iPSCs. Differentiated cells or somatic cells are characterized by expressing somatic cell-related genes such as *Thy1*, *Col5a5*, and *Fibrillin-2* and the epigenetic mechanisms such as DNA methylation [Polo et al., 2012]. Correspondingly, a study in 2008 showed that the incomplete downregulation of somatic cell-specific transcription factors inhibits establishing the endogenous pluripotency networks. However, forced suppression of somatic cell-specific transcription factors enables partially reprogramming cells to become fully reprogrammed iPSCs [Mikkelsen et al., 2008]. Similarly, early repression of *Col5a2, Fibrillin-2, and Egr1* is necessary for reprogramming [Stadtfeld et al., 2008; Worringer et al., 2014].

After the downregulation of somatic cell-related genes, pluripotency-associated marker alkaline phosphatase (AP) and SSEA-1 are activated. Both AP and SSEA1 are signs of the early events in the iPSC generation followed by the late phase marked by the expression of pluripotency gene *Oct4, Nanog, Esrrb* [Polo et al., 2012]. The downregulation of markers such as *Thy1* and CD44 is necessary for the early phase of reprogramming. However, additional somatic cell-related genes may play a vital role in maintaining somatic cell fate and restricting the progression of somatic cell reprogramming. Therefore, it is necessary to identify such highly expressed factors to sustain the somatic cell fate and become a roadblock in the early phase of the reprogramming process.

3.3. Mesenchymal-epithelial transition: a crucial event during reprogramming

Epithelial and mesenchymal cells are two major types of animal cells [Pei et al., 2019]. The two transition processes, Epithelial-mesenchymal transition (EMT) and MET, are involved in many cell fate conversions from embryonic development, tumor generation, and somatic cell reprogramming. EMT and MET are essential mechanisms in the development process

of all animals [Kalluri and Weinberg, 2009]. During gastrulation of embryonic development, EMT occurs, while the first MET occurs during preimplantation [Larue and Bellacosa, 2005]. Various cellular processes require EMT, such as embryogenesis, cell differentiation, wound healing, tissue regeneration, fibrosis, and cancer [Lamouille et al., 2014]. MET is also necessary for many development processes and cell fate changes [Brabletz, 2012]. For example, embryos undergo MET to engage in gastrulation or subsequent body formation or elongating [Pei et al., 2019].

EMT is the transition process in which epithelial cells adopt a mesenchymal form by changing their intercellular adhesion and cell-matrix adhesion, reorganizing the actin skeleton, and increasing migration, all of which may be involved in germ layer differentiation and tissue formation. For example, local wound healing involves EMT that generates fibroblast following tissue repair [Stone et al., 2016]. In model epithelial cell lines such as normal murine mammary gland (NMuMG) cells, EMT can be induced by transforming growth factor-β (TGF-β). In NMuMG cells, the effect of TGF-β is mediated by a transcription factor, *Sox4*, which causes EMT in response to TGF-β. *Sox4* directly regulates *Ehz2* expression, which encodes the polycomb repressive complex 2 (PRC2) components, the modifying chromatin structure for gene repression [Tiwari et al., 2013]. In EMT, TGF-β signaling induces transcription factors SNAIL/2, ZEB1/2, and TWIST to control the EMT regulatory network causing the downregulation of epithelial protein, including junction complexes such as *Cdh1*, *Occludin, EpCam,* and the acquisition of mesenchymal protein: SNAIL1, SNAIL2, ZEB1, ZEB2, TWIST1, TWIST2*,* N-CADHERIN *(Cdh2),* FIBRONECTIN, VIMENTIN [Kalluri and Weinberg, 2009].

In contrast, MET transitions from spindle-shaped, multipolar mesenchymal cells to round-shaped epithelial cells. MET leads to the change of gene expression by upregulation of epithelial-associated genes and downregulation of mesenchyme-associated genes [Li et al., 2010]. MET and EMT play an essential role in iPSCs generation. MET is a critical event during reprogramming; MET is also associated with other events that occur in the initial phase of reprogramming when cells show downregulation of somatic genes, metabolic switching, and epigenetic modification [Shu and Pei, 2014]. Reprograming factors (K, O, S, M) are the key regulators that trigger MET by downregulating mesenchymal and upregulating epithelial genes. In the early phase of somatic cell reprogramming, OCT4, SOX2*,* and C-MYC suppress TGF-β, which leads to downregulation of *Snai1/2, Cdh2*, *Fn1*,

Vim, *Zeb1*/2. At the same time, KLF4 triggers epithelial gene expression (*Cdh1, EpCAM, Occuludin*). Enhancing MET by activating BMP and inhibiting EMT facilitates reprogramming [Ichida et al., 2009; Li et al., 2010; Samavarchi-Tehrani et al., 2010]. MET is considered an important event during the early phase of reprogramming from MEFs [Samavarchi-Tehrani et al., 2010]. Inhibition of MET by inducing EMT with TGF-β or overexpressing *Snai1* prevents iPSC generation [Li et al., 2010]. Furthermore, several lines of evidence indicated MET as part of the mechanism when vitamin C increases the efficiency of iPSCs generation [Chen et al., 2013] and TET1 substitutes OCT4 during reprogramming [Gao et al., 2013].

Improving reprogramming efficiency could be achieved by facilitating MET and suppressing EMT. iPSCs formation from somatic cells is triggered by key transcriptional regulators that regulate the gene expression network necessary for acquiring pluripotency. The reprogramming of mouse embryonic fibroblasts (MEFs) into iPSCs requires cell morphological change from mesenchymal to epithelial stage. Therefore, transcriptional regulators associated with a mesenchymal phenotype should be downregulated during reprogramming to trigger cell morphology transition.

3.4. Sendai virus reprogramming system - the unique gene delivery approach

The original method for introducing reprogramming factors into somatic cells is an integrative system that involves the insertion of exogenous genes into the host genome (retrovirus and lentivirus) [Takahashi and Yamanaka, 2016]. However, although these vectors provide high infectivity and stable integration into the host genome, they may cause adverse consequences such as tumor formation. Therefore, alternative approaches were utilized to generate iPSCs without transgene integration, including adenovirus, transposon, episomal vectors, and Sendai virus [Takahashi and Yamanaka, 2016].

Among these alternative systems, the Sendai virus system has several benefits over other delivery systems: 1) up to four exogenous genes could be harbored in one vector; 2) exogenous genes are expressed at high levels in a wide range of host cells; 3) Sendai virus vector remains stably for a long-term expression; 4) stoichiometry of multiply expressed genes remain relatively constant; the vector remains in cytoplasm and does not integrate into the host genome [Jaenisch and Young, 2008]. Our laboratory has developed a defective and persistent Sendai virus (SeVdp)-based reprogramming system, which possesses all of these characteristics suited for producing iPSCs.

Sendai virus belongs to the Paramyxovirus family and is a negative-strand RNA virus. The C1.151 strain has mutations in the genes encoding for viral nucleocapsid (NP, P, and L proteins), enabling the long-term expression of exogenous genes. In SeVdp-derived vectors, all three structural genes (encode M, F, and HN protein) that are involved in viral particle formation then be replaced by four reprogramming factors K, O, S, M. The derived vector, packaged in cells expressing structural genes, is defective in that it can infect host cells but neither cause any host cell lysis nor produce new virus after infection [Nishimura et al., 2007, 2011].

The integration-free SeVdp system prevents the insertion of exogenous genes that cause permanent damage to the host genome and thus achieves the generation of high-quality iPSCs. Moreover, SeVdp genomic RNA could be eliminated from the host cells by suppressing L protein expression using small interfering RNA against the L gene mRNA [Nishimura et al., 2017b]. This property provides an additional advantage for SeVdp-based vectors as a safe and efficient system for somatic cell reprogramming to produce.

3.5. The hurdle of applying iPSCs and an alternative solution to enhance reprogramming

iPSC generation is an inefficient process. Therefore, there are several hurdles to generating high-quality and safe iPSCs [Ebrahimi, 2015]. Integration of reprogramming genes causes a risk of changing gene expression of host cells and leads to genetic and epigenetic abnormalities [Hussein et al., 2011]. Significantly, the continuous expression of c-MYC, one of the 4 reprogramming factors, leads to tumor formation from iPSCs transplanted to mice [Okita et al., 2007]. Therefore, incomplete suppression of exogenous reprogramming factors negatively affects the quality of iPSCs. An alternative approach to reprogramming could solve this problem by using integration-free methods for cell reprogramming, such as episomal vectors, mRNA, transposon, adenovirus, and Sendai virus. Low reprogramming efficiency is another obstacle to applying iPSCs to therapy; less than 3% of reprogrammed cells can successfully achieve fully pluripotent iPSCs [Polo et al., 2012]. Various strategies have been studied for improving reprogramming efficiency, from overexpression of enhancing factors (GLIS1, *Esrrb*, FOXH1) to utilizing small chemicals (vitamin C, TGF-β inhibitor, Wnt/β-catenin signaling) and suppression of reprogramming barriers (p53, TGF-β, MAP kinase) [Ebrahimi, 2015]. Although current knowledge about reprogramming could enhance reprogramming to some extent, safe and rapid reprogramming with higher

efficiency is still not achieved. As compared with its late phase, the early phase of reprogramming, including critical events such as increased cell proliferation and progression through MET, was less extensively studied to improve reprogramming efficiency.

Many questions related to EMT/MET during reprogramming remain unclear: First, although blockage of EMT is essential, expression of EMT-related genes such as *Tgfbr1* and *Tgfb2* impact reprogramming in both positive and negative manners [Li et al., 2010]. Second, although downregulation of mesenchymal-related genes such as *Snai1/2*, *Zeb1*/2, *Cdh2*, and *Fn1* are observed in iPSCs generation, expression of *Snai1* and downregulation of *Snai2* facilitates Nanog reprogramming [Unternaehrer et al., 2014; Gingold et al., 2014]. Third, inhibition of TGF-β signaling has various effects on reprogramming depending on its addition relative to MET during reprogramming [Li et al., 2010; Ichida et al., 2009]. Moreover, the knockdown of *Snai1/2* does not enhance reprogramming [Li et al., 2010]. These results imply that some of the effects of downregulating mesenchymal genes may not necessarily be the cause of the observed impacts on MET during reprogramming. Thus, the mechanism underlying MET and its relevant factors still remain to be more clearly defined, and more research is needed to understand the role of MET during reprogramming. Therefore, identifying new factors that inhibit reprogramming, especially during the MET stage, is expected to understand the mechanism of iPSCs generation further and devise more efficient reprogramming technologies.

4. Aims of this thesis

A substantial number of approaches have been developed to improve the efficiency of reprogramming, and many factors and mechanisms have been discovered. However, the reprogramming system still requires improvement to produce high-quality iPSCs efficiently. This study focuses on the early phase of reprogramming when cells show morphological changes (MET) by suppressing the somatic cell identity through downregulating mesenchyme-associated genes. This thesis aims to:

1. Screen for mesenchyme-associated transcriptional regulators that may delay the progression of MET during iPSCs generation.
2. Examine the effect of the above-mentioned factors on the early phase of reprogramming.
3. Investigate the mechanism underlying mesenchymal gene downregulation during reprogramming to enhance reprogramming efficiency.
4. Understand the relationship between the mesenchyme-associated transcriptional regulators and their downstream signaling pathways during reprogramming.

CHAPTER 2: MATERIALS AND METHODS

1. Cell culture

MEFs were isolated from embryos of 13.5-day-pregnant C57BL/6 mice. MEFs were cultivated in DMEM (Sigma) supplemented with 10% fetal bovine serum (FBS) (Gibco) and 100 U/mL penicillin- 100 µg/mL streptomycin (Nacalai Tesque) at 37°C, 5% CO_2, 5% O_2.

PLAT-E, BHK/T7/151M, and NIH3T3 cells were maintained in DMEM supplemented with 10% FBS and 100 U/mL penicillin- 100 µg/mL streptomycin at 37°C, 5% CO_2.

NMuMG cells were cultured in DMEM containing 10% FBS, 100 U/mL penicillin - 100 µg/mL streptomycin, and 1µg/mL insulin at 37°C with 5% CO_2. EMT was induced by 5 ng/mL of recombinant TGF-1 (BioLegend) for 3 days.

SNL feeder cells harboring puromycin resistance genes [Tran et al., 2018] were generated by treating with 10 µg/mL mitomycin C (Sigma) for 2.5 hours.

2. Vector and construction

The cDNAs encoding mesenchyme-associated transcriptional regulators fused with 3xFLAG-tag were amplified from MEF cDNA, followed by their insertion into pMCsΔYY1-IRES-Puro plasmid to construct retroviral vectors expressing each protein. Annealed DNA oligonucleotides (89 nucleotides) encoding upstream Open Reading Frame 2 (uORF2) (Table 1) were placed in pMCsΔYY1-IRES-Puro plasmid 20 nucleotides before the inserted gene to reduce expression from the retroviral vectors. For the construction of a retroviral vector expressing shRNA against *Osr2*, DNA oligonucleotides listed in Table 2 were annealed and inserted into pMXs-U6-Puro plasmid (Cambridge bioscience). Retrovirus stocks were prepared as described previously [Bui et al., 2019].

The SeVdp vector genomic cDNA was constructed as described [Nishimura et al., 2011]. The cDNA for SeVdp(KOSMaB) was produced by insertion of the blasticidin-resistant gene together with the T2A peptide sequence after the c-MYC gene of SeVdp(KOSM) [Nishimura et al., 2014].

3. Collection of SNL conditioned medium

Mitomycin-C treated SNL cells were seeded in mES1 medium (DMEM supplemented with 15% FBS, 0.1 mM non-essential amino acids (Nacalai Tesque), 100 U/mL penicillin – 100

μg/mL streptomycin, 0.055 mM 2-mercaptoethanol (Gibco), 1,000 U/mL LIF (Wako)) at 37°C, 5% CO_2. Then, SNL conditioned medium was collected 24 hours later, followed by centrifugation at 3000 rpm for 10 min at 4°C. A fresh conditioned medium was prepared each time.

4. Reprogramming

MEFs were reprogrammed by infecting with SeVdp(KOSM) or SeVdp(KOSMaB) at 32°C for 12 hours. Reprogrammed cells could be cultured on SNL feeder cells in mES2 medium (StemSure DMEM (Wako) supplemented with 15% StemSure Serum Replacement (Wako), 0.1 mM non-essential amino acids, 2 mM L-Alanyl-L-Glutamine (Nacalai Tesque), 100 U/mL penicillin – 100 μg/mL streptomycin, 0.055 mM 2-mercaptoethanol, 1,000 U/mL LIF) for first 6 days at 37°C, 5% CO_2, followed by cultivation with mES1 medium. Without SNL feeder cells, reprogrammed cells were cultured in SNL conditioned medium. For knockdown of *Osr2*, siRNA against *Osr2* (Table 3) was transfected into MEFs by using Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher) 2 days before reprogramming.

5. Retrovirus production and infection

Retroviral vector plasmids were transfected into PLAT-E cells with Lipofectamine 2000 reagent (Thermo Fisher). After 2 days of transfection, viral supernatant was collected and filtered by 0.45 μM cellulose acetate filters (Millipore), and Retrovirus was stored at -80°C.

NMuMG cells, NIH3T3 cells, and MEFs were infected with retrovirus in the presence of 8 μg/ml of polybrene (Sigma). After 48 hours of infection, the cells were treated with 2 μg/mL puromycin (Nacalai Tesque) or Neomycin (1 mg/mL) for 2 or 7 days, respectively.

6. SeVdp production

To prepare vector packaging cells, cDNA for SeVdp(KOSM) or SeVdp(KOSMaB) and the expression vector plasmids for SeV proteins (NP, P/C, M, F, HN, L) were transfected into BHK/T7/151M cells using Lipofectamine LTX with Plus Reagent (Thermo Fisher) and cultured at 32°C for 6 days. Then, the packaging cells were further transfected with the expression vector plasmids for SeV M, F, and HN proteins and cultured at 32°C for additional 4 days to rescue the SeVdp vector. The supernatant containing the SeVdp vector was filtered through a 0.45 μm cellulose acetate filter and stored at -80°C until use. Titers

of the SeVdp vectors were determined by examining NIH3T3 cells infected with a diluted vector suspension by immunostaining using an anti-SeV NP antibody [Nishimura et al., 2007].

7. Quantitative PCR

Cells were harvested using Trypsin/EDTA (Nacalai Tesque). RNA was isolated by ISOGEN (Nippon Gene) according to the manufacturer's instructions. Reverse transcripted cDNA was synthesized by SuperScript III Reverse Transcriptase (Thermo Fisher) according to the manufacturer's instructions. Gene expression was quantified by mixing the cDNA with GoTaq qPCR Master Mix (Promega) and primers listed in Table 4, followed by PCR with QuantStudio 5 Real-time PCR system (Thermo Fisher). Expression of *Tbp* was used as an internal control.

8. Immunofluorescence staining

Cells were washed with PBS twice, then fixed with 3.7% formaldehyde in PBS for 10 min., followed by permeabilization by 0.1% Triton X-100 in PBS for 5 min. The treated cells were incubated with primary antibodies at R.T. for at least 60 min, followed by anti-mouse IgG conjugated with AlexaFlour 555 (1:500, ThermoFisher) for 30 min. VECTASHIELD mounting with DAPI (Vector) was utilized to stain nuclei. Followed primary antibodies were employed in this paper: Anti-SeV NP (1:1,000), Anti-Flag-tag (1:4,000, Wako).

9. Cell proliferation assay

Cell proliferation was quantified by the XTT cell proliferation kit (Biological Industries). 500 cells were seeded into a 96-well plate and incubated at 37°C with 5% CO_2. Proliferation levels were measured at indicated time point by treatment with XTT mixture according to the manufacturer's instruction.

10. Cell migration assay

NMuMG cells were added to the top of Transwell 6.5 mm insert 24-well plate (CORNING). After 16 hours of cultivation, cells was removed from the upper side of the chamber by using a cotton swab. Cells migrating to the lower side of the transwell filter were fixed and stained with Staining solution (0.05% Crystal Violet, 1% Formaldehyde, 1% Methanol in PBS(-) for 20 min at R.T. The migrating cells were counted under a microscope.

11. Western blot analysis

NIH 3T3 cells were infected with retrovirus expressing indicated genes, followed by puromycin selection for 2 days. 4 days after infection, cells were harvested by Trypsin/EDTA and lysed in gel loading buffer (50 mM Tris-HCl (pH6.8), 2% SDS, 10% glycerol, 0.1 mg/mL Bromophenol blue, 5% 2-mercaptoethanol) incubate at 95°C for 5 min. Lysis samples were subjected to western blot analysis. Anti-Flag (1:4,000, Sigma) and anti-α-tubulin (1:10,000, Abcam) were employed as primary antibody. Protein expression levels were detected and quantified by the Fusion Fx7 Imaging system (Vilber Lourmat) and Fusion Software (Fusion Capt) v18-02.

12. Transcriptome analysis

Two replicates of total RNAs were NMuMG cells followed by library construction, and RNA-seq performed Illumina sequencing platform (Annoroad Gene Technology). Sequence reads were mapped to the mouse genome (GRCm38.p6) using STAR aligner (v2.7.3a) with default parameters (Dobin et al., 2013). The read counts were obtained by featureCounts (v2.0.0) (Liao et al., 2014) with the gene model from Ensemble (Mus_musculus GRCm38.75.gtf) at gene level. The gene count data was normalized by variance stabilizing transformations (VST) using DESeq2 (v1.26.0) (Love et al., 2014). The VST normalized value was used to show the transcriptome similarity through the heatmap with Ward's method and PCA. Differential expression was analyzed by the Likelihood ratio test using DESeq2. Pathway analysis of the differential expressed genes was performed by using WikiPathway (Slenter et al., 2018) through Enrichr (Chen et al., 2013; Kuleshov et al., 2016). The data are deposited in the Gene Expression Omnibus (GEO) with accession number GSE180471 (NMuMG experiments) and GSE180428 (Reprogramming).

13. Statistical analysis

Student’s t-tests were used to test for a statistically significant difference between data sets, except for RNA-seq data. A value of $P < 0.05$ was considered statistically significant.

CHAPTER 3: RESULTS

1. Screening for the mesenchyme-associated transcriptional regulators

The Sendai virus-based reprogramming system was proved to reprogram somatic cells into iPSCs more efficiently and speedily than the conventional retrovirus system [Nishimura et al., 2011]. A screening strategy using the MEFs and the Sendai virus-based reprogramming was designed to identify new barriers to reprogramming. In order to select the genes that show downregulation at the early phase of the MEF reprogramming, gene expression profiles in cells undergoing reprogramming on day 2, day 8, and fully reprogrammed iPSCs (Figure 1A) were compared with that in MEFs. The DNA microarray data show that 1,787 differentially expressed genes (DEGs) were reduced their expression by over 2-fold compared to MEFs. Next, 150 possible transcriptional regulators were selected based on GO terms related to transcription factor activity from the down-regulated DEGs. Then, 37 out of 150 candidates were obtained based on the possibility that they are presumably directly regulated by reprogramming factors, KLF4, OCT4, SOX2, or c-MYC, as they have been experimentally confirmed to be involved in iPSC generation or in ESCs [Sharov et al., 2008; Sridharan et al., 2009]. Because cell cycle upregulation, MET, and dedifferentiation are required for the early phase of reprogramming, finally, 10 genes associated with the reverse cellular functions include cell cycle arrest, EMT, and differentiation [Douville et al., 2011; Gan et al., 2011; Garcia-Dominguez et al., 2003; Jordan et al., 2013; Lan, 2004; Ocaña et al., 2012; Pei et al., 2019] were selected as candidates related to iPSC generation. Among the 10 candidates, 8 candidates (*Ebf1, Ebf3, Meox2, Osr2, Prrx1*, *Smarcd3,* and *Zic1*) showed significant down-regulation across the first 8 days of reprogramming (Figure 1B).

The impacts of the 8 transcriptional regulators' overexpression on MEFs viability were evaluated using a silencing-resistant retroviral vector, MCsΔYY1 [Nishimura et al., 2017a] (Figure 1C). Of these, 3 factors (*Ebf1*, *Ebf3*, and *Osr2*) showed a severe reduction of MEF viability, and the other 5 factors were unaffected (Figure 1D). It suggests that the negative impact on MEF viability by the transcriptional regulators' overexpression might affect the assessment of their function on reprogramming. As shown in Figure 1E, the expression of exogenous *Osr2* is robust, more than 500 times compared to the control, indicating that such a too high expression might be associated with the reduction of cell viability.

In order to diminish the detrimental influences of overexpression of transcriptional regulators, a new vector expressing transcriptional regulators at a lower level was newly constructed. Upstream open reading frame (uORF2) derived from the human cytomegalovirus (CMV) virion glycoprotein gpUL4 (gp48) plays a role in repressing the downstream translation of heterologous genes [Degnin et al., 1993]. To lower the expression level of the transcriptional regulators, a beneficial effect of uORF2 on inhibiting downstream translation was exploited. The DNA sequence encoding 22 amino acids for uORF2 was cloned into an MCsΔYY1 based-expression vector ahead of the protein-coding region (Figure 2A). As shown in Figure 2B, there was no significant change in mRNA level. However, protein levels of 8 transcriptional regulators in MEFs were reduced remarkably to 1~60% compared to those original vectors without uORF2. MEFs with such a reduced expression of the transcriptional regulators showed comparable cell viability and cell proliferation to mock-treated cells (Figures 2C and 2D). These results indicated that using a new vector that harbors uORF2 to express 8 transcriptional regulators on MEFs could minimize the cytotoxic effect of their overexpression.

Each transcriptional regulator was expressed individually in MEFs from the MCsΔYY1-uORF2 based-expression vector to examine the functional association between the downregulation of the 8 factors and reprogramming(Figure 3A), followed by reprogramming by SeVdp(KOSM). The efficiency of iPSC generation was evaluated by Alkaline phosphatase (AP) activity. Based on the fact that AP activity was widely used as a marker for iPSC generation in the early phase of reprogramming [Štefková et al., 2015], reprogramming efficiency was scored by a number of AP(+) colonies (Figure 3A). Among 8 transcriptional regulators, expression of *Ebf1, Ebf3,* or *Osr2* showed a reduction of AP(+) colony number on day 10 of reprogramming (Figure 3B), indicating that failure to downregulate *Ebf1, Ebf3,* and *Osr2* led to the reduction of reprogramming's efficiency. Under the observation of cell viability throughout the iPSC generation, substantial death occurred in cells expressing *Ebf1* or *Ebf3* after induction of reprogramming (Figures 3C and 3D) which led to difficulties for further functional investigation. In contrast, such a phenomenon was not observed in the cells expressing *Osr2* (Figures 3C and 3D). Therefore, the effect of *Osr2* was selected for further research.

2. Downregulation of OSR2B by reprogramming factors is predominantly required for efficient reprogramming

Based on the data from DNA microarray and quantitative RT-PCR in Figure 1B, endogenous *Osr2* expression level was reduced by the expression of reprogramming factors (KOSM). Next, how reprogramming factors downregulate *Osr2* during reprogramming need to be revealed. SeVdp vector harboring a single or combination of the transcription factors was infected to MEFs to analyze the relationship between reprogramming factors and *Osr2* expression. The qRT-PCR analysis revealed that the vector expressing KLF4 (K, KOS, KOM, and KOSM), or OCT4 and SOX2 (OS, OSM, KOS, and KOSM) downregulated *Osr2* expression in MEFs (Figure 4A). These results demonstrated that *Osr2* expression was regulated by KLF4 or the combination of OCT4 and SOX2 during reprogramming.

Osr2 is a member of the *Drosophila odd-skipped* related transcription factor family. OSR2 has two variants (OSR2A and OSR2B) due to the alternative splicing [Lan et al., 2001; Kawai et al., 2005]. OSR2A is the minor isoform that contains five zinc finger motifs, and OSR2B is the predominant isoform that contains three zinc finger motifs (Figure 4B). Because of the involvement of OSR2B in reprogramming experiments described above, the effect of OSR2A needs to be evaluated to understand the function of Osr2 on reprogramming comprehensively.

A cDNA encoding OSR2A was cloned into MCsΔYY1-uORF2 based-expression vector. Then, the reprogramming efficiency was evaluated, as shown in Figure 3A. Although both OSR2A and OSR2B lessen AP(+) colony number, the negative effect of OSR2B on reprogramming was more intense (Figure 4C). Moreover, as mentioned above, the OSR2B variant is predominantly expressed. Therefore, OSR2B was used for further analyses (referred to as OSR2 in the following parts).

3. Continuous expression of *Osr2* blunts reprogramming through inhibits MET

Given that the *Osr2* expressing cells showed low reprogramming efficiency in the early step of iPSC generation (Figure 3B), it needs to be examined whether *Osr2* affects the expression of pluripotency marker in the latter phase, SSEA1 [Mann et al., 2007], or not.

Reprogrammed cells that express exogenous *Osr2* exhibited a deficient level of SSEA1 and a remarkable reduction in the number of SSEA1 (+) colonies as well as an increase in the number of SSEA1(-) colonies (Figure 5A). Moreover, in agreement with the negative effect of *Osr2* expression during reprogramming, *Osr2*-expressing MEFs could not continue to grow after passage (Figure 5B). These results indicate that *Osr2* blocks the reprogramming process toward the fully reprogrammed stage through inhibition of early events such as induction of AP activity, SSEA1 expression, and cell proliferation.

At first, retrovirus harboring shRNA targeting *Osr2* was used for forced repression of *Osr2* expression to confirm that downregulation of *Osr2* indeed affects reprogramming (Figure 5C), and the effect of *Osr2* knock-down on reprogramming was evaluated. Compared to control (MEF cells), endogenous *Osr2* expression was reduced roughly by half, and no significant change was recorded in the AP activity (Figure 5D). Although the *Osr2* level after knocking down by shRNA might not be sufficient, siRNA for *Osr2* showed significant repression (around 10%) (Figure 5E). As expected, the number of AP(+), SSEA1(+) and Nanog (+) colonies significantly increased with the siRNA treatment (Figures 5F and 5G). Moreover, the expression of pluripotent markers (*Fbox15* and *Rex1*) was increased by *Osr2* knockdown in MEFs (Figure 5H).

Several comprehensive studies showed that *Osr2* was commonly detected in specific mesenchymal cells and tissues during embryogenesis. For example, at E14.5, *Osr2* was strongly expressed in the mesenchymal cells surrounding the tooth buds [Lan, 2004]. Notably, *Osr2* is also expressed in a subset of mesenchymal cells adjacent to the epithelial tubules in the developing metanephros [Lan et al., 2001]. In other words, *Osr2* expression correlates with sites of epithelial-mesenchymal interactions. In addition, *Osr2* was identified as a critical regulator of palatal growth and development [Fu et al., 2017; Kawai et al., 2010; Lan, 2004], where proper EMT is required for completing the process of palatal fusion [Ahmed and Nawshad, 2007; Gritli-Linde, 2007; Jalali et al., 2012]. These pieces of evidence suggest that changes in *Osr2* expression might relate to the promotion of MET during reprogramming.

For a more precise evaluation of gene expression and observation of cell morphology changes, uninfected MEFs were eliminated by using a new SeVdp(KOSMaB)

vector harboring the blasticidin S-resistant gene as well as reprogramming factors (Figure 6A). After reprogramming of MEFs by SeVdp(KOSMaB), on the day 6, while the control reprogramming displays a completed and typical iPSC-like colony in which almost all cells change to epithelial morphology, *Osr2*-expressing cells showed the small and incomplete colonies, which contained cells with an elongated shape, and lacked apparent cell-cell adhesion cells, which were similar to fibroblast's morphology (Figures 6B and 6C). Furthermore, qPCR shows that the downregulation of several mesenchymal genes such as *Snai1*, *Tgfb3*, *Cdh2*, *Fn1*, and *Vim* was far lesser in *Osr2*-expressing cells compared to control cells (Figure 6D). The epithelial genes, *Cdh1* and *Ocln,* were upregulated in the control reprogramming but not reprogramming with *Osr2*-expressing (Figure 6D). These results indicated that continuous expression of *Osr2* prevents MET at the early stage of reprogramming.

4. *Osr2* functions as a novel regulator of EMT in NMuMG cells

The NMuMG cells, which are widely used for studies of EMT [Bhowmick et al., 2001], were used to gain further insights into the relationship between *Osr2* and the transition of cellular phenotypes, MET and its reverse process EMT. The stimulation by TGF-β [Miettinen et al., 1994] and the overexpression of some regulators such as *Sox4* [Tiwari et al., 2013] trigger robust and reproducible transition NMuMG from epithelial to mesenchymal state within ten days. EMT is the process by which epithelial cells lose their cell polarity and cell-cell adhesion and gain migration properties to become mesenchymal phenotype. EMT coordinates the induction of mesenchymal genes such as *Cdh2, Vim, Tgf3, and Fn1* and the repression of epithelial genes such as *Cdh1, Epcam, and Ocln.*

Firstly, the changes in cell-cell adhesion, gene expression pattern, and cell migration abilities was evaluated by overexpression of *Osr2* in NMuMG cells (Figure 7A). Typically, during EMT, cells showed a change from epithelial to mesenchymal morphology (from round to flattened and spindle shape with loss of cell-cell adhesion), similar to the cells treated with TGF-β (Figure 7B). However, compared to TGF-β-treated cells, *Osr2*-expressed cells required a longer time for a morphological change (Figure 7B). Furthermore, the migration assay revealed that *Osr2*-expressing cells could migrate similar to TGF-β-treated cells (Figure 7C). Gene expression analyses also showed that both treatments, *Osr2*

expression and TGF-β stimulation, led to the gain of mesenchymal genes and the loss of epithelial genes (Figure 7D). Taken together, these findings reveal that *Osr2* is a novel regulator that induces EMT in NMuMG cells.

To understand by which mechanism *Osr2* induces EMT, RNA-seq analysis was performed for different samples, including (1) NMuMG cells; (2) NMuMG with TGF-β treatment for 2 days; (3) NMuMG expresses *Osr2* on day 3, and (4) NMuMG expresses *Osr2* on day 9. Hierarchical clustering and principal component analysis (PCA) illustrated that the expression pattern of *Osr2* expression for 3 or 9 days was distinct from that of TGF-β treatment (Figures 8A and 8B). Pathway analysis for DEGs was conducted to examine the differences between *Osr2*-expressing and TGF-β-treated NMuMG cells in order to find possible biological pathways involved in the EMT. Upregulated genes in both *Osr2* on day 3 and TGF-β in Cluster 3-2 are enriched in pathways related to EMT (Figures 8C and 8D). Genes in Cluster 9-3 upregulated either by *Osr2* overexpression for 9 days or TGF-β treatment showed more pronounced enrichment than Cluster 3-2 (Figures 8C and 8D). Such a difference is consistent with a slow and prolonged progression of EMT by exogenous expression (Figure 7B). Moreover, Cluster 3-2 and 9-3 includes EMT-related regulatory genes (*Snai1, Tgfb1, Wnt7a, Zeb2, Foxq1, Smad3),* as well as their downstream structural genes *(Nrp2, Pkp1, Cldn4, Fn1, Fzd1/2, and Notch2)* (Figure 8C).

Genes downregulated either by *Osr2* expression or TGF-β treatment, which are clustered in Cluster 3-6 and 9-5, are enriched in cellular metabolism, including the oxidative phosphorylation pathway, which corresponds to metabolic changes during EMT [Kang et al., 2019] (Figure 8D). Interestingly, genes in Cluster 3-1 whose expression are upregulated by *Osr2* expression (Day 3) but not TGF-β treatment are enriched in pathways related to DNA replication, transcription, and translation (Figure 8D), which are no longer enriched at day 9 (Cluster 9-4) (Figure 8D). This result is compatible well with the observation that *Osr2* promotes transient proliferation of NMuMG cells before their subsequent progression through EMT (Figure 7B). Altogether, RNA-seq analyses further confirmed the activation of EMT in NMuMG cells by *Osr2* expression, albeit in a somewhat delayed and prolonged manner compared with TGF-β-induction.

On account of *Osr2* and TGF-β playing similar roles in cell morphological change, next, I examined their association during the progression of EMT. At first, the TGF-β pathway was inhibited by two small molecule inhibitors, SB431542 [Inman et al., 2002] or RepSox [Gellibert et al., 2004; Ichida et al., 2009], during EMT induction by *Osr2*. As a result, both two small molecule inhibitors prevent *Osr2*'s function on EMT in the manner of changing cell morphology and losing cell-cell adhesion, as well as cell migration (Figures 9A and 9B). Gene expression analysis evidenced the downregulation of EMT-related genes (*Snai1*, *Snai2*, and *Zeb2*) either in *Osr2*-expressing and TGF-β-treated cells (Figure 9C). Next, we knocked down *Osr2* expression in NMuMG cells which were then treated with TGF-β. As a result, no effects on EMT were recorded in *Osr2*-knockdown cells even with TGF-β treatment (Figure 9D). Moreover, RNA-seq analysis of *Osr2*-expressing NMuMG cells confirmed that *Osr2* induces genes related to TGF-β signaling (Figure 9E). In conclusion, analyses in the EMT model conclusively documented that *Osr2* induces EMT likely through TGF-β signaling.

5. Osr2 downregulation reduces TGF-β signaling to promote MET during reprogramming.

TGF-β, a well-known inducer of EMT, has been claimed that impair reprogramming by preventing MET, and reprogramming can be promoted by inhibition of TGF-β signaling [Li et al., 2010; Maherali and Hochedlinger, 2009; Shu and Pei, 2014]. Our data indicated that continuous expression of *Osr2* blunts MET, thus hindering iPSC generation. Moreover, in NMuMG cells, *Osr2* triggers EMT through TGF-β signaling, and TGF-β inhibitors prevent EMT progression induced by *Osr2* expression. Therefore, *Osr2* might regulate TGF-β signaling, subsequent to preventing MET during reprogramming.

To examine the relationship between *Osr2* and TGF-β signaling during reprogramming, the expression of TGF-β-family genes at an early stage of reprogramming (day 3 and day 5) was analyzed. As shown in Figure 10A, control reprogramming showed the downregulation of TGF-β family genes (*Tgfb1*, *Tgfb2,* and *Tgfb3*). This pattern is consistent with the previously published report [Li et al., 2010]. However, *Osr2*-expressing cells showed significantly higher expression of these genes than the control during reprogramming (Figure 10A). About the TGF-β receptor family genes (*Tgfbr1*, *Tgfbr2*, and

Tgfbr3), the impact of *Osr2* expression was less in gene expression change (Figure 10A). Taken together, these results suggest that *Osr2* affects TGF-β-related gene expression at an early stage of reprogramming to restrict cells undergoing somatic cell reprogramming.

To deduce whether *Osr2* downregulation is required for subsequent reduction of TGF-β signaling during iPSCs generation, a small molecule, SB431542, which blocks TGF-β signaling pathway [Halder et al., 2005], was used. The addition of SB431542 has a slight impact on the number of colonies generated by control reprogramming (Figure 10B). In contrast, the addition of SB431542 on the *Osr2*-expressing cell shows the gain in the number of colonies (Figure 10B). Remarkably, most of the colonies generated from *Osr2*-expressing MEFs with SB431542 showed less tightly packed and were negative for alkaline phosphatase (AP(-)) (Figures 10B and 10C). These results suggest that SB431542 promotes colony formation in reprogramming of *Osr2*-expressing MEFs but mainly increases AP(-) colonies consisting of less pluripotent iPSCs.

TGF-β could transmit signaling via canonical Smad and non-Smad pathways [Zhang, 2017]. To examine whether *Osr2* mediates non-Smad pathway to regulate TGF-β signaling, reprogramming in the presence of SB203508 or PD0325901 [Hennig et al., 2010] (p38 or MEK inhibitor, respectively) (Figure 10D) was used separately. In contrast to SB431542, both SB203580 and PD0325901 did not show an increase in colony number (Figure 10E). It indicates that p38 and MEK in non-Smad pathways do not mediate the induction of TGF-β signaling by *Osr2* expression during iPSC generation. In summary, *Osr2* impairs iPSCs reprogramming by mediating upregulation of TGF-β signaling transmitted mainly via Smad pathways.

6. Osr2 downregulation increase Wnt signaling to facilitate reprogramming toward pluripotency

Inhibition of TGF-β signaling could only block the effect of *Osr2* partially, resulting in the formation of AP(-) colonies and incomplete reprogramming (Figures 10B and 10C). It indicates that the downregulation of *Osr2* regulates not only TGF-β signaling pathways but also other pathways. To further elucidate additional pathways that are regulated by *Osr2* expression, which is independent of TGF-β signaling during reprogramming, RNA-seq was performed for different conditions: (1) MEFs reprogrammed without SB431542, (2) *Osr2*-

expressing MEFs reprogrammed without SB431542, (3) MEFs reprogrammed with SB431542, and (4) *Osr2*-expressing MEFs reprogrammed with SB431542 (Figure 11A). Pathway analysis of DEGs with and without SB431542 indicated that the common enriched pathways include DNA replication, cell cycle, PluriNetWork, Eukaryotic Transcription Initiation, and Wnt Signaling Pathway (Figure 11B). The data in Figure 11C suggest that *Osr2* prevents the expression of Wnt-related genes: *Axin2*, *Lef1*, and *Tcf7*. Moreover, it is well known that the Wnt signaling facilitates reprogramming [Lluis et al., 2008; Marson et al., 2008]. Therefore, Wnt signaling might be regulated by *Osr2* during reprogramming.

To verify that Wnt and TGF-β signaling was regulated by *Osr2* during reprogramming, Wnt activator (CHIR99021) and TGF-β inhibitor (SB431542) were used individually or corporately on reprogramming, and then reprogramming efficiency was scored via AP activity. Treatment only with CHIR99021 did not significantly affect the reprogramming efficiency in control and *Osr2*-expressing cells (Figure 12A), in contrast with previous studies that indicate Wnt signaling can facilitate reprogramming [Zhang et al., 2014]. Impressively, both control and *Osr2*-expressing MEFs reprogrammed in the treatment with CHIR99021 and SB431542 generated a higher proportion of AP(+) colonies (Figure 12A). Gene expression profile confirmed that CHIR99021 activates Wnt-related genes (*Axin2*, *Lef1*, and *Tcf7*) and SB431542 downregulates TGF-β-related genes (*Tgfb2*, *Tgfb3,* and *Zeb2*) (Figure 12B). Indeed, SB431542 and CHIR99021 recover pluripotent gene expression in *Osr2*-expressing cells (Figures 12C and 12D). Taken together, *Osr2* downregulation is essential for the subsequent reduction of TGF-β signaling and activation of Wnt signaling, both of which contribute to facilitating the acquisition of pluripotency through MET (Figure 13).

CHAPTER 4: DISCUSSION

1. EMT in NMuMG was triggered by the expression of *Osr2* that induce TGF-β-related gene expression

In this study, *Osr2* was demonstrated to have a negative effect on MET during reprogramming at the early phase. The analysis using NMuMG cells revealed that transcription factor OSR2 induces the expression of molecules involved in the TGF-β signaling pathway to induce EMT. It is consistent with the negative function of *Osr2* on reprogramming and also with the previous studies that revealed a close relationship between *Osr2* and EMT [Fu et al., 2017; Lan, 2004]. OSR2 elicits gene expression changes that are consistent with the mesenchymal phenotype. TGF-β-related genes were also increased under the effect of OSR2. Moreover, the TGF-β inhibitors block the effect of *Osr2* on the EMT process induced by TGF-β. However, compared with TGF-β addition, *Osr2* expression required a longer time to induce EMT. Moreover, the data from RNA seq reveals that before undergoing EMT, OSR2 promotes cell proliferation which is necessary for *Osr2*-related processes: tooth development, palatal production, and patterning to increase cell mass before morphological transition.

TGF-β signaling was known as the signaling pathway involved in a broad range of biological processes. The induction of TGF-β leads to upregulation of its critical downstream gene, *Sox4,* and the effect of TGF-β is mediated by SOX4. It is demonstrated that SOX4 is the developmental transcription factor that is required for EMT. SOX4 controls the expression of the Ehz2 gene, which encodes a subunit in a polycomb complex that modifies chromatin structure for gene repression. Indeed, inhibition of *Sox4* delays EMT induction by TGF-β [Tiwari et al., 2013; Vervoort et al., 2013]. While SOX4 is located downstream of the TGF-β signaling pathway, OSR2 triggers EMT through TGF-β signaling and thus is upstream of the TGF-β signaling pathway. The result of this study reveals a new relationship between TGF-βsignaling pathways and transcription factors to control the biological processes of EMT.

2. Function of *Osr2* and TGF-β signaling in MET a required transition during MEF reprogramming

It is well established that MEFs transit to the epithelial state from the mesenchymal state in the early phase of reprogramming. From this work, OSR2 was identified as the mesenchyme-transcriptional regulator, and its suppression leads to enhanced reprogramming through reducing TGF-β signaling. This is consistent with the role of OSR2

in EMT, the reverse process of MET. However, there are reports that cells progress through not only MET but also EMT during reprogramming [Peinado et al., 2003; Liu et al., 2013]. Therefore, the molecular mechanisms between MET/EMT, TGF-β signaling, and *Osr2* during reprogramming remain to be further investigated to understand their precise causal relationships.

It is clear that MET is inhibited by inducing its reverse process, EMT, by the addition of TGF-β or expressing transcription factor SNAIL1 [Li et al., 2010]. During iPSCs generation, MET could be triggered by gene network regulation which is indicated by the loss of mesenchymal and the gain of epithelial characteristics. In the initial stages of MEF reprogramming, transcription factors SOX2, OCT4, and c-MYC suppress mesenchyme-related genes such as *Tgfb1*, *Tgfbr2, Zeb2*, and *Snai1*, whereas KLF4 induces epithelial gene expression such as *Cdh1* encoding E-cadherin [Ichida et al., 2009; Maherali and Hochedlinger, 2009]. However, MET during reprogramming is a complicated process, and the relationship between TGF-β signaling, transcription factors, and MET has remained unclear. There are apparently different or even conflicting regulatory roles of Yamanaka factors for TGF-β1 and TGF-β receptor 2 [Li et al., 2010]. In addition, OCT4 increases *Snai2* and downregulates *Cdh1* expression, whereas KLF4 and SOX2 show the opposite effects [Liu et al., 2013]. Although a report shows the adverse effects of EMT on reprogramming, sequential EMT/MET is an essential process during the reprogramming of MEFs. In fact, induction of EMT by the addition of TGF-β for a limited period increases the efficiency of subsequent reprogramming [Li et al., 2010]. In contrast, blocking EMT with a TGF-β inhibitor reduces subsequent MET and the efficiency of reprogramming [Liu et al., 2013]. Except for the expression of OCT4 and SOX2 to suppress *Zeb2* [Wang et al., 2013], the link between TGF-β signaling and MET during reprogramming remains unclear.

3. The relationship between *Osr2* and Wnt signaling during reprogramming

Recent work describes that Wnt signaling forms a part of the regulatory network that is essential for ESCs to self-renew and maintain pluripotency [Sokol, 2011]. Although Wnt signaling is critical for self-renewal and differentiation of ESCs, mechanistic analyses of reprogramming showed that Wnt signaling also functions at an early reprogramming stage [Marson et al., 2008]. In fact, small-molecule compounds that mimic the effects of Wnt promote iPSCs production [Zhang et al., 2014]. In this thesis, *Osr2* was found that regulate

Wnt signaling beside the TGF-β pathway. Such finding provides the groundwork for exploring transcription regulators play as hub genes to manipulate cell reprogramming via signaling transmission in the future.

The results in this study show that Wnt-related genes were expressed lower than the control in *Osr2* overexpressing of iPSCs generation cells. Consistent with the fact that *Osr2* inhibits Wnt signaling in iPSCs production, it was reported that *Osr2* is correlated with the inhibition of Wnt signaling. *Osr2* upregulates Wnt antagonists, *Dkk2,* and *Sfpr2* on the lingual side of the tooth bud mesenchyme [Jia et al., 2016]. Moreover, *Osr1*, another Odd skip-related family member, suppresses SOX9 and β-catenin and prevents cancer cells from proliferation and invasion [Wang et al., 2018]. Indeed, results in this study demonstrate that in mesenchymal cells such as MEF, *Osr2* expression correlates with the reduction of Wnt signaling, while in reprogramming, the reduction of *Osr2* allows Wnt signaling upregulation during reprogramming. Thus, in addition to the maintenance of iPSCs in the pluripotent state, Wnt signaling may be activated when *Osr2* expression is reduced at an earlier reprogramming stage.

Overexpression of transcription factors and small molecules that modulate signaling pathways could fully induce reprogramming [Xie et al., 2017]. Small chemicals that target intracellular signaling can function in place of transcription factors used for reprogramming [Ebrahimi, 2015]. However, functional interactions between transcription factors and signaling pathways during reprogramming needs to be better defined. This study indicated that *Osr2* downregulation promotes reprogramming through both TGF-β and Wnt signaling pathways. These results are consistent with the research on sea urchin development, which shows that networks that consist of transcription factors and signaling pathways regulate the development of an organism [Peter, 2017]. This study suggests that downregulation of *Osr2* followed by reduction of TGF-β signaling is a critical step in erasing the mesenchymal properties of MEFs. Activation of Wnt signaling after *Osr2* downregulation may promote this process.

CONCLUSION

Somatic cells can be reprogrammed to the pluripotent stage by overexpressing transcription factors or inducing small molecules that activate cellular signaling pathways. However, a

comprehensive study on the relationship between transcription factors and signaling during reprogramming remains to be investigated. This study found *Osr2* as the transcription regulator and that its downregulation promotes efficient reprogramming through regulating TGF-β and Wnt signaling pathways. *Osr2* promotes TGF-β signaling pathway, which triggers EMT inhibiting the progression of MET during iPSC generation. Furthermore, downregulation of *Osr2* allows Wnt signaling to facilitate reprogramming. In summary, *Osr2* is the key transcriptional regulator whose downregulation in MEFs promotes disintegration of their somatic cell identity and reprogramming to iPSCs.

REFERENCES

Ahmed S, Nawshad A. 2007. Complexity in interpretation of embryonic epithelial-mesenchymal transition in response to transforming growth factor-β signaling. In: Cells Tissues Organs. NIH Public Access, p 131–145.

Alberts B. 2017. Molecular Biology of the Cell. W.W. Norton, p 1–3.

Alberts B, Hopkin K, Johnson AD, Morgan D, Raff MC, Roberts K, Walter P. 2019. Essential Cell Biology. W.W. Norton, Incorporated.

Bhowmick NA, Ghiassi M, Bakin A, Aakre M, Lundquist CA, Engel ME, Arteaga CL, Moses HL. 2001. Transforming growth factor-β1 mediates epithelial to mesenchymal transdifferentiation through a RhoA-dependent mechanism. Molecular Biology of the Cell 12: 27–36.

Biehl JK, Russell B. 2009. Introduction to Stem Cell Therapy. J Cardiovasc Nurs 24: 98.

Brabletz T. 2012. EMT and MET in Metastasis: Where Are the Cancer Stem Cells? Cancer Cell 22: 699–701.

Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R. 2012. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 150: 1209–1222.

Bui PL, Nishimura K, Seminario Mondejar G, Kumar A, Aizawa S, Murano K, Nagata K, Hayashi Y, Fukuda A, Onuma Y, Ito Y, Nakanishi M, Hisatake K. 2019. Template Activating Factor-I α Regulates Retroviral Silencing during Reprogramming. Cell Reports 29: 1909-1922.e5.

Caplan AI. 1991. Mesenchymal stem cells. Journal of Orthopaedic Research 9: 641–650.

Cartwright P, McLean C, Sheppard A, Rivett D, Jones K, Dalton S. 2005. LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development 132: 885–896.

Chen J, Guo L, Zhang L, Wu H, Yang J, Liu H, Wang X, Hu X, Gu T, Zhou Z, Liu J, Liu J, Wu H, Mao SQ, Mo K, Li Y, Lai K, Qi J, Yao H, Pan G, Xu GL, Pei D. 2013. Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet 45: 1504–1509.

Cibelli JB, Stice SL, Golueke PJ, Kane JJ, Jerry J, Blackwell C, Ponce De León FA, Robl JM. 1998. Transgenic bovine chimeric offspring produced from somatic cell-derived stem-like cells. Nature Biotechnology 16: 642–646.

David L, Polo JM. 2014. Phases of reprogramming. Stem Cell Res 12: 754–761.

Degnin CR, Schleiss MR, Cao J, Geballe AP. 1993. Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. J Virol 67: 5514–5521.

Douville JM, Cheung DYC, Herbert KL, Moffatt T, Wigle JT. 2011. Mechanisms of MEOX1 and MEOX2 regulation of the cyclin dependent kinase inhibitors p21 CIP1/WAF1 and p16 INK4a in vascular endothelial cells. PLoS ONE 6.

Ebrahimi B. 2015. Reprogramming barriers and enhancers: strategies to enhance the efficiency and kinetics of induced pluripotency. Cell Regeneration 4: 4:10.

Fu X, Xu J, Chaturvedi P, Liu H, Jiang R, Lan Y. 2017. Identification of Osr2 Transcriptional Target Genes in Palate Development. J Dent Res: 22034517719749.

Gan L, Chen S, Zhong J, Wang X, Lam EKY, Liu X, Zhang J, Zhou T, Yu J, Si J, Wang L, Jin H. 2011. ZIC1 is downregulated through promoter hypermethylation, and functions as a tumor suppressor gene in colorectal cancer. PLoS ONE 6.

Gao Y, Chen J, Li K, Wu T, Huang B, Liu W, Kou X, Zhang Y, Huang H, Jiang Y, Yao C, Liu X, Lu Z, Xu Z, Kang L, Chen J, Wang H, Cai T, Gao S. 2013. Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell 12: 453–469.

Garcia-Dominguez M, Poquet C, Garel S, Charnay P. 2003. Ebf gene function is required for coupling neuronal differentiation and cell cycle exit. Development 130: 6013–6025.

Gellibert F, Woolven J, Fouchet MH, Mathews N, Goodland H, Lovegrove V, Laroze A, Nguyen VL, Sautet S, Wang R, Janson C, Smith W, Krysa G, Boullay V, de Gouville AC, Huet S, Hartley D. 2004. Identification of 1,5-naphthyridine derivatives as a novel series of potent and selective TGF-β type I receptor inhibitors. Journal of Medicinal Chemistry 47: 4494–4506.

Gingold JA, Fidalgo M, Guallar D, Lau Z, Sun Z, Zhou H, Faiola F, Huang X, Lee D-F, Waghray A, Schaniel C, Felsenfeld DP, Lemischka IR, Wang J. 2014. A Genome-wide RNAi Screen Identifies Opposing Functions of Snai1 and Snai2 on the Nanog Dependency in Reprogramming. Molecular Cell 56: 140–152.

Gritli-Linde A. 2007. Molecular control of secondary palate development. Developmental Biology 301: 309–326.

Gurdon JB. 1960. The Developmental Capacity of Nuclei Taken from Differentiating Endoderm Cells of Xenopus laevis. Development 8: 505–526.

Gurdon JB. 1962. The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. Development 10.

Halder SK, Beauchamp RD, Datta PK. 2005. A specific inhibitor of TGF-β receptor kinase, SB-431542, as a potent antitumor agent for human cancers. Neoplasia 7: 509–521.

Hanna J, Wernig M, Markoulaki S, Sun CW, Meissner A, Cassady JP, Beard C, Brambrink T, Wu LC, Townes TM, Jaenisch R. 2007. Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science (1979) 318: 1920–1923.

Hayflick L, Moorhead PS. 1961. The serial cultivation of human diploid cell strains. Experimental Cell Research 25: 585–621.

Hennig M, Yip-Schneider MT, Wentz S, Wu H, Hekmatyar SK, Klein P, Bansal N, Max Schmidt C. 2010. Targeting mitogen-activated protein Kinase Kinase with the inhibitor pd0325901 decreases hepatocellular carcinoma growth in vitro and in mouse model systems. Hepatology 51: 1218–1225.

Hescheler J. 2019. Stem cells for regenerative medicine and anti-aging. Journal of Stem Cells & Regenerative Medicine 15: 53.

Hochedlinger K, Jaenisch R. 2002. Monoclonal mice generated by nuclear transfer from mature B and T donor cells. Nature 415: 1035–1038.

Hussein SM, Batada NN, Vuoristo S, Ching RW, Autio R, Narvä E, Ng S, Sourour M, Hämälä R, Olsson C, Lundin K, Mikkola M, Trokovic R, Peitz M, Brüstle O, Bazett-Jones DP, Alitalo K, Lahesmaa R, Nagy A, Otonkoski T. 2011. Copy number variation and selection during reprogramming to pluripotency. Nature 471: 58–62.

Ichida JK, Blanchard J, Lam K, Son EY, Chung JE, Egli D, Loh KM, Carter AC, di Giorgio FP, Koszka K, Huangfu D, Akutsu H, Liu DR, Rubin LL, Eggan K. 2009. A Small-Molecule Inhibitor of Tgf-β Signaling Replaces Sox2 in Reprogramming by Inducing Nanog. Cell Stem Cell 5: 491–503.

Inman GJ, Nicolás FJ, Callahan JF, Harling JD, Gaster LM, Reith AD, Laping NJ, Hill CS. 2002. SB-431542 is a potent and specific inhibitor of transforming growth factor-β superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. Molecular Pharmacology 62: 65–74.

Jaenisch R, Young R. 2008. Stem Cells, the Molecular Circuitry of Pluripotency and Nuclear Reprogramming. Cell 132: 567–582.

Jalali A, Zhu X, Liu C, Nawshad A. 2012. Induction of palate epithelial mesenchymal transition by transforming growth factor β3 signaling. Development Growth and Differentiation 54: 633–648.

Jia S, Kwon HJE, Lan Y, Zhou J, Liu H, Jiang R. 2016. Bmp4-Msx1 signaling and Osr2 control tooth organogenesis through antagonistic regulation of secreted Wnt antagonists. Dev Biol 420: 110–119.

Jordan N v., Prat A, Abell AN, Zawistowski JS, Sciaky N, Karginova OA, Zhou B, Golitz BT, Perou CM, Johnson GL. 2013. SWI/SNF Chromatin-Remodeling Factor Smarcd3/Baf60c Controls Epithelial-Mesenchymal Transition by Inducing Wnt5a Signaling. Molecular and Cellular Biology 33: 3011–3025.

Jukes J, Both S, Post J, Van Blitterswijk C, Karperien M, De Boer J. 2008. Stem cells. Tissue Engineering: 1–26.

Kalluri R, Weinberg RA. 2009. The basics of epithelial-mesenchymal transition. Journal of Clinical Investigation 119: 1420–1428.

Kang H, Kim H, Lee S, Youn H, Youn B. 2019. Role of metabolic reprogramming in epithelial–mesenchymal transition (EMT). International Journal of Molecular Sciences 20.

Kang L, Wang J, Zhang Y, Kou Z, Gao S. 2009. iPS Cells Can Support Full-Term Development of Tetraploid Blastocyst-Complemented Embryos. Cell Stem Cell 5: 135–138.

Kawai S, Abiko Y, Amano A. 2010. Odd-skipped related 2 regulates genes related to proliferation and development. Biochem Biophys Res Commun 398: 184–190.

Kawai S, Kato T, Inaba H, Okahashi N, Amano A. 2005. Odd-skipped related 2 splicing variants show opposite transcriptional activity. Biochem Biophys Res Commun 328: 306–311.

Kawase E, Yamazaki Y, Yagi T, Yanagimachi R, Pedersen RA. 2000. Mouse embryonic stem (ES) cell lines established from neuronal cell-derived cloned blastocysts. Genesis 28: 156–63.

Keller G. 2005. Embryonic stem cell differentiation: Emergence of a new era in biology and medicine. Genes and Development 19: 1129–1155.

King TJ, Briggs R. 1955. CHANGES IN THE NUCLEI OF DIFFERENTIATING GASTRULA CELLS, AS DEMONSTRATED BY NUCLEAR TRANSPLANTATION. Proc Natl Acad Sci U S A 41: 321.

Lamouille S, Xu J, Derynck R. 2014. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol 15: 178–196.

Lan Y. 2004. Odd-skipped related 2 (Osr2) encodes a key intrinsic regulator of secondary palate growth and morphogenesis. Development 131: 3207–3216.

Lan Y, Kingsley PD, Cho ES, Jiang R. 2001. Osr2, a new mouse gene related to Drosophila odd-skipped, exhibits dynamic expression patterns during craniofacial, limb, and kidney development. Mech Dev 107: 175–179.

Larue L, Bellacosa A. 2005. Epithelial-mesenchymal transition in development and cancer: Role of phosphatidylinositol 3′ kinase/AKT pathways. Oncogene 24: 7443–7454.

Lassar AB, Paterson BM, Weintraub H. 1986. Transfection of a DNA locus that mediates the conversion of 10T1 2 fibroblasts to myoblasts. Cell 47: 649–656.

Li R, Liang J, Ni S, Zhou T, Qing X, Li H, He W, Chen J, Li F, Zhuang Q, Qin B, Xu J, Li W, Yang J, Gan Y, Qin D, Feng S, Song H, Yang D, Zhang B, Zeng L, Lai L, Esteban MA, Pei D. 2010. A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. Cell Stem Cell 7: 51–63.

Liu X, Sun H, Qi J, Wang L, He S, Liu J, Feng C, Chen C, Li W, Guo Y, Qin D, Pan G, Chen J, Pei D, Zheng H. 2013. Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT–MET mechanism for optimal reprogramming. Nature Cell Biology 2013 15:7 15: 829–838.

Lluis F, Pedone E, Pepe S, Cosma MP. 2008. Periodic activation of Wnt/beta-catenin signaling enhances somatic cell reprogramming mediated by cell fusion. Cell Stem Cell 3: 493–507.

Mackay-Sim A, Mellick G, Wood S. 2011. Stem Cell Models for Biomarker Discovery in Brain Disease. In: International Review of Neurobiology. Academic Press Inc., p 239–257.

Maherali N, Hochedlinger K. 2009. Tgfβ Signal Inhibition Cooperates in the Induction of iPSCs and Replaces Sox2 and cMyc. Current Biology 19: 1718–1723.

Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R, Plath K, Hochedlinger K. 2007. Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. Cell Stem Cell 1: 55–70.

Mann CJ, Newman ENC, Whitney DJ, Latchem NJ, Bramke I, Burke JF. 2007. Stem and iPS cell selection: quantitation of surface marker (SSEA1) and intracellular GFP. Nature Methods 2007 4:8 4: i–ii.

Marson A, Foreman R, Chevalier B, Bilodeau S, Kahn M, Young RA, Jaenisch R. 2008. Wnt signaling promotes reprogramming of somatic cells to pluripotency. Cell Stem Cell 3: 132–135.

Miettinen PJ, Ebner R, Lopez AR, Derynck R. 1994. TGF-β induced transdifferentiation of mammary epithelial cells to mesenchymal cells: Involvement of type I receptors. Journal of Cell Biology 127: 2021–2036.

Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A. 2008. Dissecting direct reprogramming through integrative genomic analysis. Nature 454: 49–55.

Moore MAS. 2014. Hematopoietic Stem Cells. Principles of Tissue Engineering: Fourth Edition: 989–1040.

Müller AM, Huppertz S, Henschler R. 2016. Hematopoietic stem cells in regenerative medicine: Astray or on the path? Transfusion Medicine and Hemotherapy 43: 247–254.

Murakami S, Kan M, McKeehan WL, De Crombrugghe B. 2000. Up-regulation of the chondrogenic Sox9 gene by fibroblast growth factors is mediated by the mitogen-activated protein kinase pathway. Proc Natl Acad Sci U S A 97: 1113–1118.

Nishimura K, Aizawa S, Nugroho FL, Shiomitsu E, Tran YTH, Bui PL, Borisova E, Sakuragi Y, Takada H, Kurisaki A, Hayashi Y, Fukuda A, Nakanishi M, Hisatake K. 2017a. A Role for KLF4 in Promoting the Metabolic Shift via TCL1 during Induced Pluripotent Stem Cell Generation. Stem Cell Reports 8: 787–801.

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

Nishimura K, Ohtaka M, Takada H, Kurisaki A, Tran NVK, Tran YTH, Hisatake K, Sano M, Nakanishi M. 2017b. Simple and effective generation of transgene-free induced pluripotent stem cells using an auto-erasable Sendai virus vector responding to microRNA-302. Stem Cell Research 23: 13–19.

Nishimura K, Sano M, Ohtaka M, Furuta B, Umemura Y, Nakajima Y, Ikehara Y, Kobayashi T, Segawa H, Takayasu S, Sato H, Motomura K, Uchida E, Kanayasu-Toyoda T, Asashima M, Nakauchi H, Yamaguchi T, Nakanishi M. 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.

Nishimura K, Segawa H, Goto T, Morishita M, Masago A, Takahashi H, Ohmiya Y, Sakaguchi T, Asada M, Imamura T, Shimotono K, Takayama K, Yoshida T, Nakanishi M. 2007. Persistent and stable gene expression by a cytoplasmic RNA replicon based on a noncytopathic variant sendai virus. Journal of Biological Chemistry 282: 27383–27391.

Ocaña OH, Córcoles R, Fabra Á, Moreno-Bueno G, Acloque H, Vega S, Barrallo-Gimeno A, Cano A, Nieto MA. 2012. Metastatic Colonization Requires the Repression of the Epithelial-Mesenchymal Transition Inducer Prrx1. Cancer Cell 22: 709–724.

Okita K, Ichisaka T, Yamanaka S. 2007. Generation of germline-competent induced pluripotent stem cells. Nature 448: 313–317.

Pei D, Shu X, Gassama-Diagne A, Thiery JP. 2019. Mesenchymal–epithelial transition in development and reprogramming. Nature Cell Biology 21: 44–53.

Peinado H, Quintanilla M, Cano A. 2003. Transforming growth factor beta-1 induces snail transcription factor in epithelial cell lines: mechanisms for epithelial mesenchymal transitions. J Biol Chem 278: 21113–21123.

Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, Bar-Nur O, Cheloufi S, Stadtfeld M, Figueroa ME, Robinton D, Natesan S, Melnick A, Zhu J, Ramaswamy S, Hochedlinger K. 2012. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 151: 1617–1632.

Prochazkova M, Chavez MG, Prochazka J, Felfy H, Mushegyan V, Klein OD. 2015. Embryonic Versus Adult Stem Cells. Stem Cell Biology and Tissue Engineering in Dental Sciences: 249–262.

Rizzoti K, Pires C, Lovell-Badge R. 2016. Perspective on stem cells in developmental biology, with special reference to neuroendocrine systems. In: Research and Perspectives in Endocrine Interactions. Springer Verlag, p 135–156.

RL D, H W, AB L. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51: 987–1000.

Rosales DW, Mullen QN. 2010. Pluripotent stem cells. Nova Science Publishers. 265 p.

Rubin LL, Haston KM. 2011. Stem cell biology and drug discovery. BMC Biology 9: 42.

Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL. 2010. Functional genomics reveals a BMP-Driven mesenchymal-to-Epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell 7: 64–77.

Shamblott MJ, Axelman J, Wang S, Bugg EM, Littlefield JW, Donovan PJ, Blumenthal PD, Huggins GR, Gearhart JD. 1998. Derivation of pluripotent stem cells from cultured human primordial germ cells. Proc Natl Acad Sci U S A 95: 13726–13731.

Sharov AA, Masui S, Sharova L v, Piao Y, Aiba K, Matoba R, Xin L, Niwa H, Ko MSH. 2008. Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. BMC Genomics 9: 269.

Shu X, Pei D. 2014. The function and regulation of mesenchymal-to-epithelial transition in somatic cell reprogramming. Current Opinion in Genetics and Development 28: 32–37.

Singh VK, Saini A, Kalsan M, Kumar N, Chandra R. 2016. Describing the Stem Cell Potency: The Various Methods of Functional Assessment and In silico Diagnostics. Frontiers in Cell and Developmental Biology 4: 134.

Smith A. 2006. A glossary for stem-cell biology. Nature 2006 441:7097 441: 1060–1060.

Smith AG. 2001. Embryo-derived stem cells: Of mice and men. In: Annual Review of Cell and Developmental Biology. Annu Rev Cell Dev Biol, p 435–462.

Sokol SY. 2011. Maintaining embryonic stem cell pluripotency with Wnt signaling. Development 138: 4341–4350.

Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K. 2009. Role of the Murine Reprogramming Factors in the Induction of Pluripotency. Cell 136: 364–377.

Stadtfeld M, Maherali N, Breault DT, Hochedlinger K. 2008. Defining Molecular Cornerstones during Fibroblast to iPS Cell Reprogramming in Mouse. Cell Stem Cell 2: 230–240.

Štefková K, Procházková J, Pacherník J. 2015. Alkaline Phosphatase in Stem Cells.

Stone RC, Pastar I, Ojeh N, Chen V, Liu S, Garzon KI, Tomic-Canic M. 2016. Epithelial-Mesenchymal Transition in Tissue Repair and Fibrosis. Cell Tissue Res 365: 495.

Tada M, Takahama Y, Abe K, Nakatsuji N, Tada T. 2001. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Current Biology 11: 1553–1558.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131: 861–872.

Takahashi K, Yamanaka S. 2016. A decade of transcription factor-mediated reprogramming to pluripotency. Nat Rev Mol Cell Biol 17: 183–193.

Takahashi K, Yamanaka S. 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126: 663–676.

Tarkowski AK. 1961. Mouse chimæras developed from fused eggs. Nature 190: 857–860.

Taylor SM, Jones PA. 1979. Multiple new phenotypes induced in and 3T3 cells treated with 5-azacytidine. Cell 17: 771–779.

Thomson JA. 1998. Embryonic stem cell lines derived from human blastocysts. Science (1979) 282: 1145–1147.

Tiwari N, Tiwari VK, Waldmeier L, Balwierz PJ, Arnold P, Pachkov M, Meyer-Schaller N, Schübeler D, van Nimwegen E, Christofori G. 2013. Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming. Cancer Cell 23: 768–783.

Tran THY, Fukuda A, Aizawa S, Bui PL, Hayashi Y, Nishimura K, Hisatake K. 2018. Live cell imaging of X chromosome reactivation during somatic cell reprogramming. Biochem Biophys Rep 15: 86–92.

Unternaehrer JJ, Zhao R, Kim K, Cesana M, Powers JT, Ratanasirintrawoot S, Onder T, Shibue T, Weinberg RA, Daley GQ. 2014. The Epithelial-Mesenchymal Transition Factor SNAIL Paradoxically Enhances Reprogramming. Stem Cell Reports 3: 691–698.

Urbach A. 2004. Modeling for Lesch-Nyhan Disease by Gene Targeting in Human Embryonic Stem Cells. Stem Cells 22: 635–641.

Urry LA, Cain ML, Wasserman SA, Minorsky P v, Orr RB, Campbell NA. 2020. Campbell Biology. Pearson.

Vervoort SJ, Lourenço AR, van Boxtel R, Coffer PJ. 2013. SOX4 Mediates TGF-β-Induced Expression of Mesenchymal Markers during Mammary Cell Epithelial to Mesenchymal Transition. PLOS ONE 8: e53238.

Wang Y, Lei L, Zheng YW, Zhang L, Li ZH, Shen HY, Jiang GY, Zhang XP, Wang EH, Xu HT. 2018. Odd-skipped related 1 inhibits lung cancer proliferation and invasion by reducing Wnt signaling through the suppression of SOX9 and β-catenin. Cancer Sci 109: 1799–1810.

Weissman IL. 2000. Stem Cells: Units of Development, Units of Regeneration, and Units in Evolution. Cell 100: 157–168.

Weissman IL, Anderson DJ, Gage F. 2001. Stem and progenitor cells: Origins, phenotypes, lineage commitments, and transdifferentiations. In: Annual Review of Cell and Developmental Biology. Annu Rev Cell Dev Biol, p 387–403.

Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KHS. 1997. Viable offspring derived from fetal and adult mammalian cells. Nature 385: 810–813.

Worringer KA, Rand TA, Hayashi Y, Sami S, Takahashi K, Tanabe K, Narita M, Srivastava D, Yamanaka S. 2014. The let-7/LIN-41 pathway regulates reprogramming to human induced pluripotent stem cells by controlling expression of prodifferentiation genes. Cell Stem Cell 14: 40–52.

Xie H, Ye M, Feng R, Graf T. 2004. Stepwise reprogramming of B cells into macrophages. Cell 117: 663–676.

Xie X, Fu Y, Liu J. 2017. Chemical reprogramming and transdifferentiation. Current Opinion in Genetics & Development 46: 104–113.

Zakrzewski W, Dobrzyński M, Szymonowicz M, Rybak Z. 2019. Stem cells: past, present, and future. Stem Cell Research & Therapy 10: 68.

Zhang P, Chang WH, Fong B, Gao F, Liu C, Alam D al, Bellusci S, Lu W. 2014. Regulation of induced pluripotent stem (iPS) cell induction by Wnt/β-catenin signaling. Journal of Biological Chemistry 289: 9221–9232.

Zhang YE. 2017. Non-Smad signaling pathways of the TGF-β family. Cold Spring Harbor Perspectives in Biology 9.

ACKNOWLEDGMENT

I am grateful to be a Ph.D. student in Professor Koji Hisatake's Lab. He gave me tremendous support to do this wonderful project on Stem cells and provided invaluable guidance when I get stuck in my study. His vast knowledge, excellent advice, and kindness to me are priceless. It was a great privilege and honor to work and study under his guidance.

Massive thanks to my advisor Professor Ken Nishimura, who has provided great feedback, positive encouragement, and wholehearted support over the past 5 years. I came to know so many new things. I am so lucky to be his student. Further, my special thanks to Prof. Aya Fukuda and Prof. Yohei Hayashi who gave me a lot of valuable instructions in my research.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Satoru Takahashi, Prof. Masafumi Muratani, Prof. Naoya Yahagi, and Prof. Yukihide Watanabe for their comments, questions, and insightful advice that support my project going on track.

I would like to thank my lab member Mrs. Tomoko Nishimura, Phuong Linh, Arun, Yen, Nhi, Thuy Linh, Aizawa, Jenny, Honda, Yuya, Yamanaka, Kishimoto, Chiu, Kei, Miho, Doi, Gonzalo, Hamzah, Norie, Rie, Zhou, Jiang, and Kato for their intense discussion and insightful comments and challenging questions to support my project.

To the many Vietnamese also international friends at the University of Tsukuba as well as in the Viet-Tsukuba futsal and Tsukuba International soccer clubs. Thank you for making my time so enjoyable. Thank you to Vu Tuan Nam, Ngoc Quynh, Phuong Linh, Hai Yen, for helping me on my very first day in Japan. We experienced and overcomes difficulties, also I will never forget our funny moments.

My funding source - MEXT scholarship have lightened my financial burden, which allows me to focus more on my study. Your generosity has inspired me to help others and give back to the community. I hope one day, I will be able to assist students to achieve their goals as you have helped me.

I would like to dedicate this work to my parents, my wife, and my son who are the most significant purpose of my life. To my mother, you have always given me unconditioned love, prayers, care, and sacrifices since I was a child. Thank my wife and my son for continually being positive reinforce in my life.

Le Phuong Hoang Anh

TABLES

Table 1. DNA oligos sequence of upstream Open Reading Frame 2 (uORF2)

Oligo	Sequences
uORF2	AATCAGTTGCCGGCCTTGTGATGCAGCCGCTGGTTCTC TCGGCGAAAAAACTGTCGTCTTTGCTGACTTGCAAATA CATCCCGCCTTAA
	TTAAGGCGGGATGTATTTGCAAGTCAGCAAAGACGAC AGTTTTTTCGCCGAGAGAACCAGCGGCTGCATCACAAG GCCGGCAACTGATT

Table 2. DNA oligos sequences of shRNA for *Osr2*

Name	Sequences
sh*Osr2*.1	GATCCGTTCACAAAACTTTACATATGTTCAAGAGACATA TGTAAAGTTTTGTGAACTTTTTTGGAAAG
	GCAAGTGTTTTGAAATGTATACAAGTTCTCTGTATACAT TTCAAAACACTTGAAAAAACCTTTCTTAA
sh*Osr2*.2	GATCCATATCCATTCCAAAGAAAAGCTTCAAGAGAGCT TTTCTTTGGAATGGATATTTTTTTGGAAAG
	GTATAGGTAAGGTTTCTTTTCGAAGTTCTCTCGAAAAGA AACCTTACCTATAAAAAAACCTTTCTTAA

Table 3. Target sequences of siRNA for *Osr2*

Name	Target sequences
siOsr2-1	CAGACACTTTACCAAATCATACA
siOsr2-2	CAGTTCACAAAACTTTACATATG

Table 4. Primer sequence for quantitative PCR analyses

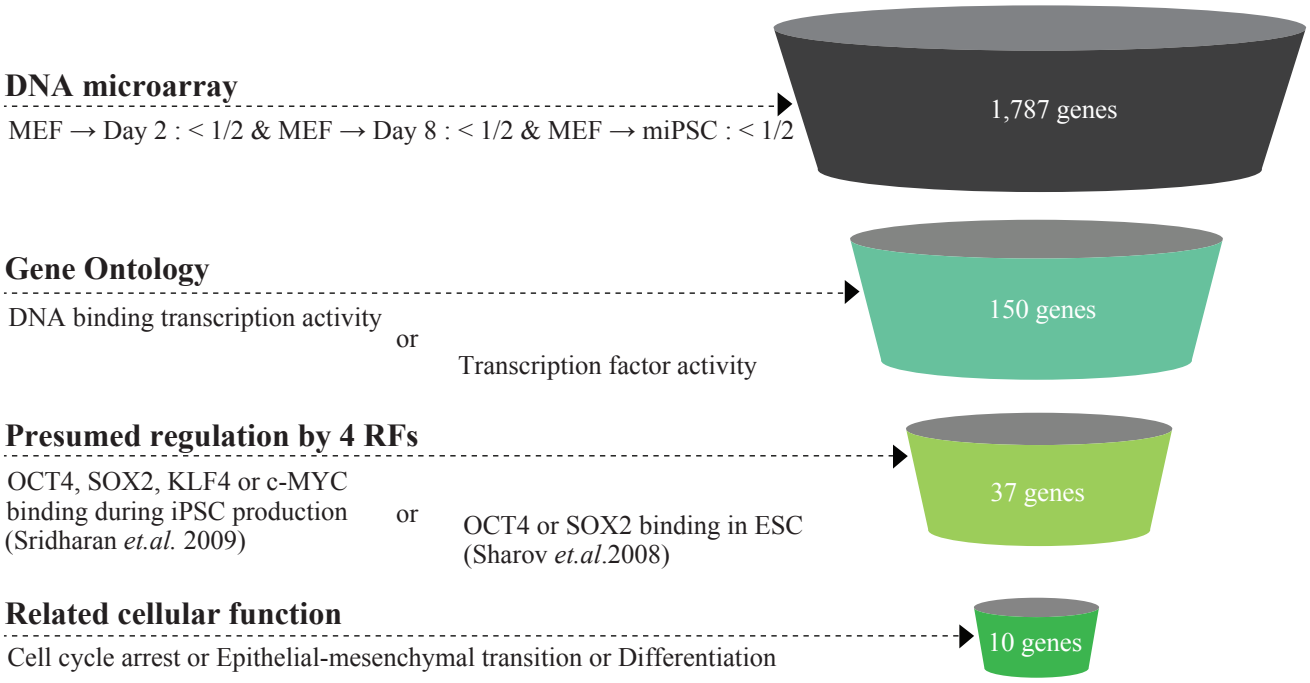
Gene	**Sequence**
Axin2	ATGGAGTCCCTCCTTACCGCAT
	GTTCCACAGGCGTCATCTCCTT
Cdh1	ACGTCCCCCTTTACTGCTG
	TATCCGCGAGCTTGAGATG
Cdh2	ATCAACCCCATCTCAGGACA
	CAATGTCAATGGGGTTCTCC
Ddit3	ACAGAGCCAGAATAACAGCCG
	GACACCGTCTCCAAGGTGAA
Ebf1	TGCTGGTCTGGAGTGAGTTG
	GATGAATCTGCCTGGTGTCC
Ebf3	AGCAATGGCGTCAGAACAG

	GCACATGATCTCGTGGGTC
endogenous Oct4	CTGTTCCCGTCACTGCTCTG
	AACCCCAAAGCTCCAGGTTC
endogenous Sox2	AGAGAAGTTTGGAGCCCGAG
	ATCTGGCGGAGAATAGTTGG
Epcam	CAAGACGACGTGGACATAGC
	GCTCTCCGTTCACTCTCAGG
Fn1	GCTCAGCAAATCGTGCAGC
	CTAGGTAGGTCCGTTCCCACT
Meox1	AAGATAGCAGCGCCTCAGGG
	GCTGTCGCTGGGAAGTCTGA
Meox2	AGGGGATTATGGCCGTCAAG
	TCCTTCCTAGGTTTGCTGTTCA
Nanog	ACCTGAGCTATAAGCAGGTTAAGAC
	GTGCTGAGCCCTTCTGAATCAGAC
Ocln	TACTGTGTGGTTGATCCCCAG
	TTTCTTCGGGTTTTCACAGC
Osr2	TGTCCAGCCACATCACATTG
	TGGTTTGATGTCCGCATCTC
Prrx1	TTCAGAACCGAAGAGCCAAG
	GATAATCGGTTGGTCTGGGA
Rex1	TTGATGGCTGCGAGAAGAG
	ACCCAGCCTGAGGACAATC
Smarca2	GGGAAGATTCAGCCAGCACA
	CTGTGGGTGTGGACATCTAGG
Smarcd3	TGACATAGATGTGGAGGTGGAG
	TATGGACTCAATCGTCTCATGG
Snai1	TGTGTCTGCACGACCTGTG
	AGTGGGAGCAGGAGAATGG
Snai2	ACACATTGCCTTGTGTCTGC
	GCCCTCAGGTTTGATCTGTC
Tbp	GGGGAGCTGTGATGTGAAGT
	CCAGGAAATAATTCTGGCTCA
Tcf7	TACTATGAACTGGCCCGCAA
	TGCATTTCTTTTTCCTCCTGTGG
Tgfb1	TGAGTGGCTGTCTTTTGACG
	GGCTGATCCCGTTGATTTC
Tgfb2	CATCATCCCGAATAAAAGCG

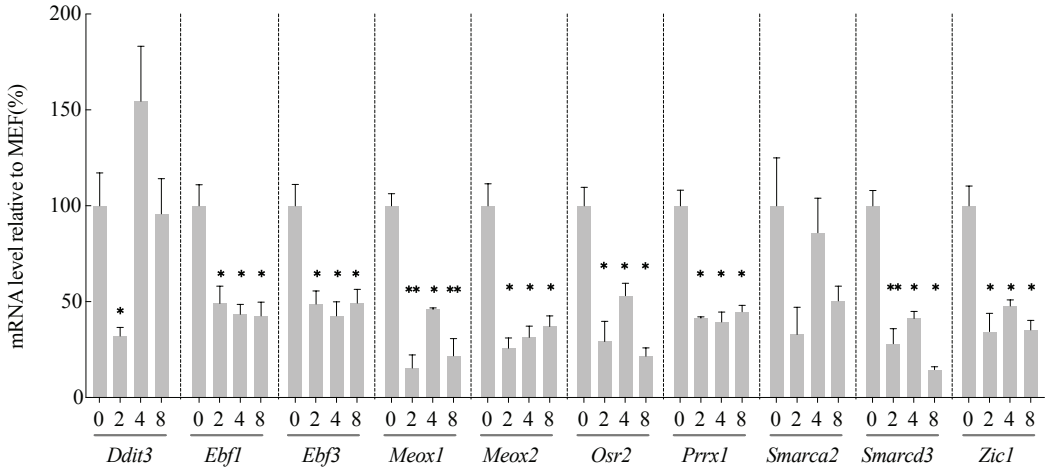
	TAGCAGGAGATGTGGGGTC
Tgfb3	GAAGGCTGCACTCAGGAGAC
	TGAGGACACATTGAAACGAA
Tgfbr1	ITGCTCCAAACCACAGAGTAGGC
	CCCAGAACACTAAGCCCATTGC
Tgfbr2	ATGAGCAACTGCAGCATCAC
	IGGCAAACCGTCTCCAGAGTA
Tgfbr3	GCCCAAAAGGAATATGGAGC
	ACTCCGCAAGGTAATTGAGC
Vim	GATGCTCCAGAGAGAGGAAGC
	TTCCGTTCAAGGTCAAGACG
Zeb2	GCTAACCCAAGGAGCAGGTAAC
	TGAACTGTAGGACCCAGAATGA
Zic1	ATGAACGTGAACATGGCTGC
	CTCCGGCTCGATCCATTTAC

FIGURES

A


DNA microarray
MEF → Day 2 : < 1/2 & MEF → Day 8 : < 1/2 & MEF → miPSC : < 1/2
1,787 genes
Gene Ontology
DNA binding transcription activity
or
Transcription factor activity
150 genes
Presumed regulation by 4 RFs
OCT4, SOX2, KLF4 or c-MYC binding during iPSC production (Sridharan et.al. 2009)
or
OCT4 or SOX2 binding in ESC (Sharov et.al.2008)
37 genes
Related cellular function
Cell cycle arrest or Epithelial-mesenchymal transition or Differentiation
10 genes

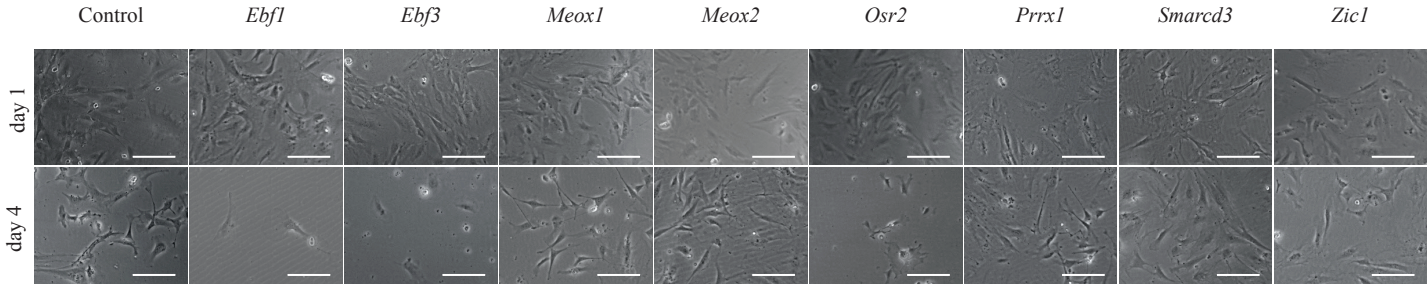
B


mRNA level relative to MEF(%)
200
150
100
50
0
0 2 4 8
Ddit3
Ebf1
Ebf3
Meox1
Meox2
Osr2
Prrx1
Smarca2
Smarcd3
Zic1

C


Retroviral vector
3xFLAG
LTR
Ψ
Gene
IRES
Puro r
LTR

D


Control
Ebf1
Ebf3
Meox1
Meox2
Osr2
Prrx1
Smarcd3
Zic1
day 1
day 4

E

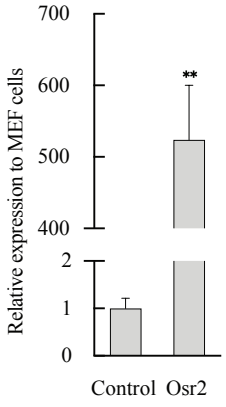
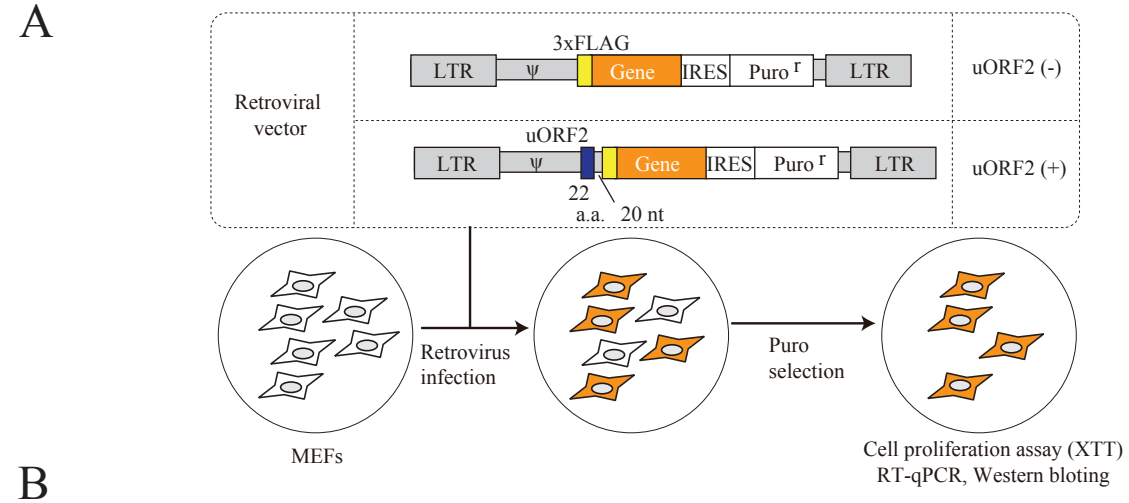

Relative expression to MEF cells
700
600
500
400
2
1
0
**
Control
Osr2

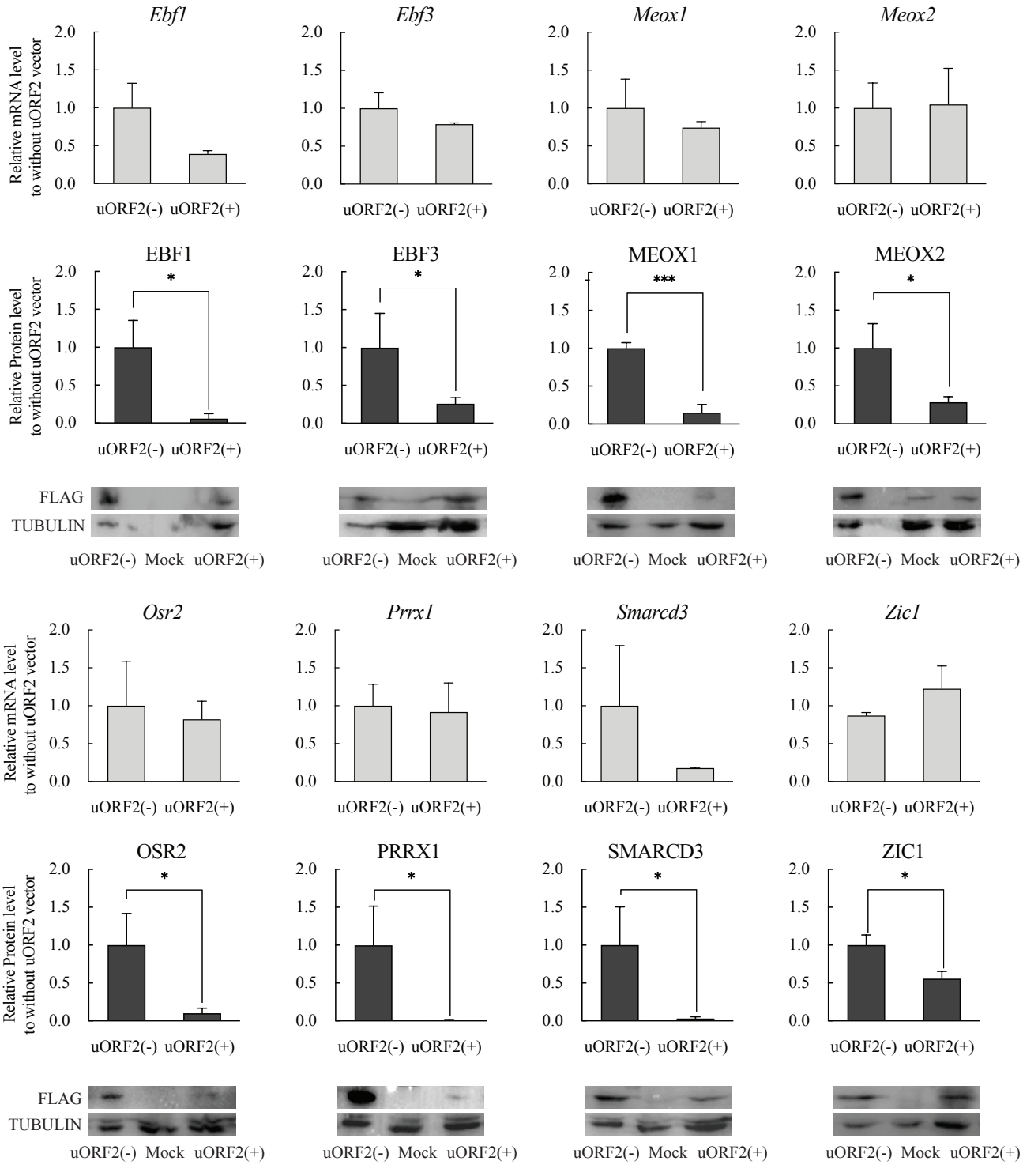
Figure 1. Screening of mesenchyme-associated transcriptional regulators that reduce the reprogramming of MEFs

(A) Flowchart of the screening mesenchyme-associated transcriptional regulators.
(B) Changes in the mRNA expression level of the 10 mesenchyme-associated transcriptional regulators during first 8 days of iPSC generation. MEFs were reprogrammed by SeVdp(KOSM) and their mRNA levels were determined at days 0, 2, 4, and 8. Data represent means ± SEM of three independent experiments. *P < 0.05, **P < 0.01 versus MEF.
(C) Structure of a retroviral vector MCsΔYY1.
(D) Growth of MEFs expressing mesenchyme-associated transcriptional regulators. MEFs were transduced with the retroviral vector expressing each transcription factor. Cells were observed at day 1 and day 4. Scalebar, 100 μm.
(E) The expression level of *Osr2*. MEFs were transduced with retroviral vector express *Osr2*, followed by puromycin selection. mRNA level of *Osr2* was determined 2 days after transduction. Data represent means ± SEM of three independent experiments. **P < 0.01, versus MEF.

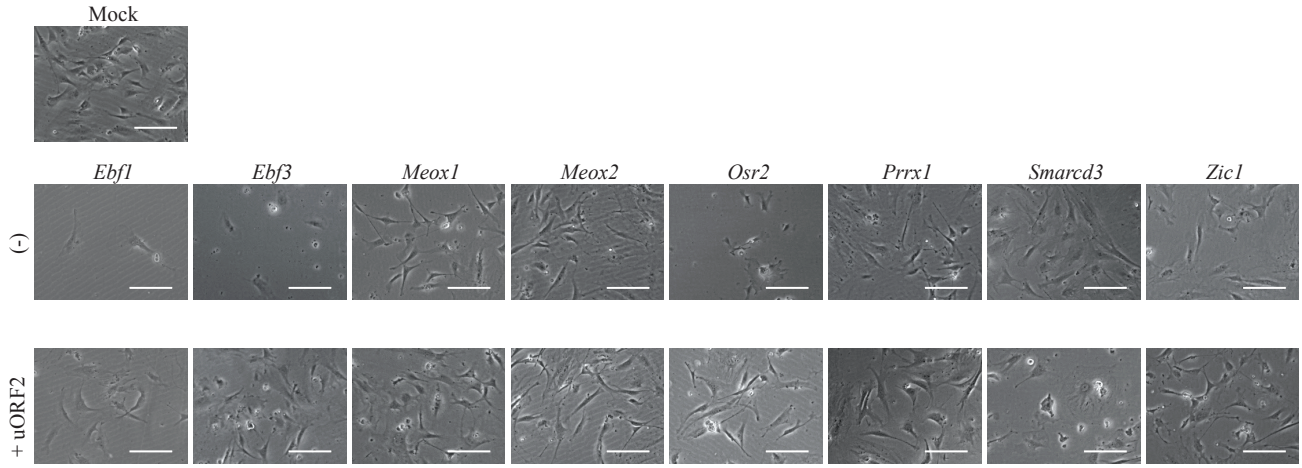
A


Retroviral vector
3xFLAG
LTR
ψ
Gene
IRES
Puro r
LTR
uORF2 (-)
uORF2
22 a.a.
20 nt
uORF2 (+)
Retrovirus infection
Puro selection
MEFs
Cell proliferation assay (XTT)
RT-qPCR, Western bloting

B


Ebf1
Ebf3
Meox1
Meox2
Relative mRNA level to without uORF2 vector
uORF2(-) uORF2(+)
EBF1
EBF3
MEOX1
MEOX2
Relative Protein level to without uORF2 vector
FLAG
TUBULIN
uORF2(-) Mock uORF2(+)
Osr2
Prrx1
Smarcd3
Zic1
OSR2
PRRX1
SMARCD3
ZIC1

C


Mock
Ebf1
Ebf3
Meox1
Meox2
Osr2
Prrx1
Smarcd3
Zic1
(-)
+ uORF2

D

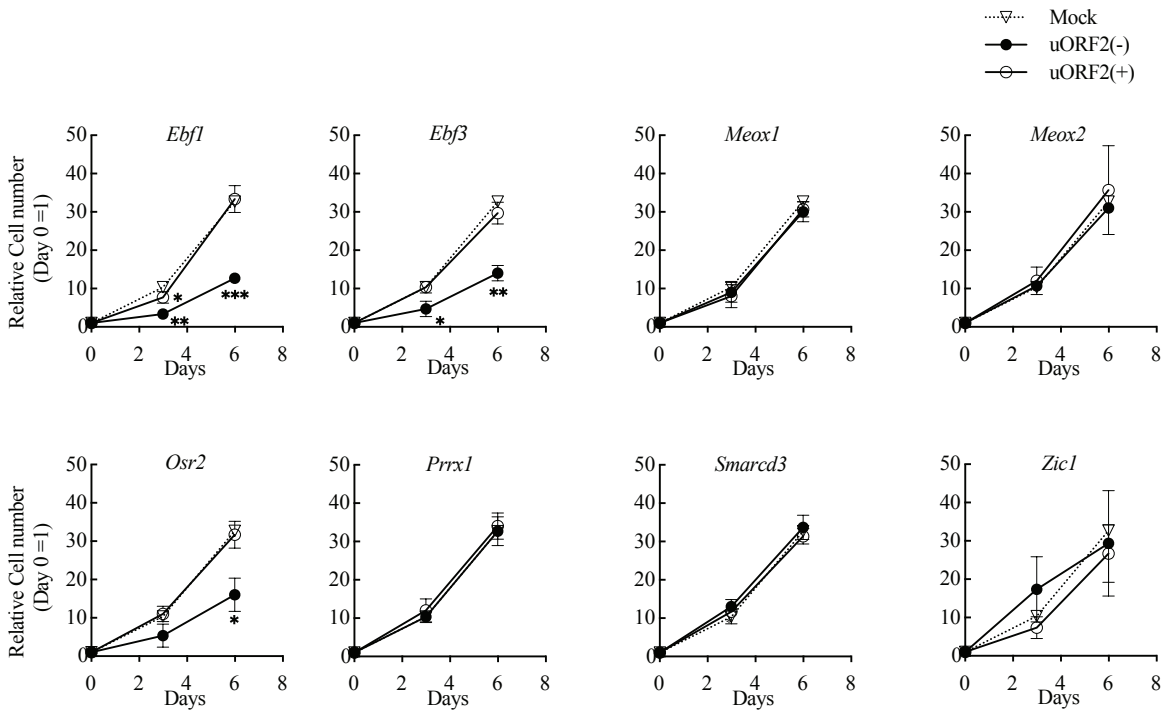
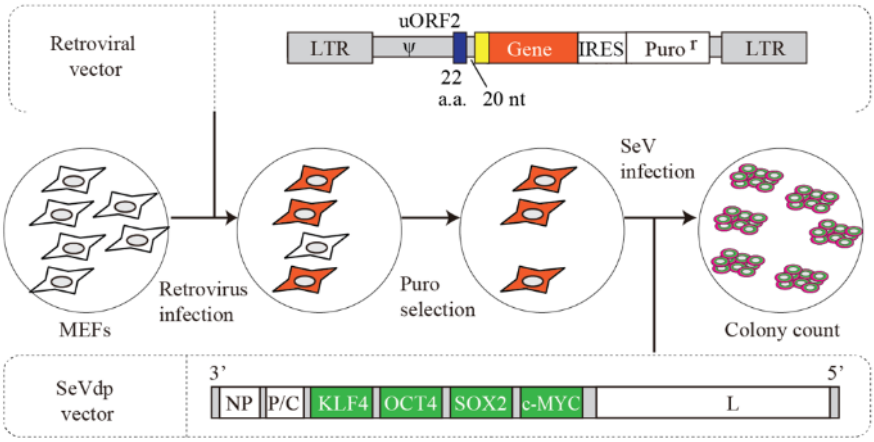

Mock
uORF2(-)
uORF2(+)
Relative Cell number (Day 0 =1)
Ebf1
Ebf3
Meox1
Meox2
Osr2
Prrx1
Smarcd3
Zic1
Days

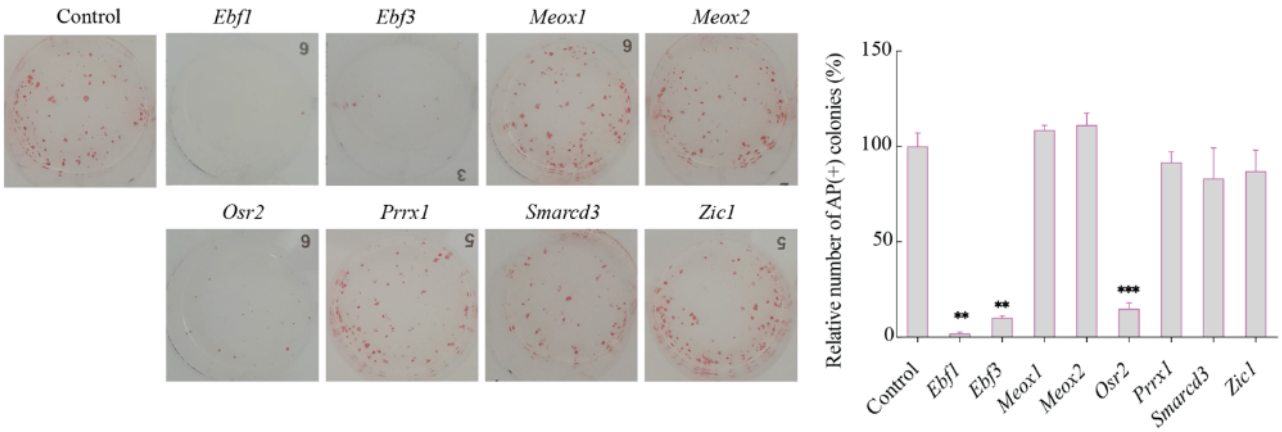
Figure 2. Reduction of toxicity by mesenchyme-associated transcriptional regulator overexpression to MEFs

(A) Experimental outline for testing the effect of expressed mesenchyme-associated transcriptional regulators on reprogramming.
(B) Changes in protein and mRNA levels by insertion of uORF2. NIH3T3 cells were transduced with standard (uORF2(-)) or uORF2-containing retroviral vector (uORF2(+)) that expresses each transcriptional regulator. Whole-cell extracts and total RNAs were prepared from NIH3T3 cells 3 days after transduction. Data represent mean + SEM from three independent experiments. *P < 0.05, ***P < 0.001 versus standard retroviral vector.
(C) Reduced cytotoxicity by lowered expression of mesenchyme-associated transcriptional regulators. MEFs were transduced with uORF2-containing retroviral vector expressing each transcriptional regulator for 2 days and selected by puromycin for 2 days. Cells were photographed after 4 days of transduction. Scale bars, 100 μm.
(D) Proliferation of MEFs expressing each mesenchyme-associated transcriptional regulator. Proliferation of cells prepared as (C) was quantified at the indicated days after transduction. Data represent mean ± SEM from three independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 versus Day 0.

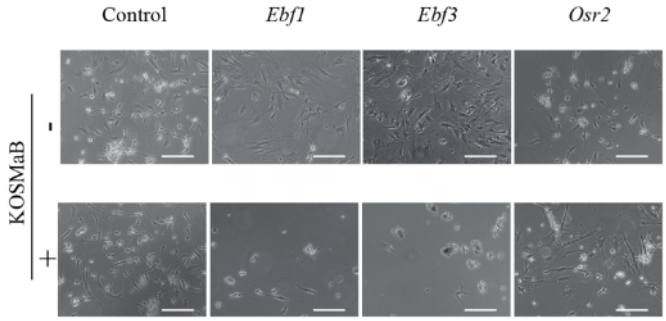
A


Retroviral vector
uORF2
LTR
ψ
Gene
IRES
Puro r
LTR
22 a.a.
20 nt
SeV infection
MEFs
Retrovirus infection
Puro selection
Colony count
SeVdp vector
3'
5'
NP
P/C
KLF4
OCT4
SOX2
c-MYC
L

B


Control
Ebf1
Ebf3
Meox1
Meox2
Osr2
Prrx1
Smarcd3
Zic1
Relative number of AP(+) colonies (%)
150
100
50
0
**
**
***

C


Control
Ebf1
Ebf3
Osr2
KOSMaB
-
+

D

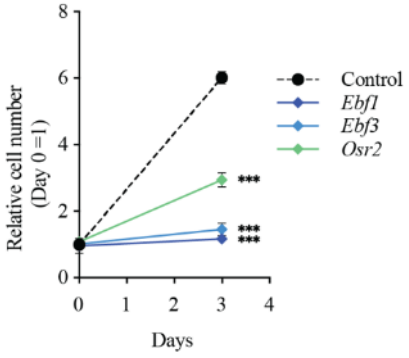

Relative cell number (Day 0 =1)
8
6
4
2
0
0
1
2
3
4
Days
Control
Ebf1
Ebf3
Osr2
***
***
***

Figure 3. Inhibitory effect of exogenous mesenchyme-associated transcription regulator's expression on iPSC generation

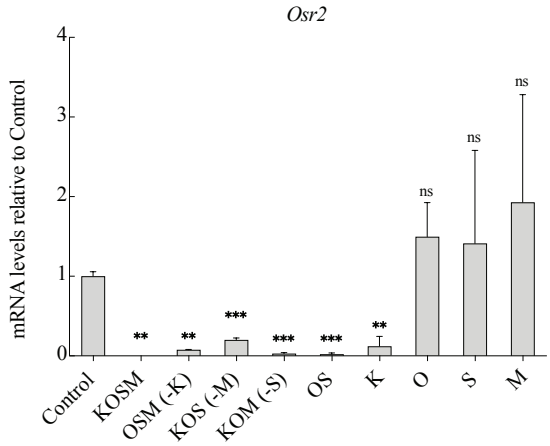
(A) Experimental outline for testing the effect of expressed mesenchyme-associated transcriptional regulators on reprogramming. ψ; Packaging signal. NP, P/C, and L; Sendai virus-derived genes.

(B) AP(+) colonies generated from reprogrammed MEFs that express a mesenchyme-associated transcriptional regulator. MEFs were transduced with uORF2-containing retroviral vector expressing each transcriptional regulator and then reprogrammed by SeVdp(KOSM), and iPSC colonies were stained for AP and counted at day 10 of reprogramming. Data represent mean ± SEM from three independent experiments. **P < 0.01, ***P < 0.001 versus control retroviral vector. Left panels; photos of iPSC colonies after AP staining.

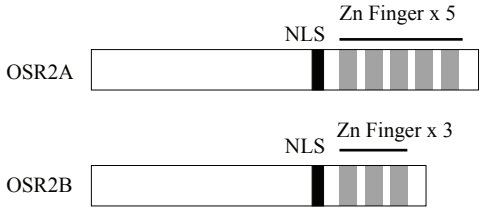
(C) Cell death of MEFs expressing *Ebf1*, *Ebf3*, or *Osr2* upon reprogramming. MEFs transduced as (B) were infected with or without SeVdp(KOSMaB) vector. Cells were observed 5 days after SeVdp vector infection. Scale bars, 100 μm.

(D) Proliferation of MEFs expressing *Ebf1*, *Ebf3*, or *Osr2* upon reprogramming. Proliferation of reprogrammed MEFs prepared as (C) was quantified before or 3 days after reprogramming. Data represent mean ± SEM from three independent experiments. ***P < 0.001 versus control retroviral vector.

A


Osr2
mRNA levels relative to Control
4
3
2
1
0
**
**
***
***
***
**
ns
ns
ns
Control
KOSM
OSM (-K)
KOS (-M)
KOM (-S)
OS
K
O
S
M

B


NLS
Zn Finger x 5
OSR2A
NLS
Zn Finger x 3
OSR2B

C

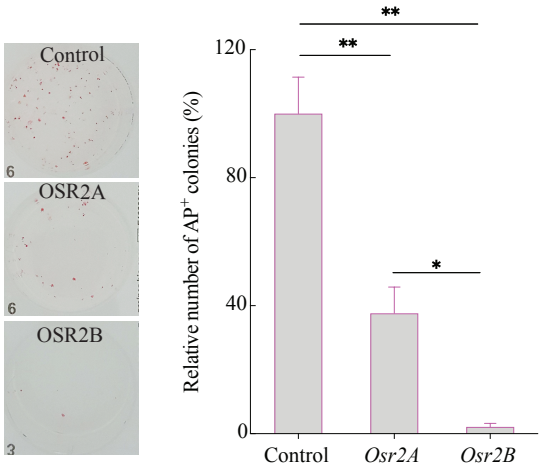

Control
6
OSR2A
6
OSR2B
3
**
**
*
120
80
40
0
Relative number of AP+ colonies (%)
Control
Osr2A
Osr2B

Figure 4. *Osr2* regulation by reprogramming factors and effect of *Osr2* isoforms on reprogramming

(A) *Osr2* regulation by different combinations of reprogramming factors. MEFs were infected with SeVdp vector expressing indicated reprogramming factors. mRNA level of *Osr2* was determined 4 days after infection. K: KLF4, O: OCT4, S: SOX2, M: c-MYC. Data represent means ± SEM of three independent experiments. **P < 0.01, ***P < 0.001 versus control MEF.
(B) Structure of OSR2 isoforms. NLS; nuclear localization signal.
(C) Induction of AP(+) colonies from MEFs expressing each OSR2 isoform. MEFs were transduced with uORF2-containing retroviral vector that expresses either OSR2A or OSR2B and then reprogrammed by SeVdp(KOSM). After 10 days of reprogramming, colonies were stained for AP and counted. Data represent mean ± SEM from three independent experiments. *P < 0.05, **P < 0.01.

Fig. 5

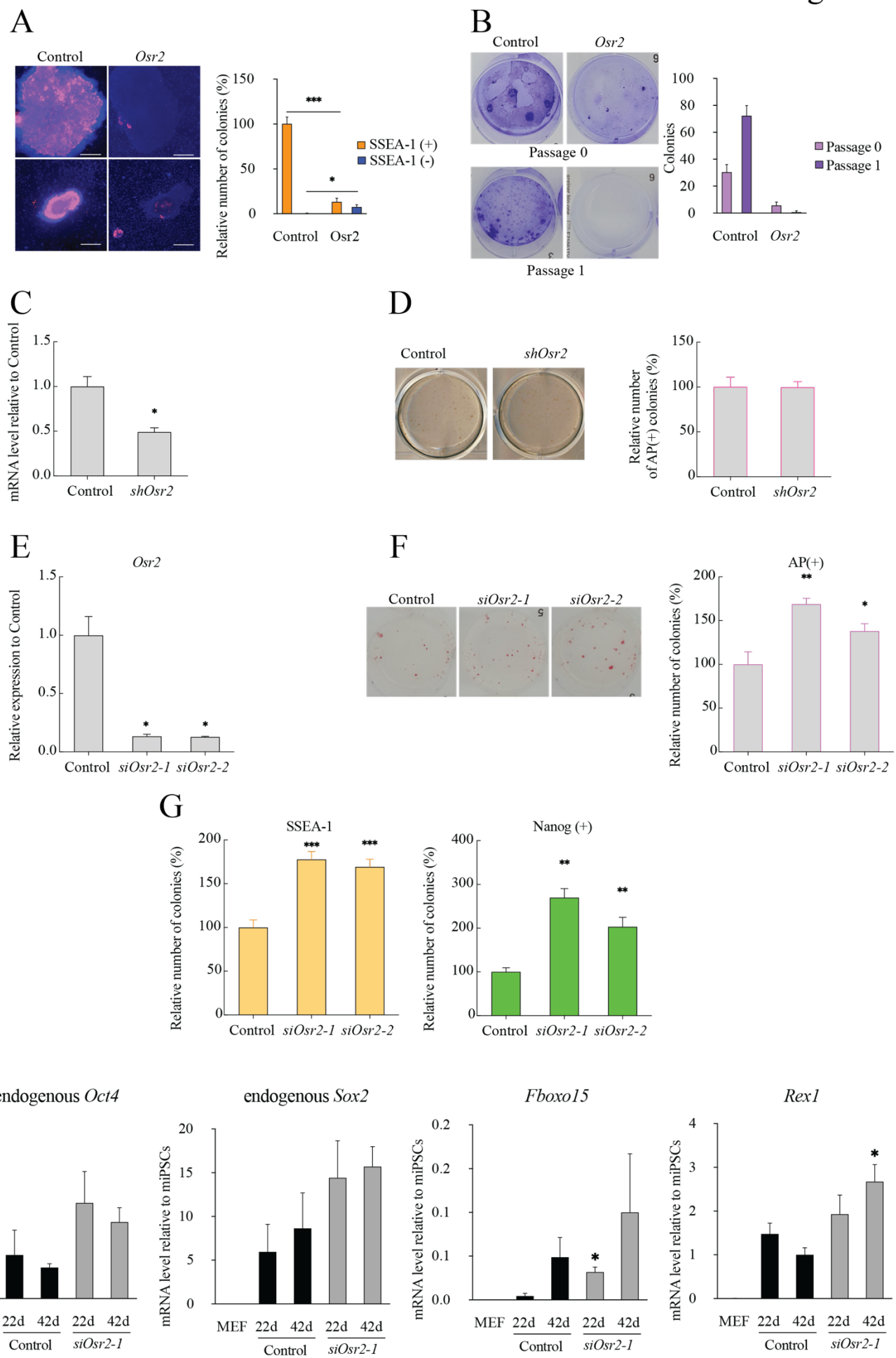

A
Control
Osr2
Relative number of colonies (%)
SSEA-1 (+)
SSEA-1 (-)
Control
Osr2
B
Control
Osr2
Passage 0
Passage 1
Colonies
Passage 0
Passage 1
C
mRNA level relative to Control
Control
shOsr2
D
Control
shOsr2
Relative number of AP(+) colonies (%)
E
Osr2
Relative expression to Control
Control
siOsr2-1
siOsr2-2
F
Control
siOsr2-1
siOsr2-2
AP(+)
Relative number of colonies (%)
G
SSEA-1
Nanog (+)
Relative number of colonies (%)
H
endogenous Oct4
endogenous Sox2
Fboxo15
Rex1
mRNA level relative to miPSCs
MEF
22d
42d
Control
siOsr2-1

Figure 5. Relationship between *Osr2* expression level and pluripotency induction

(A) Expression of SSEA1. MEFs expressing *Osr2* are reprogrammed by SeVdp(KOSM). SSEA1 was detected by immunofluorescence staining at day 12, and SSEA1-positive and -negative colonies were counted. Scale bars, 100 μm. Data represent mean ± SEM from three independent experiments. *P < 0.05, ***P < 0.001.

(B) Growth of iPSC colonies before and after passage. MEFs were reprogrammed as described in (A), and the programmed cells were stained by crystal violet before (Day 22) or after (Day 27) passage. Data represent mean ± SEM from three independent experiments. **P < 0.01.

(C) *Osr2* level after knocking down by shRNA. MEFs, transduced with retroviral vector expressing shRNA for *Osr2*, were selected with puromycin for 2 days, Total RNA was collected at day 4 after transduction. Data represent mean ± SEM from three independent experiments. *P < 0.05.

(D) Reprogramming of cells expressing shRNA for *Osr2*. Cells prepared in (C) were reprogrammed by SeVdp(KOSM), and AP(+) colonies were counted at day 10 of reprogramming. Data represent mean ± SEM from three independent experiments.

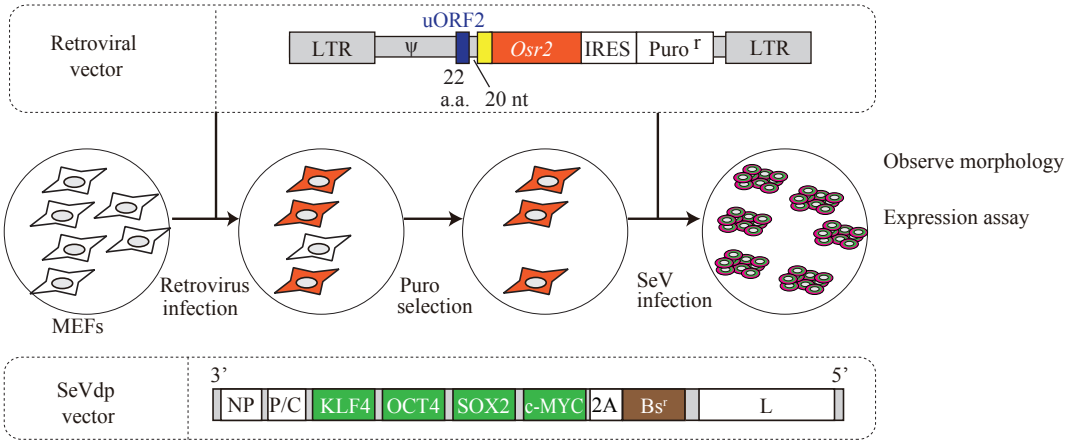
(E) *Osr2* level after knocking down by siRNA. Total RNA from MEFs transfected with siRNA for *Osr2* was collected 4 days after transduction. Data represent mean ± SEM from three independent experiments. *P < 0.05.

(F) Reprogramming of MEFs whose *Osr2* expression is knockdown by siRNA. Cells prepared in (E) were reprogrammed by SeVdp(KOSM), and AP(+) colonies were counted at day 10 of reprogramming. Data represent mean ± SEM from three independent experiments. *P < 0.05, **P < 0.01 versus control retroviral vector. Left panels; photos of iPSC colonies after AP staining.

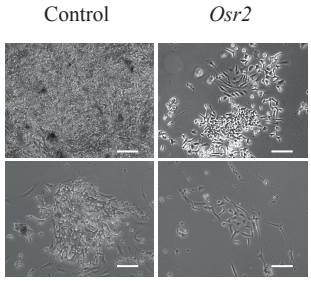
(G) Increased number of SSEA1(+) and Nanog (+) colonies by knockdown of *Osr2*. MEFs treated with siOsr2 were reprogrammed by SeVdp(KOSM). AP(+) or Nanog-GFP (+) iPSC colonies were counted at day 10 or 23 of reprogramming, respectively. All of the data represent mean ± SEM from three independent experiments. * P<0.05, **P < 0.01 versus reprogramming without *Osr2* knockdown.

(H) Changes in the mRNA expression level of pluripotency-related genes. MEFs were reprogrammed as described in (G), and then the mRNA levels of indicated genes were determined at days 22 or 42 of reprogramming. Data represent mean ± SEM from three independent experiments. *P <0.05 versus control reprogramming in each date.

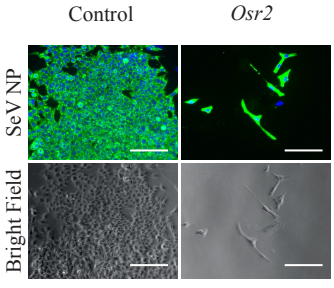
A


Retroviral vector
uORF2
LTR
ψ
Osr2
IRES
Puro r
LTR
22 a.a.
20 nt
MEFs
Retrovirus infection
Puro selection
SeV infection
Observe morphology
Expression assay
SeVdp vector
3'
5'
NP
P/C
KLF4
OCT4
SOX2
c-MYC
2A
Bsr
L

B


Control
Osr2

C


Control
Osr2
SeV NP
Bright Field

D

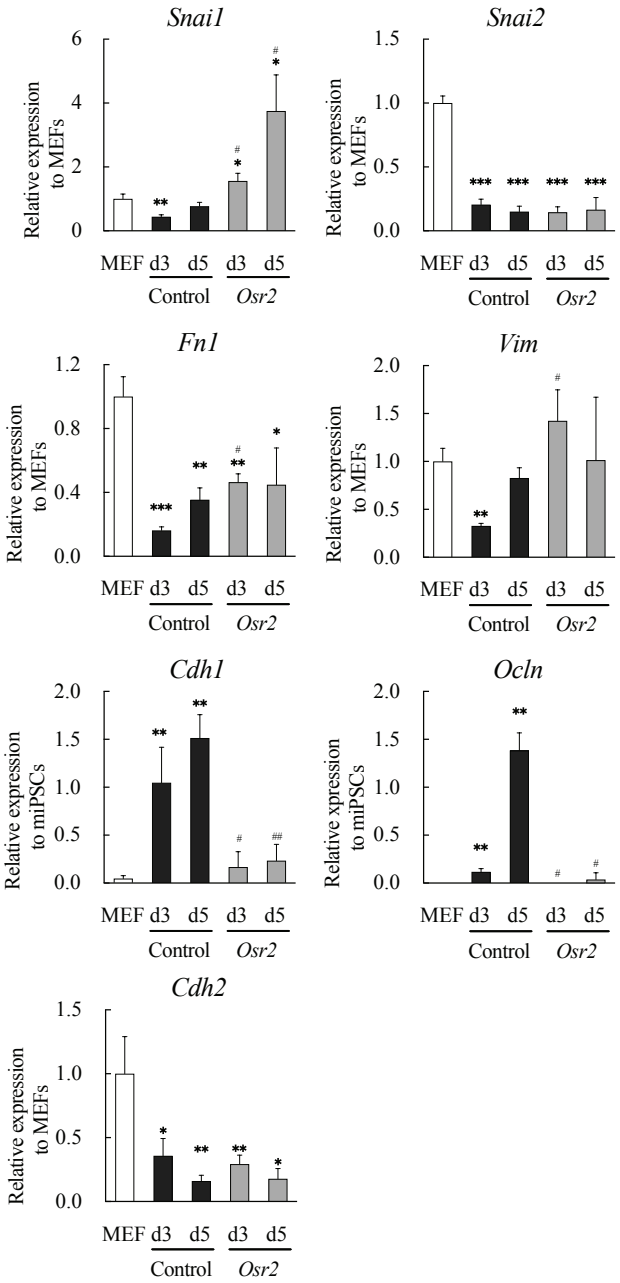

Snai1
Snai2
Fn1
Vim
Cdh1
Ocln
Cdh2
Relative expression to MEFs
Relative expression to miPSCs
Relative xpression to miPSCs
MEF d3 d5 d3 d5
Control Osr2

Figure 6. Inhibitory effect of exogenous *Osr2* expression on MET during reprogramming

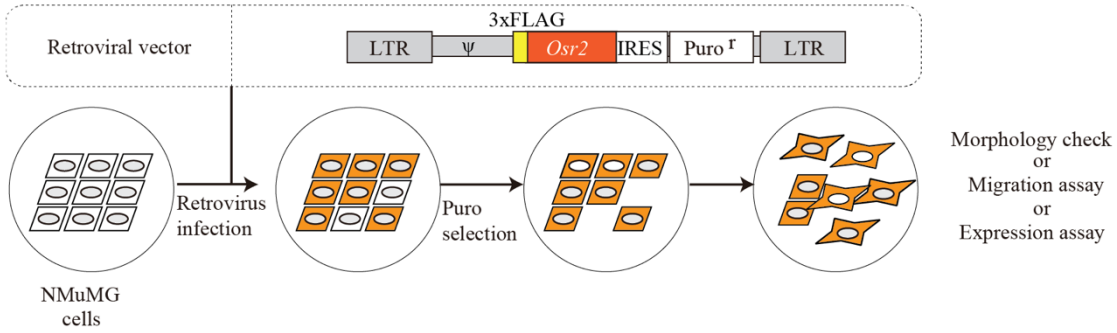
(A) Experimental outline for investigating effects of expressing *Osr2* on reprogramming. Bs^r Blasticidine S resistance gene, 2A: T2A peptide sequence.

(B) Morphology of reprogrammed cells expressing *Osr2*. *Osr2*-expressing MEFs were reprogrammed by SeVdp(KOSMaB) with Blasticidin S. Cells were observed at day 5 of reprogramming. Scale bars, 100 μm.

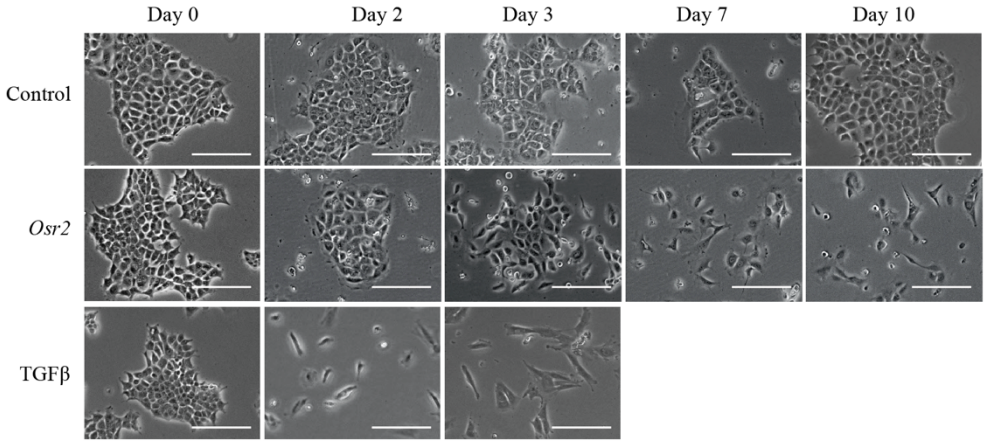
(C) SeV-NP expression in reprogrammed cells. Cells prepared as (B) were stained by anti-SeV-NP antibody at day 5 of reprogramming. Scale bars, 100 μm.

(D) Changes in the mRNA expression level of EMT-related genes. MEFs were reprogrammed as described in (B), and the mRNA levels of indicated genes were determined at days 3 and 5 of reprogramming. Data represent mean ± SEM from three independent experiments. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ versus MEF or mouse iPSCs (miPSCs). $^{\#}P < 0.05$, $^{\#\#}P < 0.01$ versus control reprogramming in each date.

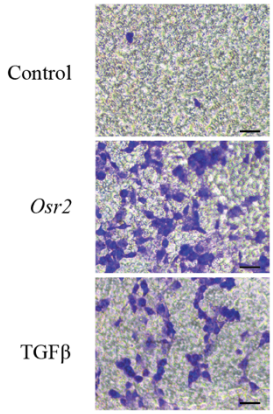
A


Retroviral vector
3xFLAG
LTR
ψ
Osr2
IRES
Puro r
LTR
Retrovirus infection
Puro selection
Morphology check
or
Migration assay
or
Expression assay
NMuMG cells

B


Day 0
Day 2
Day 3
Day 7
Day 10
Control
Osr2
TGFβ

C


Control
Osr2
TGFβ

D

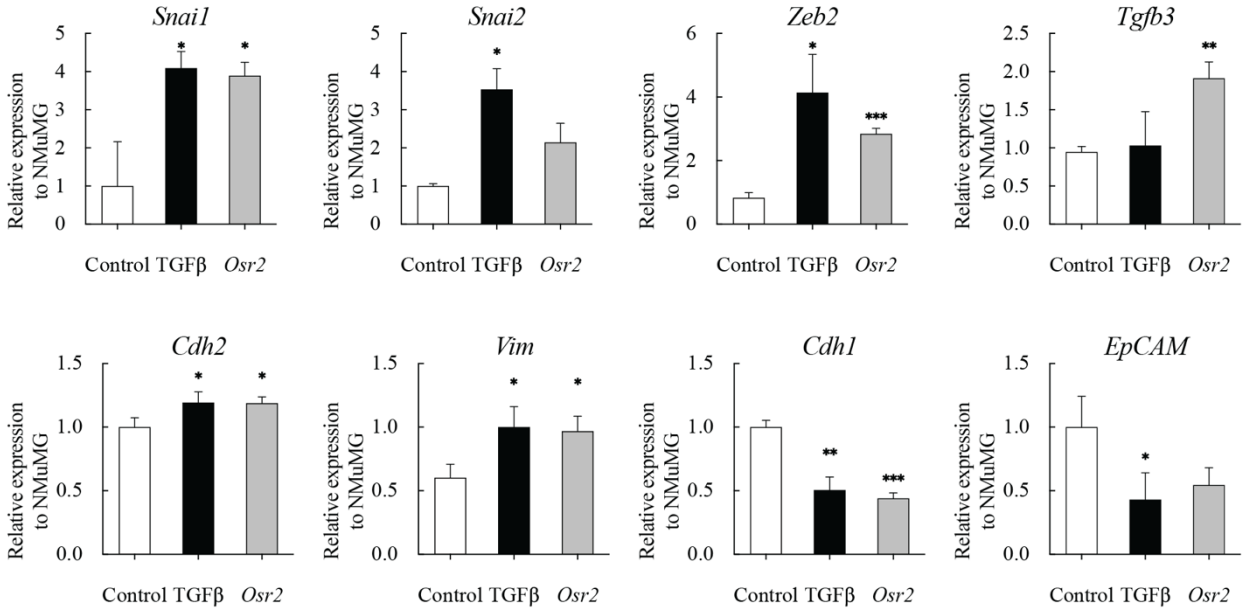

Snai1
Snai2
Zeb2
Tgfb3
Cdh2
Vim
Cdh1
EpCAM
Relative expression to NMuMG
Control TGFβ Osr2

Figure 7. Induction of EMT in NMuMG cells by OSR2

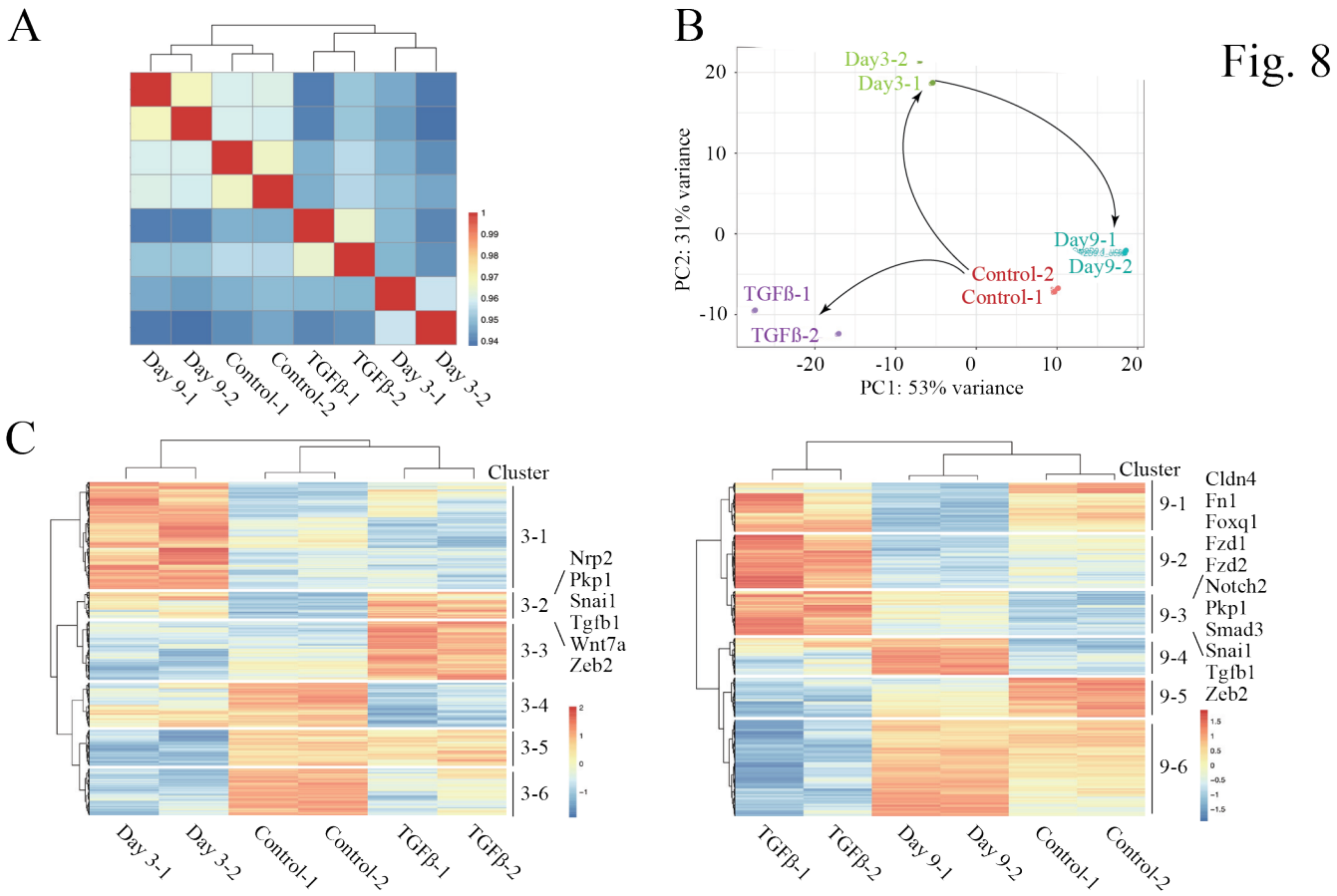
(A) Experimental outline for analyzing the effects of exogenous *Osr2* expression in NMuMG cells.

(B) Morphology of NMuMG cells expressing *Osr2*. NMuMG cells were transduced with *Osr2*-expressing retroviral vector or treated with 5 ng/mL TGF-β. Cell morphology was observed at indicated days. Scale bars, 100 μm.

(C) Cell migration assay of NMuMG cells expressing *Osr2*. NMuMG cells treated as (B) were cultured in a Transwell chamber 7 days after transduction of *Osr2*-expressing vector or one day after TGF-β treatment. Migrating cells were stained by crystal violet. Scale bars, 100 μm.

(D) mRNA expression level of EMT-related genes. NMuMG cells were treated as described in (B). mRNA levels were determined 10 days after infection or 2 days after TGF-β treatment. Data represent mean ± SEM from three independent experiments. *P < 0.05, **P <0.01, ***P < 0.001 versus control NMuMG cells.

Fig. 8



D

Osr2 Day 3

<Cluster 3-2> +*Osr2* & TGFβ: High

P-value (-log 10)

VEGFA-VEGFR2 Signaling Pathway
Focal Adhesion
MAPK Signaling Pathway
Epithelial to mesenchymal transition in colorectal cancer

<Cluster 3-1> +*Osr2* only: High

P-value (-log 10)

Retinoblastoma Gene in Cancer
Cytoplasmic Ribosomal Proteins
Translation Factors
Parkin-Ubiquitin Proteasomal System pathway
DNA Replication
Amino Acid metabolism
NRF2 pathway
DNA IR-damage and cellular response via ATR
G1 to S cell cycle control
Cell Cycle
Nucleotide Metabolism
Nuclear Receptors Meta-Pathway
Pathogenic Escherichia coli infection
Trans-sulfuration and one carbon metabolism
VEGFA-VEGFR2 Signaling Pathway
mRNA Processing

<Cluster 3-6> +*Osr2* & TGFβ: Low

p-value (-log 10)

DNA Damage Response (only ATM dependent)
Mitochondrial LC-Fatty Acid Beta-Oxidation
PTF1A related regulatory pathway

<Cluster 3-5> +*Osr2* only: Low

No highly enriched pathway

Osr2 Day 9

<Cluster 9-3> +*Osr2* & TGFβ: High

P-value (-log 10)

Focal Adhesion
Senescence and Autophagy in Cancer
TGF-B Signaling in Thyroid Cells for Epithelial-Mesenchymal Transition
Epithelial to mesenchymal transition in colorectal cancer
Mammary gland development pathway - Embryonic development
TGF-beta Receptor Signaling
Neural Crest Differentiation
Primary Focal Segmental Glomerulosclerosis FSGS
Differentiation Pathway
Nanoparticle-mediated activation of receptor signaling
TGF-beta Signaling Pathway
Osteoblast Signaling
Osteoclast Signaling
VEGFA-VEGFR2 Signaling Pathway
miR-509-3p alteration of YAP1/ECM axis
Simplified Interaction Map Between LOXL4 and Oxidative Stress Pathway
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway
miRNA targets in ECM and membrane receptors
MicroRNAs in cardiomyocyte hypertrophy
Ebola Virus Pathway on Host
Hair Follicle Development: Cytodifferentiation

<Cluster 9-4> +*Osr2* only: High

P-value (-log 10)

Matrix Metalloproteinases
Oncostatin M Signaling Pathway

<Cluster 9-5> +*Osr2* & TGFβ: Low

P-value (-log 10)

Photodynamic therapy-induced NFE2L2 (NRF2) survival signaling
Pathways in clear cell renal cell carcinoma
Selenium Micronutrient Network
Nuclear Receptors Meta-Pathway
Amino Acid metabolism
Metabolic reprogramming in colon cancer

<Cluster 9-1> +*Osr2* only: Low

P-value (-log 10)

Trans-sulfuration and one carbon metabolism
One Carbon Metabolism
ErbB Signaling Pathway
Trans-sulfuration pathway
Retinoblastoma Gene in Cancer
Cell Cycle
Gastric Cancer Network 1
Disorders of Folate Metabolism and Transport
Regulation of sister chromatid separation at the metaphase-anaphase transition
TGF-beta Signaling Pathway
Non-small cell lung cancer
Folate Metabolism

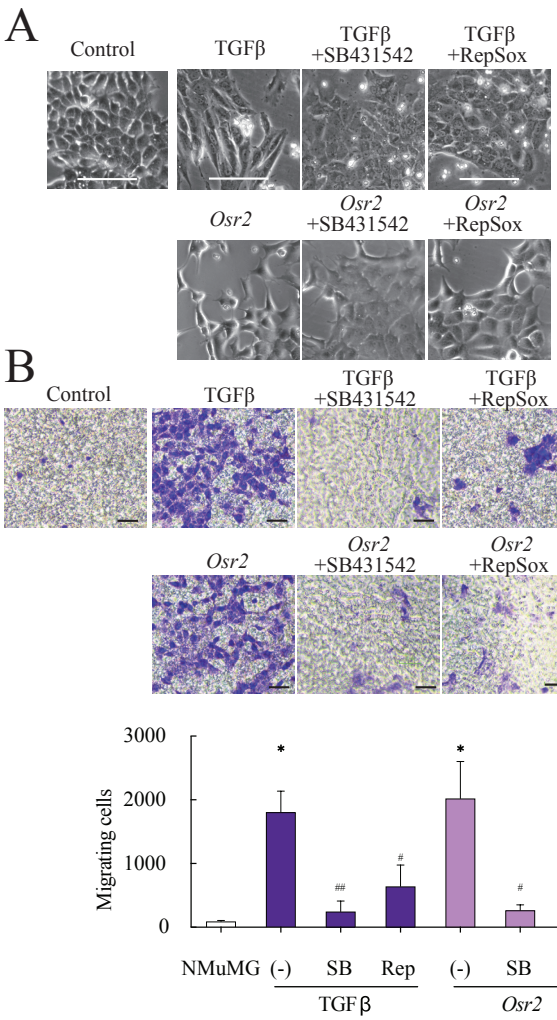
Figure 8. Global gene expression profiles of NMuMG cells expressing *Osr2*

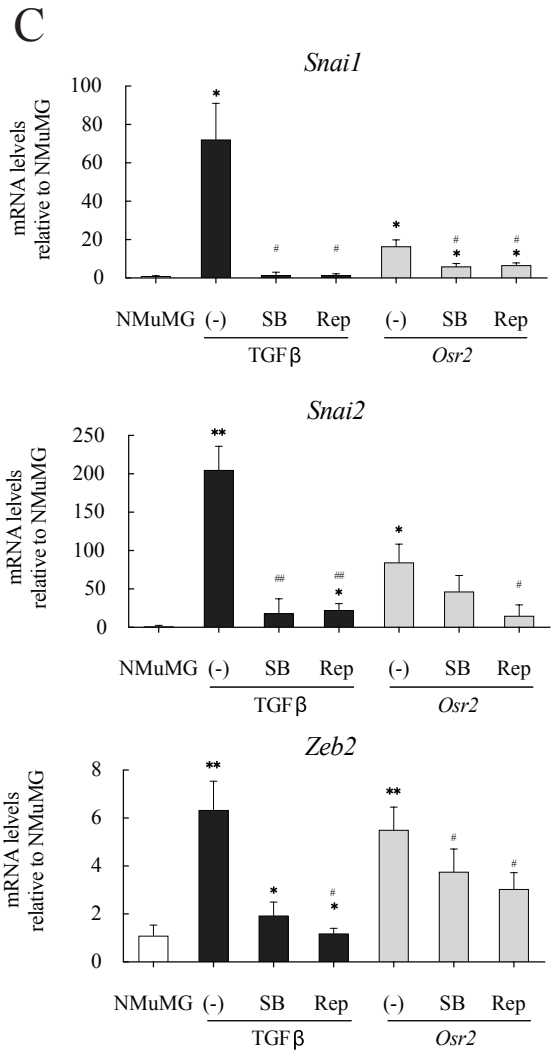
(A) Hierarchical clustering of NMuMG cells expressing *Osr2* or treated with TGF-β, based on expression profiles from RNA-seq.

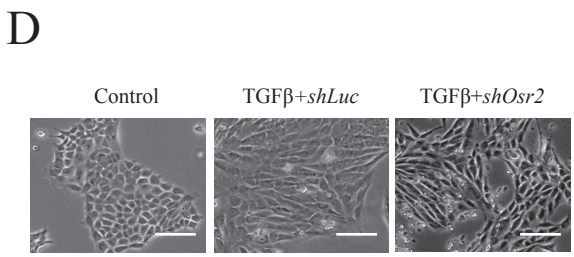
(B) PCA of NMuMG cells expressing *Osr2* or treated with TGF-β. The expression profiles of DEGs among Control, Day3 or Day9, and TGF-β, samples ($P < 0.05$) were used for the PCA.

(C) Hierarchical clustering of DEGs. DEGs among Control, Day3 or Day9, and TGF-β samples were clustered based on expression profiles. EMT-related genes induced both by exogenous *Osr2* expression and TGF-β treatment are shown on the right side of each panel.

(D) Enriched pathways in the selected cluster. Highly enriched pathways ($P < 0.05$) in clusters 3-1, 3-2, 3-5, 3-6, 9-1, 9-3, 9-4, or 9-5 in (C) are shown.


A
Control
TGFβ
TGFβ +SB431542
TGFβ +RepSox
Osr2
Osr2 +SB431542
Osr2 +RepSox
B
Control
TGFβ
TGFβ +SB431542
TGFβ +RepSox
Osr2
Osr2 +SB431542
Osr2 +RepSox
Migrating cells
3000
2000
1000
0
NMuMG (-) SB Rep (-) SB Rep
TGFβ
Osr2


C
Snai1
Snai2
Zeb2
mRNA levels relative to NMuMG
NMuMG (-) SB Rep (-) SB Rep
TGFβ
Osr2


D
Control
TGFβ+shLuc
TGFβ+shOsr2

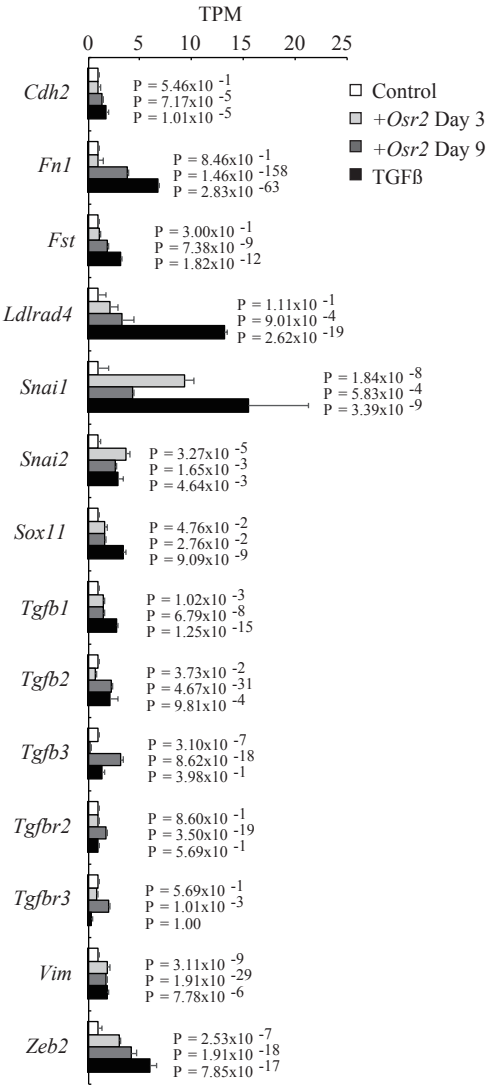

E
TPM
0 5 10 15 20 25
Control
+Osr2 Day 3
+Osr2 Day 9
TGFß
Cdh2
Fn1
Fst
Ldlrad4
Snai1
Snai2
Sox11
Tgfb1
Tgfb2
Tgfb3
Tgfbr2
Tgfbr3
Vim
Zeb2

Figure 9. EMT induced by exogenous *Osr2* expression is mediated by TGF-β signaling

(A) Morphology of *Osr2*-expressing NMuMG cells treated with TGF-β inhibitor. NMuMG cells were transduced with the retroviral vector expressing *Osr2* or treated with 5 ng/mL TGF-β1. 10 μg/mL SB431542, or 10 μg/mL. RepSox was added to the culture medium 5 days after retroviral transduction or on the same day in case of TGF-β treatment. Cell morphology was observed 2 days after treatment with the TGF-β inhibitor. Scale bars, 100 μm.

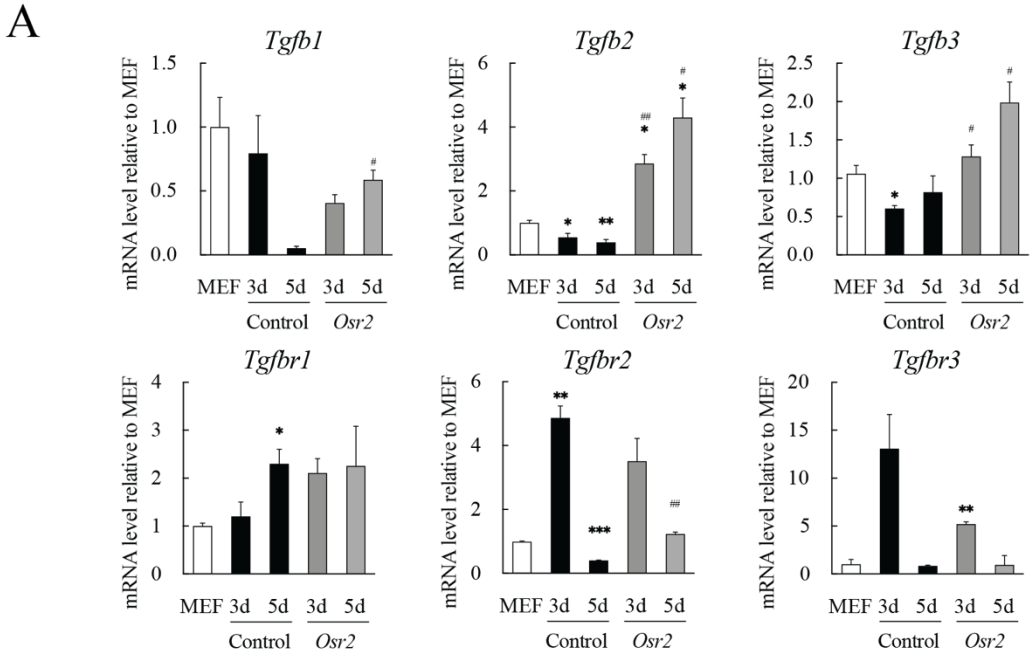
(B) Effect of TGF-β inhibitor on cell migration of *Osr2*-expressing cells. Cells prepared as described in (A) were cultured in a Transwell chamber for 16 hours. Migrating cells were stained with crystal violet and counted. Scale bars, 100 μM. Data represent mean ± SEM from three independent experiments. *P < 0.05 versus control NMuMG cells. $^{\#}$P < 0.05, $^{\#\#}$P < 0.01 versus cells without TGF-β inhibitor treatment.

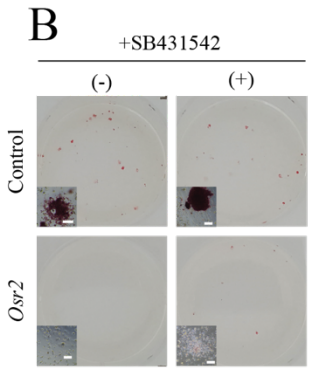
(C) mRNA expression level of EMT-related genes. mRNA levels in the cells prepared as described in (A) were determined 2 days after treatment with TGF-β inhibitor. Data represent mean ± SEM from three independent experiments. *P < 0.05, **P < 0.01 versus control NMuMG cells. $^{\#}$P < 0.05, $^{\#\#}$P < 0.01 versus cells without TGF-β inhibitor treatment.

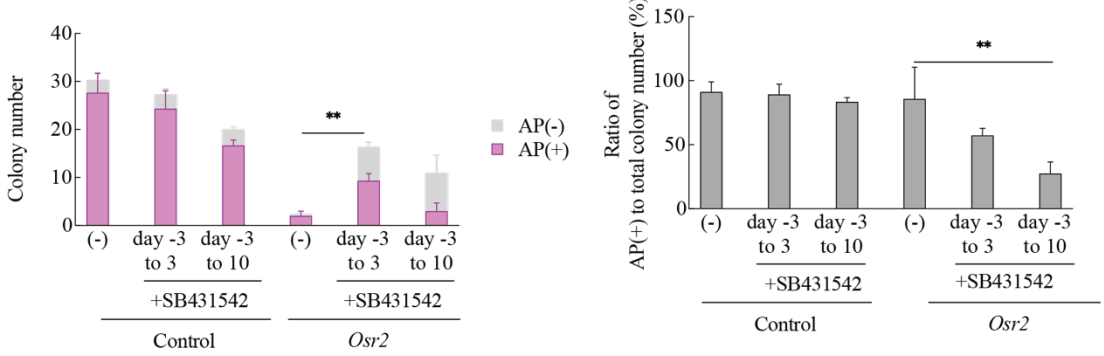
(D) Morphology of NMuMG cells treated with TGF-β after *Osr2* knockdown. NMuMG cells were transduced with retroviral vector that expresses shRNA against *Osr2* or luciferase gene and selected by puromycin. The selected cells were then treated with 5 ng/mL TGF-β1. Cell morphology was observed 2 days after TGF-β treatment. Scale bars, 100 μM.

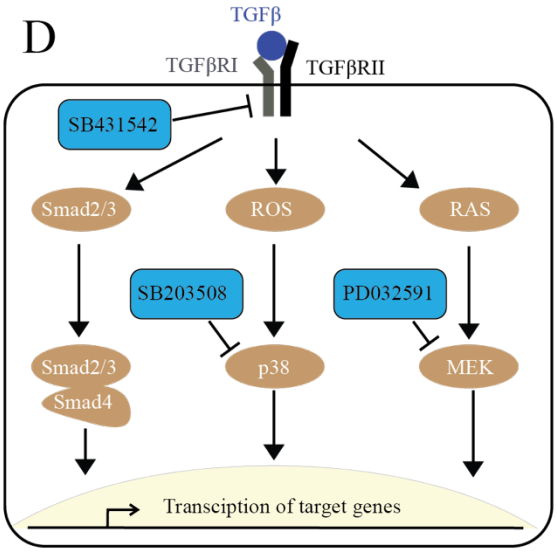
(E) Average of TPM values from RNA-seq data. mRNA levels of indicated genes were extracted from RNA-seq data in Figure 8 and normalized to those in control NMuMG cells. P-values versus control cells, calculated by DESeq2, are shown.

Fig. 10


A
Tgfb1
Tgfb2
Tgfb3
Tgfbr1
Tgfbr2
Tgfbr3
mRNA level relative to MEF
MEF 3d 5d 3d 5d
Control
Osr2


B
+SB431542
(-)
(+)
Control
Osr2


C
Colony number
AP(-)
AP(+)
(-)
day -3 to 3
day -3 to 10
+SB431542
Control
Osr2
Ratio of AP(+) to total colony number (%)


D
TGFβ
TGFβRI
TGFβRII
SB431542
Smad2/3
ROS
RAS
SB203508
PD032591
Smad2/3
Smad4
p38
MEK
Transciption of target genes

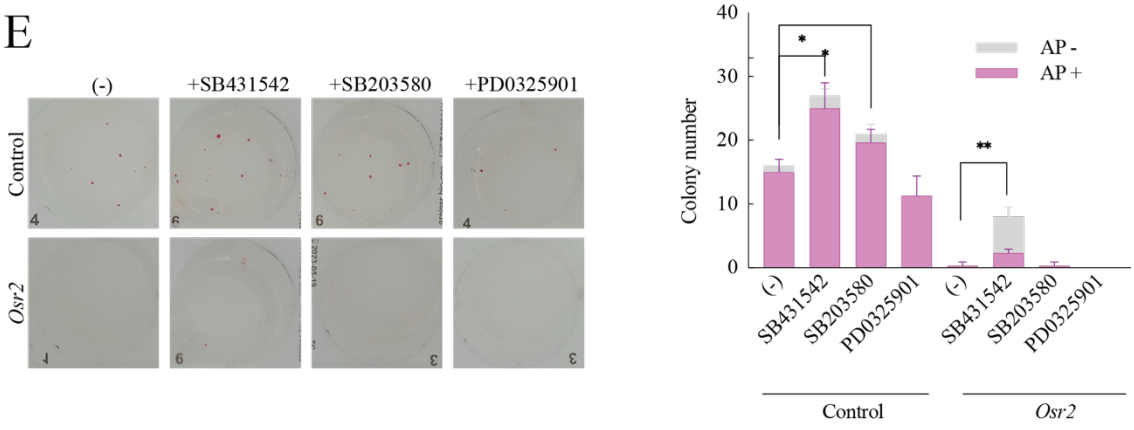

E
(-)
+SB431542
+SB203580
+PD0325901
Control
Osr2
Colony number
AP -
AP +
SB431542
SB203580
PD0325901

Figure 10. Regulation of TGF-β signaling by OSR2 during reprogramming

(A) Changes in the mRNA expression level of genes related to TGF-β signaling. *Osr2*-expressing MEFs were reprogrammed by SeVdp(KOSMaB), and, at days 3 and 5 of reprogramming, mRNA levels of indicated genes were determined. Data represent mean ± SEM from three independent experiments. *P <0.05, **P <0.01 versus MEF. $^{\#}$P<0.05, $^{\#\#}$P <0.01 versus control reprogramming at each day.

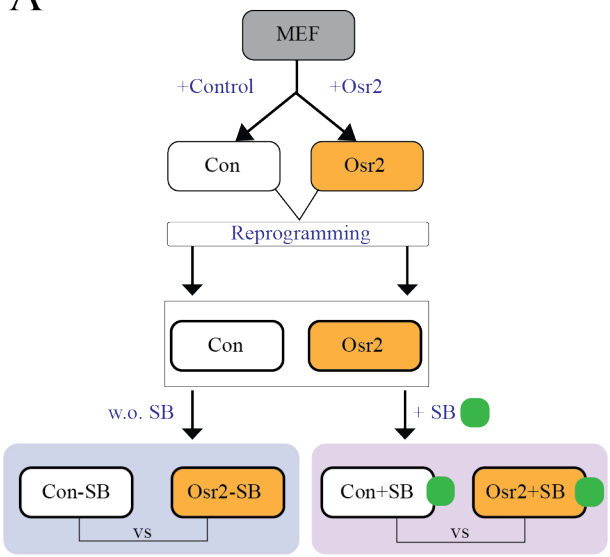
(B) Generation of iPSC colonies from MEFs expressing *Osr2* in the presence of TGF-β inhibitor. *Osr2*-expressing MEFs were reprogrammed by SeVdp(KOSM). 10 μg/mL SB431542 was added 3 days before reprogramming. iPSC colonies were stained for AP and counted at day 10 of reprogramming. Insets; enlarged images of representative colonies. Scale bars, 100 μM.

(C) AP staining of iPSC colonies with or without exogenous *Osr2* expression and TGF-β inhibitor. MEFs expressing *Osr2* were reprogrammed with or without SB431542 treatment (10 μg/mL) for indicated days. The iPSC colonies were stained for AP 10 days after reprogramming. Based on the AP staining, the number of the AP-positive and AP-negative colonies were counted separately. Bottom graph; percentage of AP(+) colonies. Data represent mean ± SEM from three independent experiments. *P < 0.05, **P < 0.01.

(D) Smad or non-Smad pathway of TGF-β signaling pathway and their inhibitors.

(E) AP staining of iPSC colonies with or without exogenous *Osr2* expression and non-Smad TGF-β inhibitors. MEFs expressing *Osr2* were treated with 10 μg/mL SB431542, 5 μM SB203580, or 1 μM PD0325901 from 3 days before reprogramming by SeVdp(KOSM). At day 10 of reprogramming, the colonies were stained for AP. Data represent mean ± SEM from three independent experiments. *P < 0.05, **P < 0.01.

A



B

Rank	Pathway	P-value
1	DNA Replication	2.9E-17
2	mRNA processing	8.2E-07
3	G1 to S cell cycle control	1.5E-06
4	Purine metabolism	2.9E-05
5	Wnt Signaling Pathway NetPath	8.7E-05
6	Lung fibrosis	6.4E-04
7	Endochondral Ossification	2.2E-03
8	Focal Adhesion	2.7E-03
9	Factors and pathways affecting insulin-like growth factor (IGF1)-Akt signaling	2.8E-03
10	PluriNetWork	3.8E-03
11	Id Signaling Pathway	4.3E-03
12	Mismatch repair	4.6E-03
13	p53 signaling	5.4E-03
14	Homologous recombination	6.1E-03
15	Spinal Cord Injury	6.1E-03
16	Hedgehog Signaling Pathway	8.8E-03
17	Serotonin and anxiety	9.3E-03
18	Nucleotide Metabolism	1.3E-02
19	Eukaryotic Transcription Initiation	1.6E-02
20	Matrix Metalloproteinases	1.7E-02

Con-SB vs Osr2-SB
>3-fold change

Rank	Pathway	P-value
1	mRNA processing	6.2E-43
2	TNF-alpha NF-kB Signaling Pathway	5.5E-18
3	Insulin Signaling	3.3E-13
4	PluriNetWork	1.4E-12
5	EGFR1 Signaling Pathway	2.5E-11
6	Proteasome Degradation	7.7E-10
7	Cytoplasmic Ribosomal Proteins	2.2E-09
8	DNA Replication	3.1E-09
9	IL-6 signaling Pathway	2.4E-08
10	G1 to S cell cycle control	1.0E-07
11	MicroRNAs in Cardiomyocyte Hypertrophy	6.5E-07
12	MAPK signaling pathway	9.4E-07
13	Eukaryotic Transcription Initiation	4.7E-06
14	Estrogen signaling	4.9E-06
15	Wnt Signaling Pathway NetPath	5.1E-06
16	TCA Cycle	9.6E-06
17	Integrin-mediated Cell Adhesion	1.2E-05
18	IL-7 Signaling Pathway	1.4E-05
19	p53 signaling	2.6E-05
20	IL-3 Signaling Pathway	3.0E-05

Con+SB vs Osr2+SB
>3-fold change

C

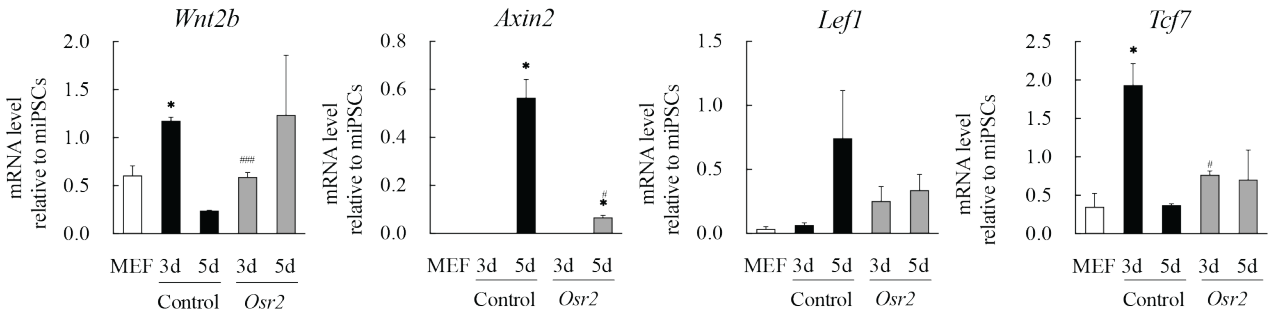
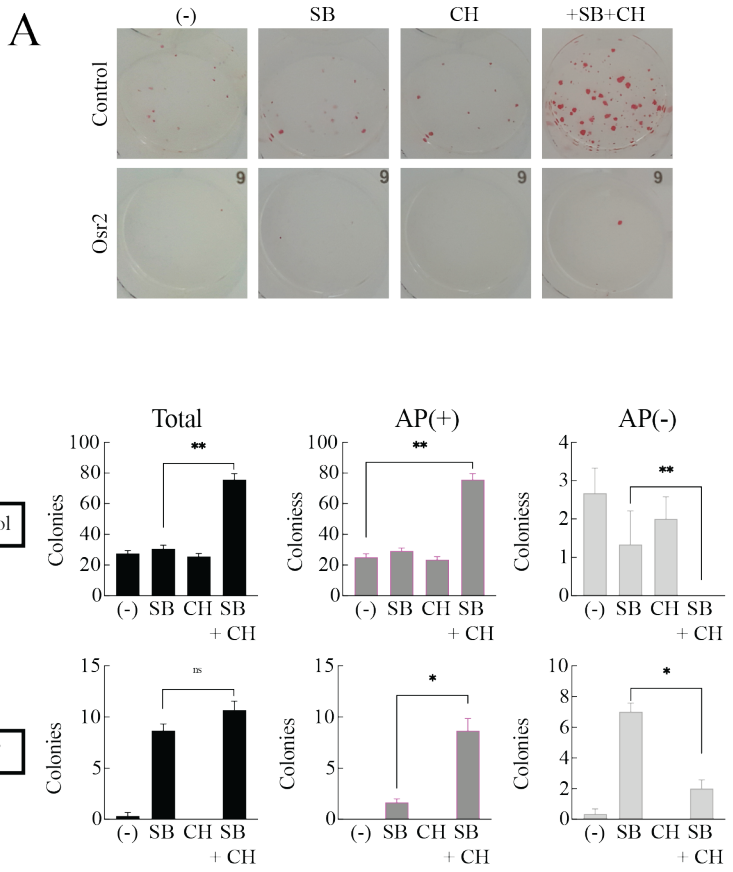


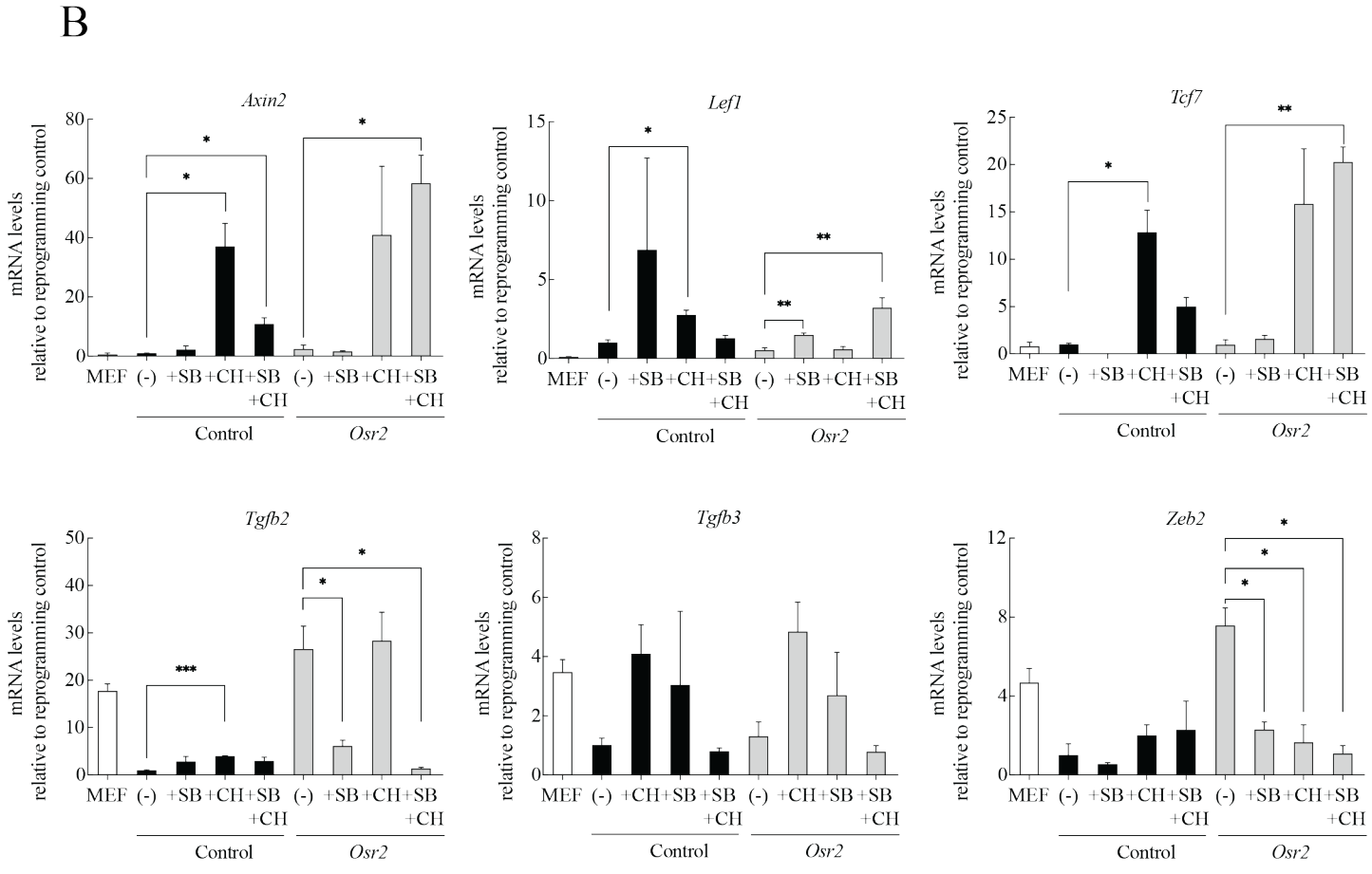
Figure 11. Global gene expression profiles of reprogramming cells expressing *Osr2*

(A) Scheme of collecting 4 kinds of cells for RNA-seq. *Osr2*-expressing or control MEFs were treated with or without 10 μg/mL SB431542 from 3 days before reprogramming by SeVdp(KOSMaB). Total RNAs were collected at day 5 of reprogramming.

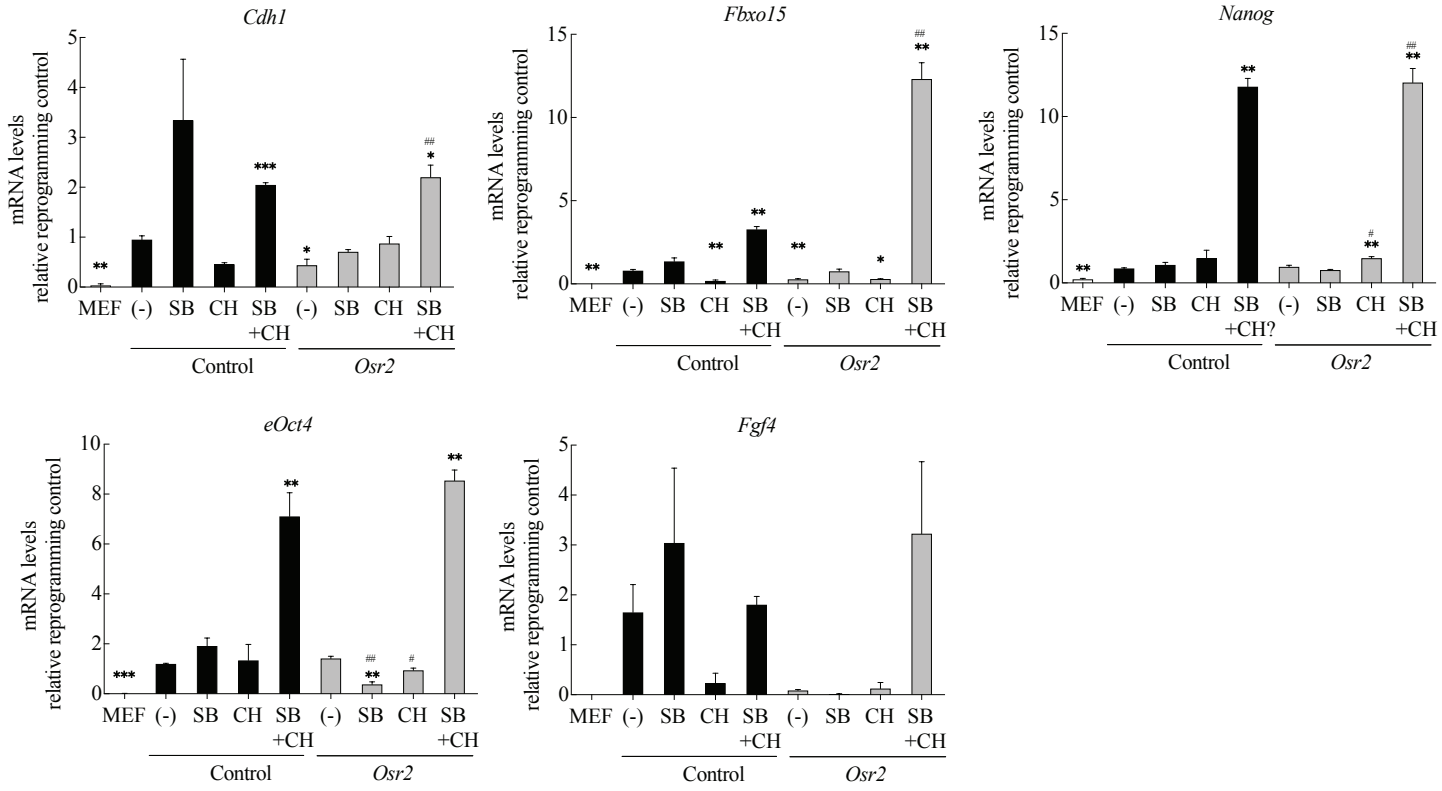
(B) Highly enriched pathways in DEGs by exogenous *Osr2* expression. The genes whose TPM changed over 3-fold by exogenous *Osr2* expression were selected as DEGs. Common pathways found in both analyses are highlighted.

(C) Changes in the mRNA expression level of genes related to Wnt signaling. *Osr2*-expressing MEFs were reprogrammed by SeVdp(KOSMaB), and, at days 3 and 5 of reprogramming, mRNA levels of indicated genes were determined. Data represent mean ± SEM from three independent experiments. *$P < 0.05$ versus MEF. $^{\#}P < 0.05$ versus control reprogramming at each day.


A
(-)
SB
CH
+SB+CH
Control
Osr2
Total
AP(+)
AP(-)
Control
Osr2
Colonies
(-) SB CH SB + CH
ns
**
*


B
Axin2
Lef1
Tcf7
Tgfb2
Tgfb3
Zeb2
mRNA levels relative to reprogramming control
MEF (-) +SB +CH +SB +CH (-) +SB +CH +SB +CH
Control
Osr2
*
**
***

C


Cdh1
Fbxo15
Nanog
eOct4
Fgf4
mRNA levels
relative reprogramming control
MEF (-) SB CH SB +CH (-) SB CH SB +CH
Control
Osr2
SB +CH?

D

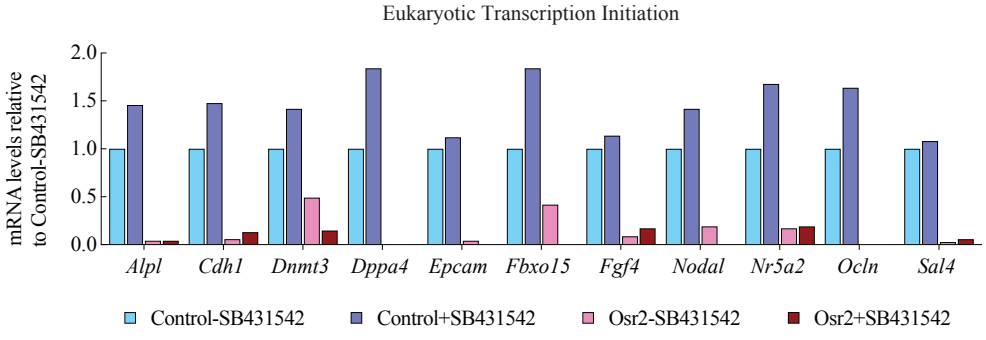

Eukaryotic Transcription Initiation
mRNA levels relative to Control-SB431542
Alpl Cdh1 Dnmt3 Dppa4 Epcam Fbxo15 Fgf4 Nodal Nr5a2 Ocln Sal4
Control-SB431542
Control+SB431542
Osr2-SB431542
Osr2+SB431542

Figure 12. Functional relevance of OSR2 to Wnt signaling during reprogramming

(A) Generation of iPSC colonies from MEFs expressing *Osr2* in the presence of TGF-β inhibitor and Wnt activator. *Osr2*-expressing MEFs were reprogrammed by SeVdp(KOSM), followed by AP staining at day 10 of reprogramming. 10 μg/mL SB431542 and/or 3 μM CHIR99021 was added from 3 days before reprogramming. Data represent mean ± SEM from three independent experiments. *P < 0.05, **P < 0.01.

(B) and (C) Changes in the mRNA expression level of Wnt-related genes (B) and pluripotent related gens (C). MEFs were reprogrammed as described in (A) using SeVdp(KOSMaB), then the mRNA levels of indicated genes were determined at day 5 (B) and day 10 (C) of reprogramming. Data represent mean ± SEM from three independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 versus control reprogramming neither with SB431542 or CHIR99021. $^{\#}$P < 0.05, $^{\#\#}$P < 0.01 versus control reprogramming under each condition.

(D) mRNA levels of indicated genes were extracted from RNA-seq data in Figure 11 and normalized to those in control reprogramming without treatment of SB431542.

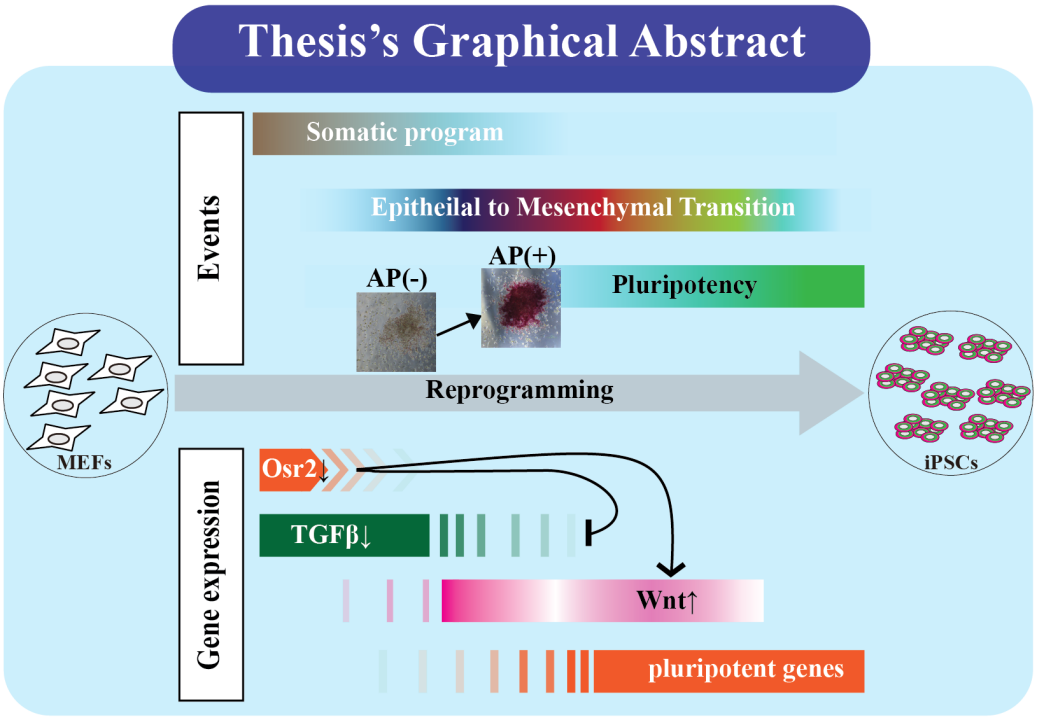


Figure 13. Thesis's graphical abstract